Imagining the future at the Sundance Film Festival p. 1188

Shrinking Colorado River flow pp. 1192 & 1252

Structure of a key SARS-CoV-2 protein p. 1260

Science

$15
13 MARCH 2020
sciencemag.org

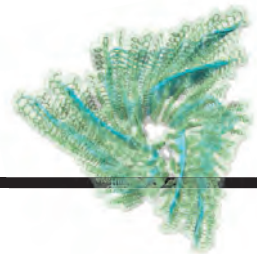
AAAS

AFTER THE FIRE

Science guides rebuilding at Notre Dame p. 1182

CONTENTS

13 MARCH 2020 • VOLUME 367 • ISSUE 6483



1230 Close-up view of a functional amyloid

1177 The limits of airport infection screening

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

BOOKS *ET AL.*

PERSPECTIVES

POLICY FORUM

LETTERS

Published by AAAS

RESEARCH

ON THE COVER

The spire of Notre Dame cathedral collapses in a cloud of smoke. The Paris landmark was nearly destroyed on 15 April 2019, after a fire swept through attic timbers, melted its lead roof, and damaged stone walls. Scientists are now guiding its restoration. By opening the cathedral up to inspection, the fire is also enabling new historical research. See page 1182.
Photo: Geoffroy Van Der Hasselt/AFP/Getty Images

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. 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: SEBASTIAN KENNERKNECHT/MINDEN PICTURES

Published by AAAS

EDITORIAL

Do us a favor

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

"Do me a favor, speed it up, speed it up." This is what U.S. President Donald Trump told the National Association of Counties Legislative Conference, recounting what he said to pharmaceutical executives about the progress toward a vaccine for severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (COVID-19). Anthony Fauci, the long-time leader of the National Institute of Allergy and Infectious Diseases, has been telling the president repeatedly that developing the vaccine will take at least a year and a half—the same message conveyed by pharmaceutical executives. Apparently, Trump thought that simply repeating his request would change the outcome.

China has rightfully taken criticism for squelching attempts by scientists to report information during the outbreak. Now, the United States government is doing similar things. Informing Fauci and other government scientists that they must clear all public comments with Vice President Mike Pence is unacceptable. This is not a time for someone who denies evolution, climate change, and the dangers of smoking to shape the public message. Thank goodness Fauci, Francis Collins [director of the U.S. National Institutes of Health (NIH)], and their colleagues across federal agencies are willing to soldier on and are gradually getting the message out.

> "...start treating science and its principles with respect."

While scientists are trying to share facts about the epidemic, the administration either blocks those facts or restates them with contradictions. Transmission rates and death rates are not measurements that can be changed with will and an extroverted presentation. The administration has repeatedly said—as it did last week—that virus spread in the United States is contained, when it is clear from genomic evidence that community spread is occurring in Washington state and beyond. That kind of distortion and denial is dangerous and almost certainly contributed to the federal government's sluggish response. After 3 years of debating whether the words of this administration matter, the words are now clearly a matter of life and death.

And although the steps required to produce a vaccine could possibly be made more efficient, many of them depend on biological and chemical processes that are essential. So the president might just as well have said, "Do me a favor, hurry up that warp drive."

I don't expect politicians to know Maxwell's equations for electromagnetism or the Diels-Alder chemical reaction (although I can dream). But you can't insult science when you don't like it and then suddenly insist on something that science can't give on demand. For the past 4 years, President Trump's budgets have made deep cuts to science, including cuts to funding for the Centers for Disease Control and Prevention and the NIH. With this administration's disregard for science of the Environmental Protection Agency and the National Oceanic and Atmospheric Administration, and the stalled naming of a director for the Office of Science and Technology Policy—all to support political goals—the nation has had nearly 4 years of harming and ignoring science.

Now, the president suddenly needs science. But the centuries spent elucidating fundamental principles that govern the natural world—evolution, gravity, quantum mechanics—involved laying the groundwork for knowing what we can and cannot do. The ways that scientists accumulate and analyze evidence, apply inductive reasoning, and subject findings to scrutiny by peers have been proven over the years to give rise to robust knowledge. These processes are being applied to the COVID-19 crisis through international collaboration at breakneck, unprecedented speed; *Science* published two new papers earlier this month on SARS-CoV-2, and more are on the way. But the same concepts that are used to describe nature are used to create new tools. So, asking for a vaccine and distorting the science at the same time are shockingly dissonant.

A vaccine has to have a fundamental scientific basis. It has to be manufacturable. It has to be safe. This could take a year and a half—or much longer. Pharmaceutical executives have every incentive to get there quickly—they will be selling the vaccine after all—but thankfully, they also know that you can't break the laws of nature to get there.

Maybe we should be happy. Three years ago, the president declared his skepticism of vaccines and tried to launch an antivaccine task force. Now he suddenly loves vaccines.

But do us a favor, Mr. President. If you want something, start treating science and its principles with respect.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

Published online 11 March 2020; 10.1126/science.abb6502

NEWS

> “Refusal of the challenge was not an option.”
>
> **Virginia seventh grader Alexander Mather**, on entering NASA’s contest to name its next Mars rover. His proposal, Perseverance, was the winner.

IN BRIEF

Edited by **Kelly Servick**

A gene-editing trial aims to restore function to light-converting retinal cells (rodlike structures, center).

BIOMEDICINE

Blindness trial marks a CRISPR milestone

CRISPR gene editing is being tried inside someone’s body for the first time. Researchers at Oregon Health & Science University (OHSU) reported last week that they’ve used the technique to treat a person who has an inherited form of blindness called Leber congenital amaurosis 10. It’s caused by a gene mutation in retinal cells that disrupts production of a protein needed to convert light to electrical signals. Previous CRISPR trials have edited cells outside the body and reinserted them, and an older gene-editing tool has been used to modify patients’ liver cells in the body. In the new trial, sponsored by Allergan and Editas Medicine, researchers injected under the patient’s retina a harmless virus carrying DNA for the CRISPR gene-editing molecules. The gene editor cuts out the disease-causing mutation and allows the cell to stitch the gene back together. OHSU researchers hope the gene will be repaired in one-tenth to one-third of the patient’s retinal cells. The modification is not passed on to offspring. The researchers expect it will take up to 1 month to determine whether the patient has any restored vision. The trial will test the therapy’s safety and preliminary effectiveness in 18 adults and children.

Coronavirus disruptions deepen

INFECTIOUS DISEASE | The spread of the novel coronavirus in the United States is scuttling plans at federal agencies and scientific societies. Travel disruptions have prompted NASA to postpone until later this year three aircraft-based research campaigns that were set to fly this spring. NASA’s Ames Research Center closed its doors to all nonessential personnel and moved its staff to telework after an employee tested positive for COVID-19, the illness caused by the coronavirus. The National Science Foundation announced that grant proposal review panels would be held virtually for at least the next 2 weeks. And the National Institutes of Health said that for 30 days, grant review panels and meetings “that are not mission critical” should be held virtually, postponed, or canceled. The Conference on Retroviruses and Opportunistic Infections, the largest annual HIV/AIDS meeting held in the United States, opted to go virtual just 2 days before its start on 8 March, noting that “infectious disease physicians are urgently needed to care for patients with COVID-19 in their own institutions.” And the American Chemical Society, which was expected to host more than 10,000 researchers at its annual meeting in Philadelphia from 22–26 March pulled the plug on the event, saying the gathering would be “inadvisable and impractical” given the virus’ spread in the area.

Virus curbs North Korean TB work

PUBLIC HEALTH | North Korea watchers warn that the country’s aggressive measures to defend itself against the coronavirus are hobbling efforts to combat tuberculosis (TB) and other infectious diseases. The incidence of COVID-19 in North Korea is unknown. But the nation has stopped the flow of most goods and people across its borders, and its state media says the restrictions will remain until the virus has stopped spreading or a vaccine is available. Three cargo containers of first-line TB drugs are among hundreds held up in the nearby port of Dalian, China, says a U.S.-based humanitarian organization official who requested anonymity. The official

PHOTO: SPL/SCIENCE SOURCE

Published by AAAS

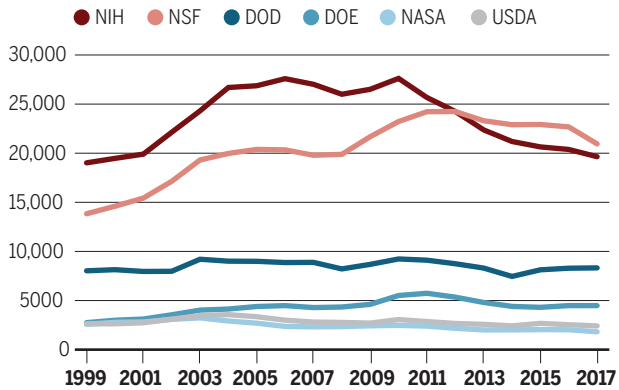
notes that North Korea's supply of such drugs is expected to run out by June, and that treatment lapses in patients infected with multidrug-resistant TB strains can lead to even more recalcitrant strains.

EPA advisers pan water rule

ENVIRONMENTAL SCIENCE | Advisers have criticized as scientifically unfounded a decision by the U.S. Environmental Protection Agency (EPA) to narrow the scope of protections under the Clean Water Act—changes that would leave about half of U.S. wetlands vulnerable to being harmed or destroyed. In a rule finalized 23 January that reversed one issued under former President Barack Obama, the Trump administration asserted that only "navigable waters" qualify for protection; the new policy omits streams that dry up occasionally and wetlands joined to lakes or rivers only via groundwater. In a 27 February letter to the agency's head, EPA's Science Advisory Board objected, citing research findings that indirect connections with larger water bodies make these smaller sites important for water quality and ecological health. The advisers also warned that irrigation canals—excluded from protection under the new rule—can contaminate crops with dangerous bacteria.

Ebb and flow of U.S. training

A decade of flat budgets caused the number of graduate students funded by the National Institutes of Health (NIH) to dip below the total for the much smaller National Science Foundation. But NIH is likely to regain its lead thanks to a string of healthy increases that began in 2016.



DRC counts down to Ebola-free

INFECTIOUS DISEASE | The 19-month Ebola outbreak in the Democratic Republic of the Congo may be nearing an end. Last week, a treatment center in Beni released its final patient, starting the 42-day countdown—two 21-day incubation periods—until officials can declare the outbreak over. Since August 2018, at least 3310 people have been infected with Ebola and 2130 have died. Attacks on treatment centers have killed another 11 health care workers and patients and injured 86. Officials are watching carefully for new infections—inaccessible, war-torn areas could still see cases, and some patients can transmit the virus months after recovery. But WHO Director-General Tedros Adhanom Ghebreyesus said the 3 March announcement was a very welcome gift on his 55th birthday.

CREDITS: (PHOTO) ANDREW TESTA/*THE NEW YORK TIMES*; (GRAPHIC) J. BRAINARD AND J. MERVIS/*SCIENCE*; (DATA) *SCIENCE & ENGINEERING INDICATORS 2020*/U.S. NATIONAL SCIENCE FOUNDATION

BIOMEDICINE

'London patient,' still HIV-free, goes public

A 40-year-old former chef, Adam Castillejo, revealed this week that he is the "London patient," the second person in the world that physicians believe has been cured of an HIV infection. "I want to be an ambassador of hope," he told *The New York Times*. Castillejo's doctor first described his case last year at an HIV/AIDS conference and in a *Nature* paper. At that point, no HIV had been detected in Castillejo's blood for 18 months after a stem cell transplant to treat a life-threatening blood cancer, but his doctor was reluctant to call him cured. A similar transplant, with bloodmaking stem cells bearing a mutation that made them resistantto HIV, led to the 2007 cure of Timothy Ray Brown, known as the "Berlin patient." That Castillejo remains HIV-free "suggests that the Berlin patient is not a one-off case," says stem cell researcher Deng Hongkui of Peking University, whose group has tried—so far unsuccessfully—to mimic the resistant mutation in donor stem cells using the genome editor CRISPR. For now, this cure isn't widely applicable because excellent HIV drugs exist and stem cell transplants have serious side effects that restrict their use to people with other serious conditions.

BY THE NUMBERS

156
kilometers

Proximity to the North Pole of the MOSAiC expedition's research icebreaker *Polarstern* as of 24 February, when it set a record for the farthest north a ship has ventured in Arctic winter.

63%

Decline in area of Myanmar covered by mangroves between 1996 and 2016 as a result of "catastrophic deforestation" to expand agriculture. (*Environmental Research Letters*)

~1000

Vultures found dead in Guinea Bissau since February, possibly because of eating poisoned animal carcasses meant to kill feral dogs. Most were hooded vultures (*Necrosyrtes monachus*), a species already critically endangered. (Vulture Conservation Foundation)

PHOTO: LIDA XING

PALEONTOLOGY

Asian amber captures the tiniest dino

This tiny head, just 7 millimeters long (about the size of a housefly), is a remnant of one of the smallest dinosaurs ever found. Entombed in amber for nearly 100 million years, it belonged to the group of dinosaurs that gave rise to modern birds. It was probably about the size of the bee hummingbird, the smallest living bird. The fossil, discovered in Myanmar, is called *Oculudentavis khaungraae*, or "eye-tooth bird." It has large eye sockets on the sides of its head like modern lizards, and its eyes have narrow openings that limit incoming light. That's a strong hint that the animal was active during the day. Its upper and lower jaws are full of sharp teeth—the most found on any ancestral bird—which implies that it was a predator that likely ate insects and other small invertebrates, researchers report this week in *Nature*. They think the species' tiny size is an example of "island dwarfism," as it likely inhabited an arc of islands that existed where Myanmar is today. Without the rest of the body, they can't tell exactly how *Oculudentavis* is related to other birdlike dinosaurs, but the researchers suspect it belongs to a group of relatively primitive birds, perhaps similar to *Archaeopteryx* and *Jeholornis*, species that lived between 150 million and 120 million years ago.

U.K. to test TB vaccine in cows

WILDLIFE MANAGEMENT | After years of controversy, the U.K. government is shifting its approach to bovine tuberculosis (bTB), a disease that leads to the culling of more than 30,000 cows per year. The disease has been hard to eliminate, in part because badgers can spread it to cattle; shooting the wild animals has led to protests. Last week, the Department for Environment, Food & Rural Affairs outlined a plan to phase out badger culling and to trap and vaccinate more badgers. Adopting recommendations from an independent scientific review of the government's 25-year strategy (*Science*, 16 November 2018, p. 729), the agency says it will conduct field trials for a more sensitive bTB test to catch outbreaks sooner. It also plans to fund testing of a cattle vaccine that might be deployed within 5 years.

MeTooSTEM backs founder

WORKPLACE | The board of MeTooSTEM last week stood behind the group's embattled founder, BethAnn McLaughlin, despite recent calls for her resignation. McLaughlin, who launched the nonprofit in 2018 to support survivors of sexual harassment in science, has been under fire from former group members who allege that she bullies colleagues and sidelines women of color. In a 2 March statement, the board said it had asked McLaughlin "to continue to serve in a leadership position." McLaughlin said she was "heartened by the board's response." The board includes McLaughlin's brother, John McLaughlin; Nobel laureate Carol Greider, a biologist at Johns Hopkins University; and Vicki Lundblad, who sued the Salk Institute for Biological Studies for gender discrimination in 2017. The lawsuit was settled out of court.

Europe aims for zero emissions

CLIMATE CHANGE | The European Commission last week unveiled legislation that would make the bloc of 27 nations climate neutral by 2050, turning into law promises made last year by every member except Poland. The proposed law, which requires approval by the European Parliament and its member states, would reduce net greenhouse gas emissions to zero by 2050 across the entire bloc; some nations could overachieve to offset laggards. The law doesn't mention targets for 2030, a more contentious issue. Activists, including Swedish teenager Greta Thunberg, have urged the European Union to improve on its current 2030 goal of 40% reductions. The commission says it intends to put forth a plan by September to raise that target to at least 50%.

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

IN DEPTH

Italy's COVID-19 outbreak has led to empty tables in St. Mark's Square in Venice.

PHOTO: REUTERS/MANUEL SILVESTRI

INFECTIOUS DISEASES

Genome analyses help track coronavirus' moves

Mutations can reveal connections between outbreaks—but it's easy to overinterpret them

By **Kai Kupferschmidt**

Immediately after Christian Drosten published a genetic sequence of the novel coronavirus online on 28 February, he issued a warning on Twitter. As the virus has raced around the world, more than 350 genome sequences have been shared on GISAID, an online platform. They offer clues to how the virus, named SARS-CoV-2, is spreading and evolving. But because the sequences represent a tiny fraction of cases and show few telltale differences, they are easy to overinterpret, as Drosten realized.

A virologist at the Charité University Hospital in Berlin, Drosten had sequenced the virus from a German patient infected in Italy. The genome looked similar to that of a virus found in a patient in Munich, the capital of Bavaria, more than 1 month earlier; both shared three mutations not seen in early sequences from China. Drosten realized the similarity could suggest the Italian outbreak was "seeded" by the one in Bavaria, which state public health officials said they had quashed by tracing and quarantining all contacts of the 14 confirmed cases. But he thought it was just as likely that a Chinese variant carrying the three mutations had taken independent routes to both countries. The newly sequenced genome "is not sufficient to claim a link between Munich and Italy," Drosten tweeted.

His warning went unheeded. A few days later, Trevor Bedford of the Fred Hutchinson Cancer Research Center, who analyzes the stream of viral genomes, tweeted that the pattern "suggested" that the outbreak in Bavaria had not been contained after all, and had touched off the Italian outbreak. The analysis spread widely on Twitter and elsewhere—this *Science* correspondent retweeted the thread as well—and some Twitter users called on Germany to apologize.

Virologist Eeva Broberg of the European Centre for Disease Prevention and Control agrees with Drosten that there are more plausible scenarios for how the disease reached northern Italy than undetected spread from Bavaria. Other scientists agree. "I have to kick [Bedford's] butt a bit for this," says Richard Neher, a computational biologist at the University of Basel who works with Bedford. "It's a cautionary tale," says Andrew Rambaut, a molecular evolutionary biologist at the University of Edinburgh. "There is no way you can make that claim just from the phylogeny alone." Bedford now acknowledges as much. "I think I should have been more careful with that Twitter thread."

It was a case study in the power and pitfalls of real-time analysis of viral genomes. "This is an incredibly important disease. We need to understand how it is moving," says Bette Korber, a biologist at the Los Alamos National Laboratory who is also studying the genome of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2). But for now, scientists who analyze genomes can only make "suggestions," she says.

The very first SARS-CoV-2 sequence, in early January, answered the most basic question about the disease: What pathogen is causing it? The genomes that followed were almost identical, suggesting the virus, which originated in an animal, had crossed into the human population just once. If it had jumped the species barrier multiple times, the first human cases would show more variety.

Some diversity is now emerging. Over the length of its 30,000-base-pair genome, SARS-CoV-2 accumulates an average of about one to two mutations per month, Rambaut says. Using these little changes, researchers draw up phylogenetic trees, much like family trees, make connections between cases, and gauge whether there might be undetected spread of the virus.

For example, the second virus genome sequenced in Washington—from a teenager diagnosed on 27 February—looked like a direct descendant of the first genome, from a case found 6 weeks earlier. Bedford tweeted that he considered it "highly unlikely" that the two genomes came from

separate introductions, and said the virus must have been circulating undetected in Washington. Both patients came from Snohomish County, making the link far more persuasive than the one Bedford drew between Bavaria and Italy, Rambaut says: "It's very unlikely that this highly related virus would travel to exactly the same town in Washington." By now the state has reported more than 160 cases, and genomes from additional patients have bolstered the link Bedford suspected.

Still, the wealth of genomes is just a tiny sample of the more than 100,000 cases worldwide, and it's uneven. On 9 March, Chinese scientists uploaded 50 new genome sequences—some of them partial—from COVID-19 patients in Guangdong province; most previous ones were from Hubei province. But overall, less than half of the published genomes are from China, which accounts for 80% of all COVID-19 cases. And sequences from around the world are still very similar, which makes drawing firm conclusions hard. "As the outbreak unfolds, we expect to see more and more diversity and more clearly distinct lineages," Neher says. "And then it will become easier and easier to actually put things together."

Scientists will also be scouring the genomic diversity for signs that the virus is getting more dangerous. There, too, caution is warranted. An analysis of 103 genomes published by Lu Jian of Peking University and colleagues on 3 March in the *National Science Review* argued they fell into one of two distinct types, named S and L, distinguished by two mutations. Because 70% of sequenced SARS-CoV-2 genomes belong to L, the newer type, the authors concluded that this type has evolved to become more aggressive and to spread faster.

"What they've done is basically seen these two branches and said, that one is bigger, [so that virus] must be more virulent or more transmissible," Rambaut says. But other factors could be at play. "One of these lineages is going to be bigger than the other just by chance." Some researchers have called for the paper to be retracted. "The claims made in it are clearly unfounded and risk spreading dangerous misinformation at a crucial time in the outbreak," four scientists at the University of Glasgow wrote on www.virological.org. In a response, Lu wrote that the four had misunderstood his study.

Most genomic changes don't alter the behavior of the virus, Drosten says. The only way to confirm that a mutation has an effect is to study it in the lab and show, for instance, that it has become better at entering cells or transmitting, he says. So far, the world has been spared that piece of bad news. ■

GLOBAL HEALTH

Airport screening is largely futile, research shows

Thermometer guns and health questionnaires may look reassuring, but very rarely catch infected travelers

By **Dennis Normile**

Those thermometer guns and health questionnaires used at many international airports to help stop COVID-19 may look reassuring. But research and recent experience shows screening of departing or arriving passengers will do very little to slow the spread of the virus; it's exceedingly rare for screeners to intercept infected travelers.

On 4 March, U.S. Vice President Mike Pence pledged "100% screening" for direct flights arriving in the United States from Italy and South Korea, which both have big outbreaks, in addition to the existing screening for travelers from China. Many other countries, including China itself—which has seen case numbers plummet—are trying to keep out new infections through airport screening.

A passenger is checked for fever at the international airport in Aceh Besar, Indonesia, on 27 January.

PHOTO: OVIYANDI EMNUR/INA PHOTO AGENCY/SIPA USA/AP IMAGES

But the data are sobering. In the first 3 weeks of screening passengers from China, the United States found only one infection among 46,016 travelers, according to a 24 February report from the Centers for Disease Control and Prevention. That clearly didn't stop the virus from entering the country from China. (The travelers were U.S. citizens, permanent residents, and their families who had been in China within the previous 14 days; anybody else who had visited the country was simply denied entry.)

China, meanwhile, didn't catch eight infected restaurant workers from Bergamo, Italy, who flew into Shanghai on 27 and 29 February and took taxis to their hometown. Shanghai Pudong International Airport screens arriving passengers using thermal imaging and requires them to report their health status; it's unclear whether any of the eight had symptoms, or what they said about their health.

There are many ways infected people can slip through the net. Thermal scanners and handheld thermometers measure skin temperature, which can be higher or lower than core body temperature, the key metric for fevers. The devices produce false positives as well as false negatives.

Passengers can also take fever-suppressing drugs or lie about their symptoms and recent whereabouts. Most important, infected people still in their incubation phase are free of symptoms, so they are often missed. For COVID-19, that period can be anywhere between 2 and 14 days.

A review published in November 2019, before the new coronavirus had emerged, confirms the dismal success rate for airport screening. Christos Hadjichristodoulou and

Varvara Mouchtouri of the University of Thessaly scrutinized 114 scientific papers and reports on screening for infections, including Ebola, severe acute respiratory syndrome (SARS), and pandemic influenza. Between August 2014 and January 2016, they found, not a single Ebola case was detected among 300,000 passengers screened before boarding flights in Guinea, Liberia, and Sierra Leone, the three countries hit hard by the West African Ebola epidemic. Yet at least four infected passengers slipped through exit screening because they didn't have symptoms yet, and flew to their destinations.

Screening of arriving passengers did no better. During the same epidemic, five countries asked incoming travelers about symptoms and possible exposure and checked for fevers. They didn't find cases either, but two Ebola-infected passengers slipped through, one in the United States and one in the United Kingdom. (The other two flew to destinations without entry screening.) During the H1N1 flu pandemic of 2009, the researchers found, China and Japan intercepted only tiny fractions of infected travelers, and both had significant outbreaks anyway. The duo doesn't expect airport screening for COVID-19 to be more effective.

Yet screening programs are costly: Canada spent an estimated $5.7 million on a fruitless SARS entry screening program in 2003 and Australia spent $50,000 per detected H1N1 case in 2009, Hadjichristodoulou and Mouchtouri say.

It's not entirely wasted money. By implementing exit screening for Ebola, the West African countries hit by the outbreak may have helped head off more draconian travel restrictions by other countries. And simply knowing that screening is in place may deter some infected or exposed people from trying to travel.

The World Health Organization (WHO) has detailed guidelines for countries that do want to screen travelers. Exit screening should start with temperature and symptom checks and interviews of passengers for potential exposure to high-risk contacts, for instance. Symptomatic travelers should be given further medical examination and testing, and confirmed cases should be moved to isolation and treatment. Entry screening is also an opportunity to gather contact information—useful in case passengers become infected during a flight—and to give travelers guidance, WHO says.

But even when done well, screening mostly serves to show governments are doing something, says epidemiologist Ben Cowling of the University of Hong Kong. At best, he says, "Measures aimed at catching infections in travelers will only delay a local epidemic and not prevent it." ■

CONSERVATION BIOLOGY

Madagascar's mysterious, murderous cats identified

Ancestors of large cats likely hopped off Arabian trading ships more than 1000 years ago

By **Joshua Sokol**

On the trail floor that day in 2009 lay the sprawled body of a white-furred sifaka, a kind of lemur. "I touched the bottom of his foot," said Michelle Sauther, a biological anthropologist at the University of Colorado, Boulder. "It was still warm." Then she heard a rustle. Looking up, she caught a glimpse of a tiger-striped feline dissolving back into the forest—one of Madagascar's "forest cats."

Cats didn't evolve on the island, and the history of these elusive felines—twice the size of house cats—has long been a mystery. Now, researchers have revealed the cats' origin story: They descend from domestic kitties that hopped off Arabian trading ships perhaps more than 1000 years ago. By pinpointing them as a separate population that has spent centuries adapting to Madagascar, the work may offer a first step toward limiting the toll these relentless hunters take on the island's rich biodiversity.

A camera trap caught this image of a Madagascar forest cat on its home turf.

PHOTO: JULIE POMERANTZ AND LUKE DOLLAR

With males averaging more than 0.6 meters long, the forest cats have striped tabby coats, straight tails, and a voracious appetite for native birds, snakes, rodents, and lemurs. They also compete with endemic carnivores like mongooses, said Zachary Farris, a biologist at Appalachian State University who was unaffiliated with the research team.

The felines could be the feral descendants of the domestic cat *Felis catus* brought to the island several hundred years ago by Europeans; if so, controlling domestic village cats might limit the population in the forest. Or they might be descendants of small wildcats "that had somehow gotten over here from mainland Africa," Sauther says.

But Sauther's team uncovered a different story when it sampled DNA from the blood of forest cats trapped using live mice or beef parts as bait. Leslie Lyons, an expert in cat genomics at the University of Missouri, Columbia, helped compare the forest cat genomes with those of cats around the world. The closest match: domestics from Arabian Sea locales such as Kuwait and Oman, the researchers reported at the end of February in the journal *Conservation Genetics*. Like other domestic cats that went wild, including Maine coons and feral cats in Australia, the Middle Eastern cats swelled in size in their new home, Lyons notes.

The Arabian origin "makes sense," said Asia Murphy, a Ph.D. student at Pennsylvania State University, University Park, who studies the fossa, an endemic carnivore that competes with forest cats. "Madagascar is a pretty special place when it comes to cultural mixing."

Linguistic and cultural evidence attests to Arabic influence on the island, linked to Indian Ocean trade routes that stretched from Arabian ports to Madagascar starting in the second millennium B.C.E. Cats employed as mousers on those ships could have deserted at port.

Another invasive species supports that scenario. Arabian ships also transported Indian civets to the island around 900 C.E. for the oil produced in their anal glands, which was used in perfumes. "Boats transporting civets [likely] were also carrying cats," Farris said.

More genomics work could tighten the timeline of when the cats arrived or tell the story of another forest cat variety, called the fitoaty, that researchers haven't yet sampled. And knowing the cats aren't just recent runaways suggests trapping in the forest, rather than simply neutering village cats, might be the quickest way to control them, Murphy said.

For Lyons, this record of Arabian Sea cats sailing to Madagascar adds to the global story of cat dispersal. "People think about dogs all the time," she says, "but the cat has been a very silent partner in our migration." ■

Joshua Sokol is a journalist in Boston.

ILLUSTRATION: DAVIDE BONAZZI/SALZMAN ART

SCHOLARLY PUBLISHING

Publishers try out alternative pathways to open access

But can they overcome free riders and price concerns?

By **Jeffrey Brainard**

In the push for "open access" (OA)—making scientific papers immediately free to everyone—it's easy to forget that publishing costs haven't vanished. They have simply shifted from subscriptions paid mostly by university librarians to fees charged to authors. Those article-processing fees (APCs), which can run several thousand dollars per paper, raise concerns of their own. Universities fear they could end up paying more to help their scientists publish their work than they do now for subscriptions. Scientists who have small research budgets fret that they won't be able to afford APCs. And some nonprofit scientific societies that publish journals worry APCs won't generate enough revenue to support other activities, such as meetings and training.

Now, two nonprofit publishers of prominent journals have debuted new ways to support OA journals without shifting the burden entirely to authors. "Everybody that we work with is watching these two [new models] closely," says Michael Clarke of the publishing consulting firm Clarke & Esposito. "There is not currently a good solution."

One approach, called Subscribe to Open and implemented this week by Annual Reviews, would transform the nature of subscriptions. To make a journal freely available, institutions would be asked for a contribution equivalent to their previous subscription—minus a 5% discount that Annual Reviews is offering to retain a critical mass of paying institutions. To deter freeloading, Annual Reviews says it will reimpose paywalls and rescind the discount if not enough subscribers renew each year. It is planning to pilot the approach in up to five of its 51 titles, many of which are widely cited.

The Association for Computing Machinery (ACM) launched a different approach earlier this year. ACM is asking the institutions that publish the most papers in its 59 journals to pay more than they do now for subscriptions—in some cases about 10 times as much, or $100,000 per year. The higher fees will allow all researchers at participating universities to publish an unlimited number of papers in ACM journals without paying APCs. The average cost per paper will beat the average market rate for APCs, the society says. ACM is betting the approach will sustain its journal revenue while it transitions to making all the 21,000 peer-reviewed papers it publishes annually free to everyone.

So far, both approaches are getting a positive response. At Annual Reviews, some 90% of subscribers have signed deals that, on 9 March, allowed the publisher to remove the paywall from the *Annual Review of Cancer Biology*, says Richard Gallagher, president and editor-in-chief. Annual Reviews could roll out OA for up to four additional journals this spring if other librarians accept the model, Gallagher says. Curtis Brundy, a library administrator at Iowa State University, which is participating in the pilot, believes that "hands down, Subscribe to Open is our best option as an alternative to APCs. It's simpler to implement, and we don't have a lot of other models."

Universities are also starting to embrace the ACM model. In January, several that produce the most ACM papers, including the Massachusetts Institute of Technology and Carnegie Mellon University, signed 3-year deals that lock in the higher payments. ACM is optimistic more universities will follow.

Still, both publishers concede there are risks. A big one is the free-rider problem. "People will start to ask themselves, especially if we experience budget cuts in a given library, 'Why am I paying for this when other people aren't?'" says Lisa Janicke Hinchliffe of the University Library at the University of Illinois, Urbana-Champaign.

Persuading some institutions to pay more than they do now is another challenge. ACM calculated that it needed to do so because two-thirds of its subscription revenue comes from about 1700 institutions that publish three papers per year or fewer in any ACM journal. The rest comes from the roughly 1000 institutions that each publish more than three papers annually. Once the journals become free, the 1700 less productive institutions will have less incentive to subscribe, and any APCs they pay won't replace the lost revenue.

Instead, under ACM's plan for OA, the 11 schools that publish more than 75 papers per year in ACM journals would pay the highest rate for unlimited publishing, $100,000 annually, a 10-fold increase over current subscriptions. In one of the lower tiers, the roughly 90 universities that publish 20 to 29 papers annually would pay $35,000 annually. Overall, ACM estimates that if half of the top 1000 institutions sign on, their journals can transition to full OA within 5 to 7 years. Whether the pricing structure works might not be known for at least 3 years, when the initial contracts expire.

Even if the models succeed, it's not clear they would spread. The Annual Reviews journals, for instance, are unusual: Although each appears just once a year and contains only review articles, many are must-haves for libraries.

The new experiments underscore that the search for viable pathways to OA publication continues. The road to success likely runs through university librarians willing to dedicate some of their tight budgets to testing new arrangements. "We want to support experimentation," says Brundy, who recently negotiated a deal under ACM's new model, which doubles what the institution was paying annually, to $17,500. "It comes down," he says, "to how much of a commitment a library has to open access." ■

Pesticide research often involves confidential health data, which can be difficult to make publicly available.

PHOTO: ISTOCK.COM/THOMAS DE WEVER

U.S. REGULATORY SCIENCE

EPA expands controversial 'transparency' plan

Critics blast revised proposal that covers a wider range of research, saying it will undermine protections

By **Erik Stokstad**

"We want to make sure we get this right," Andrew Wheeler, head of the U.S. Environmental Protection Agency (EPA), said in September 2019 about a controversial proposal to narrow the kinds of scientific studies the agency would consider when crafting regulations. Speaking to reporters, he acknowledged that the initial plan, released in 2018, had raised numerous questions about its practical implications and cost. It had also drawn withering criticism from many U.S. science and environmental groups, who saw it as a way to exclude relevant science on the health effects of pollution (*Science*, 4 May 2018, p. 472).

Now, EPA has released for public comment a document that aims to clarify the proposed rule, dubbed "Strengthening Transparency in Regulatory Science." But all is not right, the critics argue. If anything, they say, the new version is more troubling, because it greatly expands the kinds of science that the policy might put off limits. The rule also proposes to change how EPA conducts in-house studies and analyses, and it gives the politically appointed agency head a greater say in deciding which studies the agency can use.

"The [Trump] administration has doubled down," says Thomas Burke, an epidemiologist at the Johns Hopkins University Bloomberg School of Public Health and EPA science adviser for 2 years until January 2017. "This has very far reaching impacts … [and] looks like it throws an incredible monkey wrench into the agency."

The EPA proposal, first floated by former administrator Scott Pruitt, generally bars the agency from using studies that have not made their underlying data and mathematical models public. Transparency is essential, the agency says, because it allows anyone to independently "reanalyze the data and models and explore the sensitivity of the conclusions to alternative assumptions."

The original proposal appeared to limit the restriction to so-called dose-response studies, used to evaluate risks and set limits on human exposure to pollutants. In contrast, the new version applies to all "influential science," even if the agency does not directly use it to justify regulations. The rule also proposes requiring EPA's own researchers to reconsider long-standing analytical approaches. For example, rather than rely on time-tested models that assume exposure to chemicals or radiation can have some biological effect at any dose, researchers should consider whether a chemical has a hard threshold below which it's safe.

Even critics say scientific transparency is a worthy goal, as is periodically revisiting scientific methods. But they fear the true intent of EPA's rule is to enable industry groups opposed to tougher regulation to prevent the agency from using certain kinds of studies, particularly epidemiological studies that rest on confidential health information, and to entangle regulators in continual reanalyses. "This rule opens the door to endless delay and challenging of the science," Burke says.

One big question is how EPA will handle past studies based on confidential health data. Such studies have been key to setting limits on major air pollutants, including particulate matter and ozone, that the agency must review every 5 years. Researchers say it would be expensive, difficult, and in some cases impossible, to release their underlying data and still honor patient confidentiality agreements. As a result, depending on how the final rule is worded, opponents of tighter controls on air pollution and other rules might be able to challenge EPA's use of past studies when it reviews the limits. "This fundamentally changes the way the agency can use science to protect people," says Gretchen Goldman of the Union of Concerned Scientists, an advocacy organization.

In its proposal, EPA asks for comment on an approach to sharing some confidential health data. This "tiered access" system, modeled on a program at the U.S. Centers for Disease Control and Prevention (CDC), would give greater access to researchers who show they have a bona fide reason for wanting the data, and agree to protect them. But there is a difference between data at CDC and EPA, Goldman says: EPA uses pollution exposure data that includes people's locations, which can make it harder to protect their identities.

An alternate idea, EPA suggests, would be to allow regulators to consider all studies relevant to a particular regulation, regardless of their transparency, but to put a greater weight on those that have released data. But to Linda Birnbaum, former director of the National Institute of Environmental Health Sciences, that approach is "hardly a way to get the best science on which to base decisions."

The public will have 30 days to comment on the proposal after it is published in the *Federal Register*. Congress has also asked the National Academy of Sciences to review the matter. The agency hopes to finalize the rule later this year, although it is likely to face a court challenge. Meanwhile, science and environmental groups are tracking a similar scientific transparency policy under development at the Department of the Interior, which sometimes keeps confidential the locations of endangered species to protect them from poaching or collecting. Last month, the department sent its proposal to the White House for review. ■

MATERIALS SCIENCE

New electrolyzer splits water on the cheap

Devices that forgo expensive metals could turn renewable electricity into hydrogen

By **Robert F. Service**

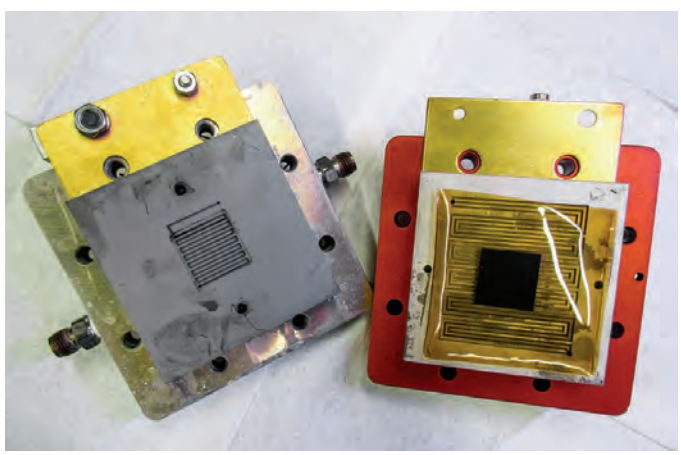
Running the world on renewable energy is simple, in principle: Harvest solar and wind energy, and use any extra to power devices called electrolyzers that split water into oxygen (O_2) and hydrogen gas. Hydrogen (H_2) can serve as a fuel; it is also a staple of the chemical industry. The trouble is that current electrolyzers are costly, requiring either expensive catalysts or pricey metal housings. Now, researchers report combining the best of both approaches to make a version that needs only cheap materials.

"I consider this a great breakthrough," says Hui Xu, a chemical engineer at Giner Inc., an electrochemistry company. Xu says he and his colleagues presented similar results at a Department of Energy meeting last year, but have not yet published them. Their work and another team's new device, described this week in *Nature Energy*, could bolster the global embrace of renewable energy, if the new electrolyzers prove to be cheap and stable during many years of operation. "We are on the cusp of getting that done," says Yushan Yan, a chemical engineer at the University of Delaware, Newark, who is working on similar technology. A handful of small companies, including one he founded, have formed to commercialize it.

Scientists have known how to split water into H_2 and O_2 for more than 200 years: Put two metal electrodes in a jar of water, apply an electrical voltage between them, and H_2 and O_2 will bubble up at separate electrodes. Because a mix of the gases can explode, today's most common setups separate the anode and cathode with a thick, porous plastic sheet. They also use metal catalysts—most often inexpensive ones such as nickel and iron—to speed the reactions.

To make the water able to better conduct ions that move through the devices, today's most common electrolyzers add high levels of potassium hydroxide (KOH) to the water. At the cathode, or negative electrode, water molecules split into H^+ and OH^- ions. The H^+ ions combine with electrons from the cathode to make H_2. The OH^- ions diffuse through the membrane to the anode, or positive electrode, where they react to generate O_2 and water.

But KOH is highly caustic, so engineers have to build their devices out of expensive inert metals such as titanium, says Yu Seung Kim, a chemist at Los Alamos National Laboratory. That drawback prompted researchers in the 1960s to develop a version of the technology known as a proton-exchange membrane (PEM) electrolyzer, in which the dividing membrane is designed to selectively allow H^+ ions through. A PEM cell's catalysts aren't on the electrodes themselves, but are tethered to opposite sides of the membrane. In this setup, catalysts on the anode side split water molecules into H^+ and OH^- ions, with the latter instantly reacting at the catalysts to form O_2 molecules. The H^+ ions then migrate through the plastic membrane to the cathode side, where catalysts tethered to the membrane turn the H^+ ions into H_2.



In this new electrolyzer, an improved ion-conducting membrane (yellow film, right) enables hydrogen generation from water without expensive catalysts.

PHOTO: YU SEUNG KIM/LOS ALAMOS NATIONAL LABORATORY

Because OH^- ions don't migrate through PEM cells, there's no need for highly alkaline conditions. The devices also typically produce hydrogen at five times the rate of the alkaline version. But these membrane cells have their own downsides: They still need some expensive corrosion-resistant metals to withstand acidic conditions produced by the proton-conducting membrane. They also require catalysts made from platinum and iridium. Those metals are expensive and rare. For example, the global production of iridium is only 7 tons. "There is simply not enough [precious metals] for large-scale hydrogen production," Xu says.

Now, Kim and his colleagues at Los Alamos, along with researchers at Washington State University, say they've combined the best of both approaches. Their new device creates a highly alkaline environment to encourage water splitting. But it does so with the PEM approach of tethering catalysts to opposite faces of an ion-conducting membrane. As with the KOH setup, catalysts on the cathode side split water molecules into H^+ and OH^- ions. The former converts to H_2, and the latter travels through the membrane, known as an anion exchange membrane (AEM). It is designed to create a highly alkaline local environment that speeds the travel of OH^- ions to the anode side, where tethered catalysts prompt them to react to make O_2.

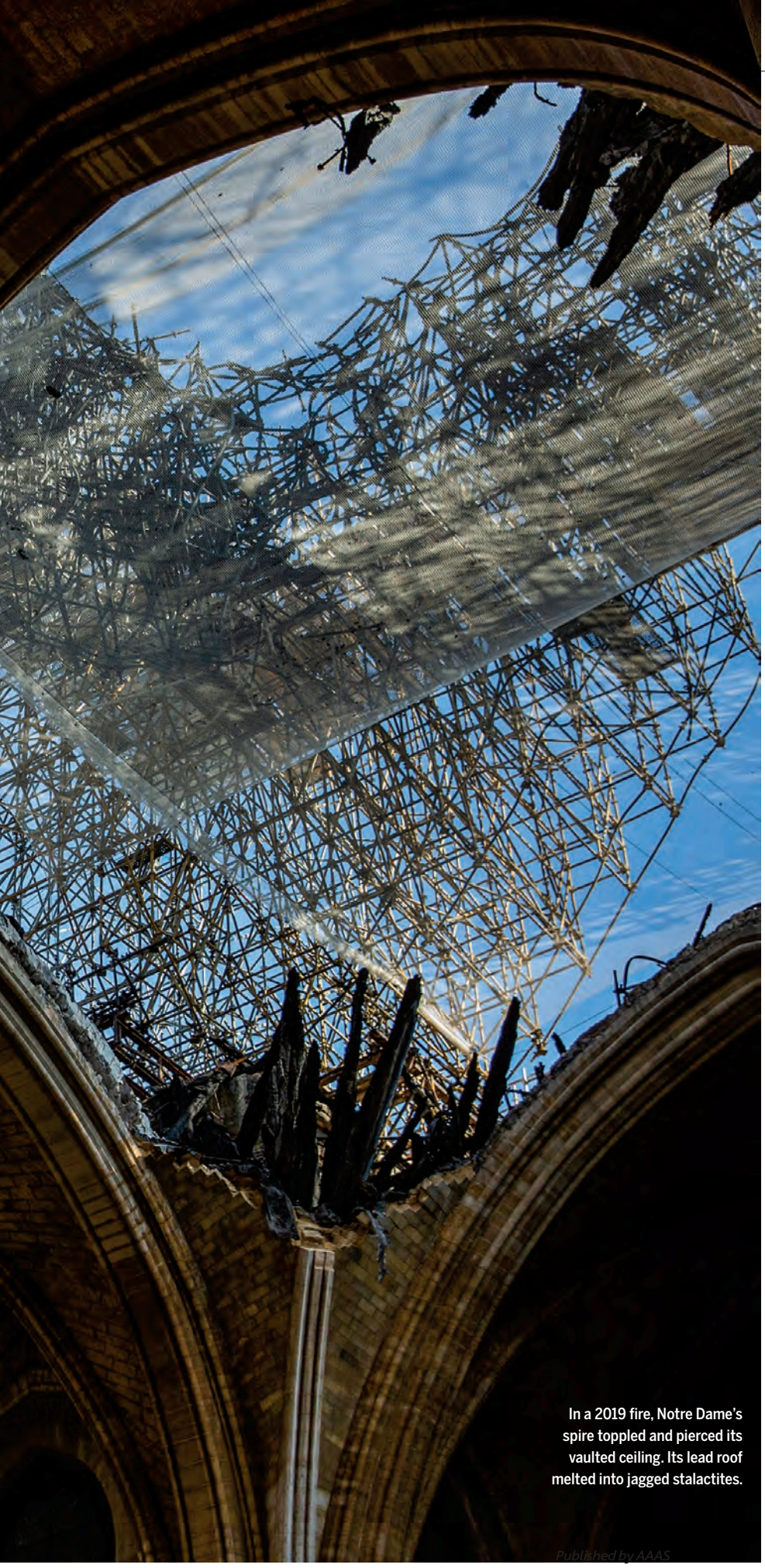
The upshot is that alkaline conditions near the membrane allow the electrolyzer to rely on cheap and abundant nickel-, iron-, and molybdenum-based catalysts to split water. Yet, because the alkalinity is localized, the electrolyzer can be built from stainless steel. The new device generates hydrogen about three times faster than conventional alkaline devices, though still more slowly than commercial PEM electrolyzers, Kim and his colleagues report. "The combination of the older alkaline technology and membrane PEM technology is the path forward," Xu says.

The new setup needs to prove its durability. Initial indications suggest the membrane begins to break down after only about 10 hours of operation. Kim says the main problem is likely that the polymer membrane readily absorbs water. Over time, this may cause the catalyst particles to come unglued and drift away. The team hopes that adding fluorine to the membrane will repel the water. With that and other fixes, Kim hopes, AEM electrolyzers could join solar cells and windmills as a key technology for a carbon-free world. ■

SAVING GRACE

Scientists are leading Notre Dame's restoration—and probing mysteries laid bare by a fire

By **Christa Lesté-Lasserre**



In a 2019 fire, Notre Dame's spire toppled and pierced its vaulted ceiling. Its lead roof melted into jagged stalactites.

Eight restoration scientists put on hard hats and heavy-duty boots and stepped inside the blackened shell of Notre Dame de Paris, the world's most famous cathedral. Ten days earlier, a fire had swept through its attic, melted its roof, and sent its spire plunging like an arrow into the heart of the sacred space. Now, it was silent but for the flutter of house sparrows. The space, normally sweet with incense, was acrid with ash and stale smoke. Light beamed through voids in the vaulted stone ceiling, cutting through the gloom and illuminating piles of debris on the marble floor.

Yet the scientists, called in by France's Ministry of Culture to inspect the damage and plan a rescue, mostly felt relief—and even hope. Rattan chairs sat in tidy rows, priceless paintings hung undamaged, and, above the altar, a great gold-plated cross loomed over the *Pietà*, a statue of the virgin Mary cradling the body of Jesus. "What matters isn't the roof and vault so much as the sanctuary they protect," says Aline Magnien, director of the Historical Monuments Research Laboratory (LRMH). "The heart of Notre Dame had been saved."

On 15 April 2019, an electrical short was the likely spark for a blaze that threatened to burn the 850-year-old cathedral to the ground. Following a protocol developed for just such a disaster, firefighters knew which works of art to rescue and in which order. They knew to keep the water pressure low and to avoid spraying stained glass windows so the cold water wouldn't shatter the hot glass.

But even though their efforts averted the worst, the emergency was far from over. More than 200 tons of toxic lead from the roof and spire was unaccounted for. And the damage threatened the delicate balance of forces between the vault and the cathedral's flying buttresses: The entire building teetered on possible collapse.

At LRMH, the laboratory tasked with conserving all the nation's monuments, Magnien and her 22 colleagues apply techniques from geology to metallurgy as they evaluate the condition of Notre Dame's stone, mortar, glass, paint, and metal. They aim to prevent further damage to the cathedral and to guide engineers in the national effort to restore it. President Emmanuel Macron has vowed to reopen Notre Dame by 2024, and he has appointed a military general to lead the operation, which involves many government agencies and has drawn philanthropic pledges of about €1 billion. But it is the LRMH scientists who lead the critical work of deciding how to salvage materials and stitch the cathedral back together. And even as they try to reclaim

Baptism of fire

The 850-year-old Notre Dame cathedral nearly burned down on 15 April 2019. Researchers are figuring out how to salvage materials and restore the Paris icon. The fire also offers a window to the cathedral's past by exposing materials that were largely off-limits to science.

Graphic by **Chris Bickel**

6:18 p.m.
Fire begins

7:50 p.m.
Spire collapses

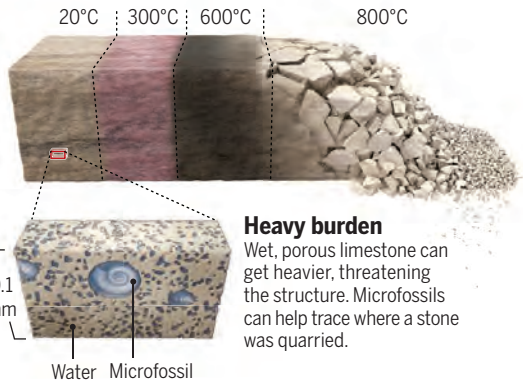
4:00 a.m.
Fire under control

Ceiling holes

Lead spire

Color code

Heat transforms iron compounds within limestone and weakens it. Associated color changes hint at whether fallen stones can be reused and where the structure is vulnerable.

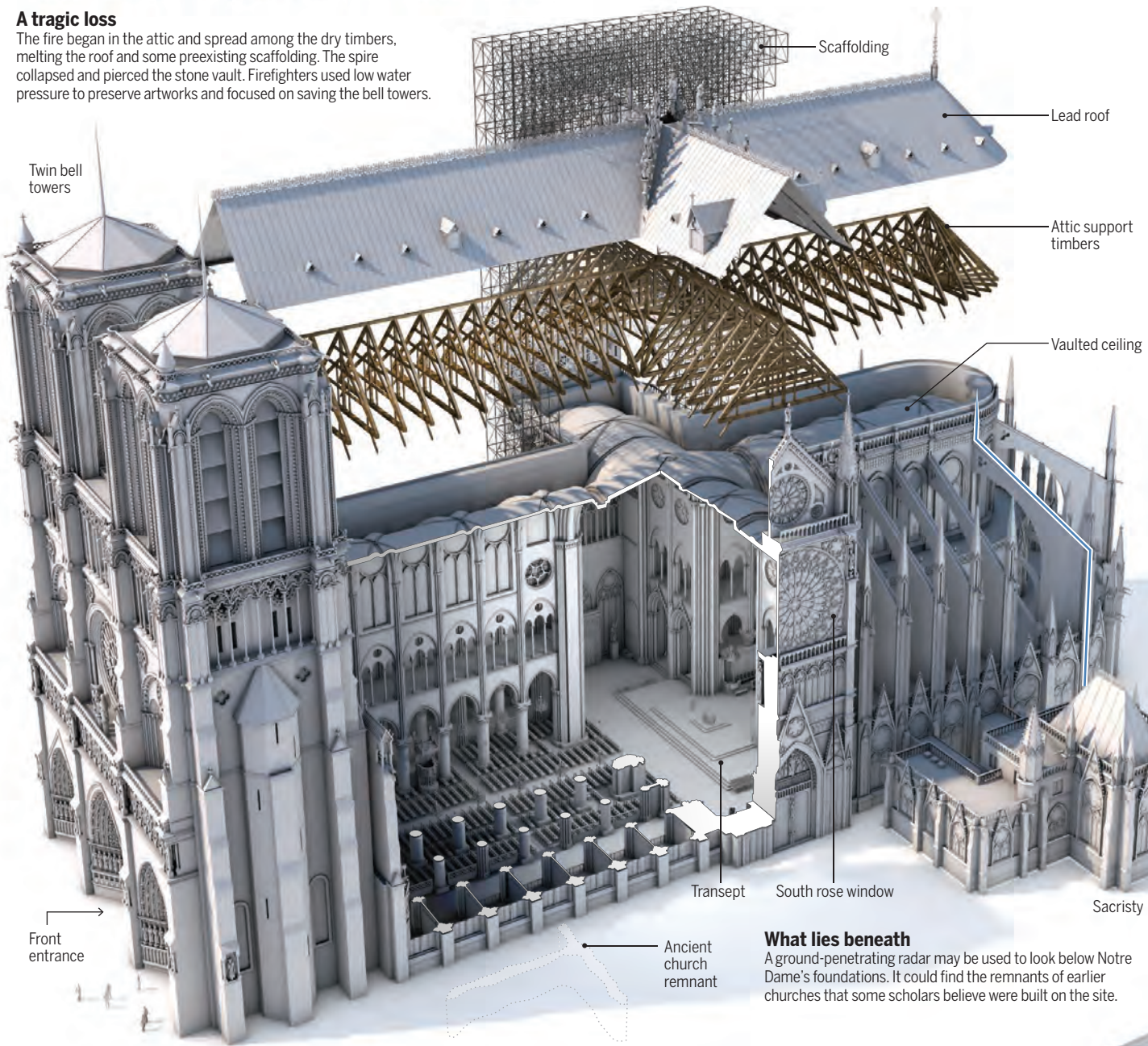


Heavy burden

Wet, porous limestone can get heavier, threatening the structure. Microfossils can help trace where a stone was quarried.

A tragic loss

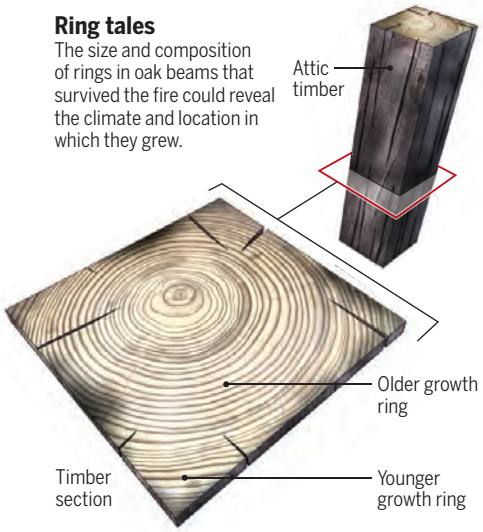
The fire began in the attic and spread among the dry timbers, melting the roof and some preexisting scaffolding. The spire collapsed and pierced the stone vault. Firefighters used low water pressure to preserve artworks and focused on saving the bell towers.



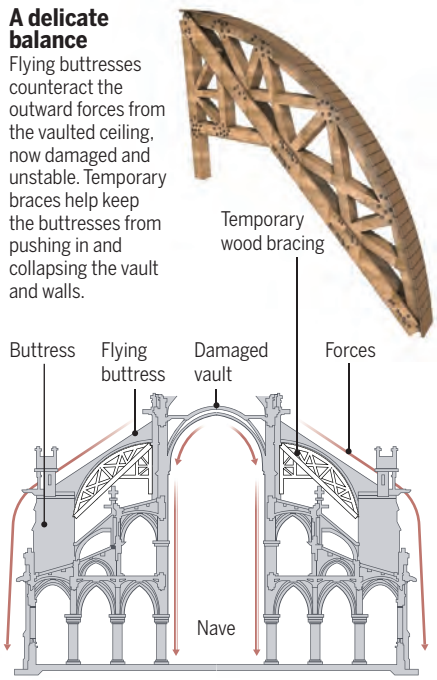
What lies beneath

A ground-penetrating radar may be used to look below Notre Dame's foundations. It could find the remnants of earlier churches that some scholars believe were built on the site.

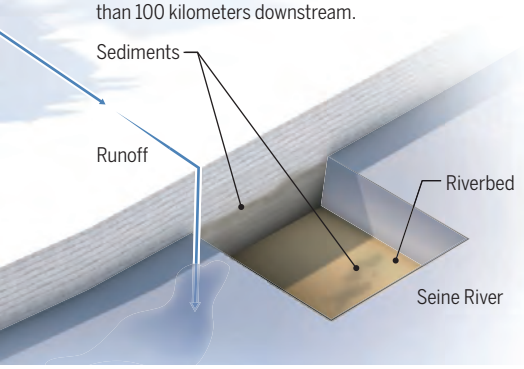
Ring tales
The size and composition of rings in oak beams that survived the fire could reveal the climate and location in which they grew.



A delicate balance
Flying buttresses counteract the outward forces from the vaulted ceiling, now damaged and unstable. Temporary braces help keep the buttresses from pushing in and collapsing the vault and walls.



A tainted legacy?
Lead aerosols from the roof and spire caked the cathedral in contamination. But runoff from the roof may have polluted the Seine River for centuries. Researchers will look for lead with Notre Dame's signature in river sediments more than 100 kilometers downstream.



what was lost, they and others are also taking advantage of a rare scientific opportunity. The cathedral, laid bare to inspection by the fire, is yielding clues to the mysteries of its medieval past. "We've got 40 years of research coming out of this event," says LRMH Assistant Director Thierry Zimmer.

THE LRMH RESEARCHERS work in the former stables of a 17th century chateau in Champs-sur-Marne, in the eastern suburbs of Paris, that once housed a horse research center. Here, they have analyzed samples from France's top monuments—the Eiffel Tower, the Arc de Triomphe—in the same rooms where some of the world's first artificial insemination experiments in horses occurred 120 years ago. The neighborhood is quiet, with a quaint brasserie and a shop offering €10 haircuts. But on a day in January, the lab is anything but sleepy. "It's an ambiance of speed!" says Zimmer, sporting a brown wool beret and a bushy mustache.

Véronique Vergès-Belmin, a geologist and head of LRMH's stone division, was sorting cathedral stones until 10 p.m. last night. This morning, she's the first to unlock the laboratory's ancient oak door.

She slips a hazmat suit over her dress clothes and slides on a respirator mask—necessary when dealing with samples contaminated with lead. In the lab's high-roofed storage hangar—once a garage for the chateau's carriages—she presents several dozen stones that fell from the cathedral's vaulted ceiling. Fallen stones hint at the condition of those still in place, which are largely inaccessible. The scientists can't risk adding their weight to the top of the vault, and debris falling near the holes in the ceiling makes it dangerous to inspect the structure from below. Many of the samples in the lab were retrieved by robots.

Heat can weaken limestone, and knowing the temperatures endured by these fallen stones can help engineers decide whether they can be reused. Vergès-Belmin has found that the stones' color can provide clues. At 300°C to 400°C, she says, iron crystals that help knit the limestone together begin to break down, turning the surface red. At 600°C, the color changes again as the crystals are transformed into a black iron oxide. By 800°C, the limestone loses all its iron oxides and becomes powdery lime. "It's an entire progressive process," she says, enunciating carefully through the muffle of the mask. "Any colored stones or parts should not be reused."

Color evaluation isn't an exact science, she says. Still, in lieu of mechanically testing each of the hundreds of thousands of stones that remain in the cathedral, color could be a useful guide to their strength.

Philippe Dillmann, an LRMH collaborator and a metal specialist with CNRS, the French national research agency, believes rust from the cathedral's iron structures can provide similar clues. At increasing temperatures, the microscopic structure of the rust changes. By investigating the cathedral's nuts and bolts—literally—as well as a "chaining" system of iron bars within and around its walls, Dillmann wants to create a heat map for the nearby stones. He says it's unknown whether these bars were used in construction and left in place or served as reinforcement. "We know they're in there, but they've never been studied," he says.

Water can also wreak havoc. Although the firefighters carefully avoided the stained glass windows, they had no choice but to drench the stone vault. The porous limestone gained up to one-third of its weight in water—and it's not set to lose it quickly. In the lab, LRMH researchers are monitoring a fallen stone, weighing it to track the drying process. When this article went to press, the stone was still losing weight.

Meanwhile, rainwater continues to fall on the roofless vault, and engineers can't install a temporary cover because of a mangled skeleton of scaffolding, set up in 2018 for long-term renovations. In January, workers began the 6-month process of removing the partly melted lattice. Because the cathedral walls support the scaffolding, it will have to be dismantled carefully, like a giant Jenga game, to prevent a collapse that could be "catastrophic," Magnien says.

Until the stones finish drying on their own, their changing weights will likely continue to have "nonnegligible" effects on the vault structure, according to Lise Leroux, a geologist in the LRMH stone division. Not only does the extra weight play with the precarious balance of forces, but when the water freezes in winter, individual stones expand or contract. "We'll start testing the mortar between the stones to see how well it's handling the strain," she says. "Now that I can get up there."

Weeks after the fire, engineers installed steel beams above the vault so technicians could rappel with ropes as they remove scaffolding and stabilize the structure. After earning a rappelling certification, Leroux last month inspected the top of the vault for the first time. She found that a plaster coating on top of the vault was still mostly intact, and had shielded many stones from fire and now rain. "It seems to have done its job," she says.

WHILE THE STONE SCIENTISTS are busy with mechanical forces, another team has concentrated on the whereabouts of the lead roof and spire. Along with grief, the fire stirred

Glass researcher Claudine Loisel found that baby wipes could sample for lead without damaging staining.

PHOTO: AP PHOTO/FRANCOIS MORI

another emotion among Parisians: fear that vaporized lead had drifted into nearby neighborhoods. In fact, Aurélia Azéma, a metallurgist who leads LRMH's metal division, and other scientists have concluded that the fire maxed out well below lead's vaporization temperature of 1700°C. Most of the lead simply melted at 300°C, pouring into the gutters and dripping into stalactites that can still be seen hanging from the vaults.

In places, however, temperatures did exceed 600°C, at which point lead oxidizes into microscopic nodules—an aerosol. "It's like hair spray," Azéma says. A yellow cloud that billowed from the cathedral during the fire showed that at least some of the lead did get hot enough to become airborne.

Sophie Ayrault, a geochemist at the French Alternative Energies and Atomic Energy Commission, wants to know where that cloud ended up. Azéma collected samples of lead dust from two surfaces that had been cleaned just before the fire—the organ bench and a drape covering a statue—and Ayrault has analyzed the signature of isotopes in the lead, a fingerprint that distinguishes Notre Dame lead from other lead sources.

She hopes to compare the cathedral fingerprint with that of dust samples from throughout the city. Some nearby schools were decontaminated after samples showed worryingly high lead levels. But it's not clear whether the lead came from the Notre Dame fire or from some other source, such as lead paint, car batteries, and leaded gasoline. "Unfortunately the testing agencies destroyed the wipes," she says, so she can't test the schools' original samples. Her lab will soon start to test new samples taken from Parisian park surfaces and compare them with the Notre Dame signature.

Ayrault also suspects corrosion from sunlight and acid rain might have been releasing lead from Notre Dame's roof for centuries. With the flying buttresses designed to carry runoff quickly into the Seine River, Ayrault says Notre Dame may have been an ongoing source of water pollution. She's starting to look for lead in sediments collected downstream in Normandy. Again, she will try to assess how much of the lead came from Notre Dame versus other sources.

MUCH OF THE LEAD mobilized by the fire remains in Notre Dame. In June 2019, when Azéma and her colleagues brought their first samples from the cathedral back to the lab, tightly sealed in plastic bags, yellow lead dust appeared to be everywhere. She unrolls small organ pipes from layers of bubble wrap, and points her gloved finger at their holes. "Even down in here," she says.

Because of lead's toxicity, especially in children, France's national health agency imposes a legal limit of 0.1 micrograms per square centimeter on the surfaces of any building, including historical monuments. "My first sample was 70 times that," says Emmanuel Maurin, a wood scientist and head of LRMH's wood division, who tested surfaces like the oak confessional and choir seats.

The scientists are largely unconcerned about their personal exposures, and blood tests have shown no significant rise in their lead levels. "It's not like we're licking the walls," Zimmer says. Nevertheless, the national work inspection agency has enforced stringent safety requirements. People entering the cathedral must strip naked and put on disposable paper underwear and safety suits before passing through to contaminated areas, where they put on €900 protective masks with breathing assistance. After a maximum of 150 minutes' exposure, they peel off the paper clothes and hit the showers, scrubbing their bodies from head to toe. "We're taking five showers a day," Zimmer says, adding that getting through the showers can be "like the Métro at rush hour."

The Ministry of Culture has charged LRMH with finding a way to cleanse the cathedral of lead without harming it. Claudine Loisel, head of the LRMH glass division, has been testing decontamination techniques for the cathedral's 113 stained glass windows, which all survived. Already blackened and sticky with soot, dust, and residue arising from millions of tourists, worshippers, and votive candles, the windows lack the yellow-powder look. But with her binocular microscope, Loisel easily detects lead oxide nodules on three panels she brought to the laboratory. "The goal right now isn't to restore, but to decontaminate," says Loisel, decked from head to toe in protective equipment.

The national health agency uses commercial wet wipes to sample surfaces and test for lead. But the wipes contain small amounts of acid that could damage the window staining, so Loisel convinced the agency to accept a compromise: "chemical-free" baby wipes from the Monoprix grocery store chain.

Although this works for small-scale testing, the scientists don't want to clean the entire cathedral with baby wipes. For most smooth surfaces—glass, metal, waxed wood, and even paint—they've found that a shop vac and cotton pads, moistened with distilled water, safely remove the lead. Raw wood surfaces require fine sanding first, Maurin says.

The porous stones call for a different approach. One possibility is plastering them with a latex "silly putty" that can be pulled off along with the lead dust, Vergès-Belmin says. A similar method uses a clay-based compress that dries and contracts, creating lead-filled "chips" that can be collected and disposed of. A third idea is to use laser cleaning. The scientists will begin to test various

methods in two of Notre Dame's chapels later this month. "We're most likely looking at a combination of techniques," she says.

AS THIS FIRST, "emergency" phase of scientific work advances, Notre Dame is slowly starting to open to "second phase" scientists—those interested in studying its history and architecture, now exposed by the fire and available to study without intruding crowds of tourists.

The Ministry of Culture and CNRS have created a dedicated science team of about 100 researchers from multiple institutions. By last month, about 10 CNRS researchers had gained some access to the cathedral for their work. The opportunities are generating palpable excitement among scientists and historians. "We're sorting all these thousands of fragments—some from our world, some from another and more ancient world—and it's like we're communicating with the Middle Ages," Dillmann says.

Yves Gallet, an art historian at Bordeaux Montaigne University, oversees a group that aims to study stones that are still in place, such as the encasements that cradle the four-story-diameter rose windows. Through detailed photographic analysis, the researchers want to understand how 13th century stonecutters designed and assembled the encasements, and the entire gothic masterpiece. Their analysis of mortar throughout the cathedral could confirm what historians believe about the order of the building's construction and repairs. "The mortar can tell us a lot about which stones were placed at the same time and what kinds of forces were anticipated in those areas," he says.

Before going that high up, Gallet hopes to use a ground-penetrating radar in a first study of what lies beneath the cathedral's ground floor. By interpreting the echoes of radar waves, Gallet and his colleagues could identify stonework that predates Notre Dame. He wonders whether he'll find remnants of the earlier churches that some scholars believe were built at the site. "We actually have no idea what's under there," he says. Ongoing cathedral activity ruled out such a study before.

Meanwhile, Leroux is eager to trace the origin of the vault stones. Many are said to have been quarried in Montparnasse, a nearby Paris district, but she thinks their origins are more diverse. "See this arrangement of plankton fossils mixed with clay and quartz?" she asks, a fallen stone in hand. "That's not from Montparnasse!" She turns to her archive drawers of 6000 samples and pulls out a stony sliver labeled "Pont d'Iena"—the Parisian bridge next to the Eiffel Tower. "This is a perfect match," she announces. "I found it." She says the bridge and vault stones both hail from a quarry in the French Vexin, a forest an hour northwest of Paris.

The charred remnants of attic timbers have stories of their own to tell, says Alexa Dufraisse, a CNRS researcher heading the wood group. Variations in thickness, density, and chemical composition of growth rings reveal climatic conditions year by year. "Wood registers absolutely everything while it's growing," she says. Notre Dame's oak beams grew in the 12th and 13th centuries, a warm period known as the Medieval Climate Optimum. By connecting the growth ring record with what's known about economic conditions at the time, researchers hope to see how climate variations affected medieval society, she says.

This gold angel once sat atop Notre Dame. It survived a fire and a fall. Researchers discovered the signature of its unknown sculptor.

The shape of the beams also intrigues the wood team. Long and narrow, they clearly grew in a dense, competitive environment, Dufraisse says. That supports the "silviculture" hypothesis, the idea that the trees were purposefully reserved or farmed for the cathedral. Their age at cutting—about 100 years old—would suggest people were planning Notre Dame several generations before construction began.

The location of that forest is another mystery Dufraisse's team is tackling, using the beams' chemical composition. The Paris area is likely, but boats might have shipped wood along the Seine from farther away. Soils contain levels of strontium and neodymium isotopes that vary from region to region, but stay constant over the centuries—especially at the depths tapped by the roots of oak trees. So her group is seeking to match the wood's isotopic makeup to that of soil in likely locations. "These questions will also be pertinent if we're looking at meeting the requirement of reconstruction that's identical to the original," she says.

As for Maurin, he's investigating the builders' marks on the roof support beams. Applied by men shaping the beams on the ground, they were meant as instructions for the assembly team working more than 30 meters above them. "It was kind of the IKEA of the Middle Ages," he says.

BEYOND THE PHYSICAL damage left by the fire is the emotional trauma suffered by thousands of Parisians and others, and CNRS researchers are also investigating this hidden aftermath. Sylvie Sagnes, a CNRS ethnologist with the Interdisciplinary Institute of Contemporary Anthropology in Paris, is part of a group that will interview tourists, locals, guides, journalists, donors, and church members to analyze the fire's emotional effect. She says people can display a powerful attachment to monuments, parks, and historical sites. When people mobilize to protect heritage, she says, it's a democratic expression—something French anthropologists studied 30 years ago during a public outcry against planned renovations of a basilica in Toulouse. In the case of Notre Dame, strong feelings are intensifying controversies around its restoration, such as whether to rebuild it exactly as it was. "Notre Dame isn't just any monument," she says. "After the fire, people remain emotionally implicated."

Valérie Tesnier, a café owner down the street from Notre Dame, says she's noticed a change in the behavior of tourists. They now solemnly watch the restoration effort before moving along—usually without stopping for food. "They don't want to stay and prolong their grief," says Tesnier, who has just sold her once-thriving business.

Across centuries marked by war and disease, Notre Dame has witnessed cycles of decline and renewal before. The LRMH scientists hope that when the vaults and buttresses are again dry and sound, the lead accounted for, and the great cathedral's history and resilience understood more deeply than before, grief and loss will once again turn to joy and gratitude.

"Certainly this is a difficult period emotionally, but there's an extraordinary unity of people coming together to not only save this monument, but to learn from it," Magnien says. "Notre Dame will be restored! Its artwork, stone, and stained glass will be cleaned; it will be more luminous and beautiful than before."

"Notre Dame will come out of this experience enriched," she says. "And so will we." ■

Christa Lesté-Lasserre is a journalist in Paris.

PHOTO: AP PHOTO/FRANCOIS MORI

INSIGHTS

BOOKS *et al.*

FILM

Science at Sundance 2020

A counterculture commune seeking a more sustainable lifestyle moves inside an airtight dome. Parents yearning to connect with their autistic children find hope in a teenager's profound testimony. The climate crisis hits home as a tight-knit California community attempts to move forward after a devastating wildfire. From a meandering love letter to an imperiled African ecosystem, to a warning about the motives that underlie social media, the science and technology stories told at this year's Sundance Film Festival were urgent, insightful, and well suited for the event's 2020 theme of "imagined futures." Read on to see what our reviewers thought of six of the festival's featured films. **—Valerie Thompson**

The Reason I Jump

Reviewed by **Robert S. Krauss**[1]

It is difficult to know the mind of another person; it is harder still when that person has autism. Yet, understandably, a fervent desire of parents of autistic people is to know the minds of their children. In 2007, Naoki Higashida, an autistic 13-year-old boy in Japan, published *The Reason I Jump*, a description of what was in his mind and why he behaved as he did. The book was translated into English by K. A. Yoshida and David Mitchell (author of *Cloud Atlas* and other celebrated novels), who together have an autistic son. Although some therapists expressed skepticism over Higashida's authorship of the book, it became a bestseller in the United States and the United Kingdom.

The Reason I Jump, a documentary from veteran director Jerry Rothwell, is based on the book, but Higashida does not appear in the film, nor is it his story. Rather, Higashida's words serve as a framing device for a portrait of the lives of five young people with autism and their families. The result is intimate and informative.

Voice-over readings from Higashida's book accompany scenes of a young Asian boy moving through a series of landscapes. These scenes, artistic and experimental, provide an impressionistic view of what a person with autism might experience. They are interspersed with straightforward documentary filmmaking.

We meet Amrit, from India, who is completely nonverbal. That Amrit has a complex interior life cannot be in doubt; she creates extraordinarily expressive drawings of people, and the film culminates in a show of her work. In the United States, Ben and Emma, friends since they were toddlers, spell out words using a board on which each letter of the alphabet is printed—as Higashida did when he wrote his book—thereby allowing them to communicate simple but profound sentiments. The family of Jestina, from Sierra Leone, faces not only the challenges of autism itself but also a stigma arising from superstitious beliefs that such children are possessed. Her parents' success in getting the government to establish a school for kids like

ILLUSTRATION: CAT FINNIE

her is inspiring. And finally, with Joss, a U.K. adolescent, we see the poignancy, worry, and commitment that accompany parenthood in the world of autism.

"I think we can change the conversation around autism by being part of the conversation," declares Ben. *The Reason I Jump*, which won an Audience Award at Sundance, succeeds in pushing the conversation forward.

The Reason I Jump, *Jerry Rothwell, director,* MetFilm Sales, 2020, 82 minutes.

The Social Dilemma

Reviewed by **DDW Cornelison**[2]

"There are only two industries that call their customers 'users': illegal drugs and software." This provocative observation, attributed to Yale computer scientist Edward Tufte, hits home in *The Social Dilemma*, as former executives from Facebook, Pinterest, Google, Twitter, and YouTube describe how they built online platforms to attract and reward our attention, with the goal of packaging and selling it to advertisers.

Just as illegal drugs hijack and overwhelm pleasure circuits in the brain, which evolved to help us survive, social media hijacks and overwhelms our prosurvival instinct to seek social connection. As the film's primary voice, Tristan Harris (formerly of Google), notes: "We evolved to care whether other people in our tribe think well of us or not, because it matters. But we were not evolved to be aware of what 10,000 people think of us; we were not evolved to have social approval dosed to us every 5 minutes." Harris and others are now raising concerns about how social media is changing how we perceive ourselves, other people, and even objective reality.

Through interviews interleaved with a narrative movie-within-a-movie, whose scenes will be familiar to anyone who has ever tried to impose a "no phones at the table" rule, the documentary describes how and why social media evolved to attract and keep our attention by gathering massive amounts of information about each of us and then using that information to target specific content to our feeds to keep us engaged. The interviewees link increases in teen self-harm and suicide, political and social polarization and isolation, outrage and self-centeredness, and even flat-Earth conspiracy theories to algorithms whose function is not necessarily to provide us with what we want or what is good for us, but to keep us scrolling and clicking. These themes are carried through the fictional narrative as well, with varying success.

One of the film's most striking interviews is with Tim Kendall, who, as director of monetization in the early days of Facebook, conceived of selling advertising to make it profitable and then helped tune the news feed to maintain engagement through intermittent positive reinforcement ("like slot machines in Vegas"). He describes how, despite knowing that he was being manipulated, he would find himself hiding in his pantry, ignoring his family, just to spend time on social media. Kendall—like the film's other subjects—has since had a change of heart. He is now CEO of Moment, a company whose app helps people spend less time on their phones.

The Social Dilemma, *Jeff Orlowski, director,* Exposure Labs, 2020, 93 minutes.

Okavango

Reviewed by **Gabrielle Kardon**[3]

Lightning crackles across the sky, lions roar as they tussle over a freshly killed waterbuck, raindrops smack the parched earth, and elephants trumpet. These are the sights and sounds of the Okavango River, the sub-

Two lionesses and their cubs cross a spillway in the Okavango River.

ject of a new documentary film by National Geographic explorers-in-residence Dereck and Beverly Joubert.

The Okavango is a distinctive river. More than 1000 km in length, it begins in the highlands of Angola, passes through Namibia, and empties into the Kalahari. It is entirely continental and never touches the ocean. Instead, when the water reaches the Kalahari, it evaporates in the heat of the desert. Using close-ups of droplets hanging on grasses growing alongside the river and expansive aerial footage, the film documents the important role of water to the Okavangan ecosystem.

The animals living along the river are the stars of the film. The story of an injured lion named Fekeetsa, for example, is interwoven with the tale of the leopardess Moporoto, who is protecting her two cubs. Meanwhile, close-up and slow-motion footage details the ever-present conflict between predators and prey. Underwater footage shows crocodiles and catfish stirring up river waters as they voraciously pursue their next meal.

The most thought-provoking scenes are those that explore how animal–river interactions shape the ecosystem. Elephants, in their quest for grasses, trudge paths through the wetlands, opening new waterways and changing the course of the Okavango. At the river's terminus, termites build towering mounds of clay that form the nucleus of islands on which trees sprout and stabilize the land, ultimately leading to the mosaic of islands that form the complex delta.

Like the river it documents, the film meanders through the Okavangan landscape. Luminous aerial images bathed in orange light, dramatic footage of lion hunts, and unusual underwater perspectives of the river draw the viewer in. The sounds—the deep rumble of lions and the snorts of baby warthogs—are some of the most surprising aspects of the film. *Okavango* is the Jouberts' love letter to the river, and Dereck's poetic narration conveys this love.

The Botswanan government recently lifted a ban on trophy hunting, endangering the inhabitants of the Okavango, particularly the region's elephant population (which is currently the world's largest). This film allows viewers to voyage to this fascinating biosphere and encourages them to advocate for its future.

Okavango: River of Dreams (Director's Cut), *Dereck Joubert and Beverly Joubert, directors,* Terra Mater Factual Studios and Wildlife Films, 2019, 94 minutes.

Spaceship Earth

Reviewed by **Michael D. Shapiro**[4]

In 1991, eight adventurers donned designer astronaut jumpsuits and began a 2-year mission in a 3-acre airtight terrarium in the Arizona desert called Biosphere 2 (Biosphere 1 being the environment the rest of us earthlings inhabit). A media circus, complete with cringeworthy celebrity cameos, surrounded the launch of the massive project that would be a model for similar ventures on other worlds. *Spaceship Earth* chronicles the fascinating history and prehistory of Biosphere 2, a $200 million earthbound space expedition that blurred the line between science and entertainment.

Biosphere 2 was the culmination of a series of ambitious projects led by the charismatic and brilliant John Allen. Allen assembled a group of followers that began as a theater troupe at the height of 1960s commune culture in San Francisco, but these self-described "Synergists" soon began experimenting with bigger projects that they were profoundly unqualified to attempt. They built a massive sailboat for the purpose of exploring Earth's biomes and left sustainable businesses in their wake, thanks to a funding partnership with Ed Bass, a rebellious Texas oil billionaire with an environmentalist streak.

As idealists but also capitalists, the Synergists were members of a remarkably functional commune. Having mastered the seas, their next move was to save the planet from ecological destruction.

Biosphere 2 was a prototype for a planetary colony. More importantly, the Synergists hoped it would teach them how to live sustainably on Earth. Everything, from water and air to nutrients, had to be recycled during the 2-year mission, so expedition members were metabolically linked to the organisms under the Buckminster Fuller–inspired enclosure. The massive scale of the project drew intense media scrutiny, for which Allen and his followers were unprepared, and raised expectations for a level of scientific rigor that they had never quite promised.

The maiden voyage of Biosphere 2 was far from a controlled experiment, and enthusiasm from scientists outside the dome dropped as precipitously as the oxygen levels on the inside. Still, some of the original Synergists look back at Biosphere 2 as a triumphant project that taught them lessons about sustainability that were not otherwise knowable. Now entering their sixth decade of collaboration, Allen's group continues to operate a sustainable ranch, and some of the businesses they established on their global voyage are still afloat.

Director Matt Wolf, interviewing key

players in the Synergist movement, recounts their 50-year wild ride, and viewers are treated to a gold mine of riveting archival footage. Wolf's subjects are wonderful storytellers, and he infuses the film with compassion and admiration for the Synergists' idealism and accomplishments.

Ultimately, the inspirational lessons of *Spaceship Earth* are that we have to push ourselves to chase important visions when moments of opportunity arise and that small collectives like the Synergists can be the engines of creative success. For Allen's group, a heavy dose of charisma and performance flair also went a long way toward seizing the moment and drawing others into their lofty, futurist goals.

Spaceship Earth, *Matt Wolf, director,* RadicalMedia and Stacey Reiss Productions, 2019, 116 minutes.

Rebuilding Paradise

Reviewed by **Michael D. Shapiro**[4]

On 8 November 2018, residents of the town of Paradise, California, evacuated through a forest of flames. Although it was past 9 a.m., the sky was black from the fire that had traveled 8 miles in just a few hours and now completely surrounded the town. The aftermath of the Camp Fire, as documented in Ron Howard's *Rebuilding Paradise*, is a portrait of staggering destruction. Empty concrete pads mark the former sites of houses, among the 18,000 structures obliterated. Sparse old-growth trees stand above the ruins of the 100-year-old town, their green crowns the only reminder that the scenes were filmed in color rather than tones of ash.

Paradise is the kind of close-knit town where everyone turns out for a parade or a funeral. Howard's privileged access and the film's immersive perspective make every new trauma feel more harrowing and every victory more ascendant as Paradise inches back toward normalcy. The film captures residents' deeply personal stories as they scatter to surrounding communities in the fire's immediate aftermath and wrestle with the decision to return or move elsewhere. Is it worth it, they ask, to rebuild in a town with toxic benzene in the water supply that will take years to purge, a century-old utility infrastructure in disrepair, and onerous government directives that nag the physically broken and financially broke community?

Howard treats the critical themes of land management and climate change with a gentler touch that reaches a crescendo late in the film. Ghosts of century-old mismanagement still haunt the forests around Paradise, and when coupled with long-term drought, they create perfect conditions for firestorms.

Rebuilding Paradise and the disaster it chronicles will deservedly get a lot of attention; Howard is a well-known filmmaker, who crafts an engrossing, personal, and emotionally raw story. Yet as the frequency of climate-fueled disasters increases worldwide, most of these stories will drift into obscurity, becoming the problems of voiceless people in distant places. *Rebuilding Paradise* challenges us to see ourselves in climate refugees and to reject the illusion that catastrophic events only happen somewhere else.

Rebuilding Paradise, *Ron Howard, director,* NatGeo, 2020, 95 minutes.

Mauny Roethler clears debris in the aftermath of the 2018 California Camp Fire.

PHOTO: NATIONAL GEOGRAPHIC/PETE MULLER

The Cost of Silence

Reviewed by **Paul L. Koch**[5]

The 2010 Deepwater Horizon explosion unleashed a catastrophic amount of oil, gas, and other toxic compounds into the Gulf of Mexico. As the massive spill overwhelmed responders, federal agencies approved the use of chemical dispersants by aerial spraying and injection into the oil plume at its source. Dispersants break up oil, which is highly toxic, into tiny droplets that are, ideally, diluted and decomposed far from beaches and marshes, and far from surface-dwelling mammals, birds, and people. Reducing the exposure of offshore responders and cleanup crews to volatile organic compounds and toxic oil was also a key consideration in 2010. However, the decision to use dispersants was controversial, because these compounds are also toxic and had never been subject to careful epidemiological study. About 3 million liters of dispersant were released—their largest application in U.S. history.

The Cost of Silence, a new documentary by director Mark Manning, offers a more nefarious reading of this decision: that it was part of a conspiracy between the U.S. government and the oil company BP to reduce the firm's liability and convince tourists and residents that the Gulf was open for business, when, in fact, a dangerous chemical stew was brewing offshore. In the film, Riki Ott, a toxicologist and environmental activist, argues that the dispersant made the oil more toxic and increased the ease with which it was taken up by people and animals. She also maintains that the tiny droplets formed clouds that wafted oil-dispersant mixtures onshore.

Whistleblowers claim that the dispersant was released too close to shore and that cleanup workers used inadequate protective gear. Over 9 years of filming, Manning interviewed offshore responders, cleanup crew members, and Gulf Coast residents who are sick and scared. Some are despondent and others defiant, but all feel abandoned and betrayed by the government.

The film has a polemical tone. Yet whether or not a viewer is convinced that the spill's impacts were worsened by dispersants, responders and cleanup crews were at the greatest risk of exposure to toxic oil and dispersant. The health of these individuals needs more study, and we need new methods for assessing exposure to spills and dispersants, as recommended by a recent report (*1*).

Physician Michael Harbut, a consultant on the BP medical settlement, argues that the potential health impacts of the Deepwater Horizon spill on coastal communities will become obvious through epidemiological studies during the next two decades. Manning's film seeks to accelerate that process and change global policy on the use of dispersants.

The Cost of Silence, *Mark Manning, director,* Conception Media, 2020, 84 minutes.

REFERENCES AND NOTES

1. National Academies of Sciences, Engineering, and Medicine, *The Use of Dispersants in Marine Oil Spill Response* (The National Academies Press, 2019).

10.1126/science.abb3608

[1]Department of Cell, Developmental, and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. Email: robert.krauss@mssm.edu [2]Division of Biological Sciences and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA. Email: cornelisond@missouri.edu [3]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA. Email: gkardon@genetics.utah.edu [4]School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA. Email: mike.shapiro@utah.edu [5]Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. Email: plkoch@ucsc.edu

PERSPECTIVES

CLIMATE CHANGE

Threatening the vigor of the Colorado River

Loss of sunlight-reflecting snow spurs evaporation and ebbs river flow

By **Mike Hobbins** and **Joseph Barsugli**

As the beating heart of the American Southwest, the Colorado River (CR) serves ~40 million people from Denver to Los Angeles and supports 16 million jobs in a $1.4 trillion regional economy (*1*). Most of the river's streamflow originates as snowpack in the mountains of the Upper Colorado River Basin (UCRB). Flow is threatened by increasing demands from a growing population, extended drought, and climate change. A 2018 workshop (*2*) convened experts in climate and hydrologic modeling, observational analysis, and paleoclimatology to address the causes of the observed declines in streamflow. All agreed that rising local temperatures were associated with drying of the basin, but varied methods yielded large differences in the magnitude of this effect. Workshop participants urged the scientific community to identify sources of these differences by deciphering the contributions of various processes to hydroclimatic changes. On page 1252 of this issue, Milly and Dunne address this challenge (*3*).

The CR is a well-studied exemplar of snowmelt-dominated rivers rising in water-rich mountains in otherwise water-limited regions. Water sources for ~10% of the global population depend on high mountain regions that are especially vulnerable to climate change (*4*). Projected reductions in CR streamflow related to regional warming range from <10 to 45% by 2050 (*5*). The higher end of projected losses would be calamitous. The design of new policies and adaptive plans requires process understanding, rigorous synthesis of observations, and modeling.

The surface water balance and surface energy balance must be resolved for accurate streamflow estimation and projection. At climate scales, these balances are affected by a tangled web of competing physical processes, but they intersect in evapotranspiration (ET). Across the UCRB streamflow-producing mountains, an uncertainty of just 5 W/m^2 of equivalent ET represents ~4 × 10^9 m^3 per year, or close to 30% of mean annual basin streamflow. In energy-limited (water-rich) areas, evaporative demand (E_0) drives ET. Thus, proper estimation of both ET and E_0 are crucial. Data for observation-based assessments of long-term hydrology are limited largely to precipitation and temperature, with temperature serving as a proxy for E_0 and the complexities of the energy balance. This poor process representation has hobbled observed analyses of trends in streamflow and land-surface aridity in two ways (*6*, *7*). A full representation of E_0 also requires data on wind speed, humidity, and solar radiation—variations in all of which drive long-term E_0 trends across the UCRB (*8*) and globally (*9*). Also, E_0 and ET should covary in a complementary fashion in water-limited areas and in parallel in energy-limited areas (*8*)—both of which are represented in the basin's complex hydroclimate. Although these issues may seem obvious, temperature-based E_0 parameterizations remain obscured in many analyses.

The relationship between long-term streamflow and regional warming drives much of the uncertainty in flow projections. Milly and Dunne call this sensitivity β (percent streamflow change per degree Celsius of warming). The analogous sensitivity to precipitation is less in dispute. The sensitivity approach provides a simple metric for comparing different analyses and models (*5*, *10*); further, temperature and precipitation are the only meteorological elements for which century-long, station-based observations are widely available in the basin. Yet, the overall sensitivity metrics can obscure important details. Large ensembles of climate-model analyses indicate that meteorological variability alone can lead to a large range in regression-based estimates (*10*, *11*), even if the underlying basin hydrologic processes remain the same. To counter this, basin sensitivity can be estimated from controlled experiments with uniform warming applied to a hydrologic model.

Milly and Dunne focused on reduced snowpack and the resulting changes in the proportion of surface-reflected sunlight (albedo). They adopted an innovative approach for measuring the energy budget with remote sensing that short-circuits the explicit calculation of various feedbacks from albedo changes. They then used their empirically calibrated model of radiative balance to derive net surface radiation from snowpack for use in the Priestley-Taylor formulation of E_0. This permitted an estimate of radiative effects of snowpack variability using temperature and precipitation records from 1920 to the present and incorporated the estimates into a surface water balance model. They calculated an overall temperature sensitivity (β) of –9.3% °C^{-1} for the basin streamflow. By examining the seasonality of precipitation in

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA, and Physical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO 80305, USA.
Email: mike.hobbins@noaa.gov

PHOTO: CHRIS WINSOR/GETTY IMAGES

Snowpack in the Colorado River basin is a solar-radiation shield that suppresses water loss through evaporation. This protection decreases as local temperatures rise.

warm versus cold years, they square observational estimates of sensitivity with substantially lower estimates from hydrologic model sensitivity studies.

The E_0 used by Milly and Dunne is more informative than approaches based only on temperature; their earlier work argues that radiation-based measures capture ET changes in "water-rich" areas better than do other methods (*12*). However, it is reasonable to ask whether the parameterization used for E_0 in the new study is the best measure for the mixed energy- and water-limited hydroclimates across the UCRB. Also, does the choice of solar radiation (and to some extent, temperature) as the driver of E_0 force the finding that albedo plays the dominant role in determining streamflow from their hydrology model?

The 9.3% loss of streamflow per degree Celsius of warming cannot reconcile all available data and model dynamics; global and regional climate models and hydrologic models also include the albedo effect yet show differing sensitivities. Other missing pieces of the puzzle include the effects of dust on snow, the direct effects of increasing CO_2 concentrations on trends in the long-wave radiation balance and on vegetation's water-use efficiency, and land-cover changes from wildfires and insect outbreaks.

The year 2020 is a momentous one for CR water policy. The interim interstate agreement on sharing water-shortage impacts will be renegotiated this year. The new, more stringent Lower Basin Drought Contingency Plan will mandate that water deliveries to states in the lower basin be reduced—a first, and unthinkable a generation ago. These adaptation strategies are difficult in a single snowmelt–driven basin in a wealthy country. How to approach such problems in similar basins worldwide is an open question. ■

GRAPHIC: A. KITTERMAN/*SCIENCE*

REFERENCES AND NOTES

1. T. James, A. Evans, E. Madly, C. Kelly, *The Economic Importance of the Colorado River to the Basin Region* (Seidman Research Institute, 2014).
2. J. J. Barsugli *et al.*, *Eos* 10.1029/2019EO117173 (2019).
3. P. C. D. Milly, K. A. Dunne, *Science* **367**, 1252 (2020).
4. H.-O. Pörtner *et al.*, *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (Intergovernmental Panel on Climate Change, 2019).
5. J. A. Vano *et al.*, *Bull. Am. Meteorol. Soc.* **95**, 59 (2014).
6. M. T. Hobbins *et al.*, *Geophys. Res. Lett.* **35**, L12403 (2008).
7. J. Sheffield *et al.*, *Nature* **491**, 435 (2012).
8. M. T. Hobbins *et al.*, *Geophys. Res. Lett.* **31**, L13503 (2004).
9. M. L. Roderick *et al.*, *Geogr. Compass* **3**, 761 (2009).
10. F. Lehner *et al.*, *Nat. Clim. Chang.* **9**, 926 (2019).
11. M. P. Hoerling *et al.*, *J. Clim.* **32**, 8181 (2019).
12. P. C. D. Milly, K. A. Dunne, *Nat. Clim. Chang.* **6**, 946 (2016).

10.1126/science.abb3624

MOLECULAR BIOLOGY

Liquid droplets in the skin

Creating enough glue to protect the body may require phase separation in skin cells

By **Arpan Rai** *and* **Lucas Pelkmans**

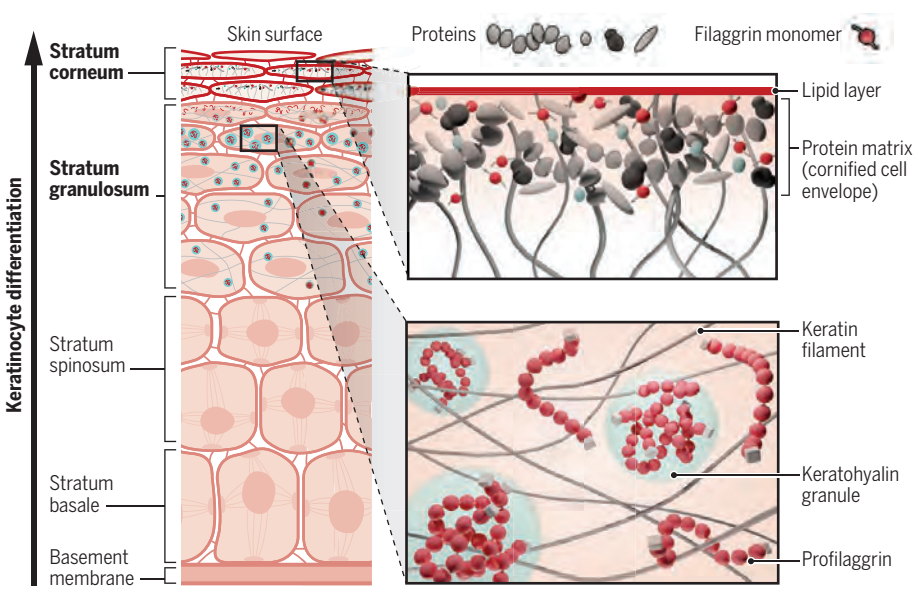
Liquid-liquid phase separation (LLPS), the unmixing of inhomogeneous fluids into two or more phases, is emerging as a paradigm for the formation of a myriad of membraneless compartments inside cells (*1*, *2*). This type of spatial organization, in contrast to membrane-bound compartmentalization, has long lacked unifying principles. However, the physiological relevance of compartmentalization through LLPS inside cells is still poorly understood and often speculative. Additionally, regulatory mechanisms through which cells control and exploit LLPS are still emerging. On page 1210 of this issue, Garcia Quiroz *et al.* (*3*) show that keratohyalin granules (KGs) that are formed during epidermal differentiation in the skin are pH-sensitive liquid-like protein condensates. Formation of KGs may be physiologically important because mutations that cause defects in this process are associated with the common skin barrier defect ichthyosis vulgaris.

The most external part of skin, the epidermis, is composed of keratinocytes, whose morphological appearance changes as they differentiate, resulting in various layers. In the stratum granulosum (granular layer), which sits just below the stratum corneum (cornified layer) in which the keratinocytes expel their nuclei and form a continuous water-impermeable protective zone, keratinocytes transiently contain KGs. These appear as electron-dense, protein-rich structures that lack a delimiting membrane (*4*). Are KGs physiologically relevant? A core component of KGs is the protein profilaggrin, which is cleaved into individual repeats (filaggrin monomers) when the stratum corneum forms. In this layer, filaggrin monomers function as part of the "glue" that forms the impermeable barrier of the skin. Mutations that result in a smaller number of filaggrin repeats lead to disappearance of KGs. The in-

Department of Molecular Life Sciences, Ernst Hadorn Chair, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Email: lucas.pelkmans@mls.uzh.ch

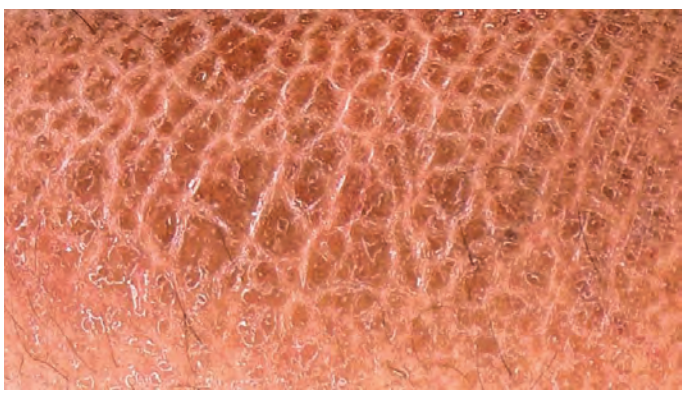
Liquid-liquid phase separation in the skin

Keratohyalin granules (KGs), formed by liquid-liquid phase separation of profilaggrin, interact with keratin filaments to organize the cytoplasm of keratinocytes during differentiation. During transition of keratinocytes from the stratum granulosum (granular layer) to the stratum corneum (cornified layer), KGs dissolve and profilaggrin is processed into monomers, which together with other proteins, contribute to the formation of a solid intracellular protein matrix.



ability to form KGs limits the expression of filaggrin monomers (*5*), which results in improper barrier formation. This is where LLPS may play a crucial role.

Garcia Quiroz *et al.* show that profilaggrin can undergo LLPS inside cultured keratinocytes and that this phenomenon depends on the number of filaggrin repeats. From synthetic systems, it is known that the concentration at which multivalent repeats undergo LLPS, as well as the composition of condensates formed, is sensitive to the valency and stoichiometry of the repeats (*6*). Because skin barrier disorder–associated profilaggrin mutations result in variable numbers of filaggrin repeats, these may reflect a disease-causing manifestation of altered multivalency of a phase-separating protein, affecting the composition and dynamics of the cellular condensates they form. Using live imaging of engineered phase separation sensors transduced in embryonic mouse skin epithelium, Garcia Quiroz *et al.* show that endogenous KGs behave like condensates, increasing in number and stiffness as keratinocytes differentiate and progress through the stratum granulosum. This may be mediated by the KGs undergoing a liquid–to–gel-like transition during differentiation, as well as by the dense network of intermediate filaments in which KGs grow and with which they interact (see the figure). Macromolecular crowding can affect phase separation properties of proteins both in vitro and inside cultured cells (*7*). Moreover, as shown in a nonbiological system, the density and elasticity of polymer networks can affect the protein solubility threshold at which LLPS occurs, as well as the size and nucleation of condensates (*8*).



The skin condition ichthyosis vulgaris is likely caused by defective liquid-liquid phase separation of profilaggrin and hence improper barrier formation in the skin.

Although aberrant liquid–to–gel-like transitions have been speculated to underly the appearance of protein aggregates in neurodegenerative diseases (*2*), the study of Garcia Quiroz *et al.* indicates that in certain contexts, such transitions are physiologically important—namely, to form a protective layer in the skin. Also during the formation of gel-like aggregates in neurodegenerative diseases, the cytoskeleton (*9*) might play a role in modulating their stiffness. In the future, it will be important to understand how condensates and cytoskeletal networks affect each other to structure the interior of the cell and how this synergy is perturbed in disease.

Garcia Quiroz *et al.* also reveal that both endogenous KGs in the skin and profilaggrin condensates in cultured cells sense pH changes, responding to the drop in pH that occurs when cells approach the stratum corneum. The lower pH triggers KG dissolution, leading to increased amounts of profilaggrin in the cytoplasm. Such environmental sensing has been shown for other phase-separating proteins, such as Sup35 and polyadenylate-binding protein (Pab1) in yeast (*10*). KG dissolution may also be aided by changes in intracellular Ca^{2+} concentrations that occur during epidermal differentiation (*11*), because keratinocytes express many Ca^{2+} binding proteins, including profilaggrin. Furthermore, although not addressed by the authors, profilaggrin becomes extensively phosphorylated in the granular layer, and subsequently dephosphorylated before being processed into monomers (*5*). Cycles of phosphorylation and dephosphorylation have been shown to regulate multiple condensates (*12*)—for instance, during stress recovery (*13*) and progression through mitosis (*14*). This provides another attractive, actively controlled mechanism by which keratinocytes could modify the critical concentration at which profilaggrin undergoes LLPS.

The processing of profilaggrin into individual filaggrin monomers is mediated by proteases, which have been shown in synthetic in vitro systems to rapidly change the valency of repeats of engineered proteins, thereby triggering the dissolution of condensates (*15*). It thus seems plausible that a multimodal mechanism is in place to ensure that condensation and subsequent dissolution of KGs in keratinocytes is robust and precisely timed during epidermal differentiation.

But why undergo LLPS to produce KGs, when they disappear and profilaggrin is cleaved into monomers in the layer above? Many functions have been proposed for condensates, including storage, modulation of signaling, environmental stress sensing, force generation, and noise buffering (*1*, *2*). Garcia Quiroz *et al.* posit that KGs, along with the keratin network, function to physically deform the nucleus, prior to enucleation. It is an attractive hypothesis, considering that the stiffness of KGs and the density of the keratin network increase during epidermal differentiation. However, it is also possible that KGs act as a storage depot for profilaggrin, protecting it from proteolytic processing. Exclusion of proteases from KGs would prevent premature processing of profilaggrin. Moreover, there are high amounts of the amino acid histidine in profilaggrin, which is metabolically converted to organic acids upon proteolytic cleavage in the stratum corneum, contributing to skin acidification. Therefore, premature processing of profilaggrin could strongly affect the intracellular milieu when it is in high abundance. Concentrating profilaggrin through LLPS and formation of KGs may thus protect the cell from possible deleterious effects of premature acidification.

Future studies are required to address whether sequestering profilaggrin or other epidermal regulators in condensates, which might otherwise have toxic effects at high cellular concentrations, allows cells to build up enough material and temporally regulate their release. The released proteins may then transition into a continuous, irreversible solid protein matrix of the stratum corneum by enzymes that introduce covalent cross-links. Multicellular systems that recapitulate the three-dimensional structure of tissues will become increasingly important to explore how LLPS is exploited and regulated in cells within the context of a tissue. By generating new properties in form and function at the supramolecular scale, LLPS may provide key insights into the mechanisms by which biological scales are connected, and how this goes wrong in disease. ■

REFERENCES AND NOTES

1. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285 (2017).
2. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).
3. F. G. Quiroz *et al.*, *Science* **367**, eaax9554 (2020).
4. K. A. Holbrook, *J. Invest. Dermatol.* **92**, S84 (1989).
5. A. Sandilands, C. Sutherland, A. D. Irvine, W. H. I. McLean, *J. Cell Sci.* **122**, 1285 (2009).
6. S. F. Banani *et al.*, *Cell* **166**, 651 (2016).
7. M. Delarue *et al.*, *Cell* **174**, 338 (2018).
8. R. W. Style *et al.*, *Phys. Rev. X* **8**, 011028 (2018).
9. J. Dubey, N. Ratnakaran, S. P. Koushika, *Front. Cell. Neurosci.* **9**, 343 (2015).
10. H. Yoo, C. Triandafillou, D. A. Drummond, *J. Biol. Chem.* **294**, 7151 (2019).
11. P. Elias, S. Ahn, B. Brown, D. Crumrine, K. R. Feingold, *J. Invest. Dermatol.* **119**, 1269 (2002).
12. J. Söding, D. Zwicker, S. Sohrabi-Jahromi, M. Boehning, J. Kirschbaum, *Trends Cell Biol.* **30**, 4 (2020).
13. F. Wippich *et al.*, *Cell* **152**, 791 (2013).
14. A. K. Rai, J.-X. Chen, M. Selbach, L. Pelkmans, *Nature* **559**, 211 (2018).
15. B. S. Schuster *et al.*, *Nat. Commun.* **9**, 2985 (2018).

10.1126/science.abb0060

NEUROPHYSIOLOGY

Opening the floodgates to the brain

Cerebrospinal fluid influx dictates early edema in response to ischemic large-vessel stroke

By **Jonathan Moss** *and* **Anna Williams**

Cerebral edema, brain swelling due to fluid influx, exacerbates the effects of ischemic large-vessel stroke, whereby a clot blocks one of the large blood vessels to the brain. Because the skull is a closed box, brain swelling cannot be accommodated, damaging other areas of the brain and leading to poor long-term prognoses (*1*). The central dogma to this fluid influx is that its source is the blood. On page 1211 of this issue, Mestre *et al.* (*2*) present compelling evidence to suggest that the initial stages of cerebral edema after ischemic stroke are driven by influx from an alternative source—the cerebrospinal fluid (CSF). This implies that targeting CSF-driven edema may offer opportunities to treat these harmful secondary effects of ischemic stroke.

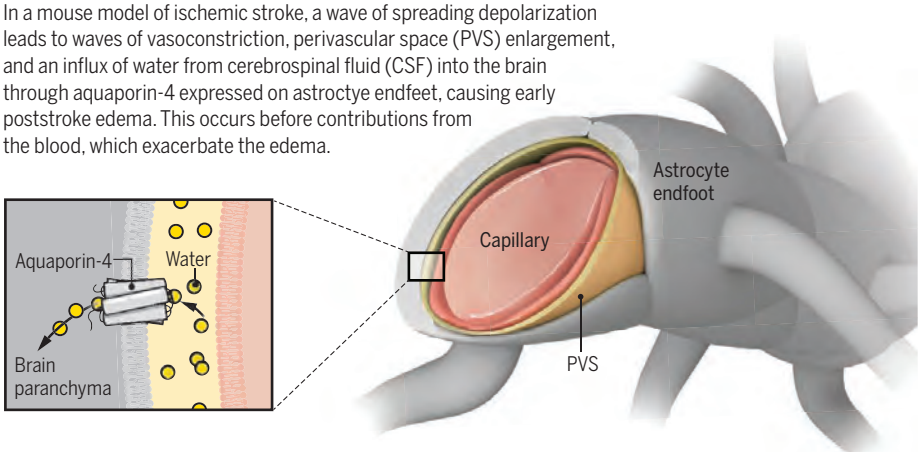
A system was previously described whereby CSF enters the brain parenchyma through the spaces around arteries and leaves through the spaces around veins (perivascular spaces), removing brain solutes (*3*). This lymphatic-like mechanism is largely dependent on water flow through aquaporin-4 (AQP4) channels into, and out of, the endfeet processes of perivascular astrocytic glial cells and was thus called the glymphatic system. The clearing of brain solutes by this system can be highly beneficial. Removal of amyloid β peptide and tau protein can reduce pathological aggregations that are implicated in neurodegeneration, Alzheimer's disease, and dementia. Indeed, when the glymphatic system is artificially perturbed by the deletion of the *Aqp4* gene in mice, clearance of these factors is suppressed and pathology worsens (*3*, *4*). But what if this clearance pathway is overwhelmed? Mestre *et al.* reveal that, in the immediate response to ischemic stroke, perivascular spaces are flooded with CSF, the physiological ebb and flow of the glymphatic system are perturbed, and fluid rushes through AQP4 channels of astrocyte endfeet into the astrocytes and the brain to cause edema.

The development of cerebral edema is typically subdivided into three distinct phases. First, in cytotoxic edema, brain cells take on fluid and swell as energy-dependent ion transport fails. Next, in ionic edema, sodium ions from the blood cross an intact blood-brain barrier and create a permissive osmotic gradient down which fluid flows into the brain. Finally, in vasogenic edema, the blood-brain barrier begins to break, allowing an influx of larger blood constituents. Using a well-characterized mouse model of unilateral ischemic stroke (affecting part of one brain hemisphere), Mestre *et al.* show that influx waves of water, at 11 s and at 5.5 min after stroke (within the window of cytotoxic edema), increased water content in the stroke-affected cortex, and this water remained for an hour. With a fluorescent CSF tracer, the authors demonstrate that CSF enters the poststroke hemisphere three times more readily than the unaffected hemisphere. Tracers in the blood or CSF (see the photo) confirmed that the CSF is the source of the increased water content, which is corroborated by edema formation along CSF flow pathways in humans.

What drives this influx of CSF into the brain? Mestre *et al.* found that a poststroke reduction in blood flow led to a hydrostatic pressure gradient to push CSF into the brain at 11 s after stroke. The larger influx of CSF occurred contemporaneously with cytotoxic edema, as seen by a decreased mobility of water molecules due to cell swelling. This decrease in mobility may be driven by spreading depolarizations—sudden waves of extensive gray matter depolarization resulting from the shutdown of transmembrane ion gradients in neurons (*5*). The authors found that the wave of depolarization spreads fast, with the wave of CSF influx following ~30 s behind, suggesting that the depolarization triggers CSF influx.

Fluid influx in poststroke edema

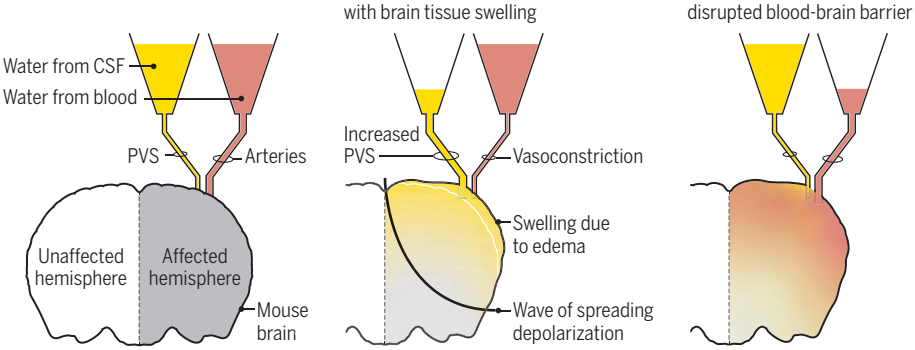
In a mouse model of ischemic stroke, a wave of spreading depolarization leads to waves of vasoconstriction, perivascular space (PVS) enlargement, and an influx of water from cerebrospinal fluid (CSF) into the brain through aquaporin-4 expressed on astroctye endfeet, causing early poststroke edema. This occurs before contributions from the blood, which exacerbate the edema.



Prestroke
Blood-brain barrier intact

Cytotoxic edema
Early stage of edema with brain tissue swelling

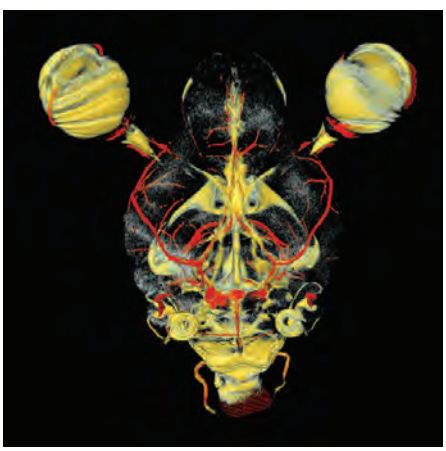
Vasogenic edema
Late stage of edema with disrupted blood-brain barrier



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, UK. Email: anna.williams@ed.ac.uk

Once initiated, why does the CSF influx wave follow the wave of spreading depolarization across the brain? Mestre *et al.* posit that this is due to spreading depolarization causing vasoconstriction of surface and penetrating arterioles (spreading ischemia) (*6*, *7*). They demonstrate that vasoconstriction in distal arteries is sufficient to increase surface arteriole perivascular spaces, which creates a pressure gradient to increase CSF flow in the perivascular spaces of penetrating arterioles. For this increased CSF flow to translate into greater influx into the brain, AQP4 channels are key. In mice engineered with a deletion of *Aqp4*, and exposed to stroke, CSF in the perivascular spaces was decreased, CSF influx into the brain was reduced, and no edema occurred within 15 min after stroke.



After stroke in live mice, cerebral spinal fluid (yellow) enters the brain along the perivascular spaces of cerebral blood vessels (red).

The overall contribution of CSF influx to early poststroke edema, and its dependence on the wave of spreading depolarization, may be relevant to other conditions displaying spreading depolarizations (*8*). For example, traumatic brain injury (TBI) can be exacerbated by cerebral edema, and many mechanisms for this have been proposed, including spreading depolarization (*9*). AQP4 expression is increased after TBI, but its polarized expression favoring astrocytic endfeet is lost, so if CSF influx is a factor in TBI edema, it may be through a different route than after stroke (*10*). Waves of spreading depolarization also underlie the aura of migraine [which doubles the risk of ischemic stroke (*11*)] and are followed by a brief increase and then prolonged decrease of blood flow, but without edema. It is possible that these waves are accompanied by more subtle CSF influxes into the brain, which are adequately removed by the glymphatic system.

It is important to remember that vasogenic edema is a substantial contributor to later edema following stroke. In their stroke-model mouse, Mestre *et al.* show that the vasogenic component of edema becomes more pronounced 24 hours after stroke (see the figure). This might be particularly relevant when considering more targeted treatments for human poststroke recovery. In spinal cord injury, where there is also edema, inhibition of AQP4 expression reduces cytotoxic edema, but increased AQP4 expression is beneficial for water reabsorption during vasogenic edema (*12*), suggesting that therapeutically targeting AQP4 after stroke for beneficial effect may be difficult.

Can CSF-induced edema in stroke be therapeutically targeted? Surgical decompression hemicraniectomy to reduce general brain edema after stroke saves lives (*13*), and cisternostomy to drain CSF has its proponents in TBI (*14*), but could a less invasive nonsurgical approach be developed? Perhaps it may be helpful to therapeutically target the spreading ischemia and/or depolarization using drugs that act on ion channels (e.g., Na^+, K^+, Cl^-, and Ca^{2+} channels) and neurotransmitter receptors [e.g., adrenergic, serotonergic, sigma-1, calcitonin gene-related peptide, γ-aminobutyric acid (GABA), and *N*-methyl-D-aspartate receptors], as has been suggested for other diseases in which this occurs (*15*). The caveat is that the time window for intervention may be short—although perhaps longer in humans than in mice. Nevertheless, Mestre *et al.* have begun decoding the mechanism by which a CSF source might contribute to cytotoxic edema in stroke, and the continuation of this work should improve the opportunities to improve long-term prognoses for the survivors of stroke by shutting the floodgates and protecting the brain. ■

REFERENCES AND NOTES

1. W. Hacke *et al.*, *Arch. Neurol.* **53**, 309 (1996).
2. H. Mestre *et al.*, *Science* **367**, eaax7171 (2020).
3. J. J. Iliff *et al.*, *Sci. Transl. Med.* **4**, 147ra111 (2012).
4. J. J. Iliff *et al.*, *J. Neurosci.* **34**, 16180 (2014).
5. J. P. Dreier, C. Reiffurth, *Neuron* **86**, 902 (2015).
6. H. K. Shin *et al.*, *J. Cereb. Blood Flow Metab.* **26**, 1018 (2006).
7. P. Dreier *et al.*, *Neurosurg.* **93**, 658 (2000).
8. D. R. Kramer *et al.*, *J. Clin. Neurosci.* **24**, 22 (2016).
9. R. M. Jha *et al.*, *Neuropharmacology* **145**, 230 (2019).
10. Z. Ren *et al.*, *J. Cereb. Blood Flow Metab.* **33**, 834 (2013).
11. T. Kurth *et al.*, *Lancet Neurol.* **11**, 92 (2012).
12. Y. Huang *et al.*, *Front. Neurosci.* **13**, 584 (2019).
13. National Institute for Health and Care Excellence (NICE), "Stroke and transient ischaemic attack in over 16s: diagnosis and initial management" (NICE, 2019); www.nice.org.uk/guidance/ng128/evidence/h-surgery-decompressive-hemicraniectomy-pdf-6777399573.
14. I. Cherian *et al.*, *J. Neurosci. Res.* **96**, 744 (2018).
15. C. Costa *et al.*, *J. Headache Pain* **62**, 14 (2013).

ACKNOWLEDGMENTS

The authors are funded by the UK Dementia Research Institute.

10.1126/science.aba8801

THERMOELECTRICS

Seeking new, highly effective thermoelectrics

Operating across a wide temperature range is a priority for thermoelectric materials

By **Yu Xiao** *and* **Li-Dong Zhao**

Thermoelectric technology can directly and reversibly convert heat to electrical energy. Although thermoelectric energy conversion will never be as efficient as a steam engine (*1*), improving thermoelectric performance can potentially make a technology commercially competitive. Thermoelectric conversion efficiency is estimated by the so-called dimensionless figure of merit, $ZT = S^2\sigma T/\kappa$, where S, σ, T, and κ denote the Seebeck coefficient, electrical conductivity, working temperature, and thermal conductivity, respectfully. These parameters are strongly coupled, and improving the final ZT is challenging as a result. Strategies for boosting thermoelectric performance include nanostructuring, band engineering, nanomagnetic compositing, high-throughput screening, and others (*2*). Many of these strategies create a high ZT in a narrow range of temperatures, limiting the overall energy conversion. Finding materials with wider operating temperature ranges may require rethinking development strategies.

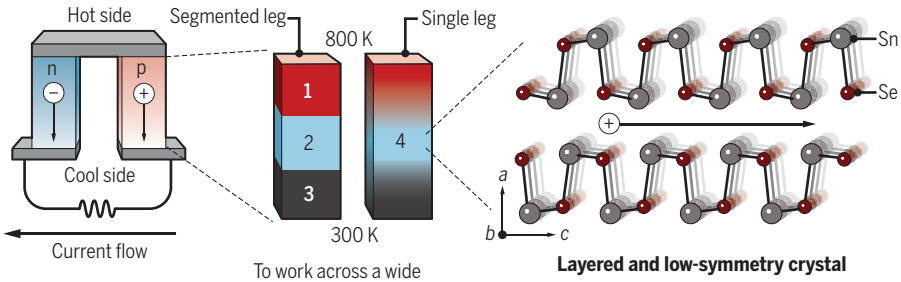
A thermoelectric device is assembled with many cascading n-type and p-type couples. The device efficiency is closely related to the performance of thermoelectric materials. Thermoelectric materials are categorized into three temperature ranges depending on their working temperatures. Bismuth tellurides are typical thermoelectric materials that operate under 400 K. Lead chalcogenides are typical for the 600 to 900 K range. Silicon-germanium and Zintl phases exhibit the best performance above 1000 K. The best thermoelectric performance at the optimal working temperature is restricted by the bandgap (E_g) owing to intrinsic excitation.

School of Materials Science and Engineering, Beihang University, Beijing 100191, China.
Email: zhaolidong@buaa.edu.cn

PHOTO: YUKI MORI

Searching for thermoelectrics

Finding thermoelectric materials that are efficient across a wide range of temperatures is a challenge. A high figure of merit (*ZT*) denotes a promising thermoelectric, but maybe only for a small range of temperatures.

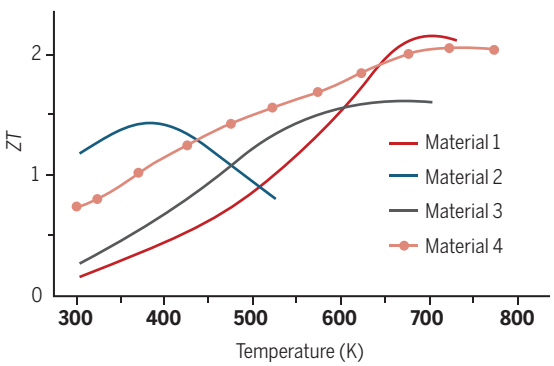


Thermoelectric devices are composed of p-type and n-type semiconducting legs. The legs combine to convert heat into current.

To work across a wide temperature range, a leg can be composed of several materials with narrow bandgaps or one material with a wide bandgap.

Layered and low-symmetry crystal

Wide-bandgap single legs [tin selenide (SnSe), for example] should be crystals that are layered and have low symmetry, which are more likely to have a promising thermoelectric performance.



Comparing the figure of merit at different temperatures
The wide-bandgap material 4 (0.86 eV) has a high *ZT* across the entire temperature range of the segmented leg, whereas each narrow-bandgap material 1 to 3 (0.13, 0.28, and 0.39 eV) works well at different temperature ranges.

This ba^{n}dg^{a}p is given by a rule $E_g = 2eS_{max}T$, where e is unit charge, S_{max} is the maximum Seebeck coefficient, and T is the temperature that corresponds to S_{max}. The Seebeck coefficient is also called the thermopower and is a measurement of the voltage produced within a temperature gradient ($S = \Delta V/\Delta T$, where V is voltage). The bandgap rule means that the most well-known thermoelectric materials are narrow-bandgap semiconductors, such as $(Bi,Sb)_2Te_3$ (E_g ~ 0.13 eV) (*3*), PbTe (E_g ~ 0.28 eV) (*4*), and GeTe-$AgSbTe_2$ (E_g ~ 0.39 eV) (*5*). The maximum *ZT* (ZT_{max}) values for these thermoelectric materials shift to higher temperatures as the bandgap increases (see the figure). To fully realize their potential, thermoelectric materials must work over the entire, several-hundred-kelvin operating range. One method for doing this is with a segmented leg (see the figure), but interfacial resistance and mismatched compatibility factors deteriorate the long-term performance under high temperature (*6*).

Wide-bandgap semiconductors could solve this temperature range issue but often have poor electrical properties, but the wide-bandgap SnSe (~0.86 eV) has proved to be an excellent thermoelectric material. Its *ZT* curve for SnSe covers several narrow-bandgap thermoelectrics (*7*–*9*). SnSe possesses attractive *ZT* values at low temperatures, which continuously increase without saturation up to 800 K. Several special features of SnSe provide some general selection rules for new thermoelectric materials that may work over a wide temperature range. First, the wide bandgap avoids the intrinsic excitation and the *ZT* values are not saturated at high temperatures. Second, layered structures can have high in-plane transport properties that circumvent the normally low carrier density that plagues wide-bandgap semiconductors. Wide-bandgap semiconductors were neglected as promising thermoelectrics because of their intrinsically low carrier density. This deviates from the optimal carrier density (n) owing to the *ZT* parameter interrelations. To achieve high electrical transport properties in materials with low carrier density, high carrier mobility (μ) can be found along an in-plane direction in layered structures, and thus layered materials can reach high electrical conductivity $\sigma = ne\mu$. Furthermore, the low carrier density allows for a high Seebeck coefficient, and, consequently, an ultrahigh power factor ($PF = S^2\sigma$) (*9*, *10*). Third, a low-symmetry structure is connected to low lattice thermal conductivity (κ_{lat}), which is a lower electronic themal conductivity (κ_{ele}) owing to low carrier density, and contributes to the total thermal conductivity ($\kappa = \kappa_{lat} + \kappa_{ele}$). Asymmetric crystal structures have strong anharmonic lattice vibrations useful for lowering thermal conductivity, and their more complex electronic band structure is also attractive for thermoelectric materials. The selection rules will not work for all materials because of the complex interplay between the *ZT* parameters but should provide at least a rough guide for candidate materials. The selection rules for identifying potentially highly effective thermoelectrics are appropriate for both n-type and p-type materials because of their similar transport principles.

Within these selection rules, some promising thermoelectrics can be identified, such as BiCuSeO (*11*), $BiSbSe_3$ (*12*), $K_2Bi_8Se_{13}$ (*13*), and $Sb_2Si_2Te_6$ (*14*). The anisotropic transport properties should lead to improved performance in crystalline forms of these materials where we expect, as for SnSe and SnS crystals, higher carrier mobility (*9*, *10*). Moreover, much more highly effective thermoelectric performance from these anisotropic thermoelectric materials could be expected through integrating present selection rules with the approach to reveal the intrinsically low thermal conductivity (*15*). Finally, it must also be mentioned that not every material with high-range *ZT* values, ZT_{ave}, is going to immediately make for a device-ready material. High ZT_{ave} thermoelectric materials may be challenging to ultimately turn into commercial devices, especially at higher temperatures. Interfacial resistivity and diffusion between the high-performance thermoelectrics and contact electrode that can degrade thermoelectric performance over time are exacerbated at high temperatures. Some of these issues may be solved by device engineering, but the present selection rules also provide a rough guide for finding different types of thermoelectric materials. ■

REFERENCES AND NOTES

1. C. B. Vining, *Nat. Mater.* **8**, 83 (2009).
2. X. Shi, L. Chen, *Nat. Mater.* **15**, 691 (2016).
3. B. Poudel *et al.*, *Science* **320**, 634 (2008).
4. P. F. Poudeu *et al.*, *Angew. Chem. Int. Ed.* **45**, 3835 (2006).
5. S. K. Placheova, *Phys. Status Solidi A Appl. Res.* **83**, 349 (1984).
6. W. Liu, S. Bai, *J. Materiomics* **5**, 321 (2019).
7. L.-D. Zhao *et al.*, *Nature* **508**, 373 (2014).
8. C. Chang *et al.*, *Science* **360**, 778 (2018).
9. L.-D. Zhao *et al.*, *Science* **351**, 141 (2016).
10. W. He *et al.*, *Science* **365**, 1418 (2019).
11. L.-D. Zhao *et al.*, *Appl. Phys. Lett.* **97**, 092118 (2010).
12. S. Wang *et al.*, *Energy Environ. Sci.* **9**, 3436 (2016).
13. D.-Y. Chung *et al.*, *Chem. Mater.* **9**, 3060 (1997).
14. Y. Luo *et al.*, *Joule* **4**, 159 (2020).
15. Y. Lee *et al.*, *Joule* **3**, 719 (2019).

ACKNOWLEDGMENTS

We acknowledge support from the National Key Research and Development Program of China (2018YFA0702100; 2018YFB0703600), National Natural Science Foundation of China (51772012; 51671015), Beijing Natural Science Foundation (JQ18004), Shenzhen Peacock Plan team (KQTD20160226195659991), National Postdoctoral Program for Innovative Talents (BX20190028), 111 Project (B17002), and Postdoctoral Science Foundation of China (2019M660399). L.-D. Z. has support from the National Science Foundation for Distinguished Young Scholars (51925101).

10.1126/science.aaz9426

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

MEDICINE

Treating sickle cell anemia

New drugs, stem cell transplants, and gene therapy show promise in treating sickle cell anemia

By **John F. Tisdale[1], Swee Lay Thein[2], William A. Eaton[3]**

Sickle cell anemia is an inherited disorder caused by a point mutation (affecting a single nucleotide) in the gene that encodes the β-globin chain of hemoglobin (Hbβ). Two β-globin chains and two α-globin chains form hemoglobin, the multisubunit protein in red blood cells that carries oxygen. The mutation results in the replacement of negatively charged glutamate by a neutral, hydrophobic valine that produces sticky patches on the protein surface. Upon delivering oxygen to the tissues, the mutant hemoglobin (HbS) polymerizes into fibers, which distort ("sickle") red blood cells and cause blockage of the circulation, resulting in acute, severe pain called a sickle cell crisis. Pauling and colleagues reported the molecular basis of sickle cell anemia in 1949, giving birth to the field of molecular medicine (*1*). Research on sickle cell anemia has again taken center stage because of new drug therapies, cures through stem cell transplantation, and the promise of gene therapy.

Current treatment options focus largely on best supportive care, including blood transfusions and pain medication. Hydroxyurea (*2*), with its proven efficacy in reducing sickle cell crises and improving survival, should also be considered standard care, but it is grossly underutilized. Hydroxyurea is the first of just two U.S. Food and Drug Administration (FDA)–approved drugs to treat sickle cell disease (SCD) by inhibiting the HbS polymerization that causes sickling. The clinical effectiveness of hydroxyurea is due to the induction of fetal hemoglobin (HbF) production by a still unknown mechanism. HbF is composed of two α-globin chains and two γ-globin chains. The amino acid sequence of HbF is sufficiently different from HbS that little or no HbF takes part in fiber formation, so the primary effect is to dilute HbS (*3*).

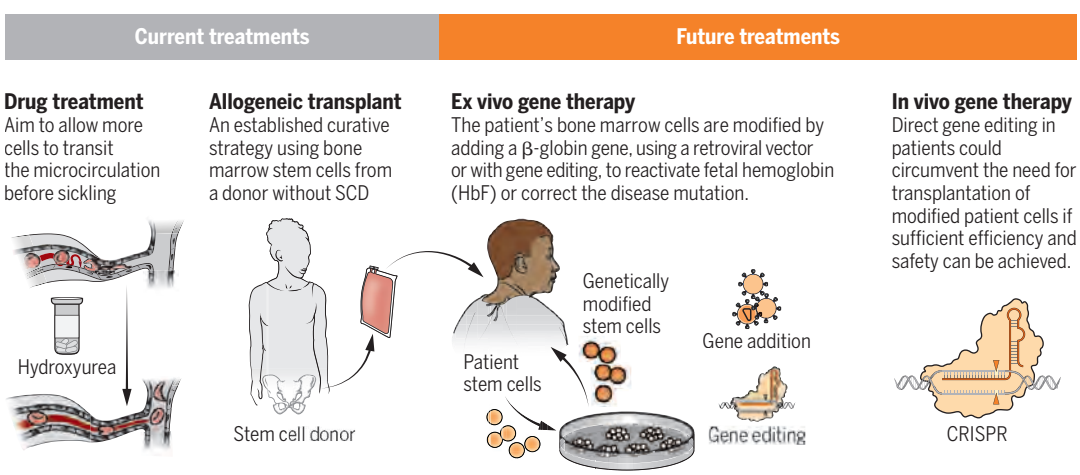
Even a small decrease in the intracellular HbS concentration is therapeutic because of the enormous sensitivity to concentration during the period before HbS fibers appear (delay time), allowing more cells to escape the capillaries of peripheral tissues, where oxygen is delivered, before sickling occurs (*3*). The rare condition of HbS with hereditary persistence of HbF (HbS/HPFH) is caused by compound heterozygous mutations in the genes encoding β-globin and γ-globin. HbF is evenly distributed in all red blood cells of individuals with HbS/HPFH, and there are no complications of SCD. The hydroxyurea-induced HbF increase is not evenly distributed among red blood cells; otherwise, it would be even more effective. The well-established clinical efficacy of hydroxyurea coupled with compelling evidence from the naturally occurring HbS/HPFH-associated mutations demonstrate that higher concentrations of HbF can alleviate clinical complications of SCD. This has motivated both pharmacological and genetic efforts to find approaches that induce HbF production in every red blood cell (*4*).

A second drug that inhibits sickling, voxelotor, was approved by the FDA in November 2019. Voxelotor preferentially binds to the high–oxygen affinity, nonpolymerizing R conformation of HbS, reducing the concentration of the polymerizing T conformation at every oxygen pressure (*3*). However, HbS molecules bound with the drug are in a conformation that delivers very little oxygen to tissues, in a disease characterized by decreased oxygen delivery. So, although patients taking voxelotor show modest increases in hemoglobin concentrations (*5*), it is not necessarily an indication of decreased anemia because the increase in hemoglobin is about the same as the concentration of the drug-bound, non–oxygen-delivering hemoglobin. Moreover, there is no current evidence of a decreased frequency of sickle cell crises, and the effects on organ damage and survival are yet to be determined. However, the increase in hemoglobin is accompanied by decreased markers of red blood cell rupture, indicating reduced sickling (*5*).

A single metric appears to be a primary determinant of SCD severity—the time taken for red blood cells to transit through the capillaries of the tissues relative to the delay time for HbS polymerization (*3*). Consequently, sickling in narrow vessels can be reduced by increasing the delay time but can also be reduced by decreasing adhesion of red blood cells to the vascular endothelium, decreasing transit times. One such agent, also approved by

Current and future treatments for sickle cell anemia

Numerous advances in the understanding of sickle cell disease (SCD) have allowed the development of curative therapies through allogenic stem cell transplanation, with the promise of gene therapy–based treatments in the future.



GRAPHIC: ADAPTED FROM ERINA HE BY N. DESAI/*SCIENCE*

[1]Cellular and Molecular Therapeutics Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA. [2]Sickle Cell Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA. [3]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. Email: johntis@nhlbi.nih.gov; eaton@nih.gov; sweelay.thein@nih.gov

the FDA in November 2019, that does reduce the frequency of sickle cell crises is crizanlizumab, an antibody that blocks the adhesion molecule P-selectin, which is expressed by red blood cells (*6*).

Correction of SCD at the molecular level can be achieved by completely replacing the patient's bone marrow, where red blood cells are produced, with bone marrow that contains red blood cell–producing stem cells with the correct β-globin (*HBB*) gene from an unaffected, tissue-matched sibling donor (see the figure). This allogeneic transplantation procedure has proven curative in ~95% of recipients, primarily in children (*7*). Using an approach that does not completely eradicate the patient's bone marrow in severely affected adults, disease reversion with minimal toxicity has also been achieved in ~90% of patients with as low as 20% replacement of the patient's bone marrow by using a tissue-matched sibling donor (*7*). Although these results are encouraging, only ~10% of patients with SCD in the United States have a tissue-matched sibling donor. Recent efforts to extend this curative matched-sibling transplantation approach to individuals with half-matched family donors are promising, allowing application to nearly every patient because the majority have a half-matched family donor available (*7*).

These allogeneic transplantation results have provided proof of concept that genetic manipulation of the defective bone marrow stem cells might be equally therapeutic. As such, genetic approaches to manipulating the patient's own stem cells and then transplanting them back into the patient (autologous transplant) have been vigorously pursued. Permanent integration of a therapeutic *HBB* gene along with key regulatory elements into the DNA of stem cells became feasible with the development of a robust gene transfer system using a modified HIV1 (*8*). This lentiviral vector system has allowed for sustained, endogenously regulated expression of therapeutic β-globin that is sufficient to revert SCD in patients (*8*–*10*). Using the same approach, Zynteglo, a gene therapy that consists of autologous transplantation of stem cells engineered with a lentiviral vector to express an *HBB* gene, has recently gained approval by the European Commission for adolescents and young adults with the SCD-related disorder, transfusion-dependent β-thalassemia.

Progress in genetic approaches aimed at HbF production has been accelerated by concomitant progress in the understanding of genetic control of the switch from HbF to adult hemoglobin that occurs at birth (hemoglobin switching). The discovery of BCL11A (B cell lymphoma/leukemia 11A) as a major repressor (among others) of the γ-globin genes, *HBG1* and *HBG2*, that compose HbF (*11*), has produced new genetic approaches to HbF production. Two strategies that target BCL11A regulation in bone marrow stem cells for autologous transplant are currently in clinical trials. One involves lentiviral vector–mediated gene transfer of a short-hairpin RNA to reduce *BCL11A* expression. The other involves disruption of discrete regulatory elements of the *BCL11A* gene with CRISPR-Cas9 gene editing (*12*). Another genetic approach uses gene editing to disrupt the DNA binding sites of BCL11A in the promoters of *HBG1* and *HBG2*, mimicking HPFH variants, but this has not reached clinical testing (*13*).

Ongoing and planned clinical trials of the resulting gene therapies designed to increase HbF in SCD have the theoretical advantage over current globin gene addition therapies of preserving the reciprocal relationship between fetal and adult globin chain expression from the endogenous locus; the increase in HbF attained with these approaches will be accompanied by a potentially therapeutic reduction in HbS.

"Research on sickle cell anemia has again taken center stage because of new drug therapies, cures through stem cell transplantation, and the promise of gene therapy."

The ultimate challenge to treat SCD is to genetically correct the HbS mutation. Although correction of the SCD mutation through gene editing is feasible in vitro (*14*), genotoxicity concerns, from off-target effects, as well as low efficiency dictate further studies before clinical application. There are safety concerns with all current therapies that involve genetic manipulation, which include vector-mediated insertional mutagenesis and off-target gene editing, as well as concerns about risks inherent to the high-dose chemotherapy required for autologous bone marrow transplantation. Furthermore, these approaches require a clinical infrastructure to provide considerable supportive care not yet widely available in areas where this disease is most prevalent, including sub-Saharan Africa. Although in vivo gene therapy does not yet currently exist, the U.S. National Institutes of Health and the Bill and Melinda Gates Foundation recently announced a collaborative effort to support the development of a curative in vivo gene therapy approach for both HIV and SCD.

The majority of SCD patients live in under-resourced countries, so an inexpensive drug that inhibits sickling is urgently needed now for these patients. There are many potential drugs in the pipeline to treat SCD, including sickling inhibitors, anti-adhesion agents, and drugs that ameliorate other deleterious sequelae of HbS polymerization, such as oxidative stress and inflammation (*15*). Therapy will not require a drug that completely inhibits sickling but one that increases the delay time to HbS polymerization, allowing more cells to escape the microcirculation and reducing the frequency of vaso-occlusion and corresponding pain. Thus, there is cause for optimism because there are already four different strategies that can increase delay times other than by increasing HbF synthesis. These are (i) increasing cell volume to decrease intracellular hemoglobin concentration, (ii) decreasing the concentration of the allosteric inhibitor 2,3-diphosphoglycerate to decrease fiber stability, (iii) shifting the allosteric equilibrium toward the nonpolymerizing R conformation, and (iv) binding to an intermolecular contact site in the fiber (*3*). Fortunately, there are now large drug libraries available for screening, such as the ReFRAME library, which contains almost 12,000 compounds that, importantly, have already been tested in humans. Compounds that show therapeutically significant effects in a pathophysiologically relevant assay at concentrations known to be nontoxic can be rapidly approved for clinical testing. ■

REFERENCES AND NOTES

1. W. A. Eaton, *Biophys. Chem.* **100**, 109 (2003).
2. O. S. Platt *et al.*, *J. Clin. Invest.* **74**, 652 (1984).
3. W. A. Eaton, H. F. Bunn, *Blood* **129**, 2719 (2017).
4. D. S. Vinjamur, D. E. Bauer, S. H. Orkin, *Br. J. Haematol.* **180**, 630 (2018).
5. E. Vichinsky *et al.*, *N. Engl. J. Med.* **381**, 509 (2019).
6. K. I. Ataga, A. Kutlar, J. Kanter, *N. Engl. J. Med.* **376**, 429 (2017).
7. M. Eapen *et al.*, *Lancet Haematol.* **6**, e585 (2019).
8. C. May *et al.*, *Nature* **406**, 82 (2000).
9. J. A. Ribeil *et al.*, *N. Engl. J. Med.* **376**, 848 (2017).
10. J. F. Tisdale *et al.*, *Blood* **132** (suppl. 1), 1026 (2018).
11. S. Menzel *et al.*, *Nat. Genet.* **39**, 1197 (2007).
12. E. Magrin, A. Miccio, M. Cavazzana, *Blood* **134**, 1203 (2019).
13. S. H. Orkin, D. E. Bauer, *Annu. Rev. Med.* **70**, 257 (2019).
14. D. P. Dever *et al.*, *Nature* **539**, 384 (2016).
15. M. J. Telen, P. Malik, G. M. Vercellotti, *Nat. Rev. Drug Discov.* **18**, 139 (2019).

ACKNOWLEDGMENTS

The authors are supported by the intramural research programs of the National Institute of Diabetes and Digestive and Kidney Diseases and the National Heart, Lung, and Blood Institute of the National Institutes of Health.

10.1126/science.aba3827

POLICY FORUM

BIODIVERSITY

Rethink the expansion of access and benefit sharing

Several UN policy processes are embracing a calcified approach to conservation and equity in science

By **Sarah Laird[1], Rachel Wynberg[2], Michelle Rourke[3,4], Fran Humphries[4], Manuel Ruiz Muller[5], Charles Lawson[6]**

Access and benefit sharing (ABS), a policy approach that links access to genetic resources and traditional knowledge to the sharing of monetary and nonmonetary benefits, first found expression in the 1992 United Nations (UN) Convention on Biological Diversity (CBD). Predicated on the sovereign rights of countries over their biodiversity and associated genetic resources and intended to harness the economic power of those resources to create incentives for and fund biodiversity conservation, the ABS transaction was conceived to foster equitable relations between those parties providing genetic resources and associated traditional knowledge and those wishing to make use of them for research and development. Yet although challenges faced within the CBD suggest that it is time to rethink ABS, several other international policy processes under the auspices of the UN have instead been embracing the ABS approach, and are doing so largely outside of mainstream scientific discourse and attention. The resulting policies could have a major impact on how genetic resources and associated information are collected, stored, shared, and used, and on how research partnerships are configured. We highlight implications for science of the recent expansion of ABS in global policy, in particular the potential incorporation of genetic sequence data.

Moving away from the notion that biodiversity was the "common heritage" of all countries, the CBD affirmed national sovereignty over genetic resources and established a framework for benefit sharing and equity associated with the collection, sharing, and use of genetic resources. The Nagoya Protocol (NP) to the CBD, which entered into force in 2014, provided more detailed mechanisms for implementing ABS and more explicitly linked the CBD's three objectives of conservation, sustainable use, and fair and equitable benefit sharing. Despite acknowledgment of the potential for multilateral approaches, both agreements embedded a bilateral approach to ABS, with the NP emphasizing contracts as a preferred benefit-sharing tool.

The goals of ABS have had broad support in the international diplomacy space, as has the innovative if unproven approach for creating incentives through ABS for biodiversity conservation. But the devil has proven to be in the details. Early on, it became clear that commercial demand for genetic resources was insufficient to incentivize biodiversity conservation. The transactions that did take place under the CBD have yet to generate substantial benefits for conservation (*1*–*3*). Domestic political imperatives have often focused benefit sharing on limited economic development rather than conservation (*4*). Advances were made toward more equitable research partnerships, and in some cases indirect benefits (such as research on threatened biodiversity) resulted, but new bureaucratic hurdles made academic and conservation research more difficult. Even streamlined approaches for noncommercial research required substantial investments of time, money, and capacity to receive permits or sign ABS agreements in countries with unclear legal and administrative structures [e.g., (*3*, *5*, *6*)].

In recent years, concerns associated with ABS policy have expanded and grown more urgent as the CBD and other processes have begun to explore the incorporation of genetic sequence data. Despite its original design as a bridge between advanced technologies and conservation, international ABS policy has focused on the collection and exchange of physical material, largely ignoring developments in biotechnology, which relies heavily on the use of genetic sequence data and information, in addition to physical samples of genetic resources (*7*, *8*). The CBD did not begin work in earnest on "digital sequence information" (DSI) until 2016 (*9*). As part of implementing the NP, the global community is tying itself in knots to retrofit an ABS mechanism designed for physical samples to DSI, and the term DSI itself remains a negotiated placeholder term, the meaning and scope of which remain in dispute.

Many in the scientific community with ABS experience are concerned that DSI might be captured by the same complex ABS policies that they currently must navigate to access physical samples (*3*, *5*). The inclusion of DSI would vastly expand the scope and impact of ABS. However, ABS is a particularly poor policy fit for regulating access to DSI. As currently conceived, ABS presumes that providers and users negotiate agreements and exchange physical material with clear provenance, ownership, and value, and that this material can usually be tracked through the research process, culminating in something of value. DSI turns most of this on its head (*9*, *10*).

Research practices and concepts of ethics and benefit sharing associated with DSI that have evolved in recent decades within the scientific community emphasize openness, transparency, networks, and free exchange. By contrast, ABS is a transactional mechanism that restricts access to genetic resources so that their use can be exchanged for benefits between identified users and providers of these resources. However, to not capture DSI would mean leaving a massive loophole in the ABS endeavor. Users might simply sidestep benefit-sharing obligations by digitizing genetic resources and synthesizing the required nucleic acid fragments with the use of openly accessible DSI. But the question remains as to whether ABS is the best and only way to benefit provider countries, science, and conservation from the use of DSI.

Despite its central role in the policy language of ABS, conservation has faded from the practice of ABS and today is a marginal concern at best. The 2019 global assessment from the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services illustrates the extent of biodiversity loss in recent decades (*11*), even as ABS came on line, but the biodiversity crisis has not triggered a reconsideration of conserva-

[1]People and Plants International, Bristol, VT, USA. [2]Department of Environmental and Geographical Science, University of Cape Town, Rondebosch, South Africa. [3]CSIRO, Synthetic Biology Future Science Platform, Canberra, ACT 2601, Australia. [4]Law Futures Centre, Griffith Law School, Griffith University, Nathan, Queensland 4111, Australia. [5]Peruvian Society of Environmental Law, Lima, Peru. [6]Environmental Futures Research Institute, Griffith Law School, Griffith University, Nathan, Queensland 4111, Australia. Email: laird@peopleandplants.org

tion within the ABS policy process. In fact, disagreements about ABS could undermine ongoing CBD negotiations for a post-2020 global biodiversity framework. At the national level, catastrophic biodiversity loss has often resulted from policies put in place by governments that, despite being keen proponents of ABS, also promote the clearing of biologically diverse habitats for industrial agriculture, mining, logging, and other resource-extractive activities.

If biodiversity conservation is removed from the ABS formulation as a desirable but unrealized objective, we are left with a policy framework that primarily navigates equity and fairness in science and technology through genetic resource exchange. For decades, ABS has provided an invaluable home for important and otherwise orphaned dialogues on ethics and equity in research, ownership, and control of genetic resources and traditional knowledge, capacity building, technology transfer, and other issues (*12*). However, their inclusion under the ABS banner means that issues relating to equity in advanced science and technology are addressed within the UN partly under the auspices of an environmental treaty, the CBD.

As a result, the Conference of Parties to the CBD makes decisions about scientific research practices that can have impacts far afield from biodiversity conservation. These meetings are generally attended by diplomats from ministries of the environment, often with little experience in genomics, bioinformatics, biotechnology, and related fields. Scientific expertise is available within these processes through expert groups, committees, and other forums, but how expertise is translated for decision-makers is a recurring challenge. Additionally, although these are constituted as scientific and technical bodies, they are nominated by governments and can be highly political and used as negotiating spaces. There are few if any forums where scientists can engage on a neutral platform.

Limited capacity within the policy-making process to understand the technical scientific issues and commercial practices underlying ABS may be one reason why—after almost 30 years, innumerable national ABS measures, and tens of millions of dollars spent discussing and developing these policies—there is relatively little to show in the way of conservation, technology transfer, capacity-building, or other monetary or nonmonetary benefits. Entrenched positions between the global North and South today add to these concerns.

ABS EXPANSION WITHIN THE UN

The challenges faced by ABS within the CBD and NP process, and the belated and still nascent consideration of DSI, might suggest that it is time to rethink ABS and learn lessons from other areas of policy. However, in the past few years, many UN forums have sought guidance from the ABS policy arena rather than from more established ethics platforms. In addition to the CBD and its supplementary NP processes, ABS has been pursued under the World Health Organization (WHO) Pandemic Influenza Preparedness (PIP) Framework for the Sharing of Influenza Viruses and Access to Vaccines and Other Benefits, the International Treaty for Plant Genetic Resources for Food and Agriculture (or "Plant Treaty"), and deliberations under the Convention on the Law of the Sea (UNCLOS).

Cross-border sharing of genetic sequence data may be made more cumbersome under certain UN policies being discussed.

PHOTO: TEK IMAGE/SCIENCE SOURCE

The WHO is perhaps the most forceful in pushing ABS as a solution to its perceived ethical and equity shortcomings on the issue of virus sharing. It originally ventured into ABS through its PIP Framework but has started exploring options for incorporating aspects of the NP into the sharing of other pathogens. Advocacy for the NP in this forum, although an admirable effort to create synergies across UN bodies and to promote the general idea of benefit sharing, ignores the very different goals of the CBD and its ABS mechanism versus the public health mandate of the WHO. Infectious disease research and the public health response require timely sharing of up-to-date pathogen samples and associated DSI, as well as global collaboration across diverse parties. In contrast, the NP emphasizes bilateral negotiation of mutually agreed terms between provider and user parties, which can take years to achieve before samples are actually shared. Already, delays in sharing influenza virus samples between national centers in Southeast Asia and South America, and between centers in France and Switzerland, have reportedly been due to NP compliance concerns (*13*, *14*).

Elsewhere in the UN, countries are negotiating a treaty under the UNCLOS to address governance gaps in the conservation and sustainable use of marine biological diversity in areas beyond national jurisdiction (ABNJ). Marine genetic resources (MGRs) found within national waters are covered by the CBD and NP, but countries are exploring how to govern MGRs found in the two-thirds of oceans outside national jurisdiction, which might have an impact on vast ex situ collections of marine samples collected since the CBD entered into force. Although the commercial potential of MGRs from ABNJ to fund conservation and capacity building is unclear, the ABS mechanism has been introduced into the negotiations as a way to redistribute wealth and technology to countries that lack the capacity to exploit and benefit from research on the high seas. Negotiating states have not yet agreed how to operationalize the ABS concept in the unique geopolitical conditions of ABNJ. There are no sovereign rights over MGRs in this area and therefore no "providers" with legal rights to share in the benefits from their use. Multilateral governance options in the draft text (November 2019) include a complex web of procedures for collecting or accessing MGRs, and options for benefit sharing draw from the CBD's approach originally designed for physical materials moving from providers to users. Most concerningly, the lack of agreement on MGR provisions is stalling crucial, and essentially agreed upon, negotiations on mechanisms for area-based management tools, environmental impact assessments, capacity building, and the transfer of marine technology.

Unlike the CBD and UNCLOS, the Plant Treaty addresses the relatively bounded field of food and agriculture, with a more clearly identified constituency, relatively actively involved in policy-making. The Plant Treaty ABS model has its own structure and logic, which evolved from long-standing tensions between developing and

developed countries about access to and ownership of plant genetic resources in ex situ collections, linked to the expansion of intellectual property over plant varieties. This, combined with concerns about ensuring uninterrupted germ plasm flows for research and innovation, led to the establishment of a multilateral mechanism for pooling benefits and sharing them through governments, based in part on a list of 64 key crops essential for food security.

Like the CBD and NP, however, the Plant Treaty is shackled by an overly contractual approach, geared toward capturing private market values for conservation, without an indication of how this can be achieved. It is also similarly struggling to address the nonmaterial aspects of plant genetic resources such as genetic sequence data (*10*, *15*). To avoid some of the complicated contracting arrangements required by the CBD and NP, the Plant Treaty has a simpler, standard-form, "take it or leave it" ABS contract (the Standard Material Transfer Agreement or SMTA), but monetary and other benefits have not met government expectations, and attempts to redraft the SMTA remain mired in controversy.

Finally, ABS has been engaged by the UN Sustainable Development Goals (SDGs), anticipating that ABS will contribute toward ending poverty and hunger (Goal 2) and protecting life on land (Goal 15). With the problems already apparent within the CBD and related processes, it appears fanciful to imagine ABS, as currently practiced, delivering on these and other SDGs.

TIME FOR A NEW APPROACH

That multiple UN forums are reaching for the ABS mechanism makes clear the urgent need for a global institutional and conceptual framework for ethical research and commercialization, and the environmental and social implications of scientific and technological advances. But ABS is not that framework. We fully support the goals of ABS and the efforts of those working within the CBD, WHO, Plant Treaty, UNCLOS, and elsewhere to make scientific research and commercialization more equitable and sustainable. There are clear inequities between the global North and South in research funding, control over resources and data, benefit sharing, and other issues that must be addressed. But ABS has calcified over the years around a bilateral transaction for physical samples that is marginal to contemporary research and development, and the dissonance between ABS and the scientific endeavor more broadly is only increasing. A new approach for ethically sharing the benefits of science and technology is sorely needed.

First, as the global community confronts massive and catastrophic biodiversity loss, the enormous sums of money and time spent nursing the hope that indirect economic incentives from high-tech sectors through ABS will conserve biodiversity should be reconsidered. It's also quite possible that the substantial funding allocated for ABS implementation has had the unintended consequence in many countries of diverting government attention away from biodiversity conservation.

Second, more researchers and their organizations, from a wider range of fields, particularly those that may be affected by the inclusion of DSI, should participate in policy processes—attend UN meetings, write background documents tailored to policy-making, and work with national delegates and focal points to develop alternative approaches to equitable research and commercialization. This will require training scientists to engage with complex policy processes and to cross disciplinary boundaries. Funders and research institutions might support the engagement of scientists in policy processes as part of grant applications and institutional policies. Correspondingly, these UN forums and Parties to Conventions must make a real effort to ensure that scientific and technical bodies comprise experts in relevant fields, that they contribute unhindered by negotiating positions, and that decision-makers are well versed in the latest scientific and technological developments.

Finally, we propose taking a step back and focusing on first principles and the foundational objectives of each respective policy process. Working from these, we can best identify how each process can contribute to biodiversity conservation, social justice, equitable research and commercialization, and public health. We can then explore legal, ethical, and policy approaches that might achieve the objectives.

New and encouraging ideas and approaches to ABS, and ethical research more broadly, have emerged in recent years, including more open-access strategies that better address science as it is increasingly practiced. Proposals include delinking access from benefit sharing for DSI, which would secure benefits while maintaining open science and generating funds from taxes, levies, or tiered approaches that feed a multilateral fund (*8*). Such funds have a poor track record to date, and if targeted to biodiversity conservation, they should also be funded by sectors destructive to biodiversity (e.g., oil, mining, logging, and industrial agriculture), not only those researching biodiversity. Streamlined multilateral systems for all genetic resources might avoid costly, duplicative, and ineffective tracking systems and could be linked to intellectual property tools to identify phases of commercial utilization that trigger benefit-sharing obligations. Efforts focused not on the monetary considerations, but on promoting more inclusive innovation and greater equity in biodiversity research and commercialization, and broader public and social benefits from the outcomes of science, are likely to have a greater impact over time.

This is a critical juncture. In the coming year, important meetings will be held in each of the policy forums discussed above, and decisions will be made on DSI and ABS that will have impacts for years to come. In many cases, the implications of these decisions have not been fully explored. In the face of rapid scientific and technological advances, and equally swift and alarming biodiversity loss, it is time to get this right. ■

REFERENCES AND NOTES

1. S. A. Laird, K. ten Kate, in *Selling Forest Environmental Services: Market-Based Mechanisms for Conservation and Development*, S. Pagiola, J. Bishop, N. Landell-Mills, Eds. (Earthscan, 2002), pp. 151–172.
2. R. Wynberg, S. A. Laird, in *Indigenous Peoples, Consent and Benefit Sharing: Lessons from the San-Hoodia Case*, R. Wynberg, D. Schroeder, R. Chennells, Eds. (Springer, 2009), pp. 69–86.
3. K. D. Prathapan, R. Pethiyagoda, K. S. Bawa, P. H. Raven, P. D. Rajan, 172 co-signatories from 35 countries, *Science* **360**, 1405 (2018).
4. R. Wynberg, *S. Afr. J. Bot.* **110**, 39 (2017).
5. F. A. Bockman *et al.*, *Science* **360**, 865 (2018).
6. D. Neumann *et al.*, *Org. Divers. Evol.* **18**, 1 (2018).
7. M. Ruiz Muller, *Genetic Resources as Natural Information: Implications for the Convention on Biological Diversity and Nagoya Protocol* (Routledge, 2015).
8. C. Lawson, F. Humphries, M. Rourke, *J. World Intellect. Prop.* **22**, 103 (2019).
9. S. A. Laird, R. Wynberg, *Fact-Finding and Scoping Study on Digital Sequence Information on Genetic Resources in the Context of the Convention on Biological Diversity and the Nagoya Protocol* (Convention on Biological Diversity, 2018).
10. C. Lawson, H. Burton, F. Humphries, *Eur. Intellect. Prop. Rev.* **40**, 243 (2018).
11. IPBES, *Global Assessment Report on Biodiversity and Ecosystem Services* (2019); https://ipbes.net/global-assessment.
12. S. A. Laird, R. P. Wynberg, *NanoEthics* **10**, 189 (2016).
13. World Health Organization, *Approaches to Seasonal Influenza and Genetic Sequence Data Under the PIP Framework* (14 December 2018); www.who.int/influenza/pip/WHA70108b_Analysis.pdf.
14. GISAID, *GISAID's Comments on the WHO Report of the Public Health Implications of Implementation of the Nagoya Protocol* (2019); www.gisaid.org/references/statements-clarifications/who-report-on-the-public-health-implications-of-nagoya-protocol-13-may-2019/.
15. E. W. Welch, M. Bagley, T. Kuiken, S. Louafi, *Potential Implications of New Synthetic Biology and Genomic Research Trajectories on the International Treaty for Plant Genetic Resources for Food and Agriculture* (Food and Agriculture Organization of the United Nations, 2018); www.fao.org/fileadmin/user_upload/faoweb/plant-treaty/GB7/gb7_90.pdf.

ACKNOWLEDGMENTS

Supported by the UK Government's Darwin Initiative and by the South African Research Chairs Initiative of the Department of Science and Innovation and National Research Foundation of South Africa.

10.1126/science.aba9609

Protecting the snow leopard's habitat, which crosses the boundaries of multiple countries, could have positive effects on both biodiversity and geopolitics.

PHOTO: SEBASTIAN KENNERKNECHT/MINDEN PICTURES

Edited by **Jennifer Sills**

Ease conflict in Asia with snow leopard peace parks

The Himalayas, the mountains of central Asia, and the mountains of southwest China—3 of the world's 34 biodiversity hotspots (*1*)—have suffered severe biodiversity loss as a result of war and related military activities (*1*, *2*). The tense political status in these areas, which share boundaries with multiple countries, hampers research into the losses and efforts to mitigate them. Establishing a peace park in this region to protect a proven flagship species for the mountain ecosystem in these countries—the snow leopard (*Panthera uncia*) (*3*)—could be an effective solution.

Peace parks are transboundary areas that promote cooperation and protect biological diversity and natural and cultural resources. Designating peace parks in central and south Asia would allow tourism, environmental restoration, and scientific research in the region (*4*). The demilitarization of the protected areas would serve as a geographic buffer zone that reinforces the existing cease-fire by physically separating the belligerents in the mountainous areas of the Himalayas, central Asia, and southwest China. (*1*). Transboundary peace parks have been established or proposed to conserve biodiversity and aid conflict resolution in several disputed and trans-border areas within the snow leopard range, including the Karakoram range (India and Pakistan), Khangchendzonga Conservation Area (India, China, and Nepal), Qomolangma/Sagarmatha Region (China and Nepal), Pamir Wakhan (Afghanistan, Pakistan, Tajikistan, and Kyrgyzstan), and Altai Mountains (Russia, China, Mongolia, and Kazakhstan) (*4*, *5*).

The demilitarized zone (DMZ) between North and South Korea, uninhabited for decades, has created a de facto 4-km-wide nature reserve spanning the Korean peninsula and harboring various endangered species (*6*). Khunjerab National Park, on Pakistan's northern border with China, serves as a good example of a transboundary protected area that already exists in the snow leopard range. The park promotes research, ecotourism, and livelihoods for local communities (*7*), and it has allowed scientists to document globally threatened biodiversity such as mountain ungulates and snow leopards in the Karakoram mountain range (*8*, *9*).

The snow leopard, which has been identified and endorsed as a flagship species by 12 range countries and the International Union for Conservation of Nature and World Commission on Protected Areas (*3*), could become an important symbol representing fragile transboundary mountain ecosystems. Transboundary habitats are multidimensional efforts with many stakeholders and joint governance. Creating snow leopard peace parks in central and south Asia could initiate a better dialogue among south Asian countries and contribute to the resolution of long-standing international conflict. Moreover, designating a protected area for the snow leopard would preserve other species in this unique high-mountain ecosystem as well. A landscape conservation approach that extends beyond the country-specific protected area network could be the way forward in this politically sensitive and ecologically fragile region.

Aishwarya Maheshwari
College of Forestry, Banda University of Agriculture and Technology, Banda-210001, Uttar Pradesh, India. Email: aishwaryamaheshwari@gmail.com

REFERENCES AND NOTES

1. T. Hanson *et al.*, *Cons. Biol.* **23**, 578 (2009).
2. T. A. Formoli, *Cons. Biol.* **11**, 309 (1995).
3. Snow Leopard Working Secretariat, "Global Snow Leopard and Ecosystem Protection Program" (Bishkek, Kyrgyz Republic, 2013).
4. S. H. Ali, *Peace Parks: Conservation and Conflict Resolution* (The MIT Press, Cambridge, Massachusetts, 2007).
5. S. H. Ali, "Environmental peace parks: Prospects for South Asia" (2004); www.envirosecurity.org/conference/presentations/session5/ESSD_Session_5_Saleem_H_Ali_PPT.ppt.
6. K. C. Kim, *Science* **278**, 242 (1997).
7. D. C. Zbicz, in *On the Frontiers of Conservation: Proceedings of the 10th Conference on Research and Resource Management in Parks and on Public Lands*, D. Harmon, H. Mich, Eds. (The George Wright Society, 1999), pp. 199–204.
8. R. Qureshi *et al.*, *Pakistan J. Bot.* **43**, 849 (2011).
9. Climate Change Division, Government of Pakistan, "Pakistan National Snow Leopard Ecosystem Protection Priorities, Global Snow Leopard Ecosystem Protection Programme, Second draft" (2013), pp. 126–145; www.globalsnowleopard.org/wp-content/uploads/2016/05/Pakistan_NSLEP.pdf.

10.1126/science.aba9882

Mass mortality of migratory birds in Iran

Since 22 January 2020, over 20,000 of an estimated more than 250,000 aquatic migratory birds have been found dead in Miankaleh International Wetland (MIW) in Iran's Mazandaran Province, and the number of mortalities is increasing (*1*–*4*). The head of the Iran Veterinary Organization has announced botulinum toxin (produced by the bacterium *Clostridium botulinum*) as the birds' cause of death (*5*). However,

some environmental groups and scientists suspect that deliberate poisoning by local hunters may be involved (*5*, *6*); such poisonings occurred at MIW in 1994 and 2003 (*5*, *7*). According to the deputy head of the Mazandaran Provincial Office of Iran's Department of Environment, birds were found dead along only 10 km of 65 km of shoreline, which suggested the cause may not be botulinum, and some of these birds were illegally brought to Fereydunkenar local market for human consumption (*8*). The Iranian government should take immediate action to clearly identify the cause (botulism or poisoning) of this mass mortality event.

If poisoning is the cause, we recommend that the government consider introducing national legislation to enhance protection of these migratory birds and increase enforcement presence in Mazandaran's wetlands—designated as Wetlands of International Importance in the Ramsar Convention on Wetlands (*9*)—to prevent illegal poisoning of birds. Irrespective of cause, we encourage the government to immediately remove any migratory birds from the Fereydunkenar local market, which remains open (*10*) and is the main market to sell migratory birds in Mazandaran Province. Without these actions, such disasters will likely be repeated, with the loss of many migratory birds and unnecessary potential risks to human health.

Jamshid Parchizadeh[1]* and Jerrold L. Belant[2]
[1]Tehran, Tehran Province, Iran. [2]Global Wildlife Conservation Center, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, USA.
*Corresponding author.
Email: jamshid.parchizadeh@gmail.com

REFERENCES AND NOTES

1. "Thousands of birds found dead in north-east Iran," BirdGuides (2020); www.birdguides.com/news/thousands-of-birds-found-dead-in-north-east-iran/.
2. "Miankaleh's dead birds were found at local markets?," Khabar Online (2020); www.khabaronline.ir/news/1349600/ [in Farsi].
3. "Migratory birds were killed in Miankaleh," Iran's Metropolises News Agency (2020); www.imna.ir/news/410799/ [in Farsi].
4. "Flocks of migratory birds perish in peninsula in northern Iran," Mehr News Agency (2020); https://en.mehrnews.com/news/155046/Flocks-of-migratory-birds-perish-in-peninsula-in-northern-Iran.
5. "Deliberate poisoning in Miankaleh," Salamat News (2020); www.salamatnews.com/news/288876/ [in Farsi].
6. "Environmental experts: The real cause of the death of the birds in Miankaleh is hidden," Deutsche Welle (2020); https://p.dw.com/p/3Xr2W [in Farsi].
7. "Seventeen thousand migratory birds were killed," Hamshahri Online (2020); www.hamshahrionline.ir/news/484647/ [in Farsi].
8. "Dead birds were brought to Fereydunkenar local market for human consumption," Iran International (2020); www.instagram.com/tv/B8Q97zKKA8O/ [in Farsi].
9. Ramsar List, "The list of wetlands of international importance" (2020); www.ramsar.org/sites/default/files/documents/library/sitelist.pdf.
10. "Fereydunkenar local market is still open!," The Islamic Republic News Agency (2020); www.irna.ir/news/83662413/ [in Farsi].

10.1126/science.abb4887

NIH progress toward inclusive excellence

In his News story "From service to science: NIH shifts focus of mentoring network aimed at boosting grantee diversity" (3 January, https://scim.ag/NIHmentoringshift), J. Mervis describes the National Research Mentoring Network (NRMN), a program designed by the National Institutes of Health (NIH) to develop mentoring and networking opportunities for biomedical researchers from diverse backgrounds, including underrepresented groups (*1*, *2*). He explains that the first phase of the program had established a successful online portal offering a variety of mentoring activities, and that the second phase of the initiative will focus on analyzing mentoring best practices, including 11 studies on mentoring (*3*). He also links to the Building Infrastructure Leading to Diversity (BUILD) program, where the "bulk of NIH's initial 5-year, $250 million investment…went." Mervis quotes one of the NRMN's leaders as hoping the NIH will not leave researchers feeling "abandoned." We were glad to see the update on these important programs, and we want to assure readers that we are committed to supporting researchers by continuing to fund programs that foster inclusion and diversity. We would also like to share some early data trends that indicate progress over the past 5 years.

After 5 years of the Diversity Program Consortium (*4*), which includes both NRMN and BUILD, progress is apparent at all career levels (*5*). NRMN grant writing programs are particularly promising: Among 540 highly diverse participants to date, 89 have been identified as project directors or principal investigators (PD/PIs) on NIH awards of about $65 million, with more than 70% funding PD/PIs from underrepresented groups and women (*5*). Although the racial gap in rates for PD/PIs on R01 awards (number of applicant PD/PIs divided by number of awards) reported in 2011 (*6*) has not been eliminated, we see good reason for optimism based on Black/African-American PD/PIs on R01 applications and awards increasing by 28.9% (425 to 548) and 117% (52 to 113), respectively (*5*). The racial gap in award rates for mentored career-development (K) awards is essentially closed with the number of Black/African-American PD/PIs on K-award applications and awards increasing by 55.5% (119 to 185) and 142% (26 to 63), respectively (*5*). We see similar trends for Hispanic/Latinx scientists with increases in applicants, awardees, and award rates (*5*).

Our biggest challenge is enhancing diversity in the pool of PD/PIs on research grant applications, but here too we see promising trends earlier in the career path. In 2017, scientists from underrepresented groups comprised 14% of Ph.D. recipients in fields relevant to biomedical research (*7*). However, fewer than 5% are hired as tenure-track faculty (*8*). To help address the climate for junior faculty, including those from underrepresented groups, NIH piloted the NIH Distinguished Scholars Program in its intramural research program, aimed at building a community of investigators committed to mentoring, diversity, and inclusion (*9*). Early indicators suggest high impact in achieving the program goals while substantially enhancing diversity among tenure-track investigators (*10*). NIH recently approved support from its Common Fund to expand this cohort model to NIH-funded institutions nationwide (*11*).

In 2020, NIH will bring to bear effective strategies that create institutional cultures where individuals with diverse perspectives can thrive and create new knowledge to improve human health. Data-driven approaches to inclusive excellence will continue to drive NIH initiatives for addressing the challenges facing all scientists, including those from underrepresented groups.

Hannah A. Valantine[1,2]* and Francis Collins[3,4]
[1]Chief Officer, Scientific Workforce Diversity, [2]Senior Investigator, Laboratory Transplantation Genomics, National Heart, Lung and Blood Institute, Bethesda, MD 20892, USA. [3]Director, National Institutes of Health, Bethesda, MD 20814, USA. [4]Senior Investigator, National Human Genome Research Institute, Bethesda, MD 20892, USA.
*Corresponding author.
Email: hannah.valantine@nih.gov

REFERENCES AND NOTES

1. C. A. Sorkness *et al.*, *BMC Proc.* **11**, 22 (2017).
2. H. P. Jones *et al.*, *BMC Proc.* **11**, 16 (2017).
3. National Institute of General Medical Sciences, National Research Mentoring Network Phase II (www.nigms.nih.gov/training/dpc/pages/nrmn.aspx).
4. S. Hurtado, *BMC Proc.* **11(Suppl 12)**, 17 (2017).
5. NIH Scientific Workforce Diversity Actions and Progress: 2014–2019 (https://acd.od.nih.gov/documents/presentations/06132019Diversity.pdf).
6. D. K. Ginther *et al.*, *Science* **333**, 1015 (2011).
7. National Science Foundation, National Center for Science and Engineering Statistics, "Doctorate recipients from U.S. universities: 2017; data table 22" (2018); https://ncses.nsf.gov/pubs/nsf19301/data.
8. K. D. Gibbs, J. Basson, I. M. Xierali, D. A. Broniatowski, *Elife* **5**, e21393 (2016).
9. "NIH Distinguished Scholars Program: A cohort model for enhancing diversity and inclusion of principal investigators in the NIH intramural research program" (https://diversity.nih.gov/programs-partnerships/dsp).
10. NIH Council of Councils, "Faculty institutional recruitment for sustainable transformation" (2020); https://dpcpsi.nih.gov/sites/default/files/CoC_Jan_2020_1115_FacultyDiversity.pdf
11. J. Mervis, *Science* 10.1126/science.abb1082 (2020).

10.1126/science.abb4619

PRIZE ESSAY

REGENERATIVE MEDICINE

Aging eyes and the immune system

Reining in age-related inflammation may enhance stem cell therapies for retinal disease

By **Joana Neves**

A central promise of regenerative medicine is the ability to repair aged or diseased organs using stem cells (SCs). This approach will likely become an effective strategy for organ rejuvenation, holding the potential to increase human health by delaying age-related diseases (*1*). The successful translation of this scientific knowledge into clinical practice will require a better understanding of the basic mechanisms of aging, along with an integrated view of the process of tissue repair (*1*).

The advent of SC therapies, now progressing into clinical trials, has made clear the many challenges limiting the application of SCs to treat disease. Our duty, as scientists, is to anticipate such limitations and propose solutions to effectively deliver on the promise of regenerative medicine.

ROADBLOCKS LIMITING THE PROGRESS OF REGENERATIVE THERAPIES

Degenerating tissues have difficulty engaging a regulated repair response that can support efficient cell engraftment and restoration of tissue function (*2*). This problem, which I encountered when trying to apply SC-based interventions to treat retinal disease, will likely be an important roadblock to the clinical application of regenerative medicine approaches in elderly patients, those most likely to benefit from such interventions. I therefore hypothesized that the inflammatory environment present in aged and diseased tissues would be a major roadblock for efficient repair and that finding immune modulators with the ability to resolve chronic inflammation and promote a prorepair environment would be an efficient approach to improve the success of SC-based therapies (*2*, *3*).

Instituto de Medicina Molecular (iMM), Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal.
Email: joana.neves@medicina.ulisboa.pt

The Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy

Immune cells, as sources and targets of inflammatory signals, emerged naturally as an ideal target for intervention. I chose to focus on macrophages, which are immune cells of myeloid origin that exist in virtually every tissue of the human body and which are able to reversibly polarize into specific phenotypes, a property that is essential to coordinate tissue repair (*3*, *4*).

EVOLUTIONARILY CONSERVED PROCESSES DRIVING NEW THERAPEUTIC APPROACHES

If there is an integral immune modulatory component to the process of tissue repair that has evolved to support the healing of damaged tissues, then it should be possible to find strategies to harness this endogenous mechanism and improve regenerative therapies. Anchored in the idea that tissue damage responses are evolutionarily conserved (*5*), I started my research on this topic using the fruit fly *Drosophila* as a discovery system.

The fruit fly is equipped with an innate immune system, which is an important player in the process of tissue repair. Using a well-established model of tissue damage, I sought to determine which genes in immune cells are responsible for their prorepair activity. MANF (mesencephalic astrocyte-derived neurotrophic factor), a poorly characterized protein initially identified as a neurotrophic factor, emerged as a potential candidate (*6*). A series of genetic manipulations involving the silencing and overexpression of MANF and known interacting partners led me to the surprising discovery that, instead of behaving as a neurotrophic factor, MANF was operating as an autocrine immune modulator and that this activity was essential for its prorepair effects (*2*). Using a model of acute retinal damage in mice and in vitro models, I went on to show that this was an evolutionarily conserved mechanism and that MANF function could be harnessed to limit retinal damage elicited by multiple triggers, highlighting its potential for clinical application in the treatment of retinal disease (*2*).

Having discovered a new immune modulator that sustained endogenous tissue repair, I set out to test my initial hypothesis that this factor might be used to improve the success of SC-based therapies applied to

GRAND PRIZE WINNER

Joana Neves

Joana Neves received undergraduate degrees from NOVA University in Lisbon and a Ph.D. from the Pompeu Fabra University in Barcelona. After completing her postdoctoral fellowship at the Buck Institute for Research on Aging in California, Neves started her lab in the Instituto de Medicina Molecular (iMM) at the Faculty of Medicine, University of Lisbon in 2019. Her research uses fly and mouse models to understand the immune modulatory component of tissue repair and develop stem cell–based therapies for age-related disease.

FINALIST

Arun Sharma

Arun Sharma received his undergraduate degree from Duke University and a Ph.D. from Stanford University. Having completed a postdoctoral fellowship at the Harvard Medical School, Sharma is now a senior research fellow jointly appointed at the Smidt Heart Institute and Board of Governors Regenerative Medicine Institute at the Cedars-Sinai Medical Center in Los Angeles. His research seeks to develop in vitro platforms for cardiovascular disease modeling and drug cardiotoxicity assessment.

www.sciencemag.org/content/367/6483/1206.1

FINALIST

Adam C. Wilkinson

Adam C. Wilkinson received his undergraduate degree from the University of Oxford and a Ph.D. from the University of Cambridge. He is currently completing his postdoctoral fellowship at the Institute for Stem Cell Biology and Regenerative Medicine at Stanford University, where he is studying normal and malignant hematopoietic stem cell biology with the aim of identifying new biological mechanisms underlying hematological diseases and improving the diagnosis and treatment of these disorders.

www.sciencemag.org/content/367/6483/1206.2

PHOTOS: (LEFT TO RIGHT) COURTESY OF J. NEVES; COURTESY OF A. SHARMA; JOYDEEP BHADURY

a degenerating retina. Indeed, the low integration efficiency of replacement photoreceptors transplanted into congenitally blind mice could be fully restored to match the efficiency obtained in nondiseased mice by supplying MANF as a co-adjuvant with the transplants (*2*). This intervention improved restoration of visual function in treated mice, supporting the utility of this approach in the clinic (*7*).

USING MANF AS AN IMMUNE MODULATOR

Next, my colleagues and I decided to address the question of whether the immune modulatory mechanism described above was relevant for aging biology and whether we could harness its potential to extend health span. We found that MANF levels are systemically decreased in aged flies, mice, and humans. Genetic manipulation of MANF expression in flies and mice revealed that MANF is necessary to limit age-related inflammation and maintain tissue homeostasis in young organisms. Using heterochronic parabiosis, an experimental paradigm that involves the surgical joining of the circulatory systems of young and old mice, we established that MANF is one of the circulatory factors responsible for the rejuvenating effects of young blood. Finally, we showed that pharmacologic interventions involving systemic delivery of MANF protein to old mice are effective therapeutic approaches to reverse several hallmarks of tissue aging (*8*).

REGENERATIVE MEDICINE AS A REJUVENATING INTERVENTION

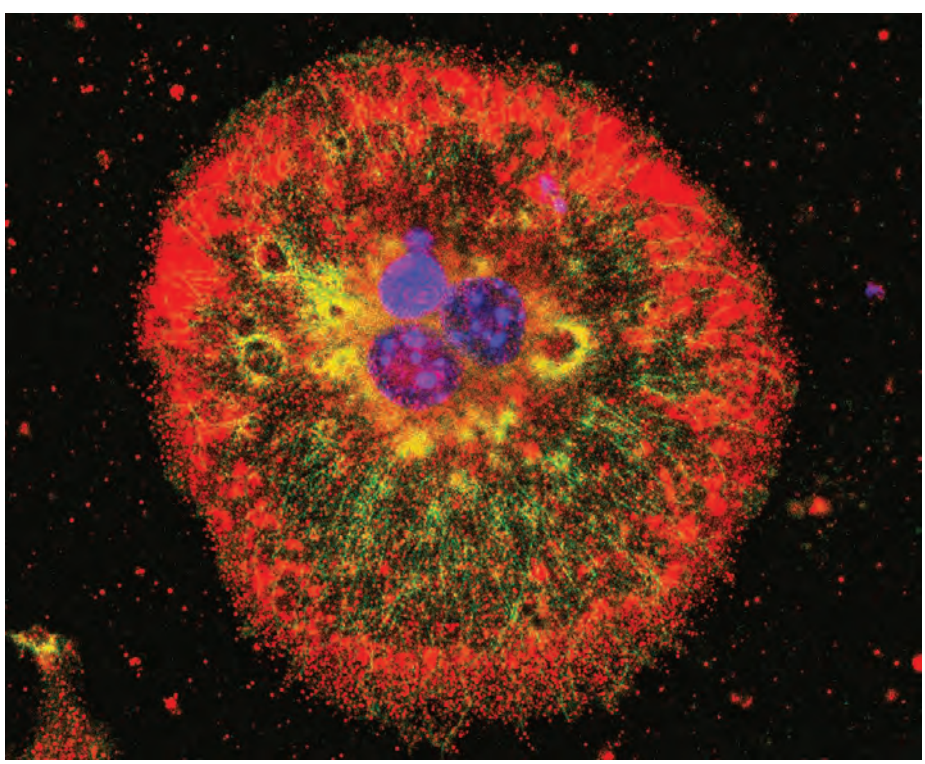
The biological process of aging is multifactorial, necessitating combined and integrated interventions that can simultaneously target several of the underlying problems (*9*). The potential of immune modulatory interventions as rejuvenating strategies is emerging and requires a deeper understanding of its underlying molecular and cellular mechanisms.

One expected outcome of reestablishing a regulated inflammatory response is the optimization of tissue repair capacity that naturally decreases during aging (*3*). Combining these interventions with SC-based therapeutics holds potential to deliver on the promise of regenerative medicine as a path to rejuvenation (*1*). ■

REFERENCES AND NOTES

1. J. Neves, P. Sousa-Victor, H. Jasper, *Cell Stem Cell* **20**, 161 (2017).
2. J. Neves *et al.*, *Science* **353**, aaf3646 (2016).
3. J. Neves, P. Sousa-Victor, *FEBS J.* **287**, 43 (2020).
4. Y. Okabe, R. Medzhitov, *Nat. Immunol.* **17**, 9 (2016).
5. J. Neves, M. Demaria, J. Campisi, H. Jasper, *Dev. Cell* **32**, 9 (2015).
6. M. Lindahl, M. Saarma, P. Lindholm, *Neurobiol. Dis.* **97** (Pt B), 90 (2017).
7. J. Neves, D. Lamba, H. Jasper, U.S. Patent US20170274048A1 (2017).
8. P. Sousa-Victor *et al.*, *Nat. Metab.* **1**, 276 (2019).
9. B. K. Kennedy *et al.*, *Cell* **159**, 709 (2014).

10.1126/science.aba6114



A confocal fluorescence microscope image of a giant macrophage shows MANF (mesencephalic astrocyte-derived neurotrophic factor) expression in red.

IMAGE: J. NEVES AND P. SOUSA-VICTOR

Published by AAAS

PRIZE ESSAY

FINALIST
Arun Sharma
Arun Sharma received his undergraduate degree from Duke University and a Ph.D. from Stanford University. Having completed a postdoctoral fellowship at the Harvard Medical School, Sharma is now a senior research fellow jointly appointed at the Smidt Heart Institute and Board of Governors Regenerative Medicine Institute at the Cedars-Sinai Medical Center in Los Angeles. His research seeks to develop in vitro platforms for cardiovascular disease modeling and drug cardiotoxicity assessment. **www.sciencemag.org/content/367/6483/1206.1**

REGENERATIVE MEDICINE

Stem cells to help the heart

Cardiac tissue derived from stem cells holds promise for modeling, screening, and therapy

By **Arun Sharma**[1,2]

Shinya Yamanaka's 2006 discovery of induced pluripotent stem cells (iPSCs) ignited a revolution in the field of stem cell biology (*1*). For the first time, nearly all human somatic tissues could be produced from iPSCs reprogrammed from blood or skin cells, in a process that took only weeks. This advance was particularly crucial for obtaining surrogate tissues from cell types that are otherwise difficult to procure and do not readily expand in vitro, such as cardiac or neural cells. Additionally, many ethical concerns are avoided, because this technology uses a patient's own genetic material to create iPSCs rather than relying on embryonic stem cells. In the aftermath of Yamanaka's discovery, entire biomedical industries have developed around the promise of using human iPSCs (hiPSCs) and their derivatives for in vitro disease modeling, drug screening, and cell therapy (*2*).

The hiPSC technology has had a particularly notable impact in cardiac regenerative medicine, a field where scientists and clinicians have been working to devise new methods to better understand how cardiovascular disease manifests and how to restore cardiovascular function after disease strikes (*3*). The heart is limited in its ability to regenerate lost cardiomyocytes (beating heart muscle cells), following an adverse event such as a heart attack (*4*). Cardiomyocytes derived from hiPSCs (hiPSC-CMs) may represent a potential replacement option for dead cells in such a scenario. However, certain issues remain to be addressed, such as whether hiPSC-CMs can integrate with host myocardial tissue in the long term (*5*).

While using hiPSC-CMs for in vivo cell therapy may become practical in the future, employing hiPSC-CMs for high-throughput drug discovery and screening is becoming a reality in the present (*6*). Cardiovascular diseases can be recapitulated "in a dish" with patient-specific hiPSC-CMs. For example, if a patient exhibits a cardiac arrhythmia caused by a genetic abnormality in a sarcomeric protein or ion channel, that same rhythm problem can be recapitulated in vitro (*7*). Thanks to advances in hiPSC differentiation protocols, hiPSC-CMs can now be mass-produced to study cardiovascular disease mechanisms in vitro (*8*).

My graduate thesis in the laboratories of Joseph Wu and Sean Wu at Stanford University focused on in vitro applications of hiPSC-CMs for cardiovascular disease modeling and for high-throughput screening of chemotherapeutic compounds to predict cardiotoxicity. I initially embarked on a project using hiPSC-CMs to model viral myocarditis, a viral infection of the heart, caused by the B3 strain of coxsackievirus (*9*). I began by demonstrating that hiPSC-CMs express the receptors necessary for viral internalization and subsequently found that hiPSC-CMs were highly susceptible to coxsackievirus infection, exhibiting viral cytopathic effect within hours of infection. I also identified compounds that could alleviate coxsackievirus infection on hiPSC-CMs, a translationally relevant finding, as there remains a shortage of treatments for viral myocarditis.

Using a genetically modified variant of coxsackievirus B3 expressing luciferase, I developed a screening platform for assessing the efficacy of antiviral compounds. Pretreatment with interferon-β, ribavirin, or pyrrolidine dithiocarbamate markedly suppressed viral replication on hiPSC-CMs by activating intracellular antiviral response and viral protein clearance pathways. These compounds alleviated viral replication in a dose-dependent fashion at low concentrations without causing cellular toxicity.

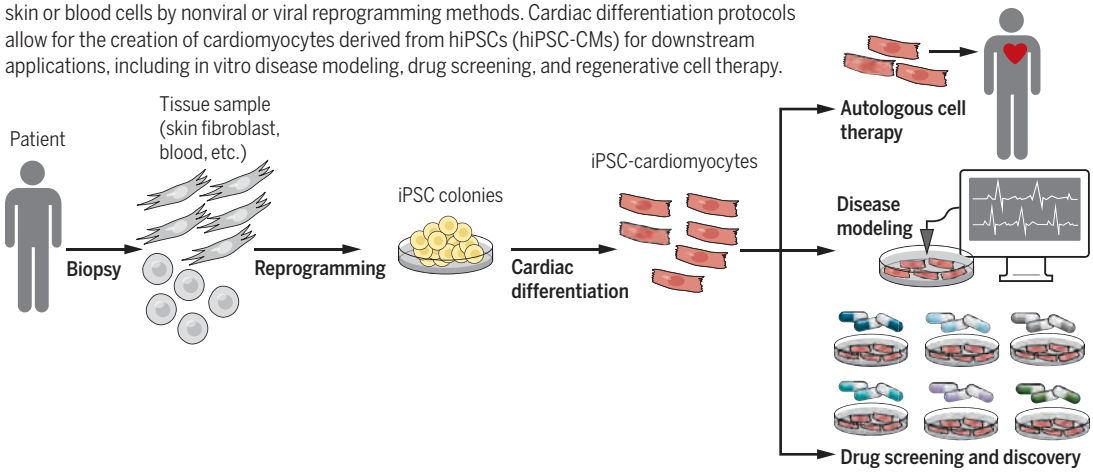
I next sought to use hiPSC-CMs to screen anticancer chemotherapeutic compounds for their off-target cardiovascular toxicities (*10*). Cardiotoxicity represents a major cause of drug withdrawal from the pharmaceutical market, and several chemotherapeutic agents can cause unintended cardiovascular damage (*11*). Using cultured hiPSC-CMs, I evaluated 21 U.S. Food and

[1]Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA. [2]Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA. Email: arun.sharma1490@gmail.com

PHOTO: COURTESY OF A. SHARMA

Applications of hiPSC-CM technology

Initially, human induced pluripotent stem cells (hiPSCs) can be produced by reprogramming skin or blood cells by nonviral or viral reprogramming methods. Cardiac differentiation protocols allow for the creation of cardiomyocytes derived from hiPSCs (hiPSC-CMs) for downstream applications, including in vitro disease modeling, drug screening, and regenerative cell therapy.



GRAPHIC: ADAPTED BY N. CARY/*SCIENCE*

Drug Administration–approved tyrosine kinase inhibitors (TKIs), commonly prescribed anticancer compounds, for their cardiotoxic potential. HiPSC-CMs express the major tyrosine kinase receptor proteins such as the insulin, insulin-like growth factor (IGF), vascular endothelial growth factor (VEGF), and platelet-derived growth factor (PDGF) receptors, lending validity to this cellular model.

With data from a battery of cellular apoptosis, contractility, electrophysiology, and signaling assays, I generated a "cardiac safety index" to help align in vitro toxicity data to clinical drug safety guidelines (*12*). From the safety index, I determined that a subclass of VEGF receptor 2/PDGF receptor–inhibiting tyrosine kinase inhibitors, some of which exhibit toxicity clinically, also elicited cardiotoxicities in hiPSC-CMs. These manifested as substantial alterations in cellular electrophysiology, contractility, and viability when administered at clinically relevant concentrations. I also discovered that cotreatment with either IGF or insulin partially rescued TKI-induced toxicity by up-regulating antiapoptotic signaling pathways. This work could prove useful for groups aiming to develop effective screening platforms to assess new chemotherapeutic compounds for cardiotoxic side effects.

I also collaborated with the Center for the Advancement of Science in Space (CASIS) to send a sample of hiPSC-CMs to the International Space Station. As humankind ventures beyond our home planet, it is imperative that we better understand how the heart functions for long periods of time in microgravity. Analysis of these hiPSC-CMs revealed microgravity-induced alterations in metabolic gene expression and calcium handling (*13*).

In recent years, the stem cell field has experienced an explosion of studies using hiPSC-CMs as a model cellular system to study cardiovascular biology. As improvements in hiPSC-CM mass production continue, we will see a rise in studies using these cells for disease modeling and drug screening. Thus, although hiPSC-CM technology is in its infancy, it holds great potential to improve cardiovascular health. ■

REFERENCES AND NOTES

1. K. Takahashi, S. Yamanaka, *Cell* **126**, 663 (2006).
2. M. Rao, *Cell Stem Cell* **12**, 149 (2013).
3. A. S. Go *et al.*, *Circulation* **127**, e6 (2013).
4. M. A. Laflamme, C. E. Murry, *Nature* **473**, 326 (2011).
5. J. J. Chong *et al.*, *Nature* **510**, 273 (2014).
6. A. Sharma, J. C. Wu, S. M. Wu, *Stem Cell Res. Ther.* **4**, 150 (2013).
7. I. Itzhaki *et al.*, *Nature* **471**, 225 (2011).
8. P. W. Burridge *et al.*, *Nat. Methods* **11**, 855 (2014).
9. A. Sharma *et al.*, *Circ. Res.* **115**, 556 (2014).
10. A. Sharma *et al.*, *Sci. Transl. Med.* **9**, eaaf2584 (2017).
11. J. D. Groarke, S. Cheng, J. Moslehi, *N. Engl. J. Med.* **369**, 1779 (2013).
12. A. Sharma *et al.*, *Nat. Protoc.* **13**, 3018 (2018).
13. A. Wnorowski *et al.*, *Stem Cell Reports* **13**, 960 (2019).

10.1126/science.aba6111

PRIZE ESSAY

FINALIST
Adam C. Wilkinson
Adam C. Wilkinson received his undergraduate degree from the University of Oxford and a Ph.D. from the University of Cambridge. He is currently completing his postdoctoral fellowship at the Institute for Stem Cell Biology and Regenerative Medicine at Stanford University, where he is studying normal and malignant hematopoietic stem cell biology with the aim of identifying new biological mechanisms underlying hematological diseases and improving the diagnosis and treatment of these disorders. **www.sciencemag.org/content/367/6483/1206.2**

REGENERATIVE MEDICINE

Hope for hematological diseases

Tweaking growth media helps cultured hematopoietic stem cells thrive

By **Adam C. Wilkinson**

The development of methodologies for growing cells ex vivo has been essential for advancing the field of regenerative medicine. The in vitro stabilization of embryonic (pluripotent) stem cells, for example, has afforded unprecedented biological insights, breakthrough technologies, and new therapeutic paradigms (*1*). Determining the equivalent culture conditions necessary for growing adult tissue stem cells holds a similar scientific and clinical potential, but this goal remains largely unmet.

During my scientific career studying hematopoiesis (blood formation), my research efforts have often been frustrated by the lack of suitable culture conditions for the stable growth of hematopoietic stem cells (HSCs). Because HSCs are a rare bone marrow cell type, this has been a major barrier to understanding the mechanism of blood formation and to developing new therapies for hematological diseases.

The Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy

A PIONEERING CELL THERAPY

Stable, ex vivo HSC expansion has the potential to revolutionize current clinical therapies. With the capacity for both self-renewal and multipotent differentiation, HSCs can reconstitute the entire blood system after transplantation and engraftment within a recipient (*2*). This remarkable activity provides the scientific basis for clinical HSC transplantation (HSCT), the pioneering cell therapy that for more than half a century has offered a cure to so many patients suffering from hematological diseases (*3*).

Although HSCT is potentially curative, it remains a risky therapy. Two of the major roadblocks to its wider and safer use are the availability of immune-compatible allogeneic donor HSCs—necessary to avoid severe graft-versus-host disease—and the need for patients to undergo genotoxic chemotherapy and/or radiation bone marrow pre-conditioning, which can cause substantial morbidity and mortality and limit patient access.

Institute for Stem Cell Biology and Regenerative Medicine and Department of Genetics, Stanford University School of Medicine, Lorry I. Lokey Stem Cell Research Building, 265 Campus Drive, Stanford, CA, USA.
Email: adamcw@stanford.edu

REIMAGINING THE HSCT THERAPEUTIC PARADIGM

The ideal HSCT therapeutic paradigm (at least for nonmalignant hematological diseases) would be the transplantation of healthy autologous patient-derived HSCs without preconditioning. CRISPR/Cas9 genome editing technologies have opened the door for efficient autologous HSC gene correction for numerous hematological diseases (*4*). Unfortunately, however, the clinical success of these new HSCT strategies (as well as existing lentivirus- or retrovirus-based HSCT gene therapies) remains challenging because of the limited window for ex vivo perturbation (usually <72 hours) and the often inadequate numbers of HSCs collectable from patients. Additionally, genotoxic preconditioning is still required to improve the chances of engraftment.

Nongenotoxic preconditioning, such as metabolic and antibody-based strategies (*5*–*7*), has been suggested for HSCT, but achieving HSC engraftment without any preconditioning would be ideal. Such a therapeutic strategy would be possible if much larger numbers of HSCs were available for HSCT because HSCs do engraft in the nonconditioned setting, but only at low efficiencies (*8*).

The existence of substantial scientific and clinical potential of ex vivo HSC expansion has stimulated an extensive search for supportive culture conditions (*9*). However, current ex vivo conditions only sustain human HSCs short term (1 to 2 weeks), during which expansion of functional HSCs is limited. Rather than displaying potent self-renewal ex vivo, HSCs rapidly differentiate and lose their "stemness." This has led to the assumption that we are still missing essential self-renewal–promoting factors. Alternatively, we proposed that HSC media contained additional contaminants that skewed HSCs to differentiation (*10*).

PHOTO: JOYDEEP BHADURY

STEPS TOWARD STABLE EX VIVO HSC EXPANSION

To test this hypothesis, I took a reductionist approach to optimize media conditions that we had previously shown could maintain mouse HSCs for 7 days and supported approximately twofold functional expansion (*11*). The results were dramatic. Using our new "cleaned-up" culture condition, we were able to grow HSCs for more than 2 months ex vivo and achieve ~900-fold expansion of functional HSCs (*10*).

The biggest factor in establishing optimal ex vivo HSC expansion conditions was the removal of serum albumin. As a major blood component, serum albumin has long been considered an essential supplement for hematopoietic cell cultures. However, the use of serum or serum albumin introduces numerous biologically active impurities (*11*). In our HSC cultures, even recombinant serum albumin contained contaminants that HSCs were highly sensitive to and that inhibited HSC self-renewal.

I discovered that serum albumin can be entirely replaced with the synthetic polymer polyvinyl alcohol within our HSC media. The use of this biologically inert albumin replacement to remove these HSC differentiation-inducing contaminants dramatically improved HSC expansion (*10*).

This substitution—combined with the right levels of stem cell factor, thrombopoietin, and fibronectin—enabled us to grow bulk populations of HSCs in long-term cultures and to expand single HSCs clonally. Using clone-splitting assays, we were even able to definitively prove ex vivo HSC self-renewal (*10*).

This new HSC culture system also provided an important clinical insight. When large numbers of ex vivo–expanded HSCs (generated from just 50 freshly isolated HSCs) were transplanted into nonconditioned recipient mice, durable engraftment was achieved. We are now working to demonstrate the efficacy of this intervention in combination with CRISPR/Cas9–based autologous HSC gene correction.

THE NEXT STEPS

Our mouse HSC expansion system offers a valuable technology with which to uncover new biological mechanisms and to develop therapeutic approaches for regenerative medicine and cell therapy. Additionally, our recent progress in expanding mouse HSCs has also provided an important roadmap with which to achieve stable ex vivo human HSC expansion, a route that will, I hope, ultimately lead to safer clinical HSCT therapies. ■

REFERENCES AND NOTES

1. M. Evans, *Nat. Rev. Mol. Cell Biol.* **12**, 680 (2011).
2. C. J. Eaves, *Blood* **125**, 2605 (2015).
3. C. Chabannon *et al.*, *Sci. Transl. Med.* **10**, eaap9630 (2018).
4. D. P. Dever, M. H. Porteus, *Curr. Opin. Hematol.* **24**, 481 (2017).
5. Y. Taya *et al.*, *Science* **354**, 1152 (2016).
6. A. C. Wilkinson, M. Morita, H. Nakauchi, S. Yamazaki, *Exp. Hematol.* **63**, 12 (2018).
7. A. Czechowicz, D. Kraft, I. L. Weissman, D. Bhattacharya, *Science* **318**, 1296 (2007).
8. D. Bhattacharya, D. J. Rossi, D. Bryder, I. L. Weissman, *J. Exp. Med.* **203**, 73 (2006).
9. S. Kumar, H. Geiger, *Trends Mol. Med.* **23**, 799 (2017).
10. A. C. Wilkinson *et al.*, *Nature* **571**, 117 (2019).
11. A. Ieyasu *et al.*, *Stem Cell Reports* **8**, 500 (2017).

ACKNOWLEDGMENTS

I thank my supportive mentors and collaborators, in particular H. Nakauchi, S. Yamazaki, K. Loh, D. Dever, and R. Ishida. I also thank my research funders, The Leukemia and Lymphoma Society, Bloodwise, and the Japan Society for the Promotion of Science.

10.1126/science.aba6108

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

Speleothems, or cave formations, in the Corchia cave system in Italy have provided precise records of the timing of deglaciations over the past million years.

PHOTO: FRANCESCO TOMASINELLI/SCIENCE SOURCE; GRAPHIC: WRAPP *ET AL.*

GLACIAL CYCLES

An underground record of past deglaciations

Understanding more exactly how the timing of deglaciations depends on changes in insolation, or the energy received by Earth from the Sun, requires precise and independent records of both environmental change and solar energy input. Bajo et al. strengthened the weak link of that two-member chain, the environmental record, by developing a precise, radiometrically dated chronology of the 11 deglaciations of the past million years derived from speleothems. This allowed them to show more clearly how the initiation and duration of glacial terminations over that period depended on solar obliquity and precession. —HJS

Science, this issue p. 1235

FRUSTRATED MAGNETISM

Looking for a crystalline 2D spin ice

Spin ices—materials in which local magnetic spins respect "ice rules" similar to those in water ice—are typically three-dimensional. Two-dimensional (2D) ice rules can also be formulated and have been found to be satisfied in engineered nanomagnetic systems, usually referred to as artificial spin ices. Zhao *et al*. used neutron scattering and thermodynamic measurements to study a crystalline candidate for a 2D spin ice, the intermetallic compound HoAgGe. They found that at low temperatures, the local spins on the distorted kagome planes of this quasi-2D material respect 2D ice rules. Increasing the temperature led to a series of transitions consistent with theoretical expectations. —JS

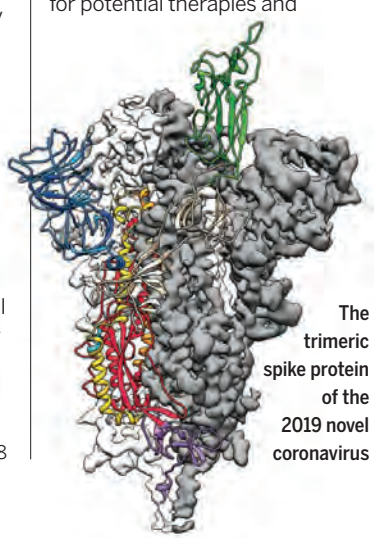
Science, this issue p. 1218

CORONAVIRUS

Structure of the nCoV trimeric spike

The World Health Organization has declared the outbreak of a novel coronavirus (2019-nCoV) to be a public health emergency of international concern. The virus binds to host cells through its trimeric spike glycoprotein, making this protein a key target for potential therapies and diagnostics. Wrapp *et al*. determined a 3.5-angstrom-resolution structure of the 2019-nCoV trimeric spike protein by cryo–electron microscopy. Using biophysical assays, the authors show that this protein binds at least 10 times more tightly than the corresponding spike protein of severe acute respiratory syndrome (SARS)–CoV to their common host cell receptor. They also tested three antibodies known to bind to the SARS-CoV spike protein but did not detect binding to the 2019-nCoV spike protein. These studies provide valuable information to guide the development of medical countermeasures for 2019-nCoV. —VV

Science, this issue p. 1260



The trimeric spike protein of the 2019 novel coronavirus

ORGANIC CHEMISTRY

Asymmetry on the plus side

Numerous positively charged metal catalysts have been paired with chiral negative ions to select for just one of two mirror-image products. Genov *et al*. now report a potentially general strategy to invert the charges in this paradigm. Because intrinsically negative metal catalysts are comparatively rare, the authors appended a sulfonate group to the common bipyridyl ligand. Iridium complexes of this ligand paired with chiral positive ions could borylate just one of two aryl rings appended to carbon or phosphorus centers with high enantioselectivity. —JSY

Science, this issue p. 1246

FUNCTIONAL AMYLOIDS

How amyloid can be a substrate of memory

Formation of memories requires changes in the molecular composition of the synapse. How these changes occur and what maintains this altered synaptic composition so

Published by AAAS

that memory can persist are unknown. Hervas *et al*. report the structure of a synaptic translation regulator called Orb2 isolated from the brains of adult fruit flies that is important for the maintenance and recall of memory. Orb2 forms an amyloid and changes its activity from a translation repressor to an activator. The amyloid core is composed of polar hydrophilic residues, as opposed to the hydrophobic ones found in nonfunctional or pathological amyloids. The structure provides insights into how amyloids could be a stable yet malleable substrate of memory. —SMH

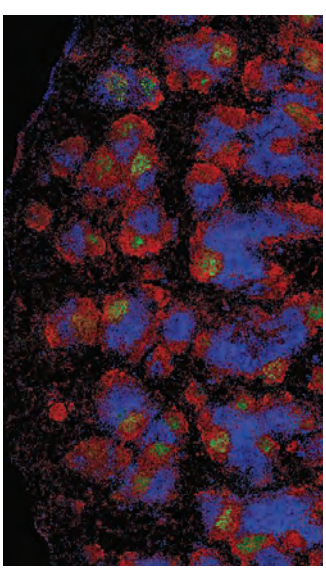
Science, this issue p. 1230

T CELLS

Revisiting memory

Certain T cell subsets express a receptor that makes them susceptible to nicotinamide adenine dinucleotide (NAD)–induced cell death (NICD), which can occur during isolation from tissues. This susceptibility has complicated our understanding of what cells are present and active both during and after the acute response. Künzli *et al*. used an NICD blocker to study the persistence of T follicular helper (T_{FH}) cells in mice after infection with a virus. They report that T_{FH} cells persisted for more than 400 days after infection and that long-lived T_{FH} cells are glycolytic and marked by high expression of folate receptor 4. Upon reinfection, these "memory" T_{FH} cells were capable of self-renewal and could also give rise to effector and central memory cells. —AB

Sci. Immunol. **5**, eaay5552 (2020).



Fluorescence microscopy image of mouse spleen stained to reveal the identities of immune cell types

TISSUE ENGINEERING

Strategic lumbar support

Diskectomy is a common treatment for herniated or slipped intervertebral disks that can help to alleviate symptoms but does not prevent reherniation or progression of disk degeneration. Sloan *et al*. developed a two-part, acellular tissue-engineered therapy to prevent degeneration after diskectomy. Injecting hyaluronic acid into the inner region of the disk and applying a photo–cross-linked collagen patch to the outer ring of fibrous tissue healed disk defects and maintained biomechanical support in the lumbar spines of sheep for 6 weeks after diskectomy. —CC

Sci. Transl. Med. **12**, eaay2380 (2020).

SPECTROSCOPY

Reading a molecule without destroying it

Achieving efficient quantum control of ultracold molecular systems may open opportunities in molecular precision spectroscopy, quantum information, and related fields. Sinhal *et al*. report a quantum-nondemolition protocol for the detection of the spin-rovibronic state of a single trapped cold molecular ion co-trapped with an atomic ion. They show that monitoring the motion of Ca^+ after coherent motional excitation of the Ca^+-N_2^+ string makes it possible to detect the N_2^+ state without destroying either the molecule or the state itself. The procedure can be repeated multiple times while preserving the high readout fidelity. —YS

Science, this issue p. 1213

IN OTHER JOURNALS

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

Experiments with subtropical forest tree species in China show that tree diversity benefits drought-sensitive species.

GENETIC DISEASE

Contracting disease-causing repeat expansions

Ongoing CAG/CTG expansions in the gene encoding huntingtin in the brains of Huntington's disease (HD) patients result in pathological accumulations of protein aggregates. It is possible that targeting these somatic expansions could be therapeutically valuable. Nakamori *et al*. investigated these genetic instabilities in a highly specific way by using a small molecule called naphthyridine-azaquinolone (NA). NA binds selectively to the unusual structures formed by the expanded DNA in the gene encoding huntingtin. NA injections into the striatum of a HD mouse model induced contractions of the expanded repeat and reduced levels of the mutant protein aggregates, with no effects genome-wide. Thus, targeting the root cause of expanded-repeat diseases is possible and could be a valuable strategy for tackling many similar diseases. —SMH

Nat. Genet. **52**, 146 (2020).

AUTOIMMUNITY

Cells gone rogue

Autoantibodies are proteins produced by the immune system that attack a person's own tissues and organs, leading to autoimmune disease. Autoantibodies can be present in the serum years before the clinical onset of autoimmunity, but it is not understood how they cause disease. Singh *et al*. used multi-omics single-cell technology to trace the evolution of "rogue" cell clones responsible for producing pathogenic autoantibodies in the blood of patients with the autoimmune disease cryoglobulinemic vasculitis. The researchers found that a benign antibody can transform into one that causes inflammation of blood vessels in the skin, kidney, nerves, and joints. The gene mutations that accumulate in the rogue cells during the early stages of autoimmune disease have also been identified in cancer cells from patients with lymphoma. —PNK

Cell **180**, 878 (2020).

CREDITS (FROM LEFT): LUDIVINE C. LITZLER; EVAN BOWEN-JONES/ALAMY STOCK PHOTO

FOREST ECOLOGY

Tree diversity relieves drought impacts

The relentless progress of current climate warming creates concerns about the impacts of drought on forest ecosystems. Fichtner *et al.* show that drought impacts on trees can be reduced when the species diversity of trees is higher. In experimental plots in a subtropical forest in China, they found that the growth of saplings of drought-sensitive tree species over a 6-year period resulted in more species-diverse neighborhoods. This effect might result from more effective partitioning of water resources where interspecific diversity is higher because of a greater variety of rooting habits and strategies. If these effects apply to forests more generally, then productivity and carbon sequestration may improve under conditions of higher local tree diversity. —AMS

J. Ecol. 10.1111/1365-2745.13353 (2020).

POLYMER SCIENCE

Tying together common plastics

Two challenges to overcome when recycling postconsumer waste are the need to sort and separate mixed waste streams and the presence of multilayer packages such as food wrappers that cannot be separated into individual plastic components. Nomura *et al.* synthesized copolymers of poly(ethylene terephthalate) (PET) and polyethylene (PE), dissimilar polymers that do not readily mix or have common solvents but which represent two of the dominant materials found in high-volume waste streams. They found that these PET-PE multiblock chains can act as an adhesive to tie together the two polymers, such that a 0.5% addition to a PET:PE (80:20) mixture generates a blended polymer with strain and stress at a break that exceeds that of the pure PET. —MSL

ACS Appl. Mater. Interfaces **12**, 9726 (2020).

AIR POLLUTION

Catalyzing heavy haze

Intense regional haze events periodically affect many heavily polluted urban areas, causing adverse effects on human health and possibly affecting local climate. The reasons that these episodes occur are not entirely clear. Zhang *et al.* present a combination of field measurements, laboratory experiments, and model simulations to show that black carbon particles are central components of haze development. They demonstrate that the frequency of heavy haze events declines significantly with the reduction of sulfur dioxide and that sulfur dioxide oxidation is efficiently catalyzed on black carbon particles in the presence of nitrogen dioxide and ammonia. This unexpected finding implies that reduction of sulfur dioxide alone will not reduce the frequency of heavy-haze events in these environments. —HJS

Proc. Natl. Acad. Sci. U.S.A. **117**, 3960 (2020).

NANOMATERIALS

An atomic view of dealloying

The dealloying of metals to create nanoporous materials with bicontinuous surfaces is often modeled as a surface dissolution and diffusion process, but many of the details lack experimental verification. Liu *et al.* used liquid-cell high-resolution scanning transmission electron microscopy to follow the nitric acid–driven dissolution of silver from gold-silver alloy ($Au_{0.2}Ag_{0.8}$) nanospheres, nanocubes, and fivefold twin nanorods. Whereas the nanoparticles had more surface defects than the nanorods and began to dissolve immediately, the nanorods had an induction period in which surface dealloying created defects that then allowed bulk silver to dealloy. Dealloying led to large decreases in particle volumes that were attributed to the inward movement of exterior atoms, creating a denser, less porous outer shell. —PDS

Nano Lett. 10.1021/acs.nanolett.9b05216 (2020).

PLANT SCIENCE

Self-organizing floral pigmentation patterns

Petals of *Mimulus* flowers are decorated with anthocyanin spots in patterns that cue pollinators. Ding *et al.* show how, following the rules of a reaction-diffusion system, a self-activating transcription factor called NECTAR GUIDE ANTHOCYANIN (NEGAN) interacts with a mobile repressor to organize pigment patterns on flower petals. Among the genes regulated by NEGAN is *RED TONGUE* (*RTO*), which encodes a transcriptional repressor. *RTO* is transcribed in anthocyanin spot cells, but the RTO protein moves to adjacent cells, where it limits anthocyanin production. Model simulations explored how NEGAN and RTO together construct a variety of pigmentation patterns on the flower petals, which in turn affect bumble bee visitations. —PJH

Curr. Biol. 10.1016/j.cub.2019.12.067 (2020).

Visibility at the Forbidden City in Beijing during, from left to right, clean or light-haze, moderate-haze, and heavy-haze periods

PHOTOS: ZHANG *ET AL.*, *PROC. NATL. ACAD. SCI. U.S.A.* **117**, 3960 (2020)

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

MEDICINE

New therapies for sickle cell disease

Sickle cell disease can be treated with matched bone marrow transplants from family members, but this treatment is not available to all patients. Sickle cell disease results from a hemoglobin subunit gene mutation and can be overcome by expression of different hemoglobin subunits. Recent developments have highlighted the possibility of gene therapy and engineered cell therapy to replace mutated bone marrow cells. In a Perspective, Tisdale *et al*. discuss the developments in anti-sickling drugs and gene and cell therapies and what is needed to treat patients effectively, including those in low- and middle-income countries. —GKA

Science, this issue p. 1198

CELL BIOLOGY

Phase separation can be skin deep

The skin's barrier arises from proliferative cells that generate a perpetual upward flux of terminally differentiating epidermal cells. Cells nearing the body surface suddenly lose their organelles, becoming dead cellular ghosts called squames. Working in mouse tissue, Garcia Quiroz *et al*. found that as differentiation-specific proteins accumulate in the keratinocytes, they undergo a vinegar-in-oil type of phase separation that crowds the cytoplasm with increasingly viscous protein droplets (see the Perspective by Rai and Pelkmans). Upon approaching the acidic skin surface, the environmentally sensitive liquid-like droplets respond and dissipate, driving squame formation. These dynamics come into play in human skin barrier diseases, where mutations cause maladapted liquid-phase transitions. —SMH

Science, this issue p. 1210; see also p. 1193

NEUROSCIENCE

Spreading edema after stroke

The brain is enveloped in a cushion of cerebrospinal fluid (CSF), which normally provides protection and helps to remove metabolic waste. CSF transport has also recently been shown to play unexpected roles in neurodegeneration and sleep. Mestre *et al*. used multimodal in vivo imaging in rodents and found that, after a stroke, an abnormally large volume of CSF rushes into the brain, causing swelling (see the Perspective by Moss and Williams). This influx of CSF is caused by constrictions of arteries triggered by a well-known propagating chemical reaction-diffusion wave called spreading depolarization. CSF transport can thus play a role in brain swelling after stroke. —SMH

Science, this issue p. 1211; see also p. 1195

COMETARY SCIENCE

Ammonium salts on comet 67P

The distribution of carbon and nitrogen in the Solar System is thought to reflect the stability of carbon- and nitrogen-bearing molecules when exposed to the heat of the forming Sun. Comets have a low nitrogen-to-carbon ratio, which is contrary to expectations because they originate in the outer Solar System where nitrogen species should be common. Poch *et al*. used laboratory experiments to simulate cometary surfaces and compared the resulting spectra with comet 67P/Churyumov-Gerasimenko. They assigned a previously unidentified infrared absorption band to nitrogen-containing ammonium salts. The salts could contain enough nitrogen to bring the comet's nitrogen-to-carbon ratio in line with the Sun's. —KTS

Science, this issue p. 1212

WATER RESOURCES

Evaporating futures

Drought and warming have been shrinking Colorado River flow for many years. Milly and Dunne used a hydrologic model and historical observations to show that this decrease is due mainly to increased evapotranspiration caused by a reduction of albedo from snow loss and the associated rise in the absorption of solar radiation (see the Perspective by Hobbins and Barsugli). This drying will be greater than the projected precipitation increases expected from climate warming, increasing the risk of severe water shortages in an already vulnerable region. —HJS

Science, this issue p. 1252; see also p. 1192

STRUCTURAL BIOLOGY

Snapshots of a rotary pump

Vesicular- or vacuolar-type adenosine triphosphatases (V-ATPases) are ATP-hydrolysis–driven proton pumps. In neurons, V-ATPase activity generates a proton gradient across the membrane of synaptic vesicles so that neurotransmitters can be loaded into the vesicles. Abbas *et al*. developed a method to purify V-ATPase from rat brain and determined the structure of the entire complex by cryo–electron microscopy. Native mass spectrometry showed that the preparation was homogeneous and complemented structural studies by confirming the subunit composition. Three rotational states were resolved at better than 4-angstrom resolution, providing insight into the conformational changes that couple ATP hydrolysis to proton pumping. —VV

Science, this issue p. 1240

CANCER

Cancer drivers converge on NOTCH

Cancer genome–sequencing projects have emphasized the handful of genes mutated at high frequency in patients. Less attention has been directed to the hundreds of genes mutated in only a few patients—the so-called "long tail" mutations. Although rare, these mutations may nonetheless inform patient care. Loganathan *et al*. developed a reverse genetic CRISPR screen that allowed them to functionally assess in mice nearly 500 long tail gene mutations that occur in human head and neck squamous cell carcinoma (HNSCC). They identified 15 tumor-suppressor genes with activities that converged on the NOTCH signaling pathway. Given that NOTCH itself is mutated at high frequency in HNSCC, these results suggest that the growth of these tumors is largely driven by NOTCH inactivation. —PAK

Science, this issue p. 1264

IMMUNOLOGY

Deadenylate or activate?

When cells are quiescent, they undergo reversible cell cycle arrest and evince low basal metabolism. Naïve T cells are normally quiescent until they recognize cognate antigens through T cell receptor–costimulatory molecule signaling. T cell quiescence appears to be an active process, but the mechanistic details are poorly understood. Hwang *et al*. report that the transcription factors BTG1 and BTG2 are selectively expressed in quiescent T cells. In mice, T cells conditionally knocked out for both factors showed enhanced proliferation and a lowered threshold of activation both in vitro and in response to *Listeria monocytogenes* infection. Deficiency of BTG1 and BTG2 resulted in increases in global messenger RNA half-life, suggesting that

messenger RNA deadenylation and degradation are important processes for maintaining T cell quiescence. —STS

Science, this issue p. 1255

CANCER

Inducing mortality by targeting mortalin

Oncogenic mutations in the kinase BRAF can enhance cell survival, even in cells exposed to targeted inhibitors. Wu *et al*. found that a heat shock protein family A (HSP70)–related chaperone protein called mortalin disrupted a protein-protein interaction that would ordinarily trigger cell death in BRAF mutant cells. Depleting mortalin or treating cells with either HSP70 derivatives or synthetic decoy peptides that mimic mortalin target sites increased the efficacy of the BRAF inhibitor vemurafenib in cells and in tumor-bearing mice. —LKF

Sci. Signal. **13**, eaay1478 (2020).

STRUCTURAL BIOLOGY

Clustering for the kill

Cluster of differentiation 20 (CD20) is a membrane protein that defines most B cell populations and is the target of therapeutic antibodies to treat malignancies and autoimmune disorders. Rougé *et al*. present the structure of CD20 bound to the antibody rituximab that activates the complement system to kill B cells. CD20 forms a dimer and each monomer binds one rituximab antigen-binding fragment (Fab) to give 2:2 stoichiometry. The compact packing between Fab arms and CD20 gives rise to circular assemblies with a diameter similar to that of antibody hexamers known to recruit the first component of the complement cascade. —VV

Science, this issue p. 1224

ECOLOGY

What becomes of fragmented forests?

Understanding how tropical forests are changing in structure and extent over time is essential to protecting these habitats. Hansen *et al.* incorporated fragmentation metrics with spatially explicit forest loss data to paint a complete picture for local communities to use in forest fragmentation monitoring. This analysis reveals higher rates of loss in highly fragmented forests. The authors suggest prioritizing conservation efforts on preventing further fragmentation of large forested areas and restoring connectivity in regions with highly fragmented forests. —SN

Sci. Adv. 10.1126/sciadv.aax8574 (2020).

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

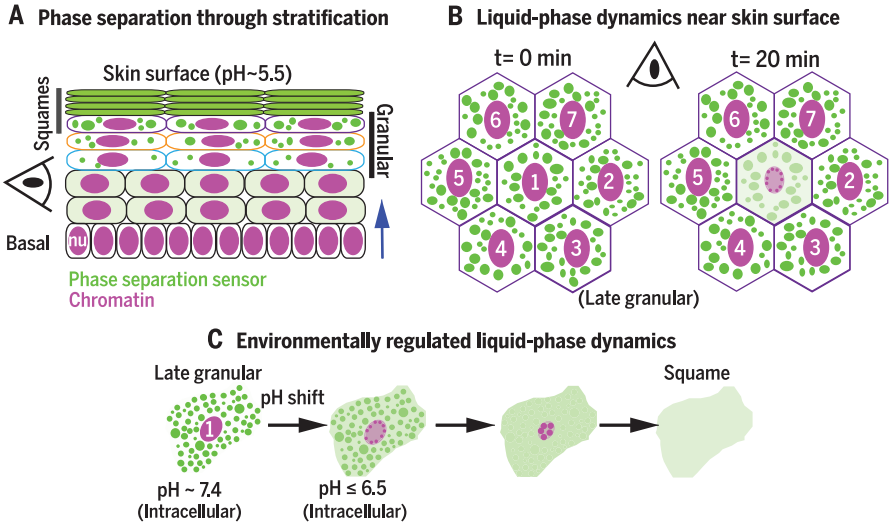
Liquid-liquid phase separation drives skin barrier formation

Felipe Garcia Quiroz, Vincent F. Fiore, John Levorse, Lisa Polak, Ellen Wong, H. Amalia Pasolli, Elaine Fuchs*

INTRODUCTION: Liquid-liquid phase separation of biopolymers has emerged as a driving force for assembling membraneless biomolecular condensates. Despite substantial progress, studying cellular phase separation remains challenging. We became intrigued by enigmatic, membraneless protein granules (keratohyalin granules, KGs) within the terminally differentiating cell layers of mammalian epidermis. As basal progenitors cease to proliferate and begin their upward journey toward the skin surface, they produce differentiation-specific proteins that accumulate within KGs. Upon approaching the surface layers, all cellular organelles and KGs are inexplicably lost, resulting in flattened, dead cellular ghosts (squames) that seal the skin as a tight barrier to the environment.

RATIONALE: In an unbiased proteome-wide in silico search for candidate phase-transition proteins, we previously identified a major KG constituent, filaggrin (FLG), whose truncating mutations are intriguingly linked to human skin barrier disorders. Using advanced tools to study phase-separation behavior in mammalian skin, we pursued the possibility that liquid-liquid phase separation might lie at the root of both epidermal differentiation and human disease.

RESULTS: We found that KGs are liquid-like condensates, which assemble as filaggrin proteins undergo liquid-liquid phase separation in the cytoplasm of epidermal keratinocytes. Disease-associated FLG mutations specifically perturbed or abolished the critical concentration for phase separation–driven assembly of KGs. By developing sensitive, innocuous phase-separation sensors that enable visualization and interrogation of endogenous liquid-liquid phase-separation processes in mice, we found that filaggrin's phase-separation dynamics crowd the cytoplasm with increasingly viscous KGs that physically affect organelle integrity. Liquid-like coalescence of KGs was restricted by surrounding bundles of differentiation-specific keratin filaments. Probing deeper, we found that as epidermal cells approached the acidic skin surface, phase-transition proteins experienced a rapid, naturally occurring pH shift and dynamically responded, causing the dissipation of their liquid-like KGs to drive squame formation.



Environmentally regulated liquid-phase dynamics drive skin barrier formation. (**A**) Using phase-separation sensors, we show that as basal progenitors flux toward the skin surface, they display phase-separation–driven assembly of liquid-like droplets. (**B**) In late-granular cells, these droplets crowd the cytoplasm and dissolve as cells (1) undergo chromatin compaction. (**C**) Near the skin surface, a sudden shift in intracellular pH regulates liquid-phase dynamics to drive squame formation.

CONCLUSION: Through the biophysical lens of liquid-liquid phase separation, our findings shed fresh light on the enigmatic process of skin barrier formation. Our design and deployment of phase-separation sensors in skin suggest a general strategy to interrogate endogenous liquid-liquid phase separation dynamics across biological systems in a nondisruptive manner.

Through engineering filaggrins, filaggrin disease–associated variants, and our phase-separation sensors, we unveiled KGs as abundant, liquid-like membraneless organelles. During terminal differentiation, filaggrin family proteins first fuel phase-separation–driven KG assembly and subsequently, KG disassembly. Their liquid-like and pH-sensitive properties ideally equip KGs to sense and respond to the natural environmental gradients that occur at the skin's surface and to drive the adaptive process of barrier formation.

Liquid-phase condensates have typically been viewed as reaction centers where select components (clients) become enriched for processing or storage within cells. Analogously, KGs may store clients, possibly proteolytic enzymes and nucleases, that are temporally released in a pH-dependent manner to contribute to the self-destructive phase of terminal differentiation. Additionally, however, we provide evidence for biophysical dynamics emerging from condensate assembly, as KGs interspersed by keratin filament bundles massively crowd the keratinocyte cytoplasm and physically distort adjacent organelles. This crowding precedes the ensuing environmental stimuli that trigger disassembly of KGs, enucleation, and possibly other cellular events linked to barrier formation. Overall, the dynamics of liquid-like KGs, actionable by the skin's varied environmental exposures, expose the epidermis as a tissue driven by phase separation.

Finally, we discovered that filaggrin-truncating mutations and loss of KGs are rooted in maladapted phase-separation dynamics, illuminating why associated skin barrier disorders are exacerbated by environmental extremes. These insights open the potential for targeting phase behavior to therapeutically treat disorders of the skin's barrier. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax9554

The list of author affiliations is available in the full article online.
*Corresponding author. Email: fuchslb@rockefeller.edu
Cite this article as F. G. Quiroz *et al.*, *Science* **367**, eaax9554 (2020). DOI: 10.1126/science.aax9554

RESEARCH ARTICLE

CELL BIOLOGY

Liquid-liquid phase separation drives skin barrier formation

Felipe Garcia Quiroz[1]*, Vincent F. Fiore[1], John Levorse[1], Lisa Polak[1], Ellen Wong[1], H. Amalia Pasolli[2], Elaine Fuchs[1]†

At the body surface, skin's stratified squamous epithelium is challenged by environmental extremes. The surface of the skin is composed of enucleated, flattened surface squames. They derive from underlying, transcriptionally active keratinocytes that display filaggrin-containing keratohyalin granules (KGs) whose function is unclear. Here, we found that filaggrin assembles KGs through liquid-liquid phase separation. The dynamics of phase separation governed terminal differentiation and were disrupted by human skin barrier disease–associated mutations. We used fluorescent sensors to investigate endogenous phase behavior in mice. Phase transitions during epidermal stratification crowded cellular spaces with liquid-like KGs whose coalescence was restricted by keratin filament bundles. We imaged cells as they neared the skin surface and found that environmentally regulated KG phase dynamics drive squame formation. Thus, epidermal structure and function are driven by phase-separation dynamics.

Liquid-liquid phase separation of biopolymers has emerged as a major driving force for assembling membraneless biomolecular condensates (*1*–*3*). Such condensates include nucleoli (*4*), receptor signaling complexes (*3*, *5*), germline granules (*1*, *6*), and stress granules (*7*). This focus on phase separation has also revealed unexpected insights into a range of biological processes, including genomic organization (*8*–*10*), RNA processing (*11*, *12*), mitosis (*13*, *14*), cell adhesion (*15*), and carbon dioxide fixation in plants (*16*). However, the study of cellular phase separation remains challenging (*17*, *18*), often relying upon truncated protein mutants, reconstituted systems in nonphysiological buffers, and overexpression or knockin of tagged fusions (*17*, *19*) that can alter a protein's phase-separation behavior.

In mammalian epidermis, a self-renewing inner (basal) layer of progenitors fuels an upward flux of nondividing keratinocytes that stratify to form the skin's surface barrier that excludes pathogens and retains body fluids (Fig. 1A) (*20*). In early spinous layers, terminally differentiating cells acquire an abundant network of K1- and K10-containing keratin filament bundles. When keratinocytes enter the granular layers, they acquire membraneless protein deposits ("keratohyalin granules," KGs) of enigmatic function (*21*). As these cells approach the surface layers, global transcription suddenly ceases and both KGs and organelles are lost, giving rise to layers of flattened, dead cellular ghosts (squames) that seal the skin as a tight barrier to the environment.

Our prior proteome-wide in silico search for candidate phase-transition proteins identified a major constituent of KGs, filaggrin (FLG) (*22*), whose truncating mutations are intriguingly linked to human skin barrier disorders (*23*) (Fig. 1B and fig. S1). Here, we asked whether liquid-liquid phase separation might lie at the root of both mammalian epidermal differentiation and human disease.

Phase-separation behavior of filaggrin and its paralogs in normal and disease states

Filaggrin and its less-studied (often less-abundant) paralogs are intrinsically disordered repeat proteins with a low-complexity (LC) sequence. Though their sequences are poorly conserved (*24*) (*25*, *26*), mouse and human filaggrin and their paralogs share similar repeat architecture, LC biases, and localization in the cell within KG-like structures (Fig. 1, B and C; figs. S2 to S4; and table S1).

Like many proteins that drive phase separation, filaggrin family proteins across species exhibit a marked bias for arginine (over similarly charged lysine) to engage in aromatic-type interactions (*22*) (Fig. 1D and fig. S3). They differ in that their only prominent aromatic residue is histidine, rather than tyrosine or phenylalanine (fig. S2). Our prior work showed that histidine-rich, intrinsically disordered proteins (IDPs) must be large to display phase-separation behaviors (*22*). Notably, both human (~435 to 504 kDa) and mouse filaggrin are among the largest proteins across these proteomes (Fig. 1E). Humans whose filaggrin variants have the greatest repeat numbers exhibit reduced susceptibility to skin inflammation and allergy (*23*).

To directly examine filaggrin and its disease-associated variants for phase-separation behavior, we first engineered expression vectors driving 1 to 16 human filaggrin repeats (humans have up to 12), each tagged with a fluorescent protein [superfolder green fluorescent protein (sfGFP) or monomeric red fluorescent protein (mRFP)] with or without the non–repeat domains (fig. S5 and table S2). When transfected into immortalized human keratinocytes (HaCATs) under conditions in which filaggrin was not expressed, a single FLG repeat displayed only diffuse cytoplasmic localization (fig. S6A). By contrast, keratinocytes transfected with genes encoding variants of ≥4 repeats efficiently formed KG-like structures. Moreover, proportional to the total repeat numbers, a monotonic increase in density within KG-like granules plateaued beyond the largest known human filaggrins, suggesting that nonphenotypic filaggrins (10 to 12 repeats) optimally define the material properties of KGs (fig. S6B).

Humans with early truncation mutations fail to generate KGs. Such mutations account for >80% of cases among northern Europeans (*27*). To quantitatively determine how disease-associated mutations alter the critical concentration for phase separation, we incorporated a self-cleavable (p2a) sequence (*28*) to express equimolar amounts of mRFP-FLG variants and H2B-GFP (as a proxy for variant concentration) (fig. S7A and table S2). Live imaging of transfected HaCATs expressing comparable nuclear GFP revealed a relationship between the number of filaggrin repeats and phase-separation propensity (Fig. 2A). Over a wide range of expression levels, disease-associated mutations with ≤4 repeats exhibited a large increase (~130 to >1500 μM) in critical concentration required for phase separation (Fig. 2B and fig. S7, B to F). By contrast, wild-type filaggrin (n = 12) phase separated at ~2 μM. These properties were confirmed by live imaging, exposing the rapid formation and growth of KG-like structures as filaggrin reached its critical concentration for phase separation (Fig. 2C).

Filaggrin and its paralogs belong to the S100-fused type protein family that feature two short "EF hand" calcium-binding motifs (~2% of the protein), N-terminal to the IDP domain. The S100 domain is known to dimerize (*29*), and when fused to filaggrin variants, it reduced the critical concentration for phase separation (Fig. 2D). Despite these favorable interactions, S100 with mut-n2 FLG mutations still failed to phase separate appreciably even at high concentration (Fig. 2D and fig. S7, B to F). Overall, when compared with mRFP fusions, sfGFP lowered the critical concentrations for phase separation of

[1]Howard Hughes Medical Institute, Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, NY 10065, USA. [2]Electron Microscopy Resource Center, The Rockefeller University, New York, NY 10065, USA.
*Present address: Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30322, USA.
†Corresponding author. Email: fuchslb@rockefeller.edu

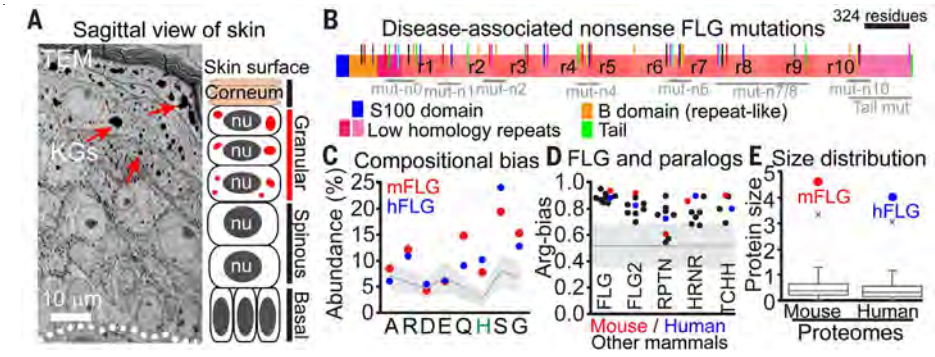


Fig. 1. Filaggrin family proteins have phase-separation characteristics, and their mutations are linked to human skin barrier disorders. (**A**) Ultrastructure and schematic of mouse skin at E17.5. Dotted lines delineate the basement membrane, where proliferative epidermal progenitors attach (basal layer). Periodically, progenitors initiate terminal differentiation, ceasing to divide, but transcribe the necessary genes for skin barrier formation as they flux upward through keratin filament bundle–rich spinous layers; keratohyalin granule (KGs, arrows)-rich granular layers; and dead, enucleated squames, which continually slough from the skin surface (corneum), replenished by differentiating cells from beneath. nu, nucleus. (**B**) Domain architecture of human FLG, the major known constituent of KGs, and location of nonsense FLG mutations (colored lines) associated with skin barrier disorders (fig. S1 shows mutants). Many mutations cluster to generate truncated variants in FLG repeat domains (labeled as mut-n0 to mut-n10). (**C**) Mouse and human FLG are histidine-rich, low-complexity (LC) proteins with identical biases in amino acid composition, but not sequence. Mean amino acid abundance across the human proteome is shown as a gray line (filled area is the standard deviation). Amino acid abbreviations: A, Ala; R, Arg; D, Asp; E, Glu; Q, Gln; H, His; S, Ser; G, Gly. (**D**) FLG and its paralogs (FLG2, RPTN, HRNR, and TCHH) share a strong preference for arginine over lysine residues [calculated as R/(R+K)], a major determinant of phase separation in LC proteins (*22*). The gray line marks the mean Arg bias across the mouse and human proteome (filled area is the standard deviation). See fig. S3 for details. (**E**) Proteome-wide distribution of protein size (unit length 1000 amino acids), underscoring the enormous size of FLG (x marks the 99th percentile).

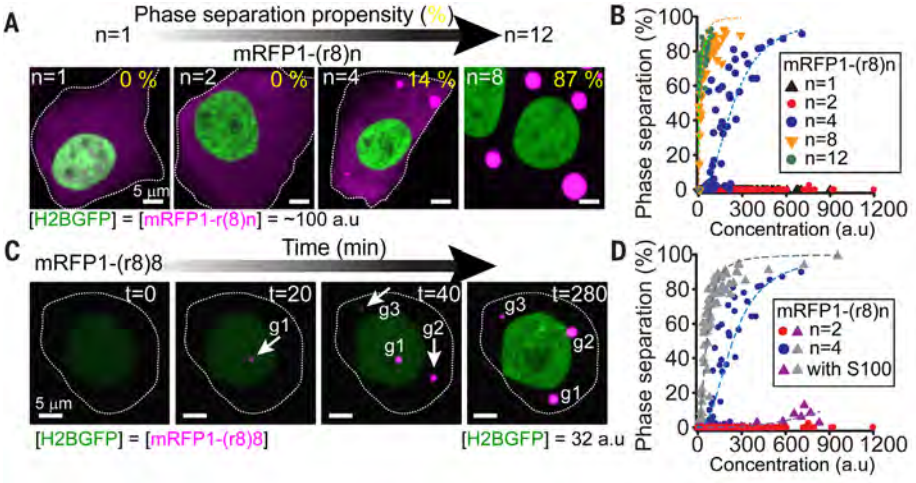


Fig. 2. Filaggrin proteins undergo liquid-liquid phase transitions that are disrupted by disease-associated filaggrin mutations. (**A**) Transfection of synthesized FLG genes into HaCATs reveals that the propensity of FLG repeat proteins to undergo phase separation is governed by the number of FLG repeats. In these experiments, genes encoding tagged-FLG variants [mRFP-(r8)n, where r8 = repeat #8 and *n* = 1 to 8 of these repeats] were fused C-terminal to a H2B-GFP-(p2a) construct. Cotranslationally, the self-cleavable (p2a) sequence (*28*) ensures that each construct generates one H2B-GFP molecule for each mRFP-(r8)n molecule. Panels show cells with the same total concentration of mRFP-(8)n. Quantitatively, phase-separation propensity was defined as the percentage of total mRFP signal within phase-separated granules. (**B**) Phase-separation propensity for FLG variants spanning the repeat distribution of truncated FLG mutants (mut-n0 to mut-n8 in Fig. 1B; WT-size is *n* =12) and across a wide range of expression levels for each variant. Dashed lines are logistic fits to data with signs of a concentration-dependent phase transition. (**C**) Time-lapse imaging of HaCATs expressing increasing amounts of mRFP-(r8)8 [related to H2BGFP via (p2a)]. Shown are the initial stages of phase separation through the formation and growth of granules (marked as g1 to g3). (**D**) The S100 (dimerization) domain of human FLG enhances the phase-separation propensity of FLG repeat proteins but fails to rescue phase behavior in disease-associated variants with ≤2 FLG repeats (mut-n0-n2 in Fig. 1B). Construct design and quantifications are as in (B). Dashed lines are logistic fits to the data. Images are maximum intensity projections.

tagged-FLG variants, although the results were consistent independent of the tag (fig. S7, G and H).

To further explore whether compromised phase separation might underlie disease severity, we next performed fluorescence recovery after photobleaching (FRAP). As expected for a diffusive process, highly truncated, smaller FLG repeat variants exhibited more rapid recovery than wild-type (WT)–sized proteins (Fig. 3A). However, even the largest FLG variants (>430 KDa) recovered fully within a few seconds (Fig. 3A). The amino and carboxy domains flanking the repeats also affected recovery time. As predicted, the S100 dimerizing domain of FLG variants increased recovery half-life after photobleaching even further, while deletion of the carboxy tail, a small truncation mutation seen in some patients, accelerated recovery (Fig. 3B and fig. S7I). Overall, the dynamic FRAP behavior established filaggrin-containing KGs as biomolecular condensates and distinguished them from mere aggregates in the cell. Moreover, because the S100 domain is cleaved during terminal differentiation, its function is likely to optimize phase separation at earlier stages when filaggrin amounts are low and KGs just begin to form.

Liquid-like behavior of filaggrin granules

Live-cell imaging revealed that HaCATs harboring our engineered filaggrins underwent granule rearrangements and fusion events that are hallmarks of liquid-like droplets (Fig. 3C and movie S1). Individual fusion events were complete within seconds (fig. S8 and movie S2).

To further probe their material properties, we used atomic-force microscopy (AFM). By applying pressure with an AFM probe directly on top of filaggrin granules, they deformed, creating liquid-like streaming around the cell's nucleus (Fig. 3D, fig. S9, and movie S3). Even unprocessed (more viscous) S100-containing filaggrin granules underwent fusion when pushed into close proximity by the AFM probe (movie S4).

Our photobleaching data suggested that the material properties of KGs may change as a function of filaggrin processing and disease-associated mutations. To test this hypothesis,

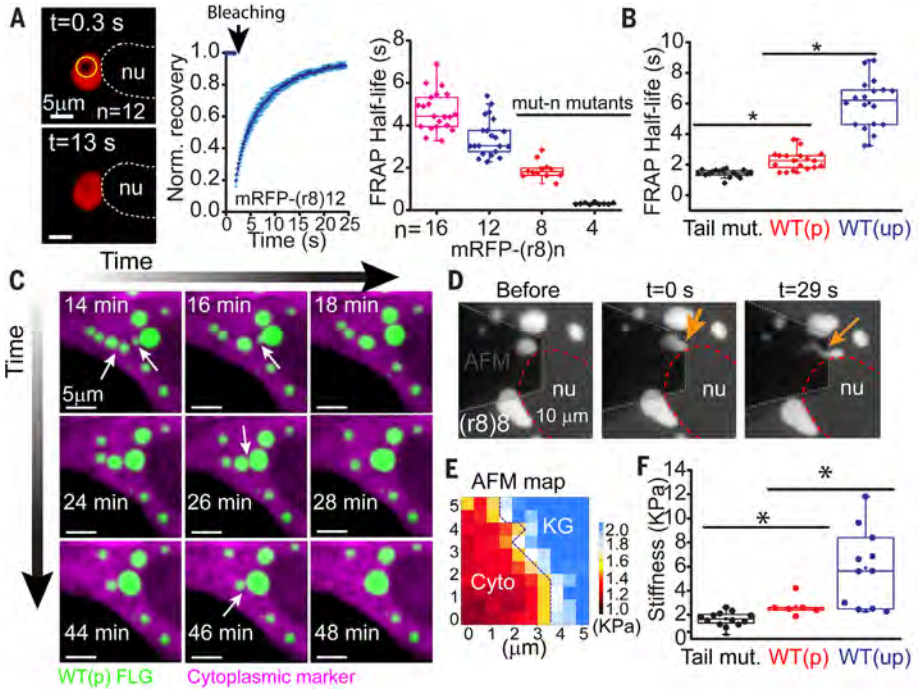


Fig. 3. Filaggrin-processing and disease-associated mutations alter the liquid-like behavior and material properties of KG-like membraneless compartments. (**A**) FRAP half-lives of granules formed de novo in immortalized human keratinocytes after transfection of indicated mRFP1-tagged FLGs with different FLG repeat truncations. Left: Representative images of a recovery event; middle: representative FRAP recovery plot (average ± SD from seven granules); right: quantifications. (**B**) FRAP half-lives after internal photobleaching of granules formed from a mRFP-FLG [WT(p), mRFP-(r8)8-Tail] in comparison to one that either lacks the 26–amino acid tail domain (Tail mut) or contains the amino (S100) domain of FLG [WT(up)]. Each symbol in (A) and (B) represents an individual FRAP half-life measurement of granules from multiple cells. Data are from ≥2 experiments. (**C**) Tagged-FLG granules undergo liquid-like fusion events. Live imaging of a cell transfected with a cytoplasmic marker (mCherry) and a WT(p) FLG [sfGFP-r(8)12-Tail]. Arrows point to granule fusion events over time (movie S1). (**D** to **F**) Atomic force microscopy (AFM) reveals liquid-like behaviors of granules. (D) Snapshots of granule (arrows) before and with pressure application reveal liquid-like streaming behavior (movie S3). (E) Representative AFM map shows that even KGs composed of the FLG tail mutant appear to be stiffer than cytoplasm (see fig. S10 for WT-type KGs data). (F) Average stiffness (Young's modulus) per granule for KGs assembled from the FLG variants described in (B). Each dot corresponds to measurements of a different granule (average of all pixels within the granule domain in the stiffness map) in a different cell. nu, nucleus; asterisks, statistically significant ($p < 0.05$).

we performed serial force-indentation measurements across the granule length in HaCATs harboring different filaggrin variants (Fig. 3E and fig. S10). Consistent with a role for the tail domain in tuning the material properties of KGs, AFM revealed a stiffening of the cellular domain spanned by tail-containing granules as compared to filaggrin counterparts that mimicked tail-deficient mutants (Fig. 3F). As suspected from the photobleaching data, unprocessed (S100-containing) filaggrin variants displayed pronounced stiffening (Fig. 3F and fig. S10). Thus, filaggrin granules are mechanically responsive, liquid-like condensates in cells.

Engineering phase separation sensors to interrogate endogenous KGs

Although our tagged filaggrin variants assembled de novo into KG-like structures, it was critical to address whether endogenous KGs in skin assemble through phase separation of filaggrin and if so, how their putative liquid-like properties contributed to epidermal differentiation. To do so, we could not use direct filaggrin tagging to label endogenous KGs, because it altered its biophysical properties (fig. S7G). Similarly, often used client proteins that directly bind to a phase-separating protein scaffold (*2*, *30*, *31*) were not suitable, because although they can be recruited to existing liquid-like condensates and be used as carriers of fluorescence, they report scaffold localization irrespective of phase separation. Moreover, with complex differentiation programs in tissues, where processing of the scaffold can occur, the caveats of conventional clients become all the more apparent (fig. S11 and supplementary text).

Thus, we sought to design different clients that would permit probing the phase-separation behavior of endogenous scaffold proteins as their concentration and processing change in living tissues. We aimed for soluble IDP clients that lack phase-separation behavior of their own but copartition efficiently and innocuously into nascent phase-separated condensates by engaging in ultraweak, phase-separation–specific (combinations of charge-charge, cation-pi, pi-pi, hydrogen-bonding, and hydrophobic) interactions with the scaffold (Fig. 4A and fig. S12A).

To engineer such "phase-separation sensors" for endogenous filaggrin, we exploited (i) the nonpathogenic behavior of human filaggrin repeat mutants that possess His:Tyr mutations (fig. S12, B and C) and (ii) the inability of a sole filaggrin repeat to drive phase separation. After documenting the tuned phase-separation characteristics of Tyr-high FLG repeat #8 (r8) variants (r8H1 and r8H2) (*22*), we then generated variants with related sequence patterns but low sequence identity (ir8H2 and pr8H2). We also engineered proteins smaller than a filaggrin repeat, but with similar compositional biases (eFlg1, ieFlg1, and eFlg2). These proteins displayed a range of phase-separation propensities (Fig. 4B; fig. S12, D to F; table S3; and supplementary text).

For live imaging, our sensors needed a fluorescent tag. Because surface charge of fusion proteins can affect IDP phase-separation behavior (*32*), we screened Tyr-high sensor variants fused to sfGFPs of varying net charges (*33*) for those that display high partition coefficients into KGs (figs. S12F and S13, A and B). We selected two +15GFP-based (sfGFP with net charge +15) sensor designs that shared little sequence identity among themselves or the native filaggrin repeat (Fig. 4C). On their own, these sensors showed no phase separation (fig. S13C), but when coexpressed with an mRFP-tagged filaggrin in HaCATs, they partitioned into the de novo assembled KGs (Fig. 4D). In particular, the relocalization of sensor A from the cytoplasm into filaggrin granules (partition coefficient $P = 21$) compared well to the behavior of filaggrin itself ($P = 125$) and remained stable over a wide range of sensor A levels (fig. S14A). This enabled faithful reporting of steep concentration gradients across membraneless granule boundaries. Notably, and in contrast to a conventional client that bound stoichiometrically to a filaggrin domain (fig. S14, B to D), FRAP dynamics of mRFP-tagged filaggrin were unaffected by sensor A (Fig. 4E).

Because of FRAP's size dependence, our studies in Fig. 3A were only suggestive that FLG truncating mutations accelerate liquid-like dynamics within KGs. With our designed sensors as internal controls, we could now accurately determine FRAP half-lives and

evaluate the liquid-like behavior of KGs formed from different-sized filaggrin mutants. As predicted, KGs assembled from truncated filaggrins differed from their full-length counterparts in displaying sensor recovery dynamics indicative of a decrease in the relative viscosity of KGs (Fig. 4F). Moreover, KGs assembled from tail mutants behaved as less viscous liquids than their tail-containing counterparts. Thus, our phase-separation sensors can integrate into and innocuously report the material properties of liquid-like KGs. Furthermore, our data suggest that patient disease phenotypes are indeed linked to shifts in the liquid-like behavior of mutant KGs.

Crowding of liquid-like KGs within skin cells in tissue

Our ultimate goal was to interrogate the dynamics of these liquid-liquid phase transitions in vivo in the skin epidermis. To this end, we used a noninvasive in utero lentiviral delivery system to selectively, efficiently, and stably transduce the single layer of embryonic day 9.5 (E9.5) mouse skin epithelium with doxycycline-inducible transgenes encoding our sensors (fig. S15). To induce expression during epidermal differentiation, we transduced embryos carrying a doxycycline-sensitive reverse tetracycline-transactivator (rtTA) driven by the human *Involucrin* (*Ivl*) promoter (*34*).

Once the skin barrier was fully mature (E18.5), doxycycline-fed embryos were subjected to live imaging and/or immunofluorescence microscopy. Sagittal confocal views revealed that bright sensor signal was confined to filaggrin-expressing granular layers, while planar views showed a robust array of sensor-labeled KGs in these cells (Fig. 5A). Moreover, and in marked contrast to conventional antibodies against filaggrin (*35*), the sensors penetrated even the large granules of the most mature (late) granular layer (fig. S16).

The marked level of KG crowding seemed incompatible with liquid-like behavior. To gain further insights, we performed live imaging and monitored keratinocyte flux through the granular layers of skin (*36*). Early granular cells displayed only a few KGs, whose numbers appeared to increase through de novo granule formation (Fig. 5B). Over a half day of imaging, occasional fusions that resolved within minutes into a round granule pointed to liquid-like behavior (fig. S17 and movie S5). Moreover, for both sensor A or B, signal recovery was rapid after photobleaching KGs within the mid-granular layer, further underscoring the liquid-like behavior of these endogenous KGs (Fig. 5, C and D).

Despite these liquid-like features, most existing KGs grew robustly without undergoing fusion (Fig. 5E and movie S6). Even KGs within the earliest granular layers in skin tissue exhibited liquid-like properties distinct from those of KG-like condensates that formed in HaCATs when transfected to express tagged filaggrin (Fig. 5F). Moreover, when the sensor's nuclear export signal was removed and sensor FRAP half-life was measured within nucleoli, only filaggrin-containing KGs in tissue appeared to be relatively more viscous than the nucleolus in keratinocytes.

Probing deeper, we noticed that granular cells exhibited substantial morphological changes as they transited through the granular layers and became increasingly crowded with KGs (Fig. 5G). Correspondingly, photobleaching these KGs within early, middle, and late granular cells revealed a gradual reduction in sensor dynamics as cells moved toward the skin surface. This increase in relative viscosity of skin KGs was also seen in stratifying cultures of primary human epidermal keratinocytes, which, unlike immortalized keratinocytes, formed endogenous KGs that were of similar size and displayed liquid-like dynamics similar to those of KGs in the early- and mid-granular layers of mouse epidermis (Fig. 5H). Thus, despite pronounced species-specific divergence in filaggrin sequence, the preservation of KG's finely tuned liquid-like behavior pointed to an underlying physiological relevance.

Stabilization of liquid-like membraneless organelles

Although the rarity of fusion events among densely packed KGs might simply reflect their apparent viscosity, it was also possible that

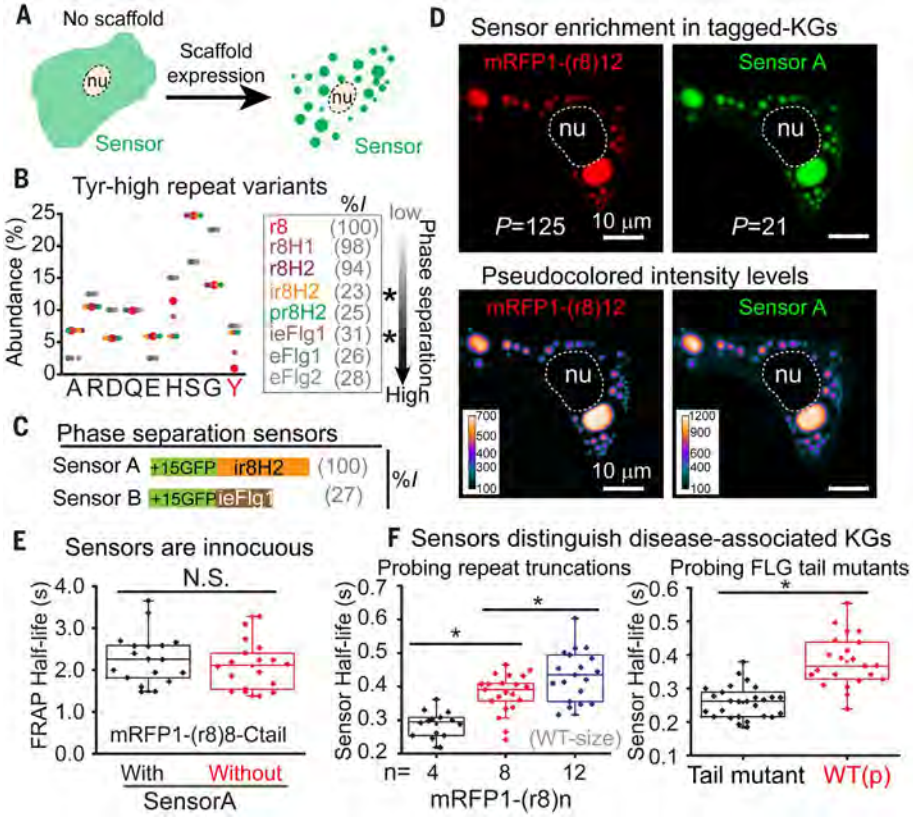


Fig. 4. Phase-separation sensors efficiently enter and detect KGs and accurately report their liquid-like properties. (**A**) Concept of a genetically encoded phase-separation sensor. (**B**) Amino acid composition of LC Tyr-high variants of a FLG repeat (repeat 8, r8), ordered at right according to phase-separation propensity. Variants were generated according to nonpathogenic residues frequently altered in FLG repeats in humans. %*I*: percent sequence identity to WT FLG repeat. Asterisks denote the two Tyr-high variants used as phase sensors in this study. Y, Tyr. (**C**) Domain architecture of the two phase-separation sensors. %*I*: percent sequence identity to sensor A. (**D**) Sensor partitioning into KGs in HaCATs expressing sensor A and indicated mRFP1-FLG. Partition coefficients (*P*, ratio of background-corrected signal inside and outside granules) reveal robust ability of sensor A to recognize FLG in its phase-separated granules (bottom row is pseudocolored to reveal the range of fluorescent intensity values). nu, nucleus. (**E**) Presence of sensor A does not alter FRAP half-life of FLG-assembled KGs in HaCATs. N.S., not statistically significant. (**F**) Sensor A recovery half-lives after photobleaching granules composed of the indicated mRFP1-tagged FLG variants that model patient mutations. Each symbol in (E) and (F) represents an individual FRAP half-life measurement of granules from multiple cells. Data are from ≥2 experiments. Asterisks, statistically significant ($p < 0.05$). See also related figs. S11 to S14.

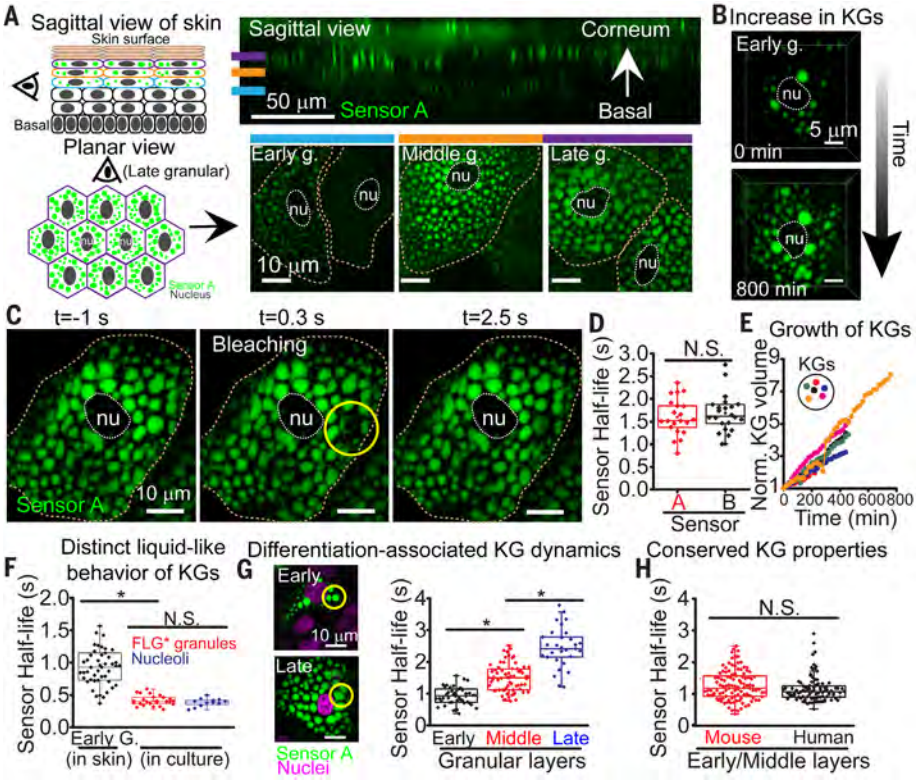


Fig. 5. Skin exhibits pronounced phase separation dynamics during barrier formation. (**A**) (Left) Schematics of sagittal and planar views. (Right) Corresponding views of fluorescent sensor A in mouse skin. Planar skin views are through early, middle, and late granular layers. nu, nucleus. Dotted lines denote cell boundaries. (**B**) Live imaging of an early granular cell over 800 min. (**C**) Example of photobleaching the sensor A signal within a KG of a granular cell in mouse skin. (**D**) Sensor recovery half-lives after photobleaching KGs across cells within mid-granular layer of transduced mouse skin (each point is from a different cell; two animals analyzed per sensor). (**E**) Quantification of changes in KG volume over time in cells as they reach the upper granular layer (related to movie S6). (**F**) Sensor A reveals distinct liquid phase properties within different biomolecular condensates and contexts [in vivo KGs versus granules generated de novo from S100-mRFP-(r8)8-Tail, expressed in cultured keratinocytes]. For nucleolar measurements, a sensor A variant lacking a nuclear export signal was used. In vivo and in vitro data from ≥2 experiments. (**G**) Sensor A detects an increase in relative KG viscosity that occurs during granular layer maturation. Shown are FRAP half-lives in KGs within different granular layers (morphological differences at left; data from three animals). (**H**) Sensor A reveals conserved liquid-like KG properties despite divergence in amino acid sequence of granule-forming proteins. Mouse KG data in (H) are same as in (G). Human KG data are from three skin equivalents and two sources of primary human keratinocytes. Asterisks, statistically significant ($p < 0.05$). N.S., not significant.

additional facets of terminal differentiation might be contributing to this puzzling behavior. Notably, the granular layer also displays an abundant network of terminal differentiation–specific keratins 1 and 10 (K1/K10) filaments. This prompted us to test whether keratin filaments might impede KGs from fusing and allow them to crowd the cytoplasm as stable organelles. When HaCATs were transduced with doxycycline-inducible human mRFP-K10 (table S4), hK10 incorporated into the endogenous network of basal K5/K14 filaments (*37*). Upon cotransfection with sfGFP-FLG to drive KG formation, many of the mRFP-tagged keratin bundles encased KGs (Fig. 6A). Live imaging showed that these KGs spent prolonged periods of seemingly inert activity. However, in regions where these KGs dislodged from filaments and became uncaged, KGs were mobile and frequently fused with other sfGFP-tagged KGs (Fig. 6B, fig. S18A, and movie S7). This may explain previous perplexing observations that unusually large KGs are produced after genetic ablation of *Krt10* in mouse skin (*38*).

Keratins possess a central coiled-coil "rod" domain that initiates heterodimer formation and forms the backbone of the 10-nm intermediate filament (*39*). Whereas K5/K14 keratins of proliferative progenitors have short amino- and carboxy-LC domains, the large LC domains of K1/K10 keratins (*40*) are thought to protrude along the outer surface of the filament and bundle into cable-like filaments.

Intrigued by the packing of K10-containing filaments around filaggrin granules, we next asked whether their distinctive features might facilitate interactions with KGs. Examining the behaviors of mCherry fused to one, both, or neither of the hK10 LC domains (table S4), we found that each was diffuse in the cytoplasm of cultured keratinocytes in the absence of KGs (Fig. 6C and fig. S18). By contrast, when sfGFP-tagged filaggrin and its KGs were present, mCherry was excluded from KGs, while both mCherry constructs with hK10 LC domains partially partitioned into KGs. Moreover, the critical concentration for phase separation of sfGFP-tagged filaggrin was reduced in the presence of the K10 keratin network without altering FLG density within KGs (Fig. 6, D and E). Thus, weak interactions between KGs and the LC domains of terminal differentiation–associated keratins may promote the caging and stabilization of KGs in skin.

To further explore this possibility, we transduced both our phase-separation sensor and suprabasal-inducible mRFP-hK10 constructs into E9.5 embryos and performed live imaging on E18.5 skin explants. Whereas early granular cells displayed small, relatively sparse KGs surrounded by a well-defined network of K10-containing filaments, mid-granular cells exhibited a denser keratin network interwoven among larger, more abundant KGs that remained caged and hence unable to fuse (Fig. 6, F and G). Our findings suggest a model whereby reciprocal density-dependent interactions between LC domains of terminal differentiation–specific keratins and KGs structure the cytoplasm to form an elaborate, interwoven network of stabilized liquid-like KGs and keratin filament bundles.

Liquid phase KG dynamics, enucleation, and environmental sensitivity

We posited that progressive crowding by keratin-stabilized KGs might distort the nucleus and other organelles in a fashion that could contribute to their destruction at the critical granular–to–stratum corneum transition. If so, this could explain why nuclei are often aberrantly retained in the outer skin layers of patients who also lack KGs (*41*).

Consistent with this notion, KGs assembled de novo in HaCATs from WT repeat filaggrins prominently deformed nuclei, whereas KGs assembled from disease-associated FLG mutants instead wetted the nuclear surface without deformation (Fig. 7A and fig. S19, A and B). Endogenous KGs in primary human keratinocytes also induced prominent nuclear deformation (fig. S19C). Similarly, when we transduced embryos with H2B-RFP and our sensors and monitored the maturation of granular cells by live imaging, we found that mature granular cell nuclei were markedly deformed as KG density increased in skin (Fig. 7B and fig. S19D).

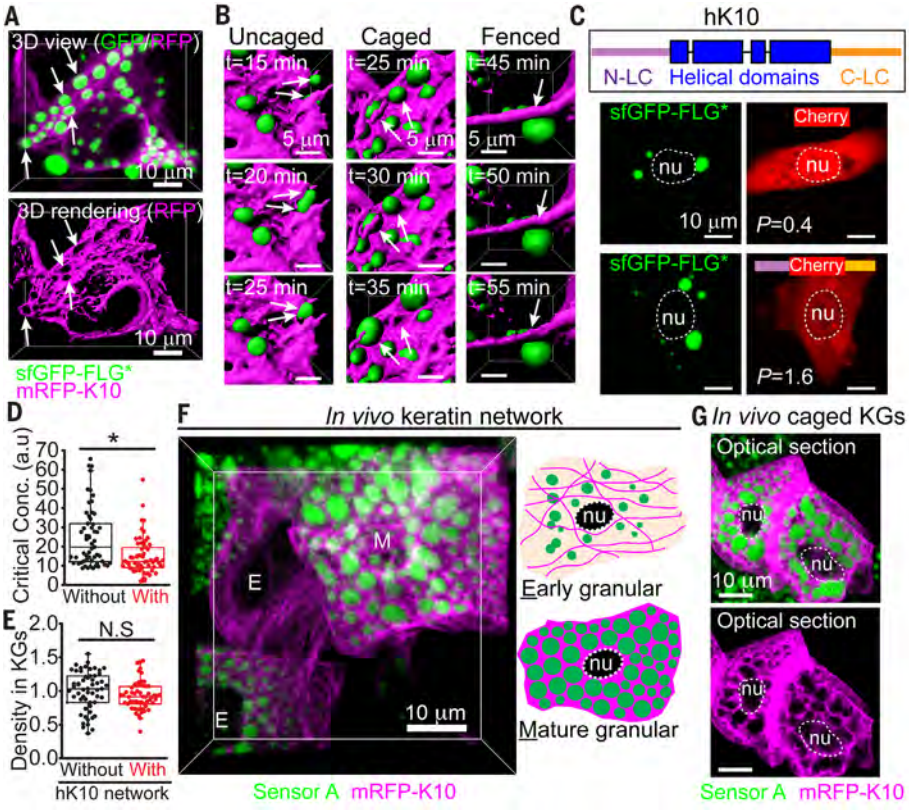


Fig. 6. Keratin-FLG interactions stabilize KGs and structure the cytoplasm in skin. (**A**) HaCATs were induced to express mRFP1-K10, which integrates into the endogenous K5/K14 filaments. Cells were then transfected with sfGFP-FLG*, which formed liquid-like KGs (arrows) interspersed within the keratin network. Top: 3D projections of GFP/RFP; bottom: surface rendering of mRFP-K10. (**B**) Live imaging of cell in (A), showing surface rendering of three different types of keratin-KG interactions (see fig. S18A for maximum intensity projections). Uncaged KGs fuse rapidly; caged KGs fuse rarely or slowly; fenced KGs are impeded from fusing. Double arrows depict temporal fusion events; single arrow denotes keratin cable preventing fusion. (**C**) When mCherry harbors hK10 LC domains, it partially partitions into KGs (P = 1.6). (**D**) Phase separation of sfGFP-(r8)4 FLG is promoted in HaCATs displaying mRFP1-K10 fibers. Critical concentrations for phase separation were estimated as in fig. S7C (data from three experiments). (**E**) FLG density within KGs assembled in (D) is similar ± an hK10 network. (**F**) Planar 3D view of E18.5 granular layer from skin of an embryo transduced in utero with a suprabasal-specific driver of mRFP1-K10 and sensor A. Accompanying cartoon depicts protein localization patterns seen in early and mature (late) granular cells. (**G**) Optical sections through mature granular cells show prominent granules encased by thick keratin bundles. Single magenta channel reveals voids where KGs reside, indicative of caged KGs. Asterisks, statistically significant ($p < 0.05$). N.S., not significant.

As granular cells move upward to complete their final stage of differentiation, they enucleate and lose their KGs to form flattened squames (Fig. 1A). These events were difficult to capture by live imaging, because the process was very rapid, occurring over 2 hours (fig. S20). However, when we captured enucleation events, they were always preceded by chromatin compaction and then chromatin loss and nuclear destruction (Fig. 7C and movie S8). Moreover, just as chromatin began to show signs of compaction, KGs began to dissolve, as revealed by a progressive shift in the phase sensor's localization from a granular-like state to being diffuse in the cytoplasm (Fig. 7C and fig. S21). This rapid shift indicated a marked change in the liquid-like properties of KGs and suggested that these terminal events may also be rooted in liquid-phase dynamics.

Probing further, we found that *Flg* knockdown in skin not only depleted KGs but also delayed the nuclear degradation process (Fig. 7D). This was accompanied by increased transepidermal water loss (TEWL) through the skin barrier. Thus, KGs accelerate loss of membrane-bound organelles, an essential feature of skin barrier formation.

Given the inherent environmental responsiveness of intrinsically disordered proteins (*42*, *43*), we wondered whether the marked shift in KG dynamics late in terminal differentiation might be fueled by the environmental changes that naturally occur near or at the skin surface. In particular, whereas proliferative basal progenitors experience physiological pH (7.4), the skin surface is acidic (pH ~5.5) (*44*). Because filaggrin is rich in histidine, whose physiological acid dissociation constant (pK_a) is ~6.1 (*45*), we posited that this natural difference in extracellular pH may also reflect intracellularly and in part be triggering the KG changes that we had detected at the granular–to–stratum corneum transition.

To detect intracellular pH shifts, we first transduced HaCATs with either mNectarine or SEpHlourin reporters, which rapidly lose fluorescence upon shifting from pH 7.4 to pH 6.3 (fig. S22). When the extracellular pH was decreased to elicit an intracellular pH shift from 7.4 to ~6.2 to 6.5, de novo–assembled KGs in HaCATs changed profoundly. In live imaging, both filaggrin and the sensor (which by design, is also rich in histidine) displayed increased cytoplasmic and diminished KG-like localization concomitant with this intracellular pH shift (Fig. 7E and fig. S23). Similar changes were seen in endogenous KGs of differentiated primary human epidermal keratinocytes when they experienced this pH shift (fig. S24A).

Given the pH sensitivity of KGs, we then turned to investigating this process in vivo. We introduced our pH reporters into mice along with either our phase-separation sensor or H2BRFP and through live imaging, monitored the natural intracellular pH shifts that we surmised would occur as granular cells approached the acidic skin surface. Over time, as each granular cell progressed to the critical granular-to-corneum transition, it experienced a sudden shift in pH, as detected by our intracellular reporters (Fig. 7F). This rapid endogenous pH shift invariably coincided with the initiation of KG dissolution (top panels) and coincided with or immediately preceded an increase in chromatin compaction (bottom panels). Moreover, within 2 hours of KG dissolution and nuclear compaction, imaged granular cells within the epidermis had undergone morphological changes characteristic of enucleation and squame formation (fig. S24B and movie S6).

Finally, we took skin explants from embryos transduced with phase sensor, H2BRFP, and either scrambled or *filaggrin* short hairpin RNAs (shRNAs) and performed live imaging immediately after shifting the extracellular pH in the medium. When the natural intracellular pH transition was accelerated, granular cell KGs showed signs of disassembly, and chromatin compaction became pronounced (Fig. 7G, top panels). This pH shift did not trigger chromatin compaction in skin devoid of KGs (bottom panels). Thus, the pH shift appears to function specifically in altering the material properties of histidine-rich KGs, which in turn promote chromatin compaction, enucleation, and skin barrier establishment. Furthermore,

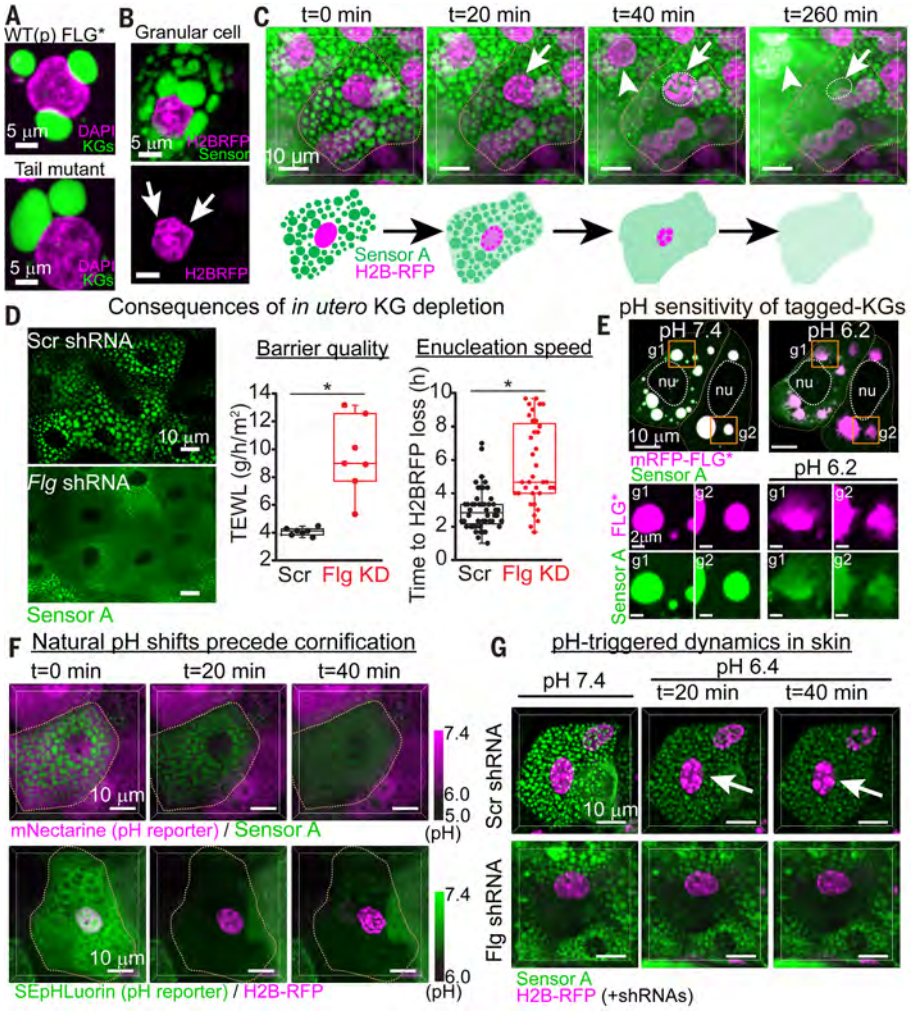


Fig. 7. Environmentally regulated KG dynamics drive skin barrier formation. (**A**) Nucleus-KG interactions in HaCATs transfected with FLG variants. (**B**) Nucleus-KG interactions in a granular cell from live imaging of E18.5 mouse skin with resolution of nuclei (H2B-RFP) and KGs (sensor A). Arrows point to KG-associated nuclear deformations. (**C**) Granular cell–to–squame transition, as depicted by live imaging (3D view) of E18.5 mouse skin (movie S8). Early signs include chromatin compaction (arrows) and diminished partitioning of sensor within KGs. Late signs include KG disassembly and enucleation. (**D**) In utero *Flg* knockdown depletes KGs, causes a delay in enucleation, and partially compromises the skin barrier. Enucleation speeds were determined by live imaging of chromatin degradation. Barrier quality was measured as transepidermal water loss (TEWL). Asterisks, statistically significant ($p < 0.05$). (**E**) Effects of shifting the intracellular pH on KG dynamics of mRFP1-tagged FLG* and sensor A, as monitored by live imaging of HaCATs (maximum intensity projections). Note the rapid (t = 5 min) pH-triggered dissolution of KG components. g1 and g2 show individual granules. Sensor A mirrored the pH-triggered drop in the phase-separation capacity of FLG, which became increasingly cytoplasmic, reflected by a decrease in its partition coefficient (P = 26 at pH 7.4 to P = 3.6 at pH 6.2). nu, nucleus. (**F**) Live imaging (3D view) of enucleation and cornification in skin of embryos transduced to express an organelle marker (top: sensorA/KGs; bottom: H2BRFP/nuclei) and a pH reporter whose fluorescence is lost below pH 6.5. mNectarine (top) shows that when the intracellular pH of granular cells drops below pH 6.5, KGs begin to disassemble. SEpHLuorin reports a similar pH drop and shows that it precedes chromatin compaction. (**G**) Effects of pH-induced KG dynamics in sensor A$^+$ skin explants transduced with H2B-RFP and either *Scr*-shRNA (top) or *Flg*-shRNA (bottom). Note that chromatin compaction is not pH-triggered if KGs are missing altogether. See also figs. S19 to S24.

enucleation events in skin are likely driven by a combination of nuclear deformation and pH-driven release of as yet undetermined KG components.

Discussion

Our design and deployment of a class of innocuous client protein provides a general strategy to analyze endogenous liquid-liquid phase separation dynamics across biological systems in a nondisruptive manner. We envision that these in vivo phase-separation sensors may be further functionalized to incorporate enzymes evolved for proximity proteomics (*46*, *47*), potentially enabling—without perturbing endogenous scaffold proteins—the molecular and biophysical interrogation of endogenous liquid-liquid phase separation in organoids, tissues, and living organisms.

We used this strategy to illuminate, through the lens of phase separation, the process of skin barrier formation, which entails the appearance of hitherto enigmatic KGs in the granular layer and then their sudden disappearance as epidermal cells undergo a poorly understood transition to the stratum corneum. These granules, long puzzling to skin biologists (*48*), had been viewed as inert, cytoplasmic aggregates of filaggrin, which eventually became cleaved into smaller fragments and amino acid derivatives to promote keratin filament bundling (*21*) and stratum corneum hydration (*24*). Despite decades of research and mutations linked to atopic dermatitis (*23*), no clear function had been established for KGs, filaggrin, or filaggrin paralogs that also accumulate as granular deposits in epithelial tissues (*49*, *50*).

Through the engineering of filaggrins and filaggrin disease-associated variants and also phase-separation sensors, we have now shown that KGs are abundant, liquid-like membraneless organelles, which, through their phase-separation–driven assembly and then disassembly, function to structure the cytoplasm and drive an environmentally sensitive program of terminal differentiation in the epidermis. By virtue of their mechanical and pH-sensitive properties, KGs are ideally equipped to confer environmental responsiveness to the rapid and adaptive process of skin barrier formation. The discovery that filaggrin-truncating mutations and loss of KGs are rooted in altered phase-separation dynamics begins to shed light on why associated skin barrier disorders are exacerbated by environmental extremes. These insights open the potential for targeting phase behavior to therapeutically treat disorders of the skin's barrier.

Liquid-phase condensates have typically been viewed as reaction centers where select components (clients) become enriched for processing or storage within cells (*2*). Analogously, KGs may store clients, possibly proteolytic enzymes and nucleases, that are timely (in a pH-dependent fashion) and rapidly released to promote the self-destructive phase of forming the skin barrier. Additionally, squame formation likely exploits general biophysical consequences of KG assembly, because KGs interspersed by keratin filament bundles massively crowd the keratinocyte cytoplasm and

physically distort adjacent organelles prior to the ensuing environmental stimuli that trigger KG disassembly. Overall, the environmentally sensitive dynamics of liquid-like KGs, actionable by the skin's varied environmental exposures, expose the epidermis as a tissue driven by phase separation.

Materials and methods

Sequence analysis of filaggrin and its paralogs

Proteomes were downloaded as FASTA files from UniProt (https://www.uniprot.org/). For the analysis of protein domains known to drive liquid-liquid phase separation (fig. S3), we downloaded the complete set (>100) of proteins from the PhaSEPro database (*51*). We implemented a script (available upon request) in MATLAB R2016a to extract protein size, amino acid abundance, and Arg-bias of all annotated proteins. Arg-bias was calculated as the total number of arginine residues relative to the total number of positively charged residues (R+K). To calculate relevant sequence parameters (length, amino acid composition, hydropathy, Arg-bias) in FLG and its paralogs across species, we implemented a script (available upon request) in MATLAB R2016a. Hydropathy was calculated as the average level across all residues in a protein, using Kyte-Doolittle's scale (*52*). Except for human Flg (and its paralogs), most mammalian Flg and Flg paralogs in mammals remain poorly annotated or poorly sequenced in publically available genome and protein databases. Table S1 shows the sequences that we used as input material and details of their manual curation. To characterize nonsynonymous mutations in human filaggrin, we downloaded known single-nucleotide polymorphism (SNPs) in the human Flg gene (not annotated in ClinVar) from NIH's dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) and from the GnomAD browser (https://gnomad.broadinstitute.org/gene/ENSG00000143631). We used custom-made MATLAB scripts (included in the supplementary text) to filter for unique SNPs corresponding to nonsynonymous mutations. This script also calculated the overall percentage of mutations assigned to each of the 20 naturally occurring amino acid residues. By generating 1000 unique Flg mutant genes through random single-nucleotide mutations in Flg cDNA (script available upon request), we also estimated the expected random mutational burden per residue. From the total SNPs, we identified 405 SNPs involving mutations of His codons. The script then identified the mutational landscape involving these SNPs and their corresponding nonsynonymous codons (encoding Asp, Leu, Asn, Pro, Gln, Arg, and Tyr).

Synthesis of repetitive DNAs encoding filaggrin and filaggrin variants

To assemble repetitive DNAs, we used recursive directional ligation by plasmid-reconstruction (PRe-RDL) (*53*), with minor modifications. Specifically, we used a modified pET-24a(+) vector as published (*53*), but eliminated the terminal Tyr-stop-stop sequence to avoid altering the hydropathy of FLG sequences. Instead, the modified vector uses a terminal Gly-stop-stop-stop sequence. Synthetic gblocks were from IDT (Integrated DNA Technologies) and encoded the eight repeat in human FLG (repeat #8, here referred as r8), sfGFP, mRFP1, and the S100 domain of human FLG. We chose r8 as this repeat is often duplicated in humans, yielding FLG variants with 11 (this is the most common of all human FLG variants) or 12 repeats. The specific choice of a repeat (among human FLG repeats 1 to 10), however, is otherwise trivial, as individual FLG repeats are nearly identical in sequence (with >90% sequence identity in humans and typically >99% in mice). We performed iterative rounds of PRe-RDL with the r8 gblock to build genes with up to 16 concatemers of r8. These genes were then modified to generate variants with the C-terminal tail domain of human FLG (table S2). DNA sequences were verified by Sanger sequencing (Genewiz, NJ) whenever possible. For long repetitive DNAs beyond the reach of Sanger sequencing, to confirm proper concatamerization of sequence-verified domains, we relied on gene size (judged by conventional DNA gel electrophoresis) and subsequent validation of expected protein properties (size and diffusion properties) upon expression in *Escherichia coli* or mammalian cells. For mammalian expression, we subcloned fully-assembled repeat genes into a modified pMAX vector (Amaxa). See table S2 for protein sequences for all constructs. To build genes with nuclear reporters of FLG concentration in the cell, we further modified our pMAX-based genes encoding FLG repeats to replace the N-terminal fluorescent protein with genes fragments encoding H2BGFP-(p2a)-mRFP, H2BRFP-(p2a)-sfGFP or H2BGFP-(p2a)-S100-mRFP (see table S2 for sequence details). (p2a) is a codon-optimized DNA sequence that self-cleaves during translation and so enables the synthesis of two proteins from a single transcript (*28*). We also built pMAX vectors encoding H2BRFP-(p2a)-H2BGFP and H2BGFP-(p2a)-H2BRFP to validate the equimolar synthesis of individual (p2a)-linked proteins.

Synthesis of phase-separation sensors

Table S3 includes the sequence information for all sensor domains reported in Fig. 4B. The rationale for the generation of these proteins is explained in detail in the supplementary text. Corresponding genes were synthesized by IDT as gblocks and cloned into modified pMAX vectors as described above for genes encoding FLG variants. We purchased additional gblocks encoding previously published supercharged variants of sfGFP: +15GFP and −20GFP. All constructs, unless indicated, include an optimized short nuclear export signal (*54*) (LELLEDLTL) as linker between the N-terminal fluorescent proteins and the sensor domain. To test the intrinsic phase separation propensity of individual sensor domains, we artificially enhanced their phase-separation capacity by synthesizing variants with a C-terminal trimerization domain. We generated constructs with one of two trimerization domains: NC1 domain from human COL18A1 (P39060, Isoform 1, residues1442-1496) (*55*) or a fibritin fragment from bacteriophage T4 (so-called foldon domain) (*56*).

Synthesis of genes encoding human K10 and its low-complexity domains

We used polymerase chain reaction (PCR) to amplify a fragment of the *Krt10* gene spanning the N-terminal LC domain and the complete central coiled-coil rod domain (forward primer: TAATCATCGATCGGATGGCTCTGTTCGATACAGCTCAAGCAAGCACTACTCTT; reverse primer: TAAGCAGGGGATCCCTCTCCTTCTAGCAGGCTGCGGTAGGTTTG) using KRT10 cDNA (NM_000421.2; from Origene). These primers added restriction sites for Pvu I and Bam HI at the N and C terminus, respectively, for seamless restriction into a pMAX vector harboring an N-terminal mRFP sequence and the C-terminal LC domain. The C-terminal LC domain was synthesized by IDT as a gblock. Similarly, we also obtained a gblock encoding the N-terminal LC domain flanked by Nhe I and Xma I sites, which we inserted into our modified pMAX vector for building a gene encoding a fusion to mCherry. This vector was further modified between Bam HI and Eco RI sites to introduce the C-terminal LC domain and generate mCherry fusions harboring both K10 LC domains. These constructs are listed in table S4. Because of Origene's third-party restrictions, we regret that we are unable to distribute our full-length mRFP-K10 construct, which contain material from Origene SC122561. However, mRFP-K10 may be reconstructed by following the protocol above and obtaining one of our LC constructs as well as Origene SC122561.

Characterization of filaggrin-like proteins and phase-separation sensors

To drive efficient expression of the relevant repetitive proteins, we transfected the corresponding pMAX plasmids into HaCATs (*57*). We routinely expanded HaCATs in low-calcium (50 μM) epidermal cell culture media (*58*) and then transfected them in glass-bottom 24-well plates containing CnT-PR media (CELLnTEC, Switzerland). Following the instructions of the manufacturer, we typically used lipofectamine 3000 (Invitrogen) to transfect cells with 0.5 to 3.5 μg of each plasmid. One day after transfection, we changed media to a prodifferentiation

media, CnT-PR-D (CELLnTEC, Switzerland) supplemented with 1.5 mM $CaCl_2$ and proceeded to use a spinning-disk confocal microscope to image cells 6 to 9 hours later. Live imaging was conducted with cells at 37°C and under a controlled CO_2 environment. To calculate the phase-separation propensity of FLG repeat proteins, we operationally defined it as the percentage of total (background-corrected) fluorescent signal residing within phase-separated granules and based on maximum intensity projections of live imaging data using ImageJ. Concentration values for FLG variants were determined from the nuclear H2B reporter signal (adjusted by total cell area) in each cell to sensitively measure protein concentration even at low expression levels when FLG proteins are diffuse. Whenever we observed a concentration-dependent increase in phase separation propensity, we applied a logistic fit [$y = (-100/(1 + (x/x0)\text{^}P)) + 100$, as expected for a phase transition] using OriginPro software (OriginLab). Using these fits, we approximated the critical concentration for phase separation as the half-maximal effective concentration (EC_{50}) of the logistic fit. The EC_{50} represents the concentration at which most cells reach a phase-separation propensity of 50% —wherein the total number of molecules in the dilute phase equals the number of molecules in the high-concentration density phase. Although phase separation happens with a given (low) probability below the EC_{50}, the concentration fluctuations that potently drive phase separation near the true critical concentration of the system become dominant near the EC_{50}, which justifies its definition as an experimental approximation to the critical value. To study protein dynamics within granules, we photobleached circular regions (0.54 μm in diameter) of interest at the center of granules and imaged the process of recovery at 200-ms intervals. For data analysis, we normalized the background-corrected fluorescence within the region of interest to the background-corrected average granule fluorescence prior to photobleaching and then corrected for loss of fluorescence in the granule area outside of the region of interest throughout the imaging process. To calculate recovery half-lives, we fitted the post-bleaching normalized data using OriginPro and a standard exponential growth curve: $f = A(1 - e\text{^}(-xt))$. Half-lives (time when $f = 0.5$) were estimated as $\mathrm{Ln}(0.5)/(-t)$. The approach to studying the behavior of phase-separation sensors is similar to the approach described for FLG variants. In addition to vectors harboring tagged-FLG proteins, the transfection mixture included a second pMAX vector encoding sensor variants (table S3). For photobleaching measurements, we first obtained photobleaching data for the mRFP1-tagged FLG protein, followed by photobleaching data for (+15GFP-tagged) sensor A in the same granule. Sensor data were processed and analyzed as for tagged-FLG proteins.

Atomic force microscopy (AFM) measurements

To enable access of the AFM probe to filaggrin granules within cells, we transfected HaCATs in 50-mm glass-bottom dishes (Fluorodish, FD5040, World Provision Instruments). The transfection mixture consisted of two pMAX vectors: one vector harbored a H2B-RFP gene and was common to all transfection reactions, whereas the second vector encoded one of the indicated FLG variants [sfGFP-(r8)8, sfGFP-(r8)8-Tail or S100-sfGFP-(r8)8-Tail; see table S2]. One day after transfection, we added pro-differentiation media (CnT-PR-D) supplemented to 1.5 mM $CaCl_2$. Cells were transported (at 37°C) soon after or up to 24 hours later to the Molecular Cytology core facility of Memorial Sloan Kettering Cancer Center (MSKCC) for AFM measurements using a microscope stage at 37°C. AFM force measurements and manual deformations of sfGFP-tagged FLG granules were performed using an MFP-3D AFM (Asylum Research) combined with an Axio Scope inverted optical microscope (Zeiss). We used silicon nitride probes with a 5-μm-diameter spherical tip (Novascan). Cantilever spring constants were measured before sample analysis using the thermal fluctuation method, with nominal values of ~100 pN/nm. 5 μm by 5 μm force maps were acquired with 10 force points per axial dimension (0.5 μm spacing) atop sfGFP-tagged FLG granules identified using the bright-field and GFP optical images. Measurements were made using a cantilever deflection set point of 10 nN and scan rate of 1 Hz. Bright-field (AFM probe), GFP (FLG variant), and H2B-RFP (nuclei) images were acquired for each cell and granule measured to enable force map and optical image co-registration. Live-video bright-field images were also taken during force map acquisition to observe granule and cellular deformations. Force-indentation curves were analyzed using a modified Hertz model for the contact mechanics of spherical elastic bodies. The sample Poisson's ratio was 0.33, and a power law of 1.5 was used to model tip geometry. To observe granule displacement and flow following force application, the AFM tip was manually placed adjacent to sfGFP-tagged FLG granules using a micrometer. During live video-rate (14 frames/s) image acquisition (bright-field and GFP), force was manually applied with the AFM probe in the absence of force set point feedback via micrometer manipulation.

Mice and lentiviral transduction

Mice were housed and cared for in an AAALAC-accredited facility, and all animal experiments were conducted in accordance with IACUC-approved protocols. We obtained hIVL-rtTA FVB mouse embryos from J. Segre at the National Institutes of Health (NIH). For rapid generation of mice with genetically modified skin, we used noninvasive, ultrasound-guided in utero lentiviral-mediated delivery of pLKO.1-based expression constructs and shRNAs (Sigma-Aldrich), which as previously published (*59*), results in selective transduction of single-layered surface ectoderm of living E9.5 mouse embryos. Lentiviral vectors with Scramble (not targeting) shRNAs and constitutive expression [phosphoglycerate kinase (PGK) promoter-driven] of H2B-RFP were previously reported (*59*). We modified these pLKO.1-based vectors to replace the PGK promoter with a newly assembled tetracycline regulatory enhancer (TRE) promoter sequence (based on TRE3G from Clontech). We cloned sensor and mRFP-K10 genes from pMAX vectors into pLKO.1-based vectors, downstream of the TRE. Using these pLKO.1 vectors, we generated high-titer viruses in 293FT cells as previously described (*59*). To induce expression of TRE-controlled genes in vivo, we fed females fostering lentivirally transduced embryos with doxycycline. For knockdown of mouse filaggrin, we identified hairpins with high intrinsic scores and no predicted off-targets using the GPP Web Portal (https://portals.broadinstitute.org/gpp/public/). We modified our lentiviral vectors harboring H2B-RFP to substitute their Scramble shRNA with hairpins against (mouse) *Flg* (#01: with target sequence ATCAATCTCACAGCTATTATT localized to the C-terminal domain, and #02: with target sequence CTCCGGATTCTACCCAGTATA within the filaggrin repeats). We tested both hairpins in mouse skin and they efficiently depleted mFLG and its KGs (Fig. 7 shows data for hairpin #02). To transduce human primary keratinocytes with lentiviral vectors harboring H2B-RFP and phase-separation sensors, we used neonatal and adult human primary keratinocytes that we purchased from Life Technologies. We transduced them by exposing them briefly to the corresponding high titer lentiviruses diluted in supplemented Epilife media (Thermo Fisher Scientific). A similar lentiviral transduction approach was used to generate HaCATs with doxycycline-inducible expression of mRFP-K10. See supplementary materials and methods for detailed protocols.

Live imaging

For live imaging of mouse skin, we harvested head skin from E18.5 mouse embryos that were in utero transduced as explained above. After removing the skin from the embryos, we gently scraped off the fat leaving the dermis intact, cut out 1-cm^2 pieces and placed them with the stratum corneum facing down on a glass-bottom 35-mm dish (MatTek) and on top of Phenol-red free growth-factor reduced matrigel (Corning). We pressed the tissue flat against the glass surface using a transparent

porous membrane (Whatman Nuclepore Track-Etched Membranes, 13-mm diameter, 5-μm pores) and a 12-mm cover glass (Fisherbrand). Once the tissue was flat, we removed the cover glass and allowed the matrigel to solidify at 37°C for 15 min. We then added CnT-Prime Airlift, Full Thickness Skin Airlift Medium (from CELLnTEC), typically supplemented with 2 μg/ml doxycycline. We imaged these samples using a spinning-disk microscope equipped with a 40× oil objective and a live imaging chamber with a constant supply of CO_2 and maintained at 37°C. We imaged with up to two lasers (488 and 561 nm, at 5.2 mW) and with exposure times of 200 ms per laser. We obtained full z stacks of the suprabasal epidermis every 20 min (to limit phototoxicity) for 16 to 20 hours. For photobleaching experiments of sensor-labeled KGs in mouse skin, we followed the procedure previously described for the analysis of sensor recovery half-lives in culture. For pH-shift experiments, we first isolated the epidermal skin layer by dispase treatment of E18.5 whole skin. The tissue was mounted with the stratum corneum facing upward in an otherwise identical manner to the approach described for imaging of whole skin. The tissue was initially imaged using CnT-Airlift media (pH 7.4) before adding an equal volume of acidic CnT-Airlift media (regular media but supplemented for buffering of intracellular pH by adding 280 mM KCl, 20 μM nigericin and HCl to reach a pH of ~3.3) to set the final media pH to ~6.2 to 6.4. Upon the pH shift, the tissue was imaged every 5 min for 50 min under usual live imaging conditions. For pH-shift experiments with primary human keratinocytes, we used the same approach as described for mouse epidermis. For HaCATs with engineered KGs, we performed the pH-shift experiments as before, but with acidic CnT-PR-D supplemented with 1.5 mM $CaCl_2$. Live imaging data of the thick epidermis were typically presented and analyzed from 3D projections of the raw (without rendering) fluorescent data. These 3D views and additional surface renderings were built using Imaris software (version 8.3.1). For live imaging of cells in culture, we typically presented (and indicated so in the legends) maximum intensity projections prepared with ImageJ.

Selection and synthesis of pH reporters

For the synthesis of genetically encoded pH reporters that sensitively respond with a pK_a near 6.5, we chose two previously published and well-characterized pH reporters: SEpHLuorin (*60*) and mNectarine (*61*). We PCR-amplified genes encoding these proteins from Addgene plasmids (#58500 and #80151, respectively) and cloned them into pMAX vectors downstream of a cytomegalovirus (CMV) promoter. For expression of pH reporters in mouse skin throughout epidermal differentiation, we subcloned genes encoding pH reporters into our TRE3G-driven pLKO.1-based vectors and lentivirally transduced embryonic mouse skin as in our previous experiments. We note that these pH reporters are not ratiometric and do not report absolute pH but rather relative changes in pH. However, because we use them for live imaging, we can confidently identify relative changes in intracellular pH by comparing changes in reporter fluorescence within individual cells over time. This approach accounts for the intrinsic limitation of non-ratiometric pH reporters—namely, that the total fluorescent signal varies based on expression levels at the single-cell level. In our approach, rapid changes in fluorescent signal are interpreted as relative changes in pH by correcting for the intensity of the reporter within each cell in time points immediately before the event. Addgene plasmid # 58500 was a gift from A. Cohen. Addgene plasmid # 80151 was a gift from S. Di Pietro.

Design and synthesis of conventional client proteins for KGs

FLG variants that are uniquely bound (with low affinity) by conventional clients were synthesized as part of pMAX vectors and as previously described for other FLG repeat proteins —see their full sequences in table S5. Briefly, these filaggrin scaffold proteins carry short unique domains recognized by the client (either the cleavage sequence for TEV protease or the murine S100 domain). Genes encoding clients were synthesized as IDT gblocks and cloned into pMAX vectors using the same cloning approach as previously described for phase separation sensors. The sequence details of each client, either a dead-variant of Tobacco Etch Virus Protease (dTEVP) or a mS100 domain, are also included in table S5. Although the dTEVP client was exclusively studied in immortalized human keratinocytes (using transfection of corresponding pMAX vectors), for the mS100-based client, which has affinity for endogenous mouse filaggrin, we also subcloned genes encoding this client into our TRE3G-driven pLKO.1-based vectors for lentiviral transduction of the embryonic murine epidermis.

Immunofluorescence of fixed cells and tissues

To prepare HaCATs for immunostaining, we fixed cultures at 37°C for 10 min using 4% paraformaldehyde in Dulbecco's phosphate-buffered saline (DPBS). Cultures were washed with DPBS and stored at 4°C before immunostaining. To prepare murine skin for whole-mount immunostaining, we treated whole skin with dispase for 30 min at 37°C to isolate the epidermis. We fixed the epidermis at 37°C for 30 min in 4% paraformaldehyde. After subsequent washes in DPBS, we stored the tissue at 4°C in DPBS before immunostaining. In all cases, we permeabilized the tissue with an antibody blocking buffer for 3 to 4 hours before overnight incubation with primary antibodies. The following primary antibodies were used: chicken anti-GFP (1:2000, Abcam), rabbit anti-RPTN (1:200, Sigma HPA030483), rabbit anti-mFLG (1:1000, Fuchs Lab), rabbit anti-mFLG (1:1000, Abcam ab24584), and goat anti-hFLG (1:200, Santa Cruz, sc-25897). After washing with DPBS, we added species-specific secondary antibodies conjugated to RRX or AF647 and incubated the cultures and tissues for 4 hours at room temperature. After washing with DAPI, the samples were mounted with ProLong Gold Antifade Mountant (Invitrogen) and cured overnight before imaging. For filaggrin immunostaining without secondary antibodies (i.e., direct detection) in mouse skin, we first conjugated anti-mFLG (abcam) to AF647 using an Alexa Fluor 647 Antibody Labeling Kit (ThermoFisher) and following the instructions of the manufacturer. Cultured cells and whole-mounted fixed tissues were imaged using a spinning-disk microscope equipped with a 40× oil objective. Images were analyzed using ImageJ and Imaris 8.3.1.

Skin barrier assay

To measure barrier quality, we obtained trans-epidermal water loss measurements (TEWL) using a Tewameter TM 300 (Courage + Khazaka electronic GmbH) on explanted neonatal back skin. Briefly, neonates were humanely sacrificed and their back skin was harvested and immediately spread over a clean surface. We collected four TEWL measurements per sample on fully acclimatized skin. The values reported by the instrument were not further processed and corresponded to grams of lost water per hour per m^2 of skin. We measured two or three animals in three independent experiments.

Statistical analyses

Whenever we indicate statistical significance, these are cases where we reject, with a confidence greater than 0.05 (i.e., $p < 0.05$) the null hypothesis that the difference in the mean values between two datasets is equal to zero. To perform this hypothesis testing, we ran two-sample *t* tests using OriginPro. In all cases, we verified that the statistical differences did not depend on the assumption of equal variance (Welch-correction) between samples.

REFERENCES AND NOTES

1. C. P. Brangwynne *et al*., Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009). doi: 10.1126/science.1172046; pmid: 19460965
2. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017). doi: 10.1038/nrm.2017.7; pmid: 28225081
3. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017). doi: 10.1126/science.aaf4382; pmid: 28935776

4. M. Feric *et al*., Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016). doi: 10.1016/j.cell.2016.04.047; pmid: 27212236
5. X. Su *et al*., Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595–599 (2016). doi: 10.1126/science.aad9964; pmid: 27056844
6. J. T. Wang *et al*., Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in C. elegans. *eLife* **3**, e04591 (2014). doi: 10.7554/eLife.04591; pmid: 25535836
7. A. Molliex *et al*., Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015). doi: 10.1016/j.cell.2015.09.015; pmid: 26406374
8. B. R. Sabari *et al*., Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018). doi: 10.1126/science.aar3958; pmid: 29930091
9. W.-K. Cho *et al*., Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018). doi: 10.1126/science.aar4199; pmid: 29930094
10. B. A. Gibson *et al*., Organization of chromatin by intrinsic and regulated phase separation. *Cell* **179**, 470–484.e21 (2019). doi: 10.1016/j.cell.2019.08.037; pmid: 31543265
11. G. Wan *et al*., Spatiotemporal regulation of liquid-like condensates in epigenetic inheritance. *Nature* **557**, 679–683 (2018). doi: 10.1038/s41586-018-0132-0; pmid: 29769721
12. Y. Lin, D. S. Protter, M. K. Rosen, R. Parker, Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015). doi: 10.1016/j.molcel.2015.08.018; pmid: 26412307
13. H. Jiang *et al*., Phase transition of spindle-associated protein regulate spindle apparatus assembly. *Cell* **163**, 108–122 (2015). doi: 10.1016/j.cell.2015.08.010; pmid: 26388440
14. A. K. Rai, J.-X. Chen, M. Selbach, L. Pelkmans, Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216 (2018). doi: 10.1038/s41586-018-0279-8; pmid: 29973724
15. O. Beutel, R. Maraspini, K. Pombo-García, C. Martin-Lemaitre, A. Honigmann, Phase Separation of Zonula Occludens Proteins Drives Formation of Tight Junctions. *Cell* **179**, 923–936.e11 (2019). doi: 10.1016/j.cell.2019.10.011; pmid: 31675499
16. E. S. Freeman Rosenzweig *et al*., The eukaryotic CO2-concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell* **171**, 148–162.e19 (2017). doi: 10.1016/j.cell.2017.08.008; pmid: 28938114
17. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019). doi: 10.1016/j.cell.2018.12.035; pmid: 30682370
18. H. B. Schmidt, A. Barreau, R. Rohatgi, Phase separation-deficient TDP43 remains functional in splicing. *Nat. Commun.* **10**, 4890 (2019). doi: 10.1038/s41467-019-12740-2; pmid: 31653829
19. D. Bracha *et al*., Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467–1480.e13 (2018). doi: 10.1016/j.cell.2018.10.048; pmid: 30500534
20. J. A. Segre, Epidermal barrier formation and recovery in skin disorders. *J. Clin. Invest.* **116**, 1150–1158 (2006). doi: 10.1172/JCI28521; pmid: 16670755
21. B. A. Dale, K. A. Resing, R. B. Presland, in *The Keratinocyte Handbook*, I. Leigh, E. B. Lane, F. M. Watt, Eds. (Cambridge Univ. Press, 1994), chap. 17, pp. 323–350.
22. F. G. Quiroz, A. Chilkoti, Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat. Mater.* **14**, 1164–1171 (2015). doi: 10.1038/nmat4418; pmid: 26390327
23. C. N. Palmer *et al*., Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat. Genet.* **38**, 441–446 (2006). doi: 10.1038/ng1767; pmid: 16550169
24. S. J. Brown, W. H. McLean, One remarkable molecule: Filaggrin. *J. Invest. Dermatol.* **132**, 751–762 (2012). doi: 10.1038/jid.2011.393; pmid: 22158554
25. D. J. Margolis *et al*., Filaggrin-2 variation is associated with more persistent atopic dermatitis in African American subjects. *J. Allergy Clin. Immunol.* **133**, 784–789 (2014). doi: 10.1016/j.jaci.2013.09.015; pmid: 24184149
26. S. Rahrig *et al*., Transient epidermal barrier deficiency and lowered allergic threshold in filaggrin-hornerin ($FlgHrnr^{-/-}$) double-deficient mice. *Allergy* **74**, 1327–1339 (2019). doi: 10.1111/all.13756; pmid: 30828807
27. X. F. C. C. Wong *et al*., Array-based sequencing of filaggrin gene for comprehensive detection of disease-associated variants. *J. Allergy Clin. Immunol.* **141**, 814–816 (2018). doi: 10.1016/j.jaci.2017.10.001; pmid: 29056476
28. C.-A. Lo *et al*., Quantification of protein levels in single living cells. *Cell Rep.* **13**, 2634–2644 (2015). doi: 10.1016/j.celrep.2015.11.048; pmid: 26686644
29. C. G. Bunick *et al*., Crystal structure of human profilaggrin S100 domain and identification of target proteins annexin II, stratifin, and HSP27. *J. Invest. Dermatol.* **135**, 1801–1809 (2015). doi: 10.1038/jid.2015.102; pmid: 25760235
30. S. F. Banani *et al*., Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016). doi: 10.1016/j.cell.2016.06.010; pmid: 27374333
31. B. S. Schuster *et al*., Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**, 2985 (2018). doi: 10.1038/s41467-018-05403-1; pmid: 30061688
32. T. Christensen, W. Hassouneh, K. Trabbic-Carlson, A. Chilkoti, Predicting transition temperatures of elastin-like polypeptide fusion proteins. *Biomacromolecules* **14**, 1514–1519 (2013). doi: 10.1021/bm400167h; pmid: 23565607
33. B. R. McNaughton, J. J. Cronican, D. B. Thompson, D. R. Liu, Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6111–6116 (2009). doi: 10.1073/pnas.0807883106; pmid: 19307578
34. J. Jaubert, S. Patel, J. Cheng, J. A. Segre, Tetracycline-regulated transactivators driven by the involucrin promoter to achieve epidermal conditional gene expression. *J. Invest. Dermatol.* **123**, 313–318 (2004). doi: 10.1111/j.0022-202X.2004.23203.x; pmid: 15245431
35. C. Bonnart *et al*., Elastase 2 is expressed in human and mouse epidermis and impairs skin barrier function in Netherton syndrome through filaggrin and lipid misprocessing. *J. Clin. Invest.* **120**, 871–882 (2010). doi: 10.1172/JCI41440; pmid: 20179351
36. P. Rompolas *et al*., Spatiotemporal coordination of stem cell commitment during epidermal homeostasis. *Science* **352**, 1471–1474 (2016). doi: 10.1126/science.aaf7012; pmid: 27229141
37. T. Kartasova, D. R. Roop, K. A. Holbrook, S. H. Yuspa, Mouse differentiation-specific keratins 1 and 10 require a preexisting keratin scaffold to form a filament network. *J. Cell Biol.* **120**, 1251–1261 (1993). doi: 10.1083/jcb.120.5.1251; pmid: 7679677
38. V. Kumar *et al*., A keratin scaffold regulates epidermal barrier formation, mitochondrial lipid composition, and activity. *J. Cell Biol.* **211**, 1057–1075 (2015). doi: 10.1083/jcb.201404147; pmid: 26644517
39. C.-H. Lee, M.-S. Kim, B. M. Chung, D. J. Leahy, P. A. Coulombe, Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. *Nat. Struct. Mol. Biol.* **19**, 707–715 (2012). doi: 10.1038/nsmb.2330; pmid: 22705788
40. M. P. Hughes *et al*., Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018). doi: 10.1126/science.aan6398; pmid: 29439243
41. S. Noda *et al*., The Asian atopic dermatitis phenotype combines features of atopic dermatitis and psoriasis with increased TH17 polarization. *J. Allergy Clin. Immunol.* **136**, 1254–1264 (2015). doi: 10.1016/j.jaci.2015.08.015; pmid: 26428954
42. T. J. Nott *et al*., Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015). doi: 10.1016/j.molcel.2015.01.013; pmid: 25747659
43. M. Dzuricky, S. Roberts, A. Chilkoti, Convergence of artificial protein polymers and intrinsically disordered proteins. *Biochemistry* **57**, 2405–2414 (2018). doi: 10.1021/acs.biochem.8b00056; pmid: 29683665
44. R. Niesner *et al*., 3D-resolved investigation of the pH gradient in artificial skin constructs by means of fluorescence lifetime imaging. *Pharm. Res.* **22**, 1079–1087 (2005). doi: 10.1007/s11095-005-5304-6; pmid: 16028008
45. J. A. Mackay, D. J. Callahan, K. N. Fitzgerald, A. Chilkoti, Quantitative model of the phase behavior of recombinant pH-responsive elastin-like polypeptides. *Biomacromolecules* **11**, 2873–2879 (2010). doi: 10.1021/bm100571j; pmid: 20925333
46. S. Alberti, Guilty by association: Mapping out the molecular sociology of droplet compartments. *Mol. Cell* **69**, 349–351 (2018). doi: 10.1016/j.molcel.2018.01.020; pmid: 29395058
47. S. Markmiller *et al*., Context-dependent and disease-specific diversity in protein interactions within stress granules. *Cell* **172**, 590–604.e13 (2018). doi: 10.1016/j.cell.2017.12.032; pmid: 29373831
48. I. Brody, An ultrastructural study on the role of the keratohyalin granules in the keratinization process. *J. Ultrastruct. Res.* **3**, 84–104 (1959). doi: 10.1016/S0022-5320(59)80018-6; pmid: 13804667
49. T. Makino, M. Takaishi, M. Morohashi, N. H. Huh, Hornerin, a novel profilaggrin-like protein and differentiation-specific marker isolated from mouse skin. *J. Biol. Chem.* **276**, 47445–47452 (2001). doi: 10.1074/jbc.M107512200; pmid: 11572870
50. P. M. Steinert, D. A. Parry, L. N. Marekov, Trichohyalin mechanically strengthens the hair follicle: Multiple cross-bridging roles in the inner root sheath. *J. Biol. Chem.* **278**, 41409–41419 (2003). doi: 10.1074/jbc.M302037200; pmid: 12853460
51. B. Mészáros *et al*., PhaSePro: The database of proteins driving liquid–liquid phase separation. *Nucleic Acids Res.* **48**, D360–D367 (2020). doi: 10.1093/nar/gkz848; pmid: 31612960
52. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982). doi: 10.1016/0022-2836(82)90515-0; pmid: 7108955
53. J. R. McDaniel, J. A. Mackay, F. G. Quiroz, A. Chilkoti, Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **11**, 944–952 (2010). doi: 10.1021/bm901387t; pmid: 20184309
54. A. C. Woerner *et al*., Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science* **351**, 173–176 (2016). doi: 10.1126/science.aad2033; pmid: 26634439
55. S. P. Boudko *et al*., Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization Fold. *J. Mol. Biol.* **392**, 787–802 (2009). doi: 10.1016/j.jmb.2009.07.057; pmid: 19631658
56. A. Ghoorchian, N. B. Holland, Molecular architecture influences the thermally induced aggregation behavior of elastin-like polypeptides. *Biomacromolecules* **12**, 4022–4029 (2011). doi: 10.1021/bm201031m; pmid: 21972921
57. N. Maas-Szabowski, A. Stärker, N. E. Fusenig, Epidermal tissue regeneration and stromal interaction in HaCaT cells is initiated by TGF-α. *J. Cell Sci.* **116**, 2937–2948 (2003). doi: 10.1242/jcs.00474; pmid: 12771184
58. J. A. Nowak, E. Fuchs, in *Stem Cells in Regenerative Medicine*. (Springer, 2009), pp. 215-232.
59. S. Beronja, G. Livshits, S. Williams, E. Fuchs, Rapid functional dissection of genetic networks via tissue-specific transduction and RNAi in mouse embryos. *Nat. Med.* **16**, 821–827 (2010). doi: 10.1038/nm.2167; pmid: 20526348
60. S. Sankaranarayanan, D. De Angelis, J. E. Rothman, T. A. Ryan, The use of pHluorins for optical measurements of presynaptic activity. *Biophys. J.* **79**, 2199–2208 (2000). doi: 10.1016/S0006-3495(00)76468-X; pmid: 11023924
61. D. E. Johnson *et al*., Red fluorescent protein pH biosensor to detect concentrative nucleoside transport. *J. Biol. Chem.* **284**, 20499–20511 (2009). doi: 10.1074/jbc.M109.019042; pmid: 19494110

ACKNOWLEDGMENTS

We thank I. Matos for assistance with the live-imaging setup; L. Hidalgo and M. Sribour for animal assistance; and S. Ellis and I. Matos for discussion and comments on the manuscript. We thank the Molecular Cytology Core Facility at MSKCC and Rockefeller University (RU) Bio-Imaging Resource Center for use of microscopes, and RU Comparative Bioscience Center (AAALAC-accredited) for care of mice in accordance with National Institutes of Health (NIH) guidelines. **Funding:** This work was funded by NIH (R01-AR27883 to E.F.) and partly supported by the Robertson Therapeutic Development Fund and by a Tri-Institutional Starr Stem Cell Scholars Fellowship at The Rockefeller University. E.F. is an HHMI investigator. F.G.Q. holds a Career Award at the Scientific Interface from Burroughs Wellcome Fund. **Author contributions:** F.G.Q. and E.F. conceived the project, designed the experiments, and interpreted the data. F.G.Q. and

V.F.F. conceived and performed AFM experiments. J.L. performed lentiviral injections. L.P. established and maintained all relevant mouse lines. E.W. participated in cloning and protein biochemistry. H.A.P. performed and interpreted ultrastructural analyses. F.G.Q. and E.F. wrote the manuscript with contributions from V.F.F. **Competing interests:** E.F. and F.G.Q. are inventors on a provisional patent application, filed by Rockefeller University, for the use of phase-separation sensors described in this work. The authors declare no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials. Plasmids described in the manuscript can be provided by E.F. or F.G.Q. pending a completed material transfer agreement. Requests for these materials should be submitted to fuchslb@rockefeller.edu.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/eaax9554/suppl/DC1
Supplementary text
Figs. S1 to S24
Tables S1 to S5
Movies S1 to S8
References (62–67)

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

7 May 2019; resubmitted 20 November 2019
Accepted 22 January 2020
10.1126/science.aax9554

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Cerebrospinal fluid influx drives acute ischemic tissue swelling

Humberto Mestre*, Ting Du*, Amanda M. Sweeney, Guojun Liu, Andrew J. Samson, Weiguo Peng, Kristian Nygaard Mortensen, Frederik Filip Stæger, Peter A. R. Bork, Logan Bashford, Edna R. Toro, Jeffrey Tithof, Douglas H. Kelley, John H. Thomas, Poul G. Hjorth, Erik A. Martens, Rupal I. Mehta, Orestes Solis, Pablo Blinder, David Kleinfeld, Hajime Hirase, Yuki Mori†, Maiken Nedergaard†

INTRODUCTION: Cerebrospinal fluid (CSF) covers and protects the brain from mechanical injury. CSF also flows along an interconnected network of perivascular spaces surrounding blood vessels and communicates with interstitial fluid permeating brain tissue, aiding in the removal of metabolic waste produced by cells. This glial cell–mediated lymphatic (glymphatic) function of CSF represents a continuous source of fluid and ions for the brain. When a cerebral artery is occluded, nearby brain tissue is abruptly deprived of blood flow, oxygen, and glucose. This process, known as acute ischemic stroke, is a leading cause of morbidity and mortality worldwide. After stroke, fluid accumulates in ischemic tissue, and the brain becomes edematous and begins to swell, a dangerous complication of the disease. In the first hours after occlusion, the degree of swelling correlates with the net gain of cations, primarily sodium, and this gain draws in fluid from surrounding sources.

RATIONALE: Because the brain is already encased by CSF, we asked if glymphatic flow could play a role in early edema formation. To test this, we evaluated CSF dynamics using in vivo magnetic resonance (MR) and multimodal optical imaging after occluding the middle cerebral artery in mice. Edema was assessed using diffusion-weighted MR, and edema fluid sources were labeled using radionuclides. Changes in the flow of CSF in perivascular spaces were explored using a network model of the mouse middle cerebral artery. Histology was used to evaluate edema formation in regions adjacent to CSF inflow routes in mouse and human autopsy tissue.

RESULTS: We found that within minutes of ischemic stroke, CSF flowed rapidly into brain tissue along perivascular spaces. Its entry coincided with the onset of swelling and increased brain water content. Radionuclides and multimodal imaging confirmed that CSF was the earliest contributor of both fluid and ions. Calcium imaging in transgenic mice expressing GCaMP7 in cortical neurons and astrocytes revealed that this process was initiated by spreading depolarizations that were triggered when tissue was deprived of blood flow. Diffusion-weighted MR imaging showed that this was the earliest phase of edema formation. This aberrant CSF inflow was found to be caused by spreading ischemia, the pathological constriction of cerebral blood vessels that follows spreading depolarizations. We present a network model that predicts that the space left unoccupied after vessels constrict would be filled by an inrush of CSF that nearly doubles flow speed. That prediction was confirmed experimentally using particle tracking velocimetry of CSF flow in live mice. Inflow depended on the aquaporin-4 water channel that is highly expressed by glial cells (astrocytes), which is a key contributor to glymphatic function. Postmortem examination of rodent and human brains showed increased fluid accumulation in tissue surrounding perivascular spaces and the cerebral ventricles compared with regions deep in the brain that were far from large CSF reservoirs.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax7171

CONCLUSION: Here, we demonstrate that CSF can provide a source of ischemic edema. Glymphatic inflow of CSF appears to be the primary initial event driving tissue swelling. This finding challenges our current understanding of edema formation after stroke and may provide a basis for treatment of acute ischemic stroke. Spreading depolarizations continue several days after stroke and are also present in many other neurological conditions, ranging from traumatic brain injury to migraine; therefore, it will be important to determine if spreading edema is also a feature of these diseases and whether CSF influx contributes to worsening at more delayed time points. It is also intriguing to speculate that abnormal CSF inflow could be a source of edema fluid in other types of chronic cerebrovascular disease, such as small-vessel disease characterized by enlarged perivascular spaces and transient accumulations of fluid in periventricular white matter. ■

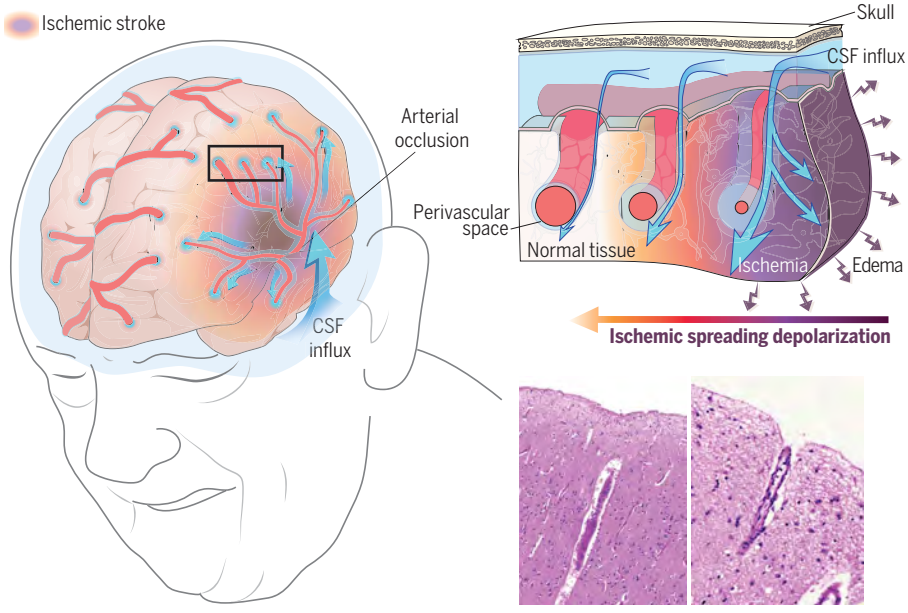


ILLUSTRATION: DAN XUE

CSF influx is responsible for early tissue swelling after stroke. (Left) Spreading ischemia accelerates CSF inflow to the region deprived of blood flow. (Top right) Spreading ischemia constricts the cortical vessels, increasing the perivascular space and resulting in CSF inflow. (Bottom right) Histology of postmortem tissue shows fluid accumulation in the ischemic human brain (diffuse white empty space, right) that is not present in a control brain (left).

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: maiken_nedergaard@urmc.rochester.edu (M.N.); yuki.mori@sund.ku.dk (Y.M.)
Cite this article as H. Mestre *et al.*, *Science* **367**, eaax7171 (2020). DOI: 10.1126/science.aax7171

RESEARCH ARTICLE

NEUROSCIENCE

Cerebrospinal fluid influx drives acute ischemic tissue swelling

Humberto Mestre[1,2]*, Ting Du[1,3]*, Amanda M. Sweeney[1], Guojun Liu[1,4], Andrew J. Samson[5], Weiguo Peng[5], Kristian Nygaard Mortensen[5], Frederik Filip Stæger[5], Peter A. R. Bork[5,6], Logan Bashford[7], Edna R. Toro[7], Jeffrey Tithof[7], Douglas H. Kelley[7], John H. Thomas[7], Poul G. Hjorth[6], Erik A. Martens[6], Rupal I. Mehta[1,2,8,9], Orestes Solis[2], Pablo Blinder[10,11], David Kleinfeld[12,13], Hajime Hirase[5,14], Yuki Mori[5]†, Maiken Nedergaard[1,5]†

Stroke affects millions each year. Poststroke brain edema predicts the severity of eventual stroke damage, yet our concept of how edema develops is incomplete and treatment options remain limited. In early stages, fluid accumulation occurs owing to a net gain of ions, widely thought to enter from the vascular compartment. Here, we used magnetic resonance imaging, radiolabeled tracers, and multiphoton imaging in rodents to show instead that cerebrospinal fluid surrounding the brain enters the tissue within minutes of an ischemic insult along perivascular flow channels. This process was initiated by ischemic spreading depolarizations along with subsequent vasoconstriction, which in turn enlarged the perivascular spaces and doubled glymphatic inflow speeds. Thus, our understanding of poststroke edema needs to be revised, and these findings could provide a conceptual basis for development of alternative treatment strategies.

Stroke is among the leading causes of death worldwide, affecting 10 million patients annually (*1*, *2*). Recent strides in treatment have improved these statistics, but stroke remains a principal cause of long-term disability (*1*, *3*). A detrimental complication is cerebral edema, the abnormal accumulation of fluid that leads to secondary ischemia, additional tissue loss, and potential death (*4*, *5*). Targeting edema represents a promising therapeutic strategy, because the severity of swelling predicts long-term functional outcomes (*6*–*8*). Although edema also develops in a range of other central nervous system diseases (e.g., trauma, tumors, and infections), treatment options remain limited, and those available are suboptimal. Traditionally, cerebral edema is divided into two distinct phases: an early cytotoxic phase and a later vasogenic phase (*9*). Cytotoxic edema occurs within minutes of an ischemic insult and is triggered by spreading depolarizations (SDs) that result from dysregulated ion homeostasis causing cell swelling (*6*, *7*, *10*–*14*). During the subsequent phase, vasogenic edema, fluid from the blood enters the brain as a result of blood-brain barrier (BBB) breakdown (*15*). Ions and fluid can cross the BBB via transcellular and paracellular routes (*12*): The transcellular pathway allows for the early entry of plasma proteins and other osmotically active solutes facilitating fluid entry, whereas the paracellular pathway takes 2 days to become active (*16*, *17*). However, edema actually develops several hours before significant BBB dysfunction (*18*). This terminology used for staging edema was defined in the 1960s, and the conceptual framework for understanding and treating edema is still based on these early observations (*8*, *9*, *19*). The term ionic edema was introduced to explain this intermediary phase, during which tissue Na^+ content increases (*5*, *18*). This generates an osmotic gradient with plasma, allowing water to move into the brain across an intact BBB (*20*, *21*). However, the source of the excess Na^+ has been attributed exclusively to influx from the intravascular compartment (*5*). Yet, the brain is bathed in cerebrospinal fluid (CSF), which has a high Na^+ concentration and accounts for almost 10% of intracranial fluid (*22*). CSF has not been recognized as a source of edema fluid (*8*), but the notion that perivascular spaces (PVSs) provide a conduit for CSF influx prompted us to reassess this idea. Glymphatic exchange of CSF with the interstitial fluid (ISF) provides a means by which CSF in large quantities may rapidly enter the brain and drive tissue swelling (*23*, *24*). Here, we found that CSF can indeed provide a key source for the initial rise in brain water content in the ischemic brain.

CSF flows into the brain after stroke, driving acute tissue swelling

To determine whether CSF contributes to edema in the early phases of acute ischemic stroke, the middle cerebral artery (MCA) was occluded by injecting macrospheres into the right common carotid artery in wild-type mice (Fig. 1, A and B) (*25*). Embolic MCA occlusion (MCAO) resulted in an immediate, steep reduction in relative cerebral blood flow (rCBF: 92.5 ± 1.2%; Fig. 1C). Mice exhibited severe neurological deficits the following day, and the infarct occupied most of the vascular territory of the MCA (fig. S1). We used this model to map influx of CSF tagged with a fluorescent tracer immediately after MCAO. The CSF tracer was injected 15 min before occlusion to ensure that it had distributed along the circle of Willis but not yet transported up along the MCA (Fig. 1D). Imaging revealed no difference in CSF tracer entry between ipsilateral and contralateral hemispheres before stroke, but MCAO resulted in a sharp threefold increase in ipsilateral CSF tracer influx compared with the contralateral hemisphere (Fig. 1, E and F). The temporal derivative of the fluorescence curves identified that CSF influx occurred at two separate time points: an early peak (11.4 ± 1.8 s) and a second peak (5.24 ± 0.48 min) after MCAO (Fig. 1, F and G). We next asked whether this CSF influx contributed to edema. The tissue water content of the ischemic cortex increased rapidly 15 min after occlusion, whereas the water content in the nonischemic hemisphere remained constant (Fig. 1H). To assess the contribution of blood versus CSF to the rapid increase in water content, the two fluid compartments were tagged by either intravenous or intracisternal injection of ^{22}Na and ^{3}H-mannitol. Early ischemic Na^+ accumulation is an indirect measure of edema (*5*, *18*, *26*); mannitol, on the other hand, is a small BBB-impermeable tracer whose accumulation reflects opening of the barrier (Fig. 1I) (*27*). ^{22}Na and ^{3}H-mannitol accumulation in the ischemic and nonischemic hemispheres was directly comparable after intravenous injection, suggesting that the vascular compartment was not the source of edema fluid in early times (Fig. 1, J and K). By contrast, both ^{22}Na and ^{3}H-mannitol accumulated to a greater degree in the ischemic hemisphere when the tracers were delivered to the CSF (Fig. 1, L and M). To determine which pool of CSF

[1]Center for Translational Neuromedicine, Department of Neurosurgery, University of Rochester Medical Center, Rochester, NY 14642, USA. [2]Department of Neuroscience, University of Rochester Medical Center, Rochester, NY 14642, USA. [3]School of Pharmacy, China Medical University, Shenyang 110122, China. [4]Department of Neurosurgery, the Fourth Affiliated Hospital of China Medical University, Shenyang 110032, China. [5]Center for Translational Neuromedicine, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. [6]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, 2800 Kgs. Lyngby, Denmark. [7]Department of Mechanical Engineering, University of Rochester, Rochester, NY 14627, USA. [8]Department of Pathology, Rush University, Chicago, IL 60612, USA. [9]Rush Alzheimer's Disease Center, Rush University, Chicago, IL 60612, USA. [10]Neurobiology, Biochemistry and Biophysics School, George S. Wise Faculty of Life Sciences, Tel Aviv University, 30 Haim Levanon St., Tel Aviv 69978, Israel. [11]Sagol School for Neuroscience, Tel Aviv University, 30 Haim Levanon St., Tel Aviv 69978, Israel. [12]Department of Physics, University of California, San Diego, La Jolla, CA 92093, USA. [13]Section of Neurobiology, University of California, San Diego, La Jolla, CA 92093, USA. [14]Laboratory for Neuron-Glia Circuitry, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.
*These authors contributed equally to this work.
†Corresponding author. Email: maiken_nedergaard@urmc.rochester.edu (M.N.); yuki.mori@sund.ku.dk (Y.M.)

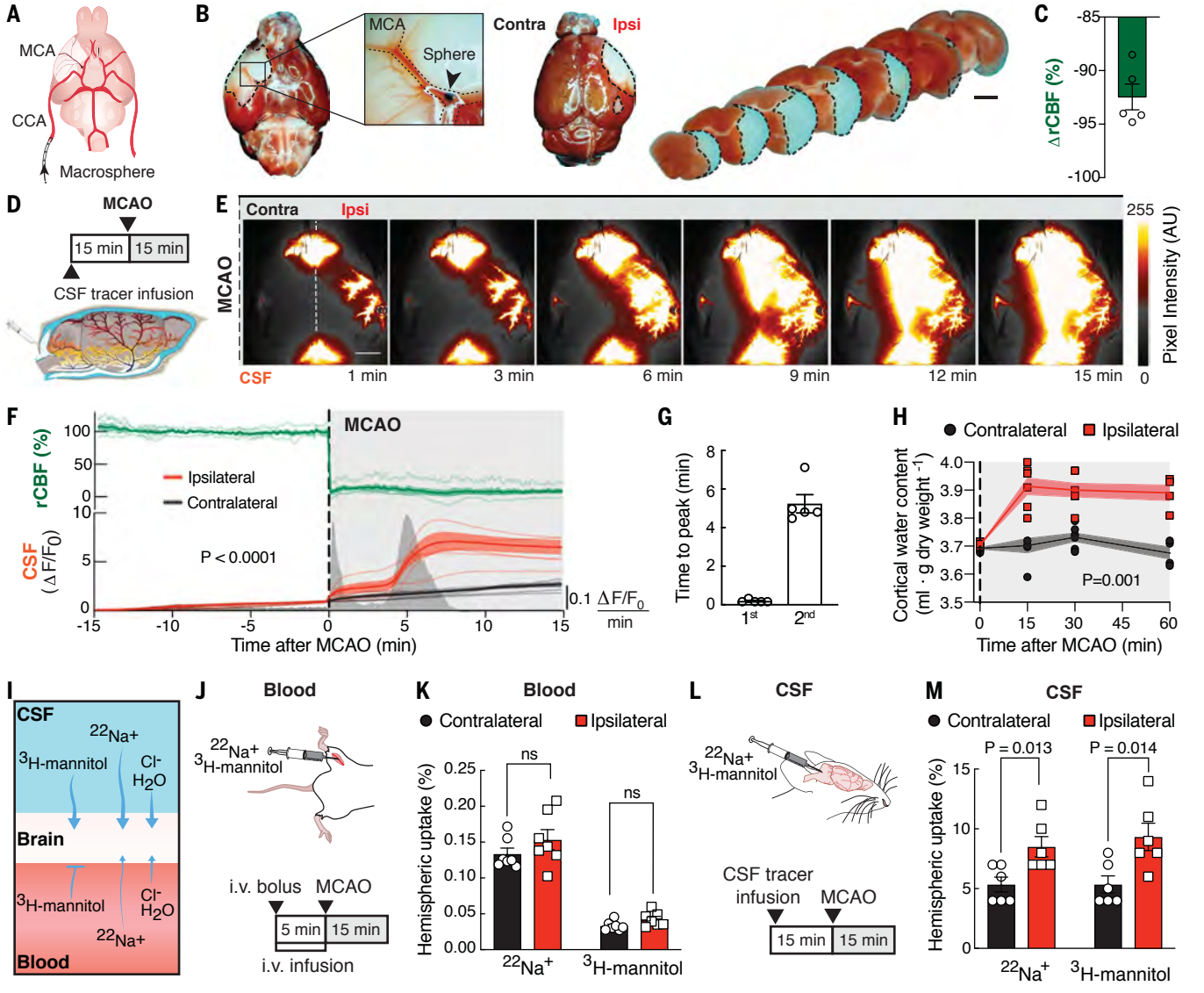


Fig. 1. CSF rapidly enters the brain after stroke, resulting in edema. (**A**) Macrospheres (arrow) were infused into the common carotid artery (CCA), occluding the MCA. (**B**) Triphenyltetrazolium chloride staining after MCAO. Contra, contralateral; Ipsi, ipsilateral. (**C**) Ipsilateral rCBF after MCAO. (**D**) Intracisternal Alexa Fluor647–conjugated bovine serum albumin (BSA-647) was infused 15 min before MCAO. (**E**) Ipsilateral and contralateral tracer influx was imaged while measuring rCBF. (The white line indicates the sagittal suture and fiber optic probe.) AU, arbitrary units. (**F**) Time series of rCBF and fluorescence intensity ($\Delta F/F_0$) in each hemisphere. Gray peaks are the temporal derivative of fluorescence intensity. Repeated measures two-way analysis of variance (ANOVA) was performed; interaction P value < 0.0001; n = 5 mice. (**G**) Timing of the first and second peaks. (**H**) Water content of the cortex. Mixed-effects repeated measures two-way ANOVA with Sidak's multiple comparisons test was performed; n = 4 to 7 mice per time point. The shaded regions above and below the plot lines indicate SEM. (**I**) The source of the edema fluid was identified by labeling either the blood or CSF compartment. (**J**) $^{22}Na^+$ and ^{3}H-mannitol were delivered intravenously (i.v.). (**K**) Percentage of the total injected radiation found in each hemisphere after intravenous delivery. Student's t tests with Holm-Sidak correction were performed; n = 7 mice. ns, not significant. (**L**) Isotopes were delivered into CSF. (**M**) Percentage of the total injected radiation in each hemisphere after intracisternal delivery. Student's t tests with Holm-Sidak correction were performed; n = 6 mice. In (B) and (E), scale bars are 2 mm. In (C), (G), (K), and (M), error bars represent SEM.

was the primary source of edema fluid, we performed CSF cisternography using three-dimensional (3D) fast imaging employing steady-state acquisition (3D-FIESTA) magnetic resonance imaging (MRI). The CSF compartment—including the lateral, third, and fourth ventricles and the cisterna magna—accounted for a total of 13 ± 1 mm^3 (Fig. 2A). Smaller collections of fluid (7.6 ± 1.3 mm^3) were observed along the PVS of the large arteries of the circle of Willis, and the rest of the intracranial volume was occupied by the brain tissue, including the intravascular compartment (509 ± 11 mm^3). After MCAO, CSF in the lateral ventricle (Fig. 2B and Movie 1) of the ischemic hemisphere and in the cisterna magna (Fig. 2C) slowly disappeared. Total CSF volume in the ventricular system and cisterna decreased by 9.0 ± 2.5% at 15 min and 17 ± 3% at 29 min poststroke (Fig. 2, D and E). PVS fluid volume began decreasing at around 15 min (Fig. 2F), several minutes after the CSF compartment started shrinking (~5 min). Combined, these results add additional support to the notion that CSF is the primary source of early edema fluid after ischemic stroke.

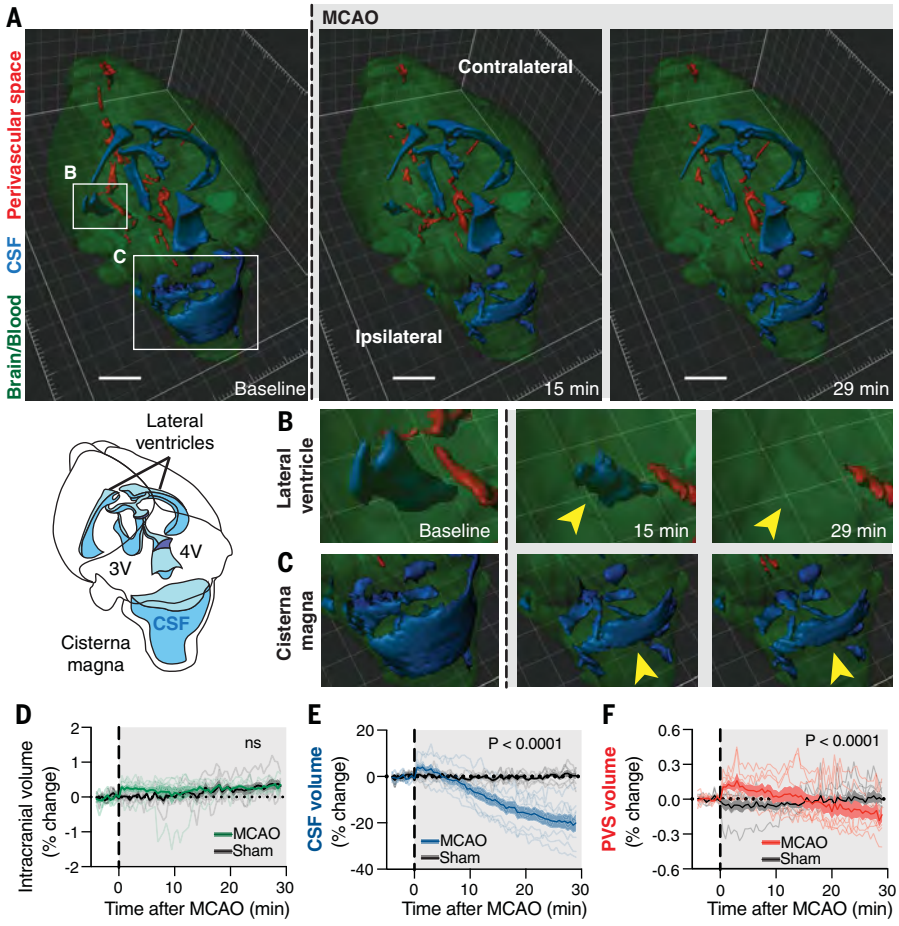
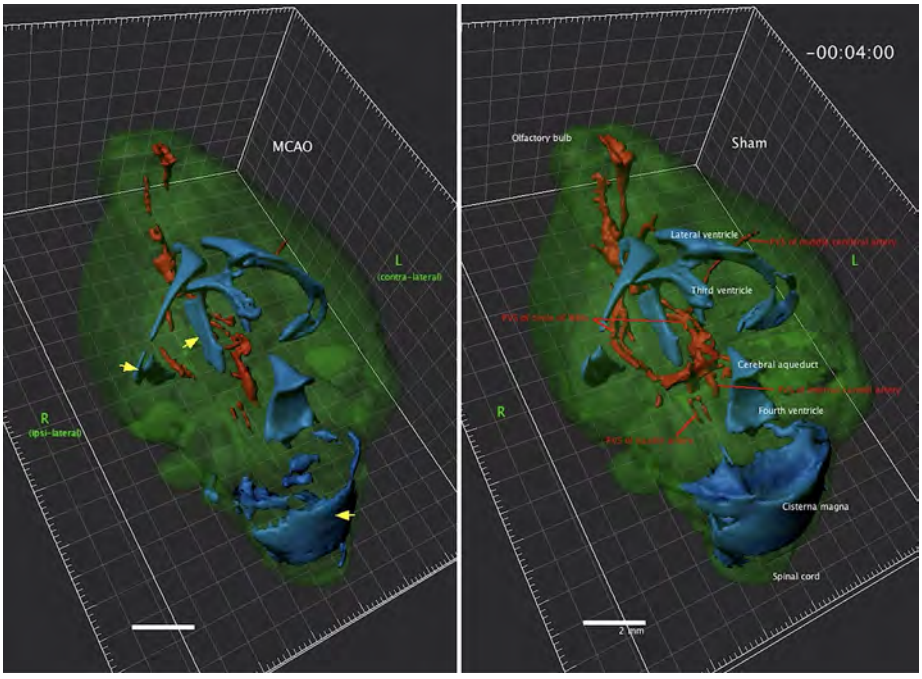


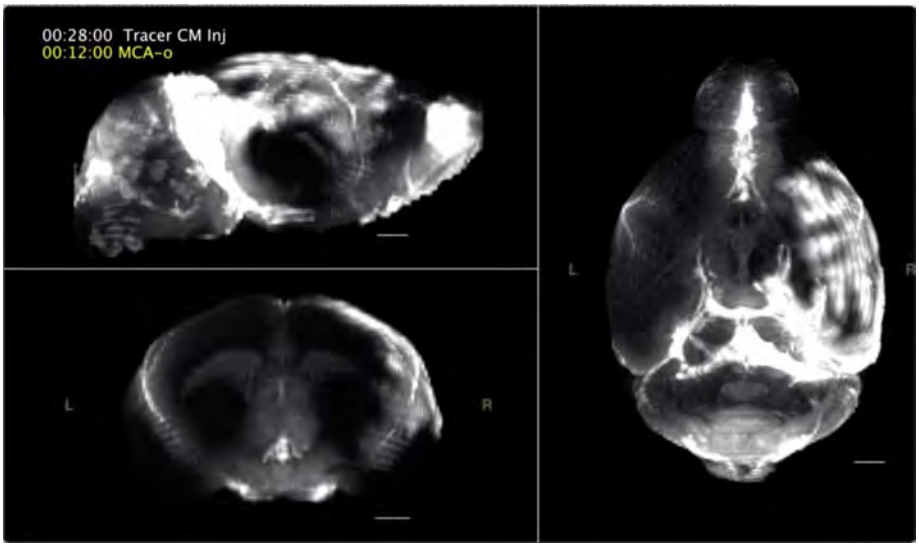
Fig. 2. After MCAO, CSF shifts into other intracranial compartments. (**A** to **C**) 3D-FIESTA shows three main compartments: (i) intraventricular CSF in the lateral, third (3V), and fourth (4V) ventricles and cisternal CSF in the cisterna magna (blue); (ii) free fluid in the PVSs (red) of the circle of Willis along the anterior, middle, posterior cerebral, and basilar arteries; and (iii) intracranial content primarily composed of brain tissue and cerebral blood volume (green). 3D-FIESTA at baseline before MCAO and 15 and 29 min later is shown. Scale bar, 2 mm. Insets in (A) of the ventral anterior horn of the ipsilateral lateral ventricle (B) and the cisterna magna (C) demonstrate the loss of free water at 15 and 29 min after MCAO, most notably in the lateral ventricles and the cisternal magna (yellow arrows). (**D**) Percent change of intracranial volume after MCAO and in sham animals. Repeated measures two-way ANOVA was performed; interaction *P* value not significant (ns). (**E**) Percent change of CSF volume after MCAO. Repeated measures two-way ANOVA was performed; interaction *P* value < 0.0001. (**F**) Percent change of PVS volume after MCAO. Repeated measures two-way ANOVA was performed; interaction *P* value < 0.0001. Time-lapse measurements from *n* = 7 mice in the sham group and 8 mice in the MCAO group were collected. The shaded regions above and below the plot lines indicate SEM.

SDs trigger CSF flow into the ischemic brain

MRI suggested that the volume of CSF rapidly declined upon artery occlusion, so we next used dynamic contrast-enhanced MRI to determine the most likely location of CSF brain entry (Movie 2 and fig. S2A). MRI analysis confirmed an abrupt increase in ipsilateral perivascular influx of a CSF-delivered contrast agent (gadobutrol) within the first 5 min after occlusion (fig. S2B). Significant contrast enhancement was found at all measured sites along the perivascular trajectory of the ipsilateral MCA: ventral segment of the MCA [region of interest 1 (ROI1): 2.9-fold compared with the contralateral hemisphere], convexity MCA (ROI2: 4.3-fold), and far from the main MCA trunk in the parenchyma (ROI3: 5.5-fold; fig. S2C). Tracer signal was seen deep below the cortical surface of the ipsilateral hemisphere, reaching 450 ± 43 μm at 14 min and 580 ± 58 μm at 28 min after MCAO, indicating that CSF enters the brain tissue (fig. S3, A to E). The entry route was along the glymphatic pathway, with tracer influx occurring primarily along penetrating arterioles (97.7%, versus 2.3% along venules; $P < 0.0001$; fig. S4, A to D) (*28*). Tracers moved along PVSs of the branching vascular network, including capillaries, and were found deep within parenchyma 30 min after MCAO (fig. S4B). Thus, several independent sets of analyses point to acceleration of glymphatic influx into the ipsilateral brain in the setting of focal ischemia (*6*, *8*). To identify the driver of CSF influx, we speculated that a loss in blood flow after stroke could cause a hydrostatic pressure gradient that would facilitate CSF flow into the brain. To test this, we measured intracranial pressure (ICP) changes during MCAO (fig. S5A). ICP dropped from 5.49 ± 0.85 to 4.63 ± 0.88 mmHg after MCAO (fig. S5, A and B), but this response was variable between mice and, in some cases, returned to baseline within a few minutes. Thus, the ICP decrease coincided with the first influx peak but could not explain the second, larger peak in CSF influx (Fig. 1F and fig. S5C). We next asked whether this large influx of CSF contributed to the cytotoxic phase of edema. To address this question, cytotoxic edema was detected as a decrease in the apparent diffusion coefficient (ADC) using diffusion-weighted MRI (*12*). A large ADC lesion first appeared in the primary somatosensory cortex after MCAO and then slowly expanded in the form of a spreading wave across the ipsilateral hemisphere (Movie 3 and fig. S2, D and E). ADC dropped by 17.5 ± 1.3% within the lesion and spread over 22 ± 2% of the brain volume within the first 4.8 ± 0.9 min after stroke (fig. S2, F and G). We generated a registered average of the contrast-enhanced and ADC datasets to compare the evolution of cytotoxic edema with CSF influx (Movie 4). The aligned contrast-enhanced and ADC data and the onset times of both indicated that the ADC changes happened simultaneously with the second peak of CSF entry ($P = 0.70$; Fig. 1, F and G). Does the ADC drop trigger the large CSF influx? The wave-like kinetics of the ADC decrease is caused by a SD (*29*). Ischemic SD consists of waves of sustained, mass depolarizations of cells in the gray matter of the central nervous system that result from the near-complete loss of cellular transmembrane ion gradients (*29*). Ischemic SD begins in the barrel cortex of the primary somatosensory cortex in rodents (Movie 3) (*30*) and is thought to trigger cytotoxic edema after ischemia (*12*). Genetically encoded calcium indicators can be used to visualize the mass depolarization caused by the SD wave (*31*). To directly test the hypothesis that SD drives glymphatic influx after stroke, we used mice that express GCaMP7 under the *Glt1* promoter (GCaMP) (*32*). GCaMP mice received an intracisternal tracer injection 15 min before MCAO and were imaged using dual-channel macroscopic imaging (Fig. 3A and Movie 5). Several minutes after MCAO, a spreading wave of GCaMP fluorescence was seen, starting at the primary sensory cortex and traveling slowly over the cortical surface, with CSF tracer following closely behind. The area covered by the SD and the expansion kinetics were consistent between mice, but the onset time varied (Fig. 3B). Aligning the data to the



Movie 1. CSF cisternography during MCAO. Mice that received a right MCAO (left) or sham controls (right) were imaged using 3D-FIESTA MRI. Cisternography revealed that ventricular and subarachnoid CSF (blue) in the ventricles and cisterna magna (yellow arrows) disappeared after stroke, primarily in the ipsilateral hemisphere. Perivascular CSF (red) volume decreased at later time points. The brain tissue and blood compartment (green) increased in volume at the expense of CSF volume decreasing, suggesting that CSF shifts into this compartment. There were no time-dependent changes in the sham mice. Scale bar, 2 mm.

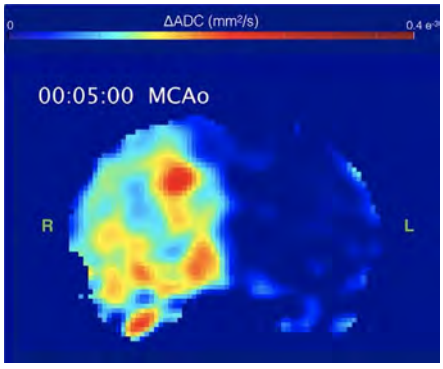


Movie 2. Dynamic contrast-enhanced MRI of intracisternal contrast after stroke. A gadolinium-based tracer (gadobutrol) was injected into the cisterna magna (CM inj) of wild-type mice 15 min before MCAO (labeled MCA-o). Whole-volume sagittal (top left), coronal (bottom left), and dorsoventral (right) projections are shown. Within the first 5 min after MCAO, the tracer is rapidly transported along the perivascular network of the MCA on the ipsilateral hemisphere. Contrast enhancement can be seen over most of the cortical surface and then is seen being distributed in the brain tissue. At later time points, contrast is found outlining the ventricular system, suggesting that labeled cisternal CSF enters the ventricles. R, ipsilateral hemisphere; L, contralateral hemisphere. Scale bar, 2 mm

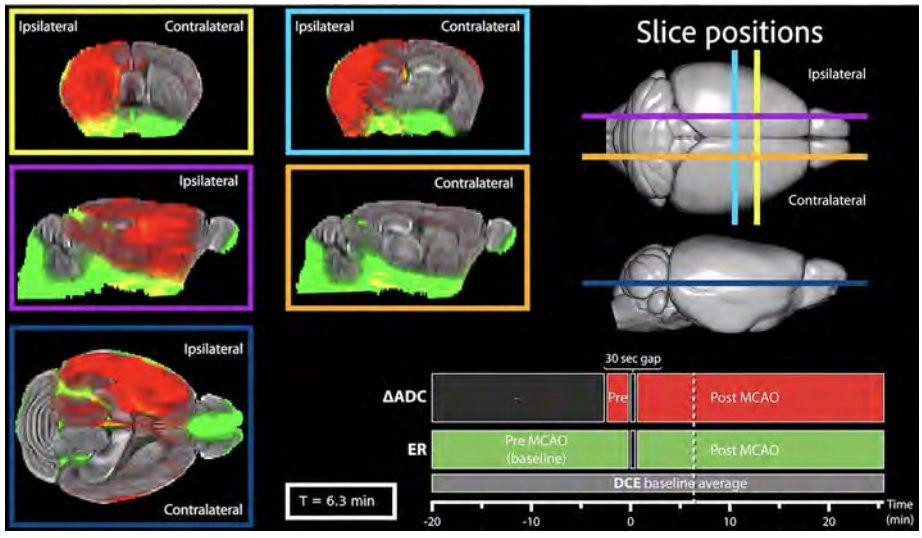
peak of GCaMP coverage revealed that the SD preceded a 5.6-fold increase in the area covered by CSF tracer (Fig. 3C). The SD peak coincided with maximal CSF influx, and SD onset time was a near-perfect predictor of the time of CSF influx [coefficient of determination (R^2) = 0.94; Fig. 3, D and E]. To gain insight into the spatiotemporal dynamics of the SD wave and CSF influx, we used a front-tracking algorithm (*33*). The SD wave propagated at a maximum speed of 36 ± 5 μm/s, consistent with an isotropic reaction-diffusion mechanism (Fig. 3, F to I) (*12*). The CSF tracer front sped up after the onset of the SD wave but reached a much slower maximum speed of 16 ± 1 μm/s, indicating that the SD might not directly drive the CSF front. To quantify this delay, we computed the time it took CSF tracer to reach each given pixel after the SD wave had passed and found that 30 ± 6 s separated the SD and the tracer fronts (Fig. 3, J and K). Thus, the ischemic SD triggers the onset of glymphatic influx after stroke.

Spreading ischemia drives a rapid increase in perivascular CSF flow

A phenomenon termed spreading ischemia (SI) has been shown to take place 10 to 30 s after SD, which would temporally coincide with glymphatic influx (*34*–*36*). Pronounced constriction of parenchymal and pial arterioles occurs after SD, owing to release of K^+ and other vasoactive substances and the depletion of nitric oxide in the endothelium (*35*, *37*). CSF flow within PVSs is highly dependent on arterial diameter changes (*38*, *39*). Could vasoconstriction drive the increase in CSF entry? To address the role of SI in CSF flow, we used two-photon (2P) microscopy (Fig. 4A and Movie 6) (*23*, *24*). Immediately after occlusion, there was a transient loss of intravascular flow as rCBF dropped (Fig. 4, B and C). Despite intravascular dye returning a few seconds later owing to collateral flow, the hemisphere remained severely hypoperfused (Fig. 1F). Several minutes after MCAO, a SD wave propagated across the imaging field, followed by a potent vasoconstriction of the pial and penetrating arterioles, culminating in an intense increase in CSF tracer signal (Fig. 4, D to F, and Movie 7). The propagation speed of the SD wave was comparable to that seen in macroscopic experiments (42 ± 2 μm/s), and the delay between SD and the maximal vasoconstriction (30 ± 2 s) was temporally consistent with the delay between the SD wave and the CSF tracer front from the previous analysis (P = 0.395; Fig. 4, G to I). The vasoconstrictive response was variable between mice and arteriolar location, with penetrating arterioles constricting more (79 ± 10%) and earlier than pial arterioles (45 ± 10%; fig. S6, A to D). The variability of this vasomotor response has been attributed to differences in the intravascular perfusion pressure and the location of vessels relative to the infarct core



Movie 3. ADC maps after focal ischemia. Delta ADC (ΔADC) maps were generated from diffusion-weighted MRI. Coronal (top), dorsal (bottom left), and sagittal projections of the right ipsilateral cortical surface (lower right) are shown. A large decrease in ADC can first be seen in the primary sensory cortex at 1:50 min after MCAO (labeled MCAo) and slowly spreads over the entire ipsilateral hemisphere by 4:00 min. The decrease in ADC marks the onset of cytotoxic edema. R, ipsilateral; L, contralateral; Ro, rostral; C, caudal.



Movie 4. Co-registered contrast-enhanced and ADC maps from group data. Owing to limitations in temporal resolution, both sequences cannot be collected simultaneously after MCAO. An average was generated from enhancement ratios (ERs) of dynamic contrast-enhanced MRI from four animals (green) and the ΔADC from six animals (red) and overlaid on a population-based average of the dynamic contrast enhancement baseline. After the spreading of cytotoxic edema, the tracer is rapidly distributed throughout the ipsilateral hemisphere. This suggests that CSF influx occurs after the onset of cytotoxic edema. Color-coded frames denote the anatomical position of the slice (top right). The 30-s gap corresponds to the angiography performed immediately after MCAO to confirm successful occlusion. High ΔADC signal change in the ventricles is due to the ventricular space shrinking and being replaced by swollen tissue in the later time points.

(*35*, *37*). Constriction of both the penetrating and pial arterioles was followed by a parallel increase in the space occupied by the PVS tracer (Fig. 4F and fig. S6, B, E, and F) (*39*). PVSs are believed to form an interconnected hydraulic network, where flow is transported by means of perivascular pumping generated by the cardiac cycle (*38*). However, we hypothesized that an increase in the size of the distal PVSs owing to vasoconstriction would create a pressure gradient that could effectively speed up perivascular flow. To test the hypothesis that diameter changes in the perivascular network could drive flow, we simulated the event using a topological network from mice pial MCAs (Fig. 5, A and B, and Movie 8) (*40*). The simulation showed that as a result of the SD, the constricting arteries increased the PVS size, drawing fluid into the tissue owing to reduced pressure in the expanding PVS. This inrush of fluid resulted in a 2.2-fold increase in baseline flow speed at the MCA, which eventually decreased (Fig. 5C). To determine whether we could detect a similar increase in CSF flow speed at the inlet of the network, we performed particle tracking velocimetry (*38*). CSF flow speed was analyzed for changes in pulsatile velocity ($v_{pulsatile}$) or net flow velocity ($v_{downstream}$) to determine if flow speed variations were caused by changes in pulsatility or by an overall pressure gradient (Movie 9 and fig. S7, A and B). After MCAO, there was a large, instantaneous increase in $v_{downstream}$ that quickly returned to baseline. But $v_{pulsatile}$ never recovered to baseline owing to the occlusion. This first component was consistent with the first peak of CSF influx. At the onset of SD (3.79 ± 0.44 min after MCAO), the large pial MCA also constricted, albeit to a lesser degree (27 ± 6%), and was accompanied by a ~2.5-fold increase in $v_{downstream}$, whereas $v_{pulsatile}$ remained below baseline (fig. S7, C and D). The degree of vasoconstriction was a significant predictor of $v_{downstream}$, with every 10% decrease in diameter causing a twofold increase in flow, but no such correlation was seen with $v_{pulsatile}$ (fig. S7E). Thus, postischemic CSF influx is not driven by changes in pulsatility but instead by an overall pressure gradient caused by vasoconstriction. The simulation also supported the notion that vasoconstriction is a sufficient mechanism to explain water accumulation in the tissue. However, this does not exclude the possibility that ionic or osmotic gradients secondary to the SD might also play a role in driving flow. In conclusion, both simulations and experimental evidence indicate that SI after SD abruptly accelerates perivascular CSF influx into the brain after stroke.

Spreading edema depends on aquaporin-4 expression

We next asked whether blocking the SD using a glutamate *N*-methyl-D-aspartate receptor antagonist (MK-801) or inhibiting SI using a cocktail of vasodilators (nimodipine, papaverine, and *S*-nitroso-*N*-acetylpenicillamine) would prevent the rapid entry of CSF (fig. S8). Both approaches have been shown to modulate either the spatiotemporal dynamics of SD or the severity of SI, but neither has been shown to inhibit completely the effects of the first depolarization after an ischemic insult (*35*, *36*, *41*–*44*). In our model, pretreatment with MK-801 or infusion of a vasodilator cocktail was ineffective in blocking the SD or SI (fig. S8, A to J). As before, MCAO was followed by a three- to fourfold increase in CSF influx in all groups (fig. S8, K to N; $P = 0.39$). However, residual CBF after SI was a significant predictor of the degree of CSF influx, suggesting that targeting reperfusion and SI might be a more feasible approach than trying to block the SD (fig. S8O). Alternatively, CSF transport into the brain has been shown to be facilitated by aquaporin-4 (AQP4) water channels expressed on the endfeet of astrocytes that form the outer wall of the penetrating PVS (*23*, *45*). Because AQP4 facilitates the transport of CSF out of the PVS, we hypothesized that AQP4-null mice ($Aqp4^{-/-}$) might have reduced CSF influx after MCAO (Fig. 6). This is in line with evidence demonstrating that knockout animals and wild types treated with AQP4 inhibitors are protected from edema after stroke (*46*–*49*). Hence, we tested the effect of AQP4 knockout on CSF tracer entry after MCAO using transcranial macroscopic imaging (Fig. 6, A and B). As with wild types ($Aqp4^{+/+}$), $Aqp4^{-/-}$ mice had a SD followed by SI after occlusion (Fig. 6C). Likewise, CSF tracer also distributed in the ipsilateral cortex but to a reduced degree in $Aqp4^{-/-}$ mice compared with wild type ($P = 0.018$; Fig. 6D). Could the absence of AQP4 at the perivascular endfoot reduce the entry of CSF to the tissue? To test this,

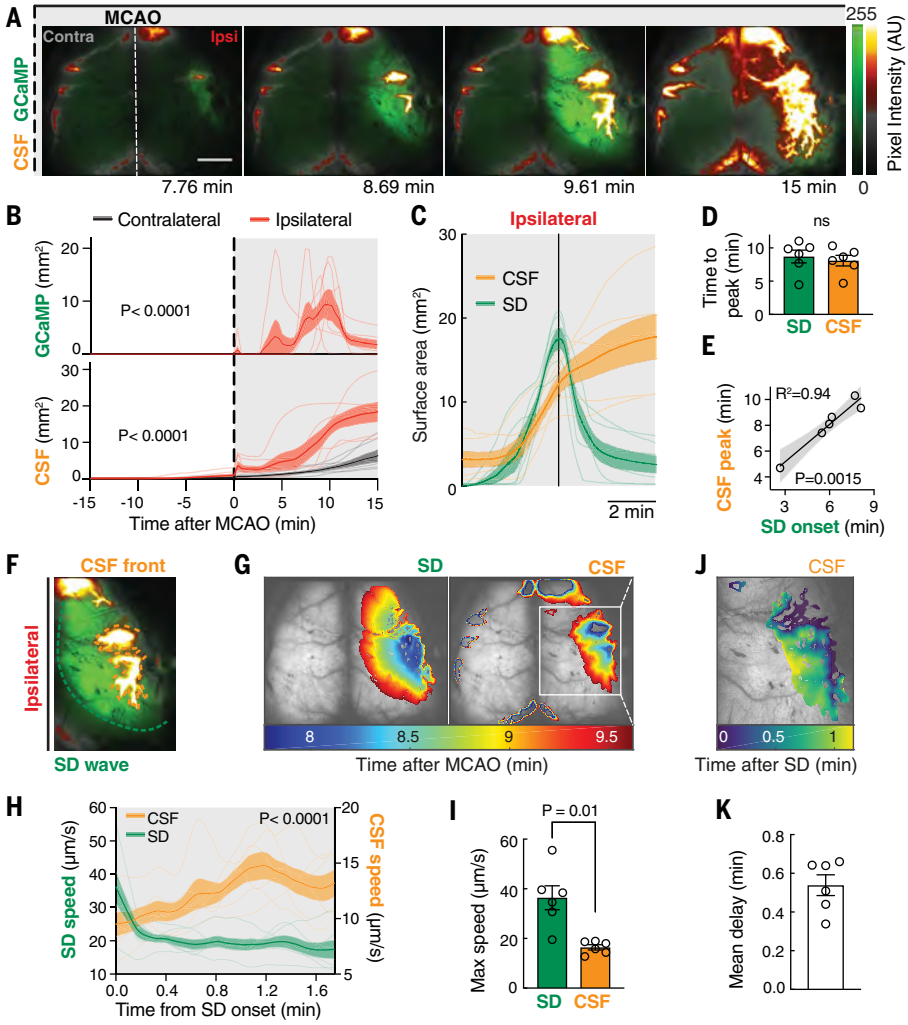
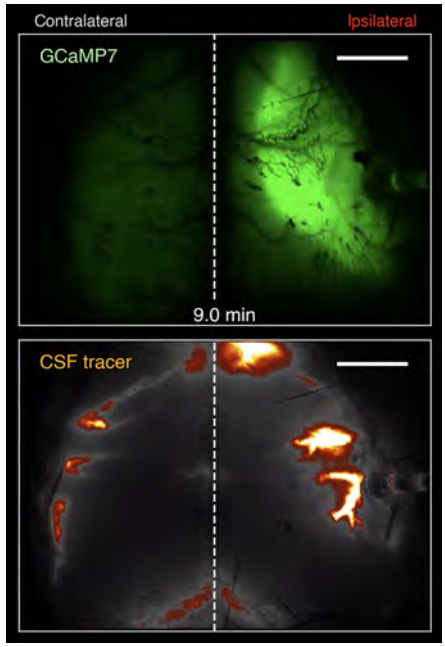


Fig. 3. CSF influx is triggered by SDs after focal ischemia. (**A**) A fluorescent CSF tracer (BSA-594) was delivered into the cisterna magna of *Glt1*-GCaMP7 mice 15 min before MCAO and imaged using a dual-channel macroscope for 15 min, as was done in Fig. 1D. Scale bar, 2 mm. (**B**) Area covered by GCaMP and CSF tracer fluorescence. Repeated measures two-way ANOVA with Sidak's multiple comparisons test was performed; interaction *P* value < 0.0001; *n* = 6 mice. (**C**) CSF tracer surface area aligned to the SD (GCaMP) peak. (**D**) Time to peak influx rate for the SD (GCaMP) and the CSF tracer. Paired Student's *t* test was performed. (**E**) Linear regression of CSF peak influx and SD onset time with 95% confidence intervals. (**F**) The fronts of the ipsilateral SD wave and the CSF tracer were tracked. (**G**) Area covered by the SD (left) and the CSF tracer (right) over time. (**H**) Front speed of the SD wave and the CSF front. Repeated measures two-way ANOVA with Sidak's multiple comparisons test was performed; interaction *P* value < 0.0001. (**I**) Maximum speed of the SD and CSF front. Paired Student's *t* test was performed. (**J**) Delay time between the SD wave and the CSF tracer front. (**K**) Mean delay between the SD and CSF front. In (B), (C), and (H), the shaded regions above and below the plot lines indicate SEM. In (D), (I), and (K), error bars represent SEM.



Movie 5. Transcranial macroscopic imaging of CSF tracer in GCaMP mice after MCAO. Dual-channel transcranial macroscopic imaging in a mouse that expresses GCaMP7 under the *Glt1* promoter (top left) after intracisternal injection of a fluorescent CSF tracer 15 min before MCAO (BSA-594; bottom left). After MCAO (dashed line), there is a rapid drop in rCBF. The surface area covered by the fluorescent front was quantified in the ipsilateral and contralateral hemispheres. Several minutes after occlusion, there is a SD in the ipsilateral hemisphere, triggering CSF influx. Scale bar, 2 mm.

we acquired 2P imaging after intracisternal infusion of a small CSF tracer (3-kDa dextran; Fig. 6E). During SI, tracer entered the penetrating PVS in large quantities, reaching depths of up to 200 μm below the cortical surface and distributing into the ISF in wild type (Fig. 6F). *Aqp4*$^{-/-}$ mice have been shown to have abnormal SD kinetics, so we first determined whether differences in the vasodynamics of penetrating arterioles during SI could be responsible for the differential response in CSF influx. Knockout animals exhibited slower arteriolar constrictions but to a similar degree as *Aqp4*$^{+/+}$ animals, which would not explain the observed results (Fig. 6, G to J). We then looked to see how much traced CSF actually reached the penetrating PVS and how much of it exchanged with the ISF. Less tracer entered the penetrating PVSs of *Aqp4*$^{-/-}$ mice compared with their wild-type counterparts ($P < 0.0001$; Fig. 6K). At the same time, less of the CSF tracer could be found in the neighboring ISF of the penetrating PVS in the knockouts, as in previous reports (Fig. 6L) (*23*). Global tracer quantification in ex vivo brain sections confirmed a significant reduction in fluorescence intensity in the ipsilateral and contralateral hemisphere of *Aqp4*$^{-/-}$ mice ($P = 0.005$ and 0.001, respectively; Fig. 6, M to O). In line with these results, AQP4 knockout animals did not develop edema within the first 15 min after MCAO, as opposed to *Aqp4*$^{+/+}$ mice (Figs. 6P and 1H). Thus, deletion of AQP4 significantly suppresses ischemia-induced CSF tracer influx.

Discussion

Historically, brain swelling has been assumed to develop entirely through the accumulation of fluid from the intravascular compartment (*15*). The present study demonstrates that CSF influx, initiated by ischemic SD, plays an important role in driving acute tissue swelling (Movie 10) (*16*, *50*–*52*). This new finding is further supported by a detailed examination of edema formation at CSF-ISF exchange zones, which we observed in cortex adjacent to leptomeningeal vessels as well as in periventricular zones 30 min after MCAO in mice and during the acute phase of human cerebral ischemic

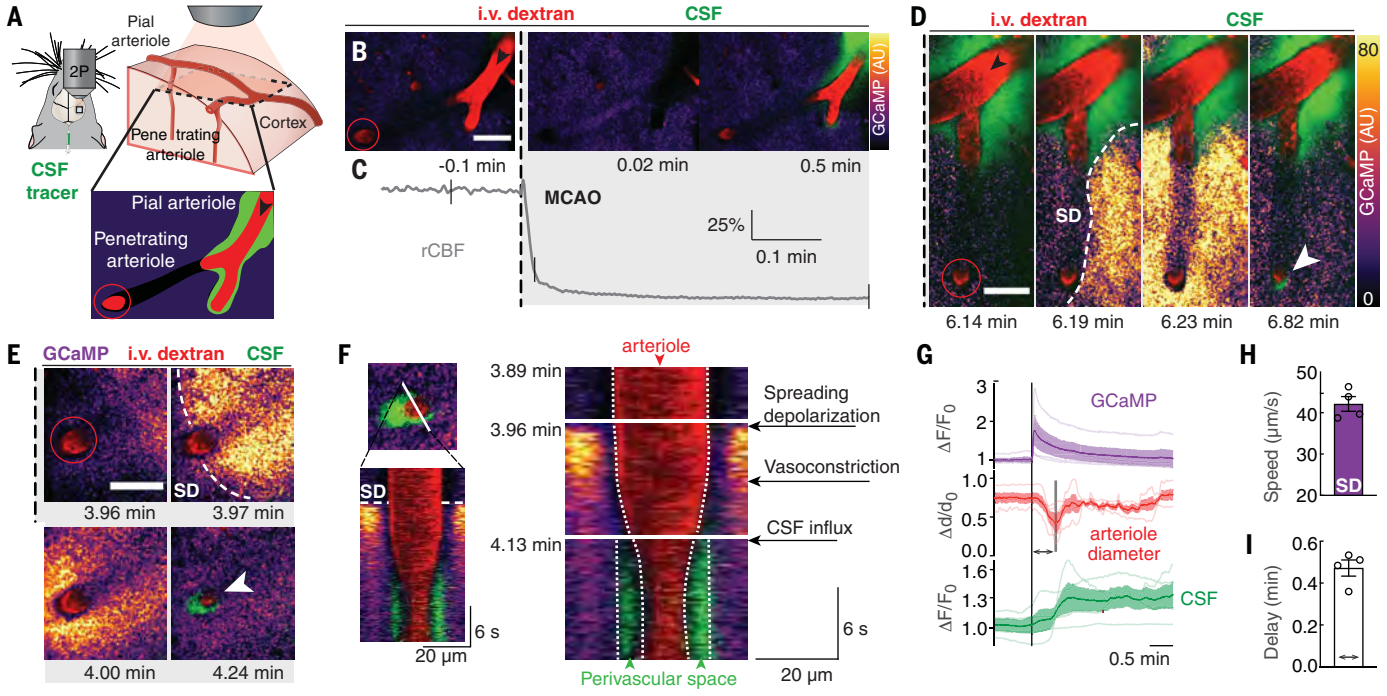


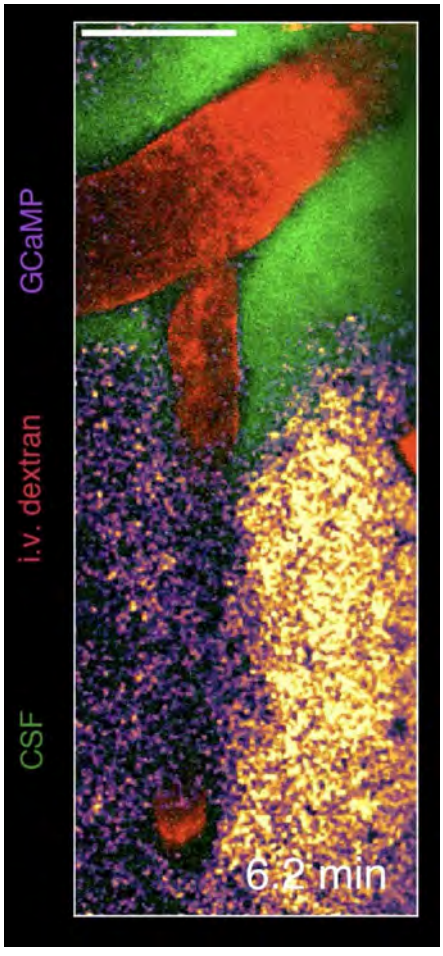
Fig. 4. SI after SD drives perivascular CSF influx. (**A**) Pial (black arrows) and penetrating arterioles (red circles, 40 to 50 µm below surface), which are branches of the ipsilateral MCA, were imaged using 2P microscopy. (**B**) Pial and penetrating arteriole (i.v. dextran) in a *Glt1-GCaMP7* mouse after receiving an intracisternal tracer injection (BSA-647) during MCAO. (**C**) rCBF measurements. The ticks align with images in (B). (**D**) SD after MCAO. Normalized GCaMP fluorescence ($\Delta F - F_{average}$) is color coded for pixel intensity and displayed in arbitrary units. (**E**) Constriction of penetrating arteries after SD causes tracer influx into the brain (white arrow). (**F**) Line scan over the penetrating arteriole in (E) depicting the appearance of the SD, the subsequent vasoconstriction, and the CSF tracer influx filling the PVS left by the constricted arteriole. (**G**) Quantification of GCaMP and CSF tracer fluorescence ($\Delta F/F_0$) and arteriole diameter ($\Delta d/d_0$) aligned to the onset of the SD; n = 4 mice. The shaded regions above and below the plot lines indicate SEM. (**H**) SD wave speed. (**I**) Delay time between SD onset and minimum arteriolar diameter. The double-headed arrow is a visual representation of the delay shown in the second plot in (G). In (H) and (I), error bars represent SEM. Scale bars in (B) to (E), 50 µm.

infarction (fig. S9 and table S1). Edema was primarily seen in brain regions that are known flow pathways for CSF-ISF rather than in deeper structures such as striatum that are farther away. Is glymphatic edema a previously unrecognized contributor to secondary ischemic injury? The entry of CSF through PVSs is likely to facilitate swelling of astrocytic endfeet and, together with pericyte constriction, would further reduce blood flow and cause infarct expansion (*53*, *54*). However, this study does not exclude the possibility that vascular fluid is a substantial source of swelling in later stages after stroke, because transvascular ion and fluid influx are increasingly important at 24 hours after MCAO when the BBB is already open (fig. S10, A to D). Several hours after CSF influx, de novo expression of the nonselective monovalent cation channel SUR1-TRPM4 in brain endothelial cells contributes to additional Na^+ influx, and blockade of these transporters with glibenclamide reduces edema (*5*, *55*). Nonetheless, CSF production continues to be a significant source of Na^+ and water after stroke (fig. S10, E to H). Our results posit that CSF is the earliest source of Na^+ and fluid, driving tissue swelling. This early phase is likely to play an additive role to the later stages of edema formation. Perhaps even more important is the fact that SDs continue for several days after stroke and are also observed in other acute conditions, including subarachnoid hemorrhage, intracerebral hemorrhage, and traumatic brain injury (*29*). Our study predicts that SDs will worsen edema and possibly explains why the frequency of depolarizations correlates with secondary injury and poor outcome (*29*, *39*, *56*–*58*). Thus, targeting glymphatic edema may offer a therapeutic strategy for treatment of a broad range of acute brain pathologies.

Materials and Methods

Animals

All experiments were approved by the University Committee on Animal Resources of the University of Rochester and the Danish Animal Experiments Inspectorate. Efforts were taken to minimize the number of animals used. Male C57BL/6 mice (Charles River) or *Glt1-GCaMP7* mice (RIKEN BioResource Research Center, Japan) on a C57BL/6 background 8 weeks of age were used for all the experiments. *Glt1-GCaMP7* mice express GCaMP7 in 95.2% of cortical astrocytes with minimal expression in other glial cell types and about 50% of cortical neurons in L4 and L6 (*32*). Male 8-week-old NG2-dsRed (Jackson Laboratory, stock no. 008241) mice were used to identify perivascular inflow routes. AQP4-knockout (*Aqp4*$^{-/-}$) mice on a C57BL/6 background between 8 and 16 weeks old were used (*45*). *Aqp4*$^{-/-}$ mice have been shown to have slower SD, delayed $[K]_e$ increase and uptake kinetics (*59*). The knockouts also exhibit shorter extracellular glutamate increases compared with wild type (*60*). Some of these studies have attributed this difference to the larger extracellular space volume observed in these mice, causing a slower increase in $[K]_e$ during depolarization (*59*). Other studies have instead credited it to AQP4-dependent changes in volume-regulated anion channels (*60*). *Aqp4*$^{-/-}$ mice also show significant differences in O_2 diffusion, especially to remote areas away from microvessels (*61*). All experiments performed in this study were done on mice anesthetized with ketamine and xylazine (100 and 10 mg/kg, intraperitoneally). The vasomotor response to SDs in naïve cortex is different in mice compared with other species (*29*). However, in ischemic cortex, as seen in our model of MCAO, SI is

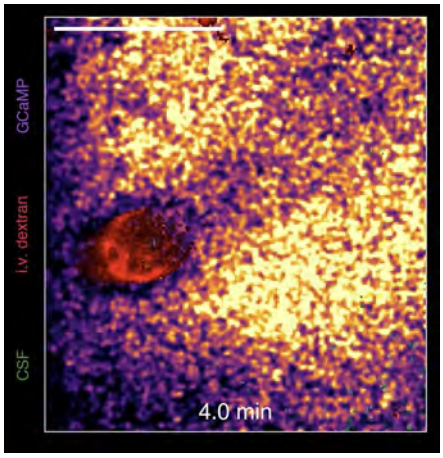


Movie 6. 2P imaging of a SD at a pial and penetrating arteriole after stroke. An anesthetized *Glt1-GCaMP7* (purple) mouse received intravenous tetramethylrhodamine isothiocyanate (TRITC)–dextran injection (i.v. dextran, 2000 kDa, red) and intracisternal BSA-647 tracer (CSF, green). At early time points, CSF tracer can first be seen around the pial arterioles. At ~6.2 min after MCAO, a SD can be seen crossing over the imaging field (GCaMP fluorescence, $\Delta F/F_0$). Then at ~6.5 min, there is a pronounced vasoconstriction of the penetrating arteriole, $\Delta d/d_0$, and a parallel increase in the amount of tracer in the penetrating PVS. At ~6.8 min, there is a smaller constriction of the pial arteriole. Both vasomotor events cause the CSF tracer fluorescence to intensify several minutes after MCAO (CSF tracer fluorescence, $\Delta F/F_0$). Scale bar, 50 μm.

the predominant vasomotor response after SD across a variety of diseases and animal models [i.e., mice (*35*), rats (*62*), cats (*34*), swine (*63*), and humans (*64*–*67*)].

Intracisternal injections

Animals were fixed in a stereotaxic frame. A 30-gauge needle was connected to PE-10 tubing filled with artificial CSF (aCSF) into the cisterna magna as described here (*68*). For intracisternal



Movie 7. 2P imaging of SI at a penetrating arteriole after stroke. An anesthetized *Glt1-GCaMP7* (purple) mouse received intravenous TRITC-dextran injection (i.v. dextran, 2000 kDa, red) and intracisternal BSA-647 tracer (CSF, green). At ~3.9 min after MCAO, a SD covers the field of view (GCaMP fluorescence, $\Delta F/F_0$). Several seconds later, the penetrating arteriole constricts to ~80% of its pre-MCAO diameter, $\Delta d/d_0$). In response to this, CSF tracer fills the penetrating PVS (CSF tracer fluorescence, $\Delta F/F_0$). Scale bar, 50 μm.

injections, 10 μl of CSF tracer was injected at a rate of 2 μl/min over 5 min with a syringe pump (Harvard Apparatus).

CSF tracers

Alexa Fluor647– or Alexa Fluor594–conjugated bovine serum albumin (BSA-647 or BSA-594, 66 kDa, Invitrogen) and Cascade Blue-conjugated 3-kDa dextran (Invitrogen) were diluted in aCSF (0.5 or 1% m/v) and used as a fluorescent CSF tracer. Radio-labeled ^{22}Na (0.5 μCi, Perkin Elmer) and ^{3}H-mannitol (1 μCi, American Radiolabeled Chemicals) were dissolved in aCSF. For particle tracking velocimetry experiments, red fluorescent polystyrene microspheres (FluoSpheres 1.0 μm, excitation 580 nm, emission 605 nm, 0.25% solids in aCSF, Invitrogen) were briefly sonicated and infused as before (*38*).

Drugs

MK-801 (5 mg/kg in 0.9% NaCl; Tocris) was administered by intraperitoneal injection 15 min before MCAO. The vasodilator cocktail contained nimodipine (2 μg/kg/min; Tocris), papaverine (3 mg/kg/min; Sigma Aldrich) and *S*-nitroso-*N*-acetylpenicillamine (3 μmol/kg/min in 0.9% NaCl; Sigma-Aldrich) and was delivered intravenously through a PE-10 tubing catheter in the femoral vein using an infusion pump (Harvard Apparatus).

Physiological recordings

Heart rate and respiratory rate were acquired using an animal physiological monitoring device (Harvard Apparatus). ICP was measured through a 30-gauge needle connected to a PE-10 catheter filled with aCSF into the cisterna magna. The line was connected to a pressure transducer and monitor (World Precision Instruments). rCBF was measured using laser Doppler flowmetry (PF5010 Laser Doppler Perfusion Module, PR 418-1, Perimed). The fiber optic probe (MT500-0, Perimed) was fixed onto the skull above the MCA vascular territory (5 mm lateral and 1 mm posterior to bregma) on the right hemisphere with cyanoacrylate glue. For 2P experiments, rCBF was measured at the lateral aspect of the skull over the right temporal bone in order to place the fiber optic probe below the headplate and shield it from laser irradiation. All the signals were collected using a 1440A digitizer and AxoScope software (Axon Instruments) and analyzed using Matlab.

MCAO

The anesthetized animal was laid on its back on a heating pad. A surgical midline incision was made from the clavicle to the chin to expose the right common carotid artery (CCA), internal carotid artery (ICA), and external carotid artery (ECA). Then the distal side of ECA and proximal side of CCA were ligated by 7-0 silk suture. The distal side of the ICA was clamped using a micro clamp (B-1, Fine Science Tools). A small incision was subsequently made in the proximal portion of the CCA. A PE-10 polyethylene tube (PE-10, Beckton-Dickinson) filled with heparinized saline and six macrospheres, 200 μm in diameter, was inserted into the CCA and the clamp was subsequently removed. The tubing was secured in position by 7-0 silk suture. The macrospheres were advanced into the ICA via an injection of 0.02 ml heparinized saline. Ischemia was confirmed by laser Doppler flowmetry. The incision was closed with a 5-0 suture. All mice were monitored using laser Doppler flowmetry during the macrosphere infusion. If rCBF did not decrease >70% of baseline or if rCBF returned to baseline, the animal was excluded.

Behavior testing

Twenty-four hours after MCAO or sham, mice were placed in a 45-cm–by–61-cm open-field apparatus. Movement of the mice was recorded for 30 min and then analyzed on ANY-maze video tracking software. MCAO and sham mice were placed on an accelerating rota-rod (Ugo Basile) that accelerated from 5 to 40 rpm. The mice were allowed to remain on the rota-rod until they could no longer continue, and a final time was recorded once they fell off the rotating cylinder. MCAO and sham mice were allowed to grasp a wire bar suspended 43 cm above the surface with only their forepaws before being released. Each mouse was tested for three 20-s trials and given a score from 0 to 3 depending on how well it was able to stay on the string.

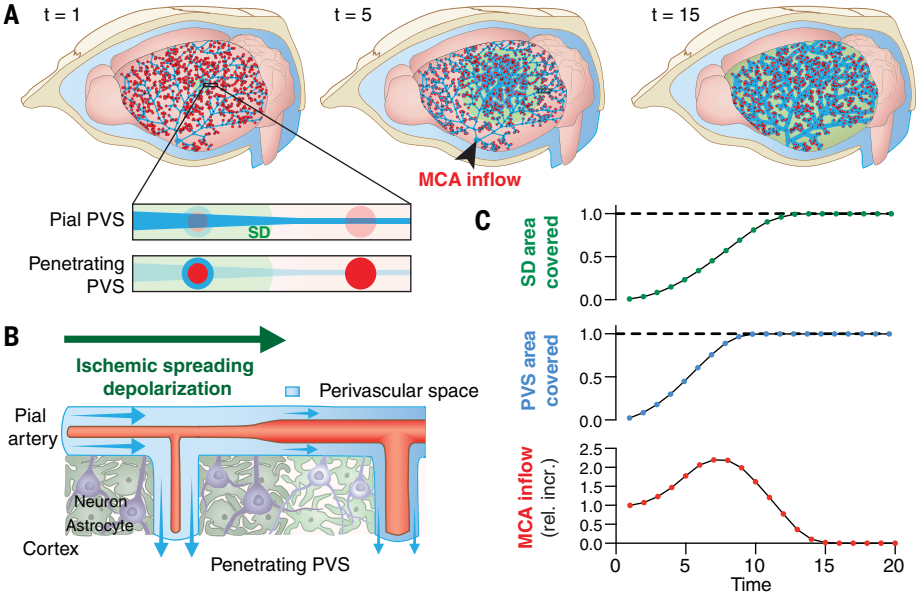
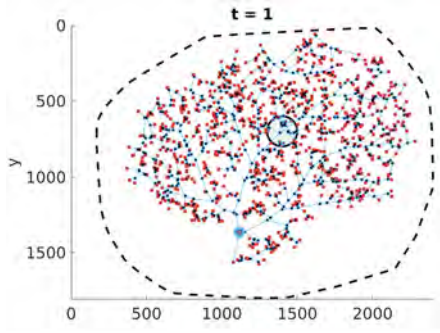
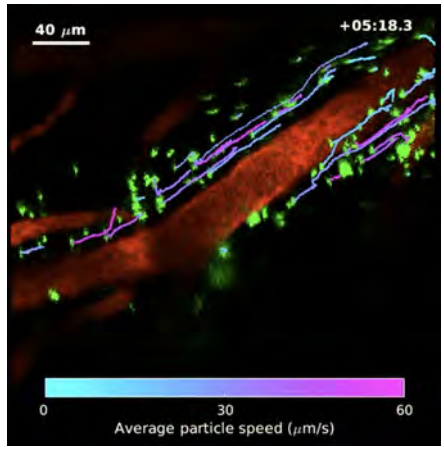


Fig. 5. Topological glymphatic network model of PVSs around the mouse MCA. (**A**) Network model representing a system of interconnected PVSs surrounding the mouse pial MCA at different time points (t) during an ischemic SD (green). Pial PVSs are depicted as blue lines that get wider during SI. Penetrating arteries are depicted as red circles, and the size of the surrounding blue circle is proportional to the flow rate into the penetrating PVS as the arteriole constricts after SD. The simulation evaluated the relative increase in baseline flow at the inlet of the MCA (MCA inflow; black arrow). (**B** and **C**) Pial and penetrating PVS area increases as the arteries constrict owing to the passage of a SD, thus increasing the fluid volume in the network (B). Conservation of mass controls the resulting MCA inflow (C), resulting in a net increase in fluid volume in the network. The dashed line represents the tissue border of the cortical surface. The SD travels over the entire cortex, spanning an area larger than that covered by the MCA network. rel. incr., relative increase.



Movie 8. Topological glymphatic network simulation of perivascular CSF influx. A network was generated from a mouse pial MCA. Nodes were located at bifurcations (black squares) and penetrating arteries (red squares). Edges represent the perivascular flow pathways following the trajectory of the MCA (blue). The SD (green) starts at the proximal MCA (blue circle) and spreads outward toward the cortical tissue border (dashed line). As the SD propagates isotropically, there is a delayed vasoconstrictive response of the pial and penetrating arterioles, causing the area of the PVSs to increase and fluid volume within the network to follow. This spreading vasoconstriction causes flow speed at the MCA inlet to increase.



Movie 9. Particle tracking velocimetry in the PVS of the MCA. Fluorescent 1-μm microspheres were injected into the cisterna magna of an anesthetized wild-type mouse after labeling the vasculature with a fluorescent dextran [fluorescein isothiocyanate (FITC)–dextran, 2000 kDa, i.v.]. 2P imaging of the particles in the PVS of the MCA (left) is shown. Particle tracks are color coded to the average particle speed (μm/s). The average speed of all the particles (v_{mean}) is plotted simultaneous to quantification of $v_{downstream}$ and $v_{pulsatile}$, rCBF (in percent pressure units, p.U.), and artery diameter (μm). Around 300 s after MCAO, there is marked vasoconstriction and CSF flow speed increases. Time denotes minutes after MCAO. Scale bar, 40 μm.

A score of 3 was given to mice who would immediately fall from the string when released, a 2 if the mouse was able to hang on the string and attempted to climb, and a 1 was given to mice that were able to put both forepaws and one or both hind paws onto the string; if the mouse placed all four paws and the tail onto the string and made lateral movement, a score of 0 was given (*69*).

Transcranial optical imaging

Mice were fixed to either a MAG-1 or MAG-2 stereotaxic holding device (Narishige) and injection, via a cisterna magna cannula, of 10 μl of BSA-conjugated Alexa Fluor-594 or -647 (Invitrogen) in aCSF, occurred over a period of 5 min using a syringe pump (Harvard Apparatus). Whole-cortical macroscopic imaging was performed using either an Olympus MVX10 with a PRIOR Lumen LED and Hamamatsu ORCA-Flash4.0 V2 Digital CMOS camera using Metamorph software or a Leica M205 FA fluorescence stereomicroscope equipped with an Xcite 200DC light source and A12801-01 W-View GEMINI (Hamamatsu) for simultaneous GFP (excitation 480 nm, emission 510 nm, Leica) and Cy3 (excitation 560 nm, emission 630 nm, Leica) acquisition. Images were acquired with a Hamamatsu ORCA-Flash4.0 V2 Digital CMOS camera using LAS X Leica software, with a 512-pixel–by–512-pixel, 16-bit resolution, at 20 Hz.

MRI scans

MRI was performed using a 9.4-tesla animal scanner (BioSpec 94/30 USR, Bruker BioSpin, Ettlingen, Germany) equipped with a cryogenically cooled quadrature-resonator (CryoProbe, Bruker BioSpin, Ettlingen, Germany). Mice were placed on a magnetic resonance (MR)–compatible stereotactic holder with ear bars to minimize head movement during scanning. Body temperature was maintained at 37°C and monitored along with the breathing rate by a remote monitoring system (SA Instruments, NY, USA). All required adjustments and T2-weighted scans [2D TurboRARE: TR/TE: 8000/36 ms, Matrix 384 by 256, field of view (FOV) 19.2 mm by 12.8 mm, NEX 1, 48 horizontal slices, slice thickness = 0.2 mm] were performed within 15 min after CCA catheter placement for visualization of the geometry before the baseline dynamic contrast enhancement (DCE), diffusion-weighted imaging (DWI), or cisternography scans. Ischemia was confirmed by flow-compensated 2D time-of-flight MR angiography scans (2D-GEFC: TR/TE 10/2.4 ms, Matrix 192 by 128, FA 80, FOV 19.2 mm by 12.8 mm, NEX 1, 10 horizontal slices, slice

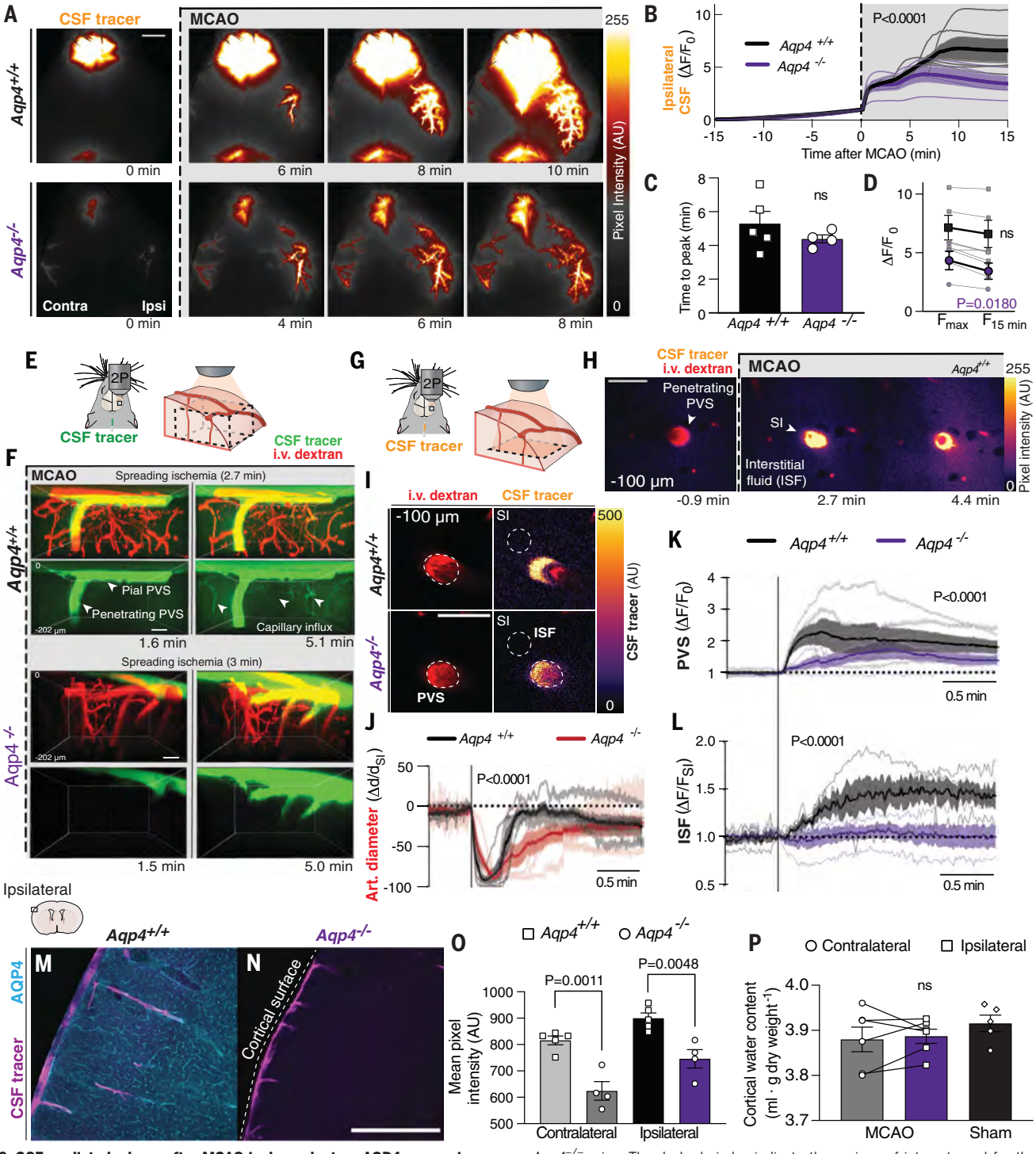
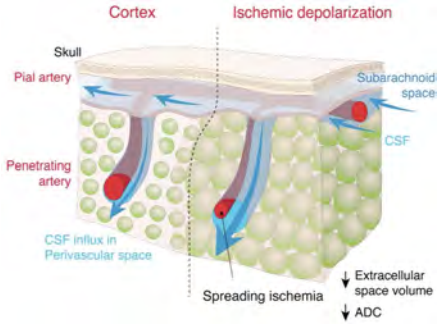


Fig. 6. CSF-mediated edema after MCAO is dependent on AQP4 expression. (**A**) Transcranial imaging after intracisternal tracer injection (BSA-647) in *Aqp4*$^{+/+}$ and *Aqp4*$^{-/-}$ mice. (**B**) Quantification of ipsilateral CSF tracer influx ($\Delta F/F_0$). Two-way repeated measures ANOVA was performed; *P* value from interaction term < 0.0001; *n* = 4 or 5 mice per genotype. (**C**) Time of CSF influx peak. Unpaired Student's *t* test was performed. (**D**) Change in maximum $\Delta F/F_0$ (F_{max}) to 15 min after MCAO ($F_{15\ min}$). Paired Student's *t* test was performed. (**E** and **F**) Volumetric 2P imaging shows CSF tracer (3-kDa dextran) entering the brain via the surface and penetrating PVSs in *Aqp4*$^{+/+}$ and *Aqp4*$^{-/-}$ mice. After SI, the tracers were also found surrounding capillaries only in *Aqp4*$^{+/+}$. (**G**) Time-lapse 2P imaging of a penetrating arteriole 100 µm below the cortical surface. (**H**) The tracer enters the ISF after SI. (**I**) CSF tracer entered the penetrating PVS of both *Aqp4*$^{+/+}$ and *Aqp4*$^{-/-}$ mice. The dashed circles indicate the regions of interest used for the quantifications in (K) and (L). (**J** to **L**) Penetrating arteriolar (art.) diameter changes after MCAO (J). Tracer influx was quantified in the penetrating PVS (K) and the ISF neighboring the same PVS (L). One PVS per mouse was analyzed; *n* = 4 or 5 mice per genotype. (**M** and **N**) Immunohistochemical labeling for AQP4 from ipsilateral dorsal cortex of *Aqp4*$^{+/+}$ (M) and *Aqp4*$^{-/-}$ (N) after CSF tracer injection. (**O**) Mean pixel intensity from six coronal sections from 4 or 5 mice per genotype. (**P**) Cortical water content 15 min after MCAO or bilateral cortices from sham-treated *Aqp4*$^{-/-}$ mice. Paired Student's *t* test was performed; *n* = 5 or 6 mice per group. Scale bars, 2 mm (A), 50 µm (F) to (I), and 500 µm (M) and (N). In (B), (J), (K), and (L), the shaded regions above and below the plot lines indicate SEM. In (C), (O), and (P), error bars represent SEM.



Movie 10. Spreading glymphatic edema animation. SDs propagate over the cortex, followed by SI. As the penetrating and pial arterioles constrict, the volume of the perivascular network increases, driving flow into the parenchyma. Cytotoxic edema and CSF influx cause cells to swell and the extracellular space to shrink, seen as a decrease in the ADC.

thickness = 0.3 mm) 30 s immediately after macrosphere infusion: If angiography did not show occlusion of MCA, the animal was excluded.

DCE MRI

Pre- and postcontrast T1-weighted (T1W) images were collected with 3D-FLASH sequence (TR/TE 17.7/3.1 ms, FA = 15°, Matrix 192 by 128 by 96, FOV 19.2 mm by 12.8 mm by 9.6 mm, NEX 1). T1W imaging of the entire mouse brain was performed in 2 min at an isotropic spatial resolution of 100 μm as a T1-enhancing contrast agent, gadobutrol (Gadovist, Bayer Pharma AG, Leverkusen, Germany), was injected into cisterna magna. The time series T1W scanning protocol was composed of three baseline scans (6 min) followed by intracisternal infusion of 12.5 mM gadobutrol at a constant rate of 1.0 μl/min for 10 min. Considering the dead space within the catheter, the actual infused volume of the contrast was 9 μl. Scans continued over 50 measurements, and MCAO started at 24 min after the first T1W scan.

DWI

DWI was performed with 2D single-shot echo-planar DWI sequence with parameters set as follows: TR/TE 2500/20 ms, NEX 1, FOV 18 mm by 15 mm, Matrix 108 by 96, eight axial slices, slice thickness 0.8 mm. Z-direction motion probing gradient was acquired with multiple b values (0, 100, 400, and 1000 s/mm^2). For the DWI study, 30 baseline images before MCAO were acquired, followed by a series of images acquired over time after MCAO. Scans continued over 360 measurements, with each measurement lasting 10 s. Maps of the mean diffusivity (ADC map) were derived using the standard algorithm of Paravision 6.0.1 software (Bruker).

MR CSF cisternography

CSF cisternography was performed with 3D-FIESTA sequence (TR/TE 17.7/3.1 ms, FA = 15°, Matrix 192 by 128 by 96, FOV 19.2 mm by 12.8 mm by 9.6 mm, NEX 1). For CSF volumetric measurements, 10 baseline images were acquired before MCAO, followed by a series of images acquired over time after MCAO. Scans continued over 60 measurements, with each measurement lasting 30 s. Ventricular compartments were segmented, and temporal volume change was measured by IMARIS (v. 9.2.1, Bitplane, Concord, MA, USA).

In vivo 2P laser scanning microscopy

A cranial window was prepared over the right anterolateral parietal bone above the MCA vascular territory. The dura mater was left intact and, to prevent intracranial depressurization, the window was sealed with agarose (1.1% at 37°C) and a glass coverslip (8 mm diameter). Unsealed craniotomies have been shown to affect glymphatic function (*45*). 2P imaging was performed using a resonant scanner Bergamo scope (Thorlabs) and a Chameleon Ultra II laser (Coherent) with a water-immersion 20× objective (1.0 NA, Olympus). Intravascular FITC-dextran (2000 kDa, 2.5%, Sigma-Aldrich) and either red microspheres or BSA-647 were excited at an 820-nm wavelength. In GCaMP mice, vessels were labeled with TRITC-dextran (2000 kDa) and CSF with BSA-647 and excited at 860 nm. Emission was filtered at 525, 607, and 647 nm. Images were acquired using ThorImage software and synchronized with physiological recordings (3 kHz, ThorSync software).

Quantification of CSF production rate

CSF production rate was quantified as previously described in rats (*70*, *71*). Anesthetized mice were fixed on a stereotactic apparatus, and a 0.5 mm burr hole was made over the left lateral ventricle (anterior-posterior = −0.1 mm, medial-lateral = −0.85 mm]. A 30-gauge needle connected to PE-10 tubing was lowered through the burr hole to −2.00 mm dorsal-ventral. The mouse then received a MCAO as before. The rCBF drop was confirmed with laser Doppler flowmetry for 5 mins, and then the mouse was placed back in a stereotactic head frame and the cisterna magna was surgically exposed with the mouse's neck flexed at 90°. A 30-gauge needle connected to PE10 tubing filled with mineral oil (Sigma Aldrich, M5904) was inserted into the cisterna magna and advanced gently 2 mm through the foramen of Magendie into the fourth ventricle. One microliter of mineral oil was infused at a rate of 1 μl/min for 1 min with a syringe pump (Harvard Apparatus) to block outflow from the ventricular system. The location of CSF within the intraventricular PE-10 tubing was marked at 10-min intervals. The volume of CSF was calculated as: CSF volume = π × (internal radius: 0.14 mm)^2 × length. The rate of CSF production (μl/min) was calculated as the slope of the linear regression from each mouse.

Brain water content measurements

Anesthetized animals were decapitated immediately, and the cortex was dissected and weighed (W_{wet}; g). The tissue was dried at 65°C until it reached a constant weight (~48 hours) and brains were reweighed (W_{dry}). For calculations of the tissue water content (ml/g dry weight), the following formula was used: $(W_{wet} - W_{dry})/W_{dry}$.

Radioisotope influx

To evaluate CSF influx into the brain, radiolabeled tracers ^{22}Na (0.5 μCi, Perkin Elmer) and ^{3}H-mannitol (1 μCi, Perkin Elmer) were injected into the cisterna magna at 2 μl/min for 5 min in aCSF and allowed to circulate for 15 min before MCAO. After 15 min, the animals were rapidly decapitated, the skull and dura were removed, and the brain was harvested. Brains were cut into six sections for processing. In a separate set of experiments, the radiolabeled tracers were injected intravenously first as a 2 μl bolus and then a 2 μl/min infusion for 5 min. Immediately after the end of the infusion, MCAO was induced, and 15 min later, animals were rapidly decapitated, the skull and dura were removed, and the brain was harvested. All brain tissue was weighed and solubilized in 0.5 ml of tissue solubilizer (Solvable, PerkinElmer) overnight. Upon solubilization, 5 ml of scintillation cocktail was added (Ultima Gold, PerkinElmer). The injectate controls were treated in the same way as the tissue samples. All samples were analyzed by liquid scintillation spectrometry using a scintillation counter (LS 6500 Multipurpose Scintillation Counter, Beckman Coulter). The radioactivity (counts per minute) remaining in the brain after injection (percentage of injected dose) was determined as $R_b/R_i \times 100$, where R_b is the radioactivity remaining in the brain at the end of the experiment and R_i is the radioactivity in the injectate controls for each experiment. Only cerebrum was considered, and cerebellum was excluded from analysis because ischemia does not reach this region. All samples were background subtracted to blank samples with 0.5 ml of tissue solubilizer and 5 ml of scintillation cocktail.

Triphenyltetrazolium chloride staining

Triphenyltetrazolium chloride (TTC) was prepared at 2% in phosphate-buffered saline (PBS) and kept in a 37°C water bath. MCAO and sham mice were anesthetized and perfused with 5 to 10 ml of ice-cold PBS. Mice were immediately decapitated, and the brains were extracted and put into TTC. Brains remained in the TTC solution for 15 min and then imaged under an Olympus SZX12 bright-field microscope. In a separate experiment, the brains were sectioned into slices 1 mm thick on a vibratome (Leica VT1200S) in ice-cold PBS. The coronal slices were placed into TTC for

15 min and then 4% paraformaldehyde (PFA) for an additional 30 min. The coronal sections were then imaged under the microscope for infarct volume quantification. The area of the infarct volume was calculated in each slice by tracing the infarct core using the polygon selection tool in Fiji. Area (mm^2) of the infarct was then converted to volume (mm^3) accounting for the 1 mm slice thickness. The sum of all the coronal infarct volumes was calculated to give whole brain infarct volume. The edema-corrected infarct volume was calculated as previously described (*72*) and expressed as a percentage of the contralateral hemisphere.

Arteriole and venule labeling experiments

NG2-dsRed mice were anesthetized and received an intravenous injection of 0.2 ml of 1 mg/ml wheat-germ agglutinin lectin (Invitrogen Alexa Fluor 647 conjugated) into the saphenous vein. Next, a fluorescent CSF tracer (0.5% 3-kDa dextran Cascade Blue; Invitrogen) was infused into the cisterna magna and allowed to circulate for 15 min before MCAO. Thirty min later, the mouse was perfusion-fixed with PBS (100 μg lectin) followed by 4% PFA, and then the brains were harvested. The following day, the brains were sectioned into 100-μm slices using a vibratome (Leica VT1200S) and mounted onto slides with ProLong Gold antifade mounting medium (Invitrogen). Images were acquired using a Leica TCS SP8 confocal system. A total of 8 to 10 *Z*-stacks at 20× magnification from cortex were taken from four to six coronal sections from each mouse. The percentage of tracer-positive arterioles and venules was calculated from the total number of blood vessels with CSF tracer labeling (42 to 49 vessels per mouse).

Tissue histology

Mouse brains were obtained after sacrifice, after sham or MCAO, and fixed in 4% PFA. Patients who died with acute cerebral ischemia documented between January 2018 and December 2018 and who underwent autopsy were identified retrospectively by reviewing the autopsy records of the Department of Pathology at the University of Rochester Medical Center. With these constraints, five focal cerebral infarct specimens from patients were identified. As controls, brain samples were also evaluated from adult patients with documented absence of an ischemic brain lesion who died from acute cardiorespiratory collapse. In all cases, the presence or absence of an ischemic brain lesion was confirmed by a neuropathologist. Standard tissue fixation protocols (7 to 10 days in formalin) were applied. Paraffin-embedded tissue blocks from frontal cortex, basal ganglia, and periventricular regions were obtained in each case by an expert neuropathologist (RIM). Blocks were sectioned at 6 μm thickness, stained with hematoxylin and eosin (HE), and imaged at 1000× (oil immersion) using a Nikon NiU Microscope and Nikon Microscope Solutions Imaging Software (NIS-Elements AR Version 4.30.01). Edema area was quantified by a blinded rater using Fiji. HE images were automatically thresholded to include the highest-intensity pixels (Otsu's method), which consisted of the white pixels of the fluid accumulation. The edema area was measured as a percentage of the total tissue area from a representative image from each animal or subject. In the case of mouse cortex (fig. S9B), each biological replicate was an average from four separate images from the same mouse. For coronal sections in Fig. 6, M to O, brains were sliced into 100-μm sections using a vibratome (Leica VT1200S) and mounted onto slides with ProLong Gold antifade mounting medium (Invitrogen). Images were acquired with an Olympus MVX10 stereomacroscope and a PRIOR Lumen LED and Hamamatsu ORCA-Flash4.0 V2 Digital CMOS camera using Metamorph software. Images were analyzed as previously described (*33*). Mean pixel intensity for six sections starting at the anterior aspect of the corpus callosum and skipping 400-μm intervals moving posteriorly was quantified using ROIs for the ipsilateral and contralateral hemispheres in Fiji. An average of the six coronal sections was computed for each mouse.

Transcranial optical imaging analysis

An ROI was manually drawn around the skull, and fluorescence intensity of CSF tracer influx was quantified using Fiji software (*33*). Mean fluorescence intensity was normalized to the time of MCAO (F_0), and the derivative of the curve was computed to calculate the mean rate of change in fluorescence intensity over time. The time at which a peak in the rate of change occurred was determined to calculate time to peak influx. To extract more quantitative information, we tracked fronts using an automated Matlab algorithm previously developed at the University of Rochester (*33*, *73*). Fronts are curves that separate bright regions from dark regions in videos of murine brains. Local front speeds quantify the local speeds of CSF influx and SD. We tracked fronts only within the brain hemispheres, sometimes treating the ipsilateral and contralateral hemispheres separately. We located those hemispheres by calculating the time-averaged GCaMP brightness of each video, then finding the two largest bright regions. Front tracking requires choosing a brightness threshold: Each front separates a region brighter than the threshold from a region dimmer than the threshold. To calculate thresholds, we calculated the mean brightness of each channel in each hemisphere, varying over time. We chose the thresholds as the brightness values halfway between the initial, dim values and the final, saturated values. From the thresholded dataset, we calculated the surface area in mm^2 over time in an ipsilateral and contralateral hemisphere ROI. The surface area data for the CSF tracer and the GCaMP channels were aligned in time using the maximum area of the GCaMP channel. We next calculated the mean rate of change in the surface area covered by CSF tracer over time. The maximal rate of change was used to compute the time to CSF influx peak and was compared with the time at which maximal GCaMP surface area occurred. SD onset time was measured as the first nonzero value of the thresholded GCaMP channel after MCAO. We next measured front speeds in a smaller ROI over the MCA territory. The average front speed for all pixels was plotted over time. To reduce noise caused by outlier pixels, we smoothed front speeds with a sliding line fit and a 20-s smoothing window. The maximum value of front speeds after the onset of the SD was considered the max speed. To determine the delay time between the SD and the arrival of CSF tracer, we evaluated the time elapsed between a pixel having GCaMP fluorescence and CSF tracer fluorescence; an average of all the pixels in the ROI was calculated to obtain mean delay time.

MRI postprocessing and analysis

Both DCE and ADC-map time series data were motion corrected using Advanced Normalization Tools normalization software (*74*, *75*). DCE-MRI time series were motion corrected and converted into percent change from baseline-time series, calculated as the percent signal change from the averaged signal of baseline images. To avoid biases in tracer calculations from availability of tracer to the glymphatic system, we then normalized percent change maps to the peak average. A population-based average of the baseline scans was created by an iterative registration process with two rigid, two affine, and five nonlinear registration-and-averaging steps. To allow unbiased ipsilateral-contralateral comparisons, original and left-right–flipped images were both used in the template creation. Registration was carried out using Advanced Normalization Tools v. 2.1.0. The average template was skull-stripped manually using ITK-SNAP and was then segmented into brain structures by registering the Waxholm Space Atlas of the C57BL/6J Mouse Brain (*76*) to the template. Normalized percent-change maps were finally transformed to the average space for analysis. We the drew lines extending from the MCA at the level of the temporal ridge and 1 mm orthogonal to the brain surface in the coronal slice. Tracer penetration depth was defined as the distance from the MCA to the deepest voxel with normalized tracer signal >1. For DWI analysis, edema was defined as the ADC value in a pixel dropping three standard deviations below baseline. Analysis of images was performed in ITK-SNAP

(version 3.6.0, www.itksnap.org), ImageJ (version 2.00, NIH, Bethesda, MD, USA), and Matlab (version 9.3.0, MathWorks, Natick, MA, USA). Circular ROIs were drawn in the images both for the DCE enhancement ratio calculations and ADC calculations. 3D maximum intensity projection and multiplanar reformation for DCE data visualization was generated using IMARIS (v. 9.2.1, Bitplane, Concord, MA, USA). To create the registered average of the DCE and ADC time series, the time series were registered to the average template described above, and from those, different slices of the volumes were processed into videos using Python 3.6.

2P imaging analysis

Images obtained were 16 bit, with either two or three channels depending on the experiment, each with spatial dimensions of 512 by 512 and an average frame rate of 15 Hz. The 525-nm channel captured the vascular FITC-dextran or GCaMP7 fluorescence and the 607-nm channel the vascular Texas Red–dextran or the red microspheres in CSF for particle tracking, whereas the 647-nm channel captured the fluorescent BSA CSF tracer. Images were analyzed using Fiji and Matlab. The GCaMP channel was processed by converting it to 8 bit and generating a *Z*-projected average intensity of the entire imaging session and subtracting the average image from each frame ($\Delta F - F_{\text{average}}$). A Gaussian blur filter with a sigma of 1 was applied to the resulting stack and was subsequently color coded using the lookup table mpl-inferno. A mask of the CSF tracer and vascular channel was applied, and then all channels were merged. Line scans of the XYT data were generated using Fiji. For quantification purposes, GCaMP and CSF tracer fluorescence were quantified as $\Delta F/F_0$, where F_0 is the first frame of the experiment before MCAO. Wave speed was calculated by drawing two ROIs 100 μm apart, along the direction of wave propagation. The time between the GCaMP fluorescence appearing in the first and second ROIs was considered the speed of propagation. PVSs were defined as any fluid compartment following a blood vessel. Pial PVS followed leptomeningeal vessels at the brain surface, and penetrating PVS dove into cortical parenchyma with penetrating arterioles. PVS area was compared with arteriolar area for both pial and penetrating PVS. Images for both the CSF and intravenous dextran channel were thresholded using an automated method (Otsu), and the diameter of the vessel was measured. In early time points after intracisternal infusion, tracers had not reached the penetrating arteriole PVS, so CSF tracer area was normalized to the penetrating arteriolar area ($\Delta A/A_{\text{art}}$). Tracer was found in the pial PVS starting at 15 min after infusion and was therefore normalized to the area covered by tracer before MCAO before normalizing to the pial arteriolar area ($\Delta A_{\text{norm}}/A_{\text{art}}$). For experiments in Fig. 6, E to G, time-lapse *Z*-stacks were acquired and processed using IMARIS (v. 9.2.1, Bitplane, Concord, MA, USA). To quantify the arteriolar diameter changes in *Aqp4*$^{-/-}$ and *Aqp4*$^{+/+}$ mice, the diameter was normalized to the beginning of the SI-induced constriction ($\Delta d/d_{\text{SI}}$). An ROI was placed over the penetrating arteriole and the surrounding PVS, and fluorescence intensity was normalized to the background intensity before MCAO ($\Delta F/F_0$). For ISF tracer quantification, an ROI was placed next to the PVS, making sure to avoid any surrounding capillaries or the overlying pial arteriole. The same size ROI was used for all animals. To determine the amount of PVS tracer that entered the ISF, fluorescence intensity was normalized to the time of onset of the SI ($\Delta F/F_{\text{SI}}$).

Particle tracking velocimetry

To directly measure the speed of CSF flowing through PVSs, we performed particle tracking velocimetry using time series of images obtained from in vivo 2P laser scanning microscopy. Our procedure is similar to one described previously (*38*). The particle tracking velocimetry analysis is performed using automated Matlab software (*77*, *78*), in which tens of thousands of microspheres are tracked with subpixel accuracy through the time series of images to obtain spatially and temporally resolved velocities. We use an improved algorithm in which dynamic masking is performed to remove stagnant particles from the measurements. Specifically, before identifying and tracking particles, we first subtract a unique background image from each frame that is obtained by averaging the nearest 300 frames (150 before and after). To calculate heatmaps of average flow speed (fig. S7A), the spatial domain was divided into 5-pixel–by–5-pixel bins, the velocities were time-averaged in each bin using measurements from 1 min or less (centered on the indicated time), and the speed was computed from the average velocity. To calculate the time series of $v_{\text{downstream}}$, first the downstream velocity component of each particle in each frame is computed as $\boldsymbol{u} \cdot \hat{\boldsymbol{u}}_{\text{avg}}$, where $\boldsymbol{u}$ is the instantaneous particle velocity and $\hat{\boldsymbol{u}}_{\text{avg}}$ is a field of unit vectors computed from time-averaging the velocity field over the entire time series. We then obtain $v_{\text{downstream}}$ by calculating the spatial average of all the downstream velocity components in 1-s bins and normalizing the time series by the baseline value. The baseline value is the average value of $v_{\text{downstream}}$ over the initial 2 min of the experiment before the MCAO occurs. To calculate $v_{\text{pulsatile}}$, we first obtained the time series of v_{mean} by computing the spatially averaged speed for each frame. We then defined the time series of $v_{\text{pulsatile}}$ as max (v_{mean}) − min (v_{mean}) over each cardiac cycle, divided by the baseline value of this quantity. The cardiac cycle is defined using sequential peaks of the R wave in synchronized electrocardiogram measurements. The normalized artery diameter time series was measured using custom software developed in Matlab. The measurements were performed by subsampling the time series of images obtained from in vivo 2P laser scanning microscopy by a factor of 10 and analyzing a user-defined ROI centered on a straight segment of the artery in each frame using an automated algorithm. The algorithm first separates the artery segment from the background by binarizing the ROI based on a user-defined threshold. The major and minor axes of the artery segment are then measured using the built-in Matlab function "regionprops." By ensuring that the segment length is longer than the artery diameter, the major and minor axes then correspond to the segment length and artery diameter, respectively. The normalized artery diameter time series then corresponds to the length of the minor axis in each frame, divided by the baseline value. The baseline value is the average artery diameter over the initial 2 min of the experiment before the MCAO occurs.

Glymphatic network model

Transport of water follows a network composed of PVSs, bifurcations, and terminal nodes (penetrating PVS). In mathematical language, we call the pial PVS the "edges" along the network, collected in the set E, and the terminals and bifurcations are collected in the set of "nodes," N. Thus, the network is given by nodes $i \in N$ and edges $(i, j) \in E$. The edges of the network have the property of an inverse resistance, or conductance C_{ij}. To simplify the model, we assume that each PVS is a straight pipe segment of length L_{ij} with cross-sectional area A_{ij}. The conductance is then

$$C_{\text{ij}} = \frac{kA_{\text{ij}}^2}{L_{\text{ij}}} \tag{1}$$

where $k = \pi/(8\mu)$ and μ is the dynamic viscosity, in agreement with Hagen-Poiseuille flow.

We assume that the SD wave propagates from (x_0, y_0) across a two-dimensional domain on the cortical surface with radius $R(t)$ growing at a constant speed c, so that $R(t) = ct$. The presence of the wave at node i with coordinates $(x_{\text{i}}, y_{\text{i}})$ is then defined as $w(\delta) = 1$ for $\delta \le 1$ or $w(\delta) = 0$ otherwise, where the relative distance from node to the wave's origin is $\delta_{\text{i}} := \sqrt{(x_{\text{i}} - x_0)^2 + (y_{\text{i}} - y_0)^2}/R(t)$, or from the farthest end of an edge to the wave's origin, $\delta_{\text{i,j}} = \max(\delta_{\text{i}}, \delta_{\text{j}})$.

The arrival of the SD wave at a penetrating arteriole located at site $i \in N$ triggers a dilation process

$$\frac{\mathrm{d}}{\mathrm{d}t} V_{\text{i}} = k_V \cdot w(\delta_{\text{i}}) \cdot V_{\text{i}} \cdot \left(1 - \frac{V_{\text{i}}}{V_{\text{i}}(\infty)}\right) \tag{2}$$

where $V_i(\infty)$ are the final PVS volumes. The wave covering pial surface vessels $(i, j) \in E$ triggers an analogous dilation response

$$\frac{\mathrm{d}}{\mathrm{d}t}C_{ij} = k_C \cdot w(\delta_{ij}) \cdot C_{ij} \cdot \left(1 - \frac{C_{ij}}{C_{ij}(\infty)}\right) \quad (3)$$

leading to an increase of the PVS's conductance C_{ij}, and $C_{ij}(\infty)$ are the final cross-sectional PVS conductances. Mass balance demands

$$\rho \frac{\mathrm{d}}{\mathrm{d}t}S_i + \sum\nolimits_j Q_{ij}(t) = q_i(t) \quad (4)$$

where ρ is the mass density, q_i is the flow rate in or out of a terminal node, Q_{ij} is the flow rate along an edge (i, j), and $S_i = 1/2\sum_j A_{ij}L_{ij}$ is the volume of the PVS connected to the bifurcating node i. Summation of eq. 4 over all nodes i yields the following relation: $\rho \frac{\mathrm{d}}{\mathrm{d}t}\sum_i S_i(t) + \sum_{i,j} Q_{ij}(t) = \sum_i q_i(t) = \sum_{i \in P} q_i(t) + Q_{\mathrm{MCA}}$, in which the sum over the edge flows cancels (i.e., over forward and backward directions). Furthermore, the pial surface volume is defined as $S_{\mathrm{pial}} = \sum_j S_j$ and $P \subseteq N$ is defined as the subset of nodes corresponding to penetrating arterioles which we use to split up the sum over terminal nodes into inflows (Q_{MCA}) and outflows. Thus, we can compute the inflow at the MCA

$$Q_{\mathrm{MCA}} = -\sum\nolimits_i q_i(t) + \rho \frac{\mathrm{d}}{\mathrm{d}t}S_{\mathrm{pial}} \quad (5)$$

To determine this flow rate, we first need to compute $\frac{\mathrm{d}}{\mathrm{d}t}S_{\mathrm{pial}}$. To obtain an expression for the rate of change of the cross-sectional area A_{ij}, we simply differentiate eq. 1, resulting in $\frac{\mathrm{d}}{\mathrm{d}t}A_{ij} = \frac{1}{2}k^{-1} \cdot \frac{L_{ij}}{A_{ij}} \cdot \frac{\mathrm{d}}{\mathrm{d}t}C_{ij} = \frac{1}{2}k^{-1/2} \cdot L_{ij}^{1/2} \cdot C_{ij}^{-1/2} \cdot \frac{\mathrm{d}}{\mathrm{d}t}C_{ij}$, where we in the last step again use eq. 1 or $A_{ij} = k^{-1/2} \cdot L_{ij}^{1/2} \cdot C_{ij}^{-1/2}$. Second, we need to express the terminal flow rates $q_i(t)$. The following relation holds: $q_i + \rho \frac{\mathrm{d}}{\mathrm{d}t}S_i = 0$. Hence, we can fully determine the value of eq. 5 through computation of eqs. 3 and 2. We set the initial values $V_i(t = 0) = V_0 > 0$, $S_i(t = 0) = S_0 > 0$ and $C_{ij}(t = 0) = C_0 > 0$, consistent with a constant downstream flow (inflow-outflow of the network) that terminates at the time of occlusion. For the network graph, we used data from (*40*) and data to reflect realistic vessel diameters, adapted algorithmically to optimize power dissipation losses inside the network (*79*). The algorithm was implemented in Matlab using a simple Euler forward method.

Estimated diffusion calculations

The diffusion of tracer presented in fig. S3E was estimated as previously described (*80*, *81*). To our knowledge, the diffusion coefficient of gadobutrol in live rodent brain has not been reported in the literature, so we estimated the degree of penetration of a similar sized tracer (3-kDa dextran). Calculations were done using the error function solution for plane diffusion into a half-space: $C = C_0\mathrm{erfc}\,[x/2\mathrm{sqrt}\,(D^*t)]$ and an effective diffusivity $D^* = 5.36 \times 10^{-7}$ cm^2/s (*82*). Because the D^* value was calculated for the cortex of live normoxic rats, we also used a value $D^* = 0.284 \times 10^{-7}$ cm^2/s after 1 min of terminal ischemia induced by intracardiac 1M KCl to better reflect the reduced extracellular space after the SD, as seen in fig. S2, D to F.

Statistical analysis

All statistical analyses were done in GraphPad Prism 8. Data in all graphs are plotted as mean ± standard error of the mean (SEM) over the individual data points and lines from each mouse. Parametric and nonparametric tests were selected based on normality testing and are reported in the figure legends. Normality tests were chosen depending on the sample size (D'Agostino Pearson omnibus test where possible and Shapiro-Wilk if the n was too small). Sphericity was not assumed; in all repeated measures, two-way ANOVAs and a Geisser-Greenhouse correction were performed. All hypothesis testing was two-tailed, and significance was determined at $\alpha = 0.05$.

REFERENCES AND NOTES

1. E. J. Benjamin *et al*., Heart disease and stroke statistics—2019 update: A report from the American Heart Association. *Circulation* **139**, e56–e528 (2019). doi: 10.1161/CIR.0000000000000659; pmid: 30700139
2. V. L. Feigin *et al*., Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: The GBD 2013 Study. *Neuroepidemiology* **45**, 161–176 (2015). doi: 10.1159/000441085; pmid: 26505981
3. C. Iadecola, J. Anrather, Stroke research at a crossroad: Asking the brain for directions. *Nat. Neurosci.* **14**, 1363–1368 (2011). doi: 10.1038/nn.2953; pmid: 22030546
4. R. L. Rungta *et al*., The cellular mechanisms of neuronal swelling underlying cytotoxic edema. *Cell* **161**, 610–621 (2015). doi: 10.1016/j.cell.2015.03.029; pmid: 25910210
5. J. M. Simard, T. A. Kent, M. Chen, K. V. Tarasov, V. Gerzanich, Brain oedema in focal ischaemia: Molecular pathophysiology and theoretical implications. *Lancet Neurol.* **6**, 258–268 (2007). doi: 10.1016/S1474-4422(07)70055-8; pmid: 17303532
6. D. Liang, S. Bhatta, V. Gerzanich, J. M. Simard, Cytotoxic edema: Mechanisms of pathological cell swelling. *Neurosurg. Focus* **22**, E2 (2007). doi: 10.3171/foc.2007.22.5.3; pmid: 17613233
7. A. J. Hansen, M. Nedergaard, Brain ion homeostasis in cerebral ischemia. *Neurochem. Pathol.* **9**, 195–209 (1988). pmid: 3247069
8. J. A. Stokum, V. Gerzanich, J. M. Simard, Molecular pathophysiology of cerebral edema. *J. Cereb. Blood Flow Metab.* **36**, 513–538 (2016). doi: 10.1177/0271678X15617172; pmid: 26661240
9. I. Klatzo, Neuropathological aspects of brain edema. *J. Neuropathol. Exp. Neurol.* **26**, 1–14 (1967). doi: 10.1097/00005072-196701000-00001; pmid: 5336776
10. A. Van Harreveld, Changes in the diameter of apical dendrites during spreading depression. *Am. J. Physiol.* **192**, 457–463 (1958). doi: 10.1152/ajplegacy.1958.192.3.457; pmid: 13520934
11. A. van Harreveld, S. Ochs, Cerebral impedance changes after circulatory arrest. *Am. J. Physiol.* **187**, 180–192 (1956). doi: 10.1152/ajplegacy.1956.187.1.180; pmid: 13362612
12. J. P. Dreier, C. L. Lemale, V. Kola, A. Friedman, K. Schoknecht, Spreading depolarization is not an epiphenomenon but the principal mechanism of the cytotoxic edema in various gray matter structures of the brain during stroke. *Neuropharmacology* **134**, 189–207 (2018). doi: 10.1016/j.neuropharm.2017.09.027; pmid: 28941738
13. J. P. Dreier, The role of spreading depression, spreading depolarization and spreading ischemia in neurological disease. *Nat. Med.* **17**, 439–447 (2011). doi: 10.1038/nm.2333; pmid: 21475241
14. M. Nedergaard, J. Astrup, Infarct rim: Effect of hyperglycemia on direct current potential and [^{14}C]2-deoxyglucose phosphorylation. *J. Cereb. Blood Flow Metab.* **6**, 607–615 (1986). doi: 10.1038/jcbfm.1986.108; pmid: 3760045
15. I. Klatzo, Pathophysiological aspects of brain edema. *Acta Neuropathol.* **72**, 236–239 (1987). doi: 10.1007/BF00691095; pmid: 3564903
16. D. Knowland *et al*., Stepwise recruitment of transcellular and paracellular pathways underlies blood-brain barrier breakdown in stroke. *Neuron* **82**, 603–617 (2014). doi: 10.1016/j.neuron.2014.03.003; pmid: 24746419
17. E. J. Kang *et al*., Blood-brain barrier opening to large molecules does not imply blood-brain barrier opening to small ions. *Neurobiol. Dis.* **52**, 204–218 (2013). doi: 10.1016/j.nbd.2012.12.007; pmid: 23291193
18. W. Young, Z. H. Rappaport, D. J. Chalif, E. S. Flamm, Regional brain sodium, potassium, and water changes in the rat middle cerebral artery occlusion model of ischemia. *Stroke* **18**, 751–759 (1987). doi: 10.1161/01.STR.18.4.751; pmid: 3603602
19. T. W. Battey *et al*., Brain edema predicts outcome after nonlacunar ischemic stroke. *Stroke* **45**, 3643–3648 (2014). doi: 10.1161/STROKEAHA.114.006884; pmid: 25336512
20. S. Hatashita, J. T. Hoff, S. M. Salamat, Ischemic brain edema and the osmotic gradient between blood and brain. *J. Cereb. Blood Flow Metab.* **8**, 552–559 (1988). doi: 10.1038/jcbfm.1988.96; pmid: 3392116
21. S. Hatashita, J. T. Hoff, Brain edema and cerebrovascular permeability during cerebral ischemia in rats. *Stroke* **21**, 582–588 (1990). doi: 10.1161/01.STR.21.4.582; pmid: 1691534
22. A. S. Thrane, V. Rangroo Thrane, M. Nedergaard, Drowning stars: Reassessing the role of astrocytes in brain edema. *Trends Neurosci.* **37**, 620–628 (2014). doi: 10.1016/j.tins.2014.08.010; pmid: 25236348
23. J. J. Iliff *et al*., A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. *Sci. Transl. Med.* **4**, 147ra111 (2012). doi: 10.1126/scitranslmed.3003748; pmid: 22896675
24. L. Xie *et al*., Sleep drives metabolite clearance from the adult brain. *Science* **342**, 373–377 (2013). doi: 10.1126/science.1241224; pmid: 24136970
25. Y. Inoue *et al*., Detection of necrotic neural response in super-acute cerebral ischemia using activity-induced manganese-enhanced (AIM) MRI. *NMR Biomed.* **23**, 304–312 (2010). pmid: 19950123
26. O. Gotoh, T. Asano, T. Koide, K. Takakura, Ischemic brain edema following occlusion of the middle cerebral artery in the rat. I: The time courses of the brain water, sodium and potassium contents and blood-brain barrier permeability to 125I-albumin. *Stroke* **16**, 101–109 (1985). doi: 10.1161/01.STR.16.1.101; pmid: 3966252
27. W. B. Sisson, W. H. Oldendorf, Brain distribution spaces of mannitol-3H, inulin-14C, and dextran-14C in the rat. *Am. J. Physiol.* **221**, 214–217 (1971). doi: 10.1152/ajplegacy.1971.221.1.214; pmid: 5555788
28. J. J. Iliff *et al*., Brain-wide pathway for waste clearance captured by contrast-enhanced MRI. *J. Clin. Invest.* **123**, 1299–1309 (2013). doi: 10.1172/JCI67677; pmid: 23434588
29. J. P. Dreier, C. Reiffurth, The stroke-migraine depolarization continuum. *Neuron* **86**, 902–922 (2015). doi: 10.1016/j.neuron.2015.04.004; pmid: 25996134
30. V. B. Bogdanov *et al*., Susceptibility of primary sensory cortex to spreading depolarizations. *j. neurosci.* **36**, 4733–4743 (2016). doi: 10.1523/JNEUROSCI.3694-15.2016; pmid: 27122032
31. R. Enger *et al*., Dynamics of ionic shifts in cortical spreading depression. *Cereb. Cortex* **25**, 4469–4476 (2015). doi: 10.1093/cercor/bhv054; pmid: 25840424
32. H. Monai *et al*., Calcium imaging reveals glial involvement in transcranial direct current stimulation-induced plasticity in mouse brain. *Nat. Commun.* **7**, 11100 (2016). doi: 10.1038/ncomms11100; pmid: 27000523
33. B. A. Plog *et al*., Transcranial optical imaging reveals a pathway for optimizing the delivery of immunotherapeutics to the brain. *JCI Insight* **3**, 120922 (2018). doi: 10.1172/jci.insight.120922; pmid: 30333324
34. A. J. Strong *et al*., Peri-infarct depolarizations lead to loss of perfusion in ischaemic gyrencephalic cerebral cortex. *Brain* **130**, 995–1008 (2007). doi: 10.1093/brain/awl392; pmid: 17438018
35. H. K. Shin *et al*., Vasoconstrictive neurovascular coupling during focal ischemic depolarizations. *J. Cereb. Blood Flow Metab.* **26**, 1018–1030 (2006). doi: 10.1038/sj.jcbfm.9600252; pmid: 16340958

36. J. P. Dreier *et al*., Nitric oxide scavenging by hemoglobin or nitric oxide synthase inhibition by *N*-nitro-L-arginine induces cortical spreading ischemia when K^+ is increased in the subarachnoid space. *J. Cereb. Blood Flow Metab.* **18**, 978–990 (1998). doi: 10.1097/00004647-199809000-00007; pmid: 9740101
37. J. Chuquet, L. Hollender, E. A. Nimchinsky, High-resolution in vivo imaging of the neurovascular unit during spreading depression. *J. Neurosci.* **27**, 4036–4044 (2007). doi: 10.1523/JNEUROSCI.0721-07.2007; pmid: 17428981
38. H. Mestre *et al*., Flow of cerebrospinal fluid is driven by arterial pulsations and is reduced in hypertension. *Nat. Commun.* **9**, 4878 (2018). doi: 10.1038/s41467-018-07318-3; pmid: 30451853
39. A. J. Schain, A. Melo-Carrillo, A. M. Strassman, R. Burstein, Cortical spreading depression closes paravascular space and impairs glymphatic flow: Implications for migraine headache. *J. Neurosci.* **37**, 2904–2915 (2017). doi: 10.1523/JNEUROSCI.3390-16.2017; pmid: 28193695
40. P. Blinder, A. Y. Shih, C. Rafie, D. Kleinfeld, Topological basis for the robust distribution of blood to rodent neocortex. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12670–12675 (2010). doi: 10.1073/pnas.1007239107; pmid: 20616030
41. J. P. Dreier *et al*., Ischaemia triggered by spreading neuronal activation is inhibited by vasodilators in rats. *J. Physiol.* **531**, 515–526 (2001). doi: 10.1111/j.1469-7793.2001.0515i.x; pmid: 11230523
42. T. Iijima, G. Mies, K. A. Hossmann, Repeated negative DC deflections in rat cortex following middle cerebral artery occlusion are abolished by MK-801: Effect on volume of ischemic injury. *J. Cereb. Blood Flow Metab.* **12**, 727–733 (1992). doi: 10.1038/jcbfm.1992.103; pmid: 1506440
43. R. Gill, P. Andiné, L. Hillered, L. Persson, H. Hagberg, The effect of MK-801 on cortical spreading depression in the penumbral zone following focal ischaemia in the rat. *J. Cereb. Blood Flow Metab.* **12**, 371–379 (1992). doi: 10.1038/jcbfm.1992.54; pmid: 1314840
44. M. Lauritzen, A. J. Hansen, The effect of glutamate receptor blockade on anoxic depolarization and cortical spreading depression. *J. Cereb. Blood Flow Metab.* **12**, 223–229 (1992). doi: 10.1038/jcbfm.1992.32; pmid: 1312539
45. H. Mestre *et al*., Aquaporin-4-dependent glymphatic solute transport in the rodent brain. *eLife* **7**, e40070 (2018). doi: 10.7554/eLife.40070; pmid: 30561329
46. X. Yao, N. Derugin, G. T. Manley, A. S. Verkman, Reduced brain edema and infarct volume in aquaporin-4 deficient mice after transient focal cerebral ischemia. *Neurosci. Lett.* **584**, 368–372 (2015). doi: 10.1016/j.neulet.2014.10.040; pmid: 25449874
47. G. T. Manley *et al*., Aquaporin-4 deletion in mice reduces brain edema after acute water intoxication and ischemic stroke. *Nat. Med.* **6**, 159–163 (2000). doi: 10.1038/72256; pmid: 10655103
48. H. Igarashi, V. J. Huber, M. Tsujita, T. Nakada, Pretreatment with a novel aquaporin 4 inhibitor, TGN-020, significantly reduces ischemic cerebral edema. *Neurol. Sci.* **32**, 113–116 (2011). doi: 10.1007/s10072-010-0431-1; pmid: 20924629
49. I. Pirici *et al*., Inhibition of aquaporin-4 improves the outcome of ischaemic stroke and modulates brain paravascular drainage pathways. *Int. J. Mol. Sci.* **19**, 46 (2017). doi: 10.3390/ijms19010046; pmid: 29295526
50. A. L. Betz, R. F. Keep, M. E. Beer, X. D. Ren, Blood-brain barrier permeability and brain concentration of sodium, potassium, and chloride during focal ischemia. *J. Cereb. Blood Flow Metab.* **14**, 29–37 (1994). doi: 10.1038/jcbfm.1994.5; pmid: 8263055
51. S. Ishimaru, K. A. Hossmann, Relationship between cerebral blood flow and blood-brain barrier permeability of sodium and albumin in cerebral infarcts of rats. *Acta Neurochir. Suppl.* **51**, 216–219 (1990). doi: 10.1007/978-3-7091-9115-6_73; pmid: 2089898
52. W. D. Lo, A. L. Betz, G. P. Schielke, J. T. Hoff, Transport of sodium from blood to brain in ischemic brain edema. *Stroke* **18**, 150–157 (1987). doi: 10.1161/01.STR.18.1.150; pmid: 3810748
53. U. Ito, Y. Hakamata, E. Kawakami, K. Oyanagi, Temporary cerebral ischemia results in swollen astrocytic end-feet that compress microvessels and lead to delayed focal cortical infarction. *J. Cereb. Blood Flow Metab.* **31**, 328–338 (2011). doi: 10.1038/jcbfm.2010.97; pmid: 20588315
54. L. Khennouf *et al*., Active role of capillary pericytes during stimulation-induced activity and spreading depolarization. *Brain* **141**, 2032–2046 (2018). doi: 10.1093/brain/awy143; pmid: 30053174
55. J. M. Simard *et al*., Glibenclamide in cerebral ischemia and stroke. *Neurocrit. Care* **20**, 319–333 (2014). doi: 10.1007/s12028-013-9923-1; pmid: 24132564
56. T. Gaberel *et al*., Impaired glymphatic perfusion after strokes revealed by contrast-enhanced MRI: A new target for fibrinolysis? *Stroke* **45**, 3092–3096 (2014). doi: 10.1161/STROKEAHA.114.006617; pmid: 25190438
57. J. Lückl *et al*., Peri-infarct flow transients predict outcome in rat focal brain ischemia. *Neuroscience* **226**, 197–207 (2012). doi: 10.1016/j.neuroscience.2012.08.049; pmid: 22986160
58. J. A. Hartings *et al*., The continuum of spreading depolarizations in acute cortical lesion development: Examining Leão's legacy. *J. Cereb. Blood Flow Metab.* **37**, 1571–1594 (2017). doi: 10.1177/0271678X16654495; pmid: 27328690
59. X. Yao *et al*., Aquaporin-4 regulates the velocity and frequency of cortical spreading depression in mice. *Glia* **63**, 1860–1869 (2015). doi: 10.1002/glia.22853; pmid: 25944186
60. R. Enger *et al*., Deletion of aquaporin-4 curtails extracellular glutamate elevation in cortical spreading depression in awake mice. *Cereb. Cortex* **27**, 24–33 (2017). doi: 10.1093/cercor/bhw359; pmid: 28365776
61. A. S. Thrane *et al*., In vivo NADH fluorescence imaging indicates effect of aquaporin-4 deletion on oxygen microdistribution in cortical spreading depression. *J. Cereb. Blood Flow Metab.* **33**, 996–999 (2013). doi: 10.1038/jcbfm.2013.63; pmid: 23611872
62. Z. Bere, T. P. Obrenovitch, F. Bari, E. Farkas, Ischemia-induced depolarizations and associated hemodynamic responses in incomplete global forebrain ischemia in rats. *Neuroscience* **260**, 217–226 (2014). doi: 10.1016/j.neuroscience.2013.12.032; pmid: 24365459
63. J. A. Hartings *et al*., Subarachnoid blood acutely induces spreading depolarizations and early cortical infarction. *Brain* **140**, 2673–2690 (2017). doi: 10.1093/brain/awx214; pmid: 28969382
64. J. Woitzik *et al*., Propagation of cortical spreading depolarization in the human cortex after malignant stroke. *Neurology* **80**, 1095–1102 (2013). doi: 10.1212/WNL.0b013e3182886932; pmid: 23446683
65. J. M. Hinzman *et al*., Inverse neurovascular coupling to cortical spreading depolarizations in severe brain trauma. *Brain* **137**, 2960–2972 (2014). doi: 10.1093/brain/awu241; pmid: 25154387
66. J. P. Dreier *et al*., Cortical spreading ischaemia is a novel process involved in ischaemic damage in patients with aneurysmal subarachnoid haemorrhage. *Brain* **132**, 1866–1881 (2009). doi: 10.1093/brain/awp102; pmid: 19420089
67. J. Lückl *et al*., The negative ultraslow potential, electrophysiological correlate of infarction in the human cortex. *Brain* **141**, 1734–1752 (2018). doi: 10.1093/brain/awy102; pmid: 29668855
68. A. L. R. Xavier *et al*., Cannula implantation into the cisterna magna of rodents. *J. Vis. Exp.* **2018**, 57378 (2018). doi: 10.3791/57378; pmid: 29889209
69. O. Albayram *et al*., *Cis* P-tau is induced in clinical and preclinical brain injury and contributes to post-injury sequelae. *Nat. Commun.* **8**, 1000 (2017). doi: 10.1038/s41467-017-01068-4; pmid: 29042562
70. J. K. Karimy *et al*., Inflammation-dependent cerebrospinal fluid hypersecretion by the choroid plexus epithelium in posthemorrhagic hydrocephalus. *Nat. Med.* **23**, 997–1003 (2017). doi: 10.1038/nm.4361; pmid: 28692063
71. J. K. Karimy *et al*., A novel method to study cerebrospinal fluid dynamics in rats. *J. Neurosci. Methods* **241**, 78–84 (2015). doi: 10.1016/j.jneumeth.2014.12.015; pmid: 25554415
72. D. W. McBride, D. Klebe, J. Tang, J. H. Zhang, Correcting for brain swelling's effects on infarct volume calculation after middle cerebral artery occlusion in rats. *Transl. Stroke Res.* **6**, 323–338 (2015). doi: 10.1007/s12975-015-0400-3; pmid: 25933988
73. T. D. Nevins, D. H. Kelley, Front tracking for quantifying advection-reaction-diffusion. *Chaos* **27**, 043105 (2017). doi: 10.1063/1.4979668; pmid: 28456164
74. B. B. Avants *et al*., A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011). doi: 10.1016/j.neuroimage.2010.09.025; pmid: 20851191
75. B. Avants *et al*., Multivariate analysis of structural and diffusion imaging in traumatic brain injury. *Acad. Radiol.* **15**, 1360–1375 (2008). doi: 10.1016/j.acra.2008.07.007; pmid: 18995188
76. G. A. Johnson *et al*., Waxholm space: An image-based reference for coordinating mouse brain research. *Neuroimage* **53**, 365–372 (2010). doi: 10.1016/j.neuroimage.2010.06.067; pmid: 20600960
77. N. T. Ouellette, H. T. Xu, E. Bodenschatz, A quantitative study of three-dimensional Lagrangian particle tracking algorithms. *Exp. Fluids* **40**, 301–313 (2006). doi: 10.1007/s00348-005-0068-7
78. D. H. Kelley, N. T. Ouellette, Using particle tracking to measure flow instabilities in an undergraduate laboratory experiment. *Am. J. Phys.* **79**, 267–273 (2011). doi: 10.1119/1.3536647
79. E. A. Martens, K. Klemm, Transitions from trees to cycles in adaptive flow networks. *Front. Phys.* **5**, 62 (2017). doi: 10.3389/fphy.2017.00062
80. E. Syková, C. Nicholson, Diffusion in brain extracellular space. *Physiol. Rev.* **88**, 1277–1340 (2008). doi: 10.1152/physrev.00027.2007; pmid: 18923183
81. G. Ringstad *et al*., Brain-wide glymphatic enhancement and clearance in humans assessed with MRI. *JCI Insight* **3**, e121537 (2018). doi: 10.1172/jci.insight.121537; pmid: 29997300
82. R. G. Thorne, C. Nicholson, In vivo diffusion analysis with quantum dots and dextrans predicts the width of brain extracellular space. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5567–5572 (2006). doi: 10.1073/pnas.0509425103; pmid: 16567637

ACKNOWLEDGMENTS

We thank D. Xue for assistance with the illustrations. **Funding:** National Institute of Neurological Disorders and Stroke and the National Institute on Aging (U.S. National Institutes of Health: R01NS100366 to M.N.; RF1AG057575 to M.N., D.H.K., and J.H.T.; and K08NS089830 to R.I.M.), NIH-NINDS (R35 NS097265 to D.K.), the U.S. Army Research Office (grant no. MURI W911NF1910280 to M.N., D.H.K., and J.H.T.), Fondation Leducq Transatlantic Networks of Excellence Program, Novo Nordisk and Lundbeck Foundations, and the EU Horizon 2020 research and innovation program (grant no. 666881; SVDs@target). The views and conclusions contained in this manuscript are solely those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Institutes of Health, Army Research Office, or the U.S. government. The U.S. government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein. **Author contributions:** Conceptualization: H.M., T.D., Y.M., M.N.; Investigation: H.M., T.D., A.M.S., G.L., A.J.S., W.P., O.S., Y.M.; Methodology: H.M., T.D., P.A.R.B., J.T., D.H.K., P.G.H., E.A.M., R.I.M., P.B., D.K., H.H., Y.M., M.N.; Formal analysis: H.M., T.D., A.M.S., K.N.M., F.F.S., P.A.R.B., L.B., E.R.T., J.T., D.H.K., J.H.T., P.G.H., E.A.M., R.I.M., Y.M.; Writing–original draft: H.M., M.N.; Writing–review and editing: H.M., T.D., A.M.S., A.J.S., F.F.S., P.A.R.B., J.H.T., R.I.M., H.H., Y.M., M.N.; Funding acquisition: M.N., D.H.K., J.H.T.; **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Glt1-GCaMP7 mice (no. RBRC09650) are available through the RIKEN BioResource Center repository under a material transfer agreement with RIKEN BRC.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/eaax7171/suppl/DC1
Figs. S1 to S10
Table S1
References

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

22 June 2019; resubmitted 16 December 2019
Accepted 17 January 2020
Published online 30 January 2020
10.1126/science.aax7171

RESEARCH ARTICLE SUMMARY

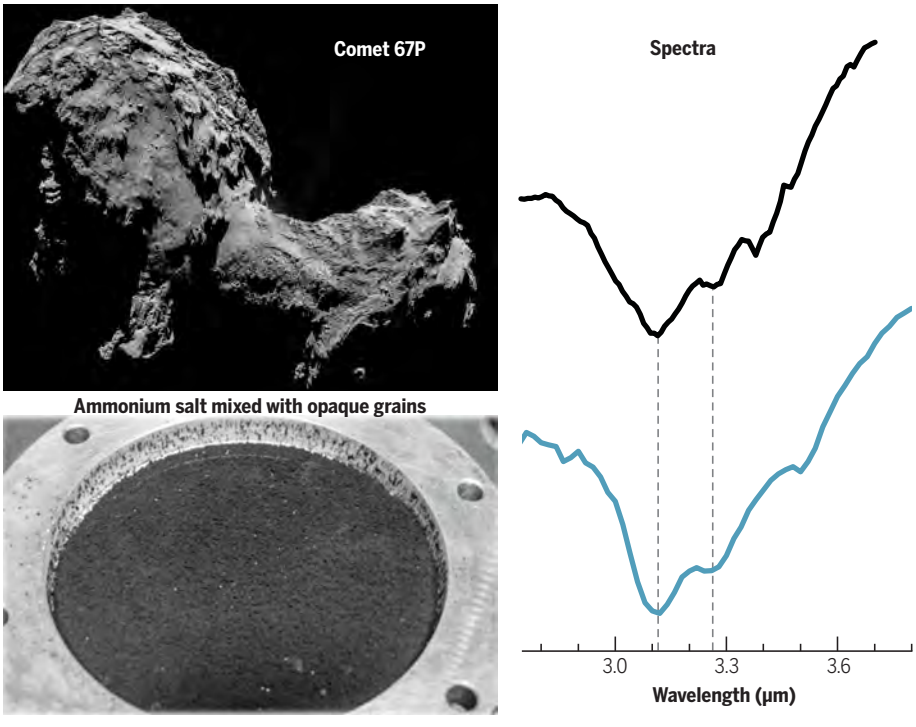
COMETARY SCIENCE

Ammonium salts are a reservoir of nitrogen on a cometary nucleus and possibly on some asteroids

Olivier Poch*, Istiqomah Istiqomah, Eric Quirico, Pierre Beck, Bernard Schmitt, Patrice Theulé, Alexandre Faure, Pierre Hily-Blant, Lydie Bonal, Andrea Raponi, Mauro Ciarniello, Batiste Rousseau, Sandra Potin, Olivier Brissaud, Laurène Flandinet, Gianrico Filacchione, Antoine Pommerol, Nicolas Thomas, David Kappel, Vito Mennella, Lyuba Moroz, Vassilissa Vinogradoff, Gabriele Arnold, Stéphane Erard, Dominique Bockelée-Morvan, Cédric Leyrat, Fabrizio Capaccioni, Maria Cristina De Sanctis, Andrea Longobardo, Francesca Mancarella, Ernesto Palomba, Federico Tosi

INTRODUCTION: Comets and asteroids preserve information on the earliest stages of Solar System formation and on the composition of its building blocks. The nature of their solid material can be investigated by analyzing the sunlight scattered by their surfaces. The nucleus of comet 67P/Churyumov-Gerasimenko (hereafter 67P) was mapped by the Visible and InfraRed Thermal Imaging Spectrometer, Mapping Channel (VIRTIS-M) on the Rosetta spacecraft from 2014 to 2015. The nucleus appeared almost spectrally uniform from 0.4 to 4 μm, characterized by a low reflectance of few percent, a reddish color, and an unidentified broad absorption feature around 3.2 μm, which was ubiquitous throughout the surface. The darkness and the color of comet 67P could be due to a mixture of refractory organic molecules and opaque minerals. Although water ice may contribute to the 3.2-μm absorption, it cannot explain the entire feature.

RATIONALE: Semivolatile compounds of low molecular weight, such as carboxylic (–COOH)–bearing molecules or ammonium (NH_4^+) ions, have been proposed as potential carriers of the 3.2-μm absorption feature. To test these hypotheses, we performed laboratory experiments to measure the reflectance spectra of these compounds mixed in a porous matrix of submicrometric opaque mineral grains, under simulated comet-like conditions (170 to 200 K, $<10^{-5}$ mbar).



Comparison of ammonium formate spectrum with the average spectrum of comet 67P. The average reflectance spectrum of comet 67P (black line) and the spectrum of a mixture of ammonium formate (NH_4^+ $HCOO^-$) with opaque grains measured in the laboratory under comet-like conditions (blue line). Also shown are views of the 4-km-diameter comet nucleus (Credit: ESA/Rosetta/NAVCAM–CC BY-SA IGO 3.0; http://creativecommons.org/licenses/by-sa/3.0/igo) and the 48-mm-diameter laboratory sample.

RESULTS: The 3.2-μm absorption feature is consistent with ammonium salts mixed with the dark cometary surface material. We attribute additional absorption features to carbonaceous compounds and traces of water ice. Several ammonium salts can match the absorption feature equally well: ammonium formate, ammonium sulfate, or ammonium citrate. A mixture of different ammonium salts could be present.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw7462

Ammonium salts at the surface of comet 67P could have been synthesized through acid-base reactions of ammonia (NH_3) with the corresponding acid molecules in solid ices. That reaction may have occurred in the interstellar medium, in the protoplanetary disk, or during the sublimation of the ices in the cometary nucleus.

The depth of the band suggests that the cometary surface contains an upper limit of ~40 weight % (wt %) of ammonium salts, but the exact concentration remains unknown. If the amount of ammonium salts is higher than ~5 wt %, they constitute the dominant reservoir of nitrogen in the comet, containing more nitrogen than the refractory organic matter and the volatile species, such as NH_3 and N_2. Consequently, the abundance of nitrogen in this comet is closer to that of the Sun than previously thought.

CONCLUSION: Ammonium salts may dominate the reservoir of nitrogen in comets. Their presence in cometary dust may explain increases of gas-phase NH_3 and HCN observed in some comets when close to the Sun, which could be caused by the thermal dissociation of ammonium salts. Several asteroids in the Main Belt, Jupiter's Trojan asteroids, and its small moon Himalia have similar spectra to that of comet 67P, with a broad spectral absorption feature at 3.1 to 3.2 μm, which we suggest could also be due to ammonium salts. The dwarf planet Ceres has ammoniated phyllosilicates on its surface, which may have formed from ammonium ions inherited from outer Solar System objects with compositions similar to that of comet 67P. The presence of these salts on comet 67P, and possibly on other primitive Solar System bodies, suggests a compositional link between asteroids, comets, and the protosolar nebula. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: olivier.poch@univ-grenoble-alpes.fr
Cite this article as O. Poch *et al.*, *Science* **367**, eaaw7462 (2020); DOI: 10.1126/science.aaw7462

RESEARCH ARTICLE

COMETARY SCIENCE

Ammonium salts are a reservoir of nitrogen on a cometary nucleus and possibly on some asteroids

Olivier Poch[1]*, Istiqomah Istiqomah[1], Eric Quirico[1], Pierre Beck[1,2], Bernard Schmitt[1], Patrice Theulé[3], Alexandre Faure[1], Pierre Hily-Blant[1], Lydie Bonal[1], Andrea Raponi[4], Mauro Ciarniello[4], Batiste Rousseau[1]†, Sandra Potin[1], Olivier Brissaud[1], Laurène Flandinet[1], Gianrico Filacchione[4], Antoine Pommerol[5], Nicolas Thomas[5], David Kappel[6,7], Vito Mennella[8], Lyuba Moroz[7], Vassilissa Vinogradoff[9], Gabriele Arnold[7], Stéphane Erard[10], Dominique Bockelée-Morvan[10], Cédric Leyrat[10], Fabrizio Capaccioni[4], Maria Cristina De Sanctis[4], Andrea Longobardo[4,11], Francesca Mancarella[12], Ernesto Palomba[4], Federico Tosi[4]

The measured nitrogen-to-carbon ratio in comets is lower than for the Sun, a discrepancy which could be alleviated if there is an unknown reservoir of nitrogen in comets. The nucleus of comet 67P/Churyumov-Gerasimenko exhibits an unidentified broad spectral reflectance feature around 3.2 micrometers, which is ubiquitous across its surface. On the basis of laboratory experiments, we attribute this absorption band to ammonium salts mixed with dust on the surface. The depth of the band indicates that semivolatile ammonium salts are a substantial reservoir of nitrogen in the comet, potentially dominating over refractory organic matter and more volatile species. Similar absorption features appear in the spectra of some asteroids, implying a compositional link between asteroids, comets, and the parent interstellar cloud.

The composition of comets and asteroids can be investigated from the light scattered by their surfaces. For objects of which the visible to near-infrared wavelength range shows no, or only weak, spectral features, analysis of the 3-μm region (between roughly 2.4 and 3.6 μm) can be used to investigate volatile and organic compounds present on their surfaces (*1*). The Visible and InfraRed Thermal Imaging Spectrometer, Mapping Channel (VIRTIS-M) instrument (*2*) on the Rosetta spacecraft observed the nucleus of comet 67P/Churyumov-Gerasimenko (hereafter 67P) in the spectral range of 0.2 to 5.1 μm (*3*). The surface imaged by VIRTIS-M appears almost spectrally uniform (*4*), characterized by a very low reflectance [geometric albedo of 6% at 0.55 μm (*5*)], positive (red) visible and infrared spectral slopes, and a broad absorption feature from 2.8 to 3.6 μm, centered at 3.2 μm (*3*, *4*).

This absorption band, which has not been detected on other comets, is observed on all types of surface terrains and was persistently observed from August 2014 when comet 67P was 3.6 astronomical units (au) from the Sun and cometary activity was weak until just before the comet reached its closest point to the Sun and experienced maximum activity in May 2015 at 1.7 au, for as long as the VIRTIS-M infrared channel could record measurements (*6*). Analyses of the VIRTIS-M reflectance spectra ascribed the darkness and slope to a refractory polyaromatic carbonaceous component mixed with opaque minerals (anhydrous Fe-sulfides and Fe–Ni alloys), but the carrier of the 3.2-μm feature remained unknown (*3*). Water ice contributes to this absorption on some parts of the surface (*6*–*9*), causing a broadening and deepening of the absorption feature from 2.7 to 3.1 μm, but cannot explain the entire feature (*3*, *10*). Except in specific ice-rich areas, the surface of the comet nucleus is uniform in composition, with a predominance of non-ice materials (*9*). Semivolatile materials of low molecular weight have been proposed as carriers of the 3.2-μm feature, with carboxylic (–COOH)–bearing molecules or NH_4^+ ions being the most plausible candidates (*11*). However, a lack of reference spectral data for these compounds has prevented a firm attribution of the feature.

Spectral identification of ammonium salts

We conducted laboratory experiments to produce analogs of cometary surface material and measured their reflectance spectra under comet-like conditions (low temperature and high vacuum) (*12*). Cometary dust is known to consist of aggregated submicrometer-sized grains (*13*), and opaque iron sulfides are probable contributors to the low albedo of comet nuclei (*11*, *14*). We therefore used submicrometer-sized grains of pyrrhotite ($Fe_{1-x}S$, with $0 < x < 0.2$) mixed with different candidate compounds (carboxylic acid, ammonium salts) to test for the 3.2-μm feature. The pyrrhotite grains and candidate compounds were mixed in liquid water then frozen to obtain ice-dust particles (*12*). By sublimating these particles in a thermal vacuum chamber, we formed very porous mixtures made of submicrometer-sized grains (hereafter "sublimate residues") (figs. S1 and S2). This is representative of the process we expect at the surface of a comet nucleus, and the resulting textures strongly influenced the band depths of the reflectance spectra (*15*), which we attempted to reproduce to allow quantification of the components of the cometary surface (*12*).

Shown in Fig. 1A is the reflectance spectrum of the sublimate residue made of pyrrhotite grains mixed with ≲17 weight % (wt %) [≲43 volume % (vol %)] ammonium formate ($NH_4^+ HCOO^-$). Also shown is an average spectrum of 67P from a combination of VIRTIS-M observations taken between August to September 2014, when Rosetta was 50 to 350 km from the nucleus and 67P was 3.6 to 3.3 AU from the Sun (*12*). The position of the cometary absorption band, its asymmetric shape, and the minima at 3.1 and 3.3 μm all match the absorption bands owing to the N–H vibration modes of NH_4^+ in ammonium formate (Fig. 1A and fig. S5). The spectral resolving power $RP \equiv \lambda/\Delta\lambda_{resolution}$ (where λ is the wavelength and $\Delta\lambda_{resolution}$ is the spectral resolution) at 3.5 μm is $RP = 90$ for the laboratory spectrum and 233 for the VIRTIS spectrum. The residual cometary spectrum, obtained by dividing the comet spectrum by the ammonium salt spectrum, is flat from 3.05 to 3.35 μm (Fig. 1A), but features in the range of 3.35 to 3.60 μm indicate the additional presence of C–H stretching modes in carbonaceous compounds (*10*). The proximity of these C–H modes, the limited spectral resolution, and the limited spectral sampling impede a search for a weaker N–H mode of ammonium salts centered around 3.50 μm (fig. S5 and table S2). Other differences between these spectra are due to the contribution of additional compounds (possibly traces of water ice around 3.0 μm)

[1]Université Grenoble Alpes, Centre National de la Recherche Scientifique (CNRS), Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), 38000 Grenoble, France. [2]Institut Universitaire de France (IUF), Paris, France. [3]Aix-Marseille Université, CNRS, Centre National d'Etudes Spatiales (CNES), Laboratoire d'Astrophysique de Marseille (LAM), Marseille, France. [4]Istituto di Astrofisica e Planetologia Spaziali (IAPS), Istituto Nazionale di Astrofisica (INAF), 00133 Rome, Italy. [5]Physikalisches Institut, Sidlerstrasse 5, University of Bern, CH-3012 Bern, Switzerland. [6]Institute of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany. [7]Institute for Planetary Research, German Aerospace Center (DLR), 12489 Berlin, Germany. [8]Istituto Nazionale di Astrofisica (INAF)–Osservatorio Astronomico di Capodimonte, Napoli, Italy. [9]CNRS, Aix-Marseille Université, Laboratoire Physique des Interactions Ioniques et Moléculaires (PIIM), Unité Mixte de Recherche (UMR) CNRS 7345, 13397 Marseille, France. [10]Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique (LESIA), Observatoire de Paris, Université Paris Sciences et Lettres (PSL), CNRS, Sorbonne Université, Université de Paris, 92195 Meudon, France. [11]Dipartimento di Scienze e Tecnologie (DIST), Università Parthenope, 80143 Napoli, Italy. [12]Dipartimento di Matematica e Fisica "E. De Giorgi," Università del Salento, Lecce, Italy.
*Corresponding author. Email: olivier.poch@univ-grenoble-alpes.fr
†Present address: Istituto di Astrofisica e Planetologia Spaziali (IAPS), Istituto Nazionale di Astrofisica (INAF), 00133 Rome, Italy.

and/or differing properties of the salts present on the cometary surface (such as concentration, mixing, and counter-ions). We investigated other candidate compounds—fine water ice grains, carboxylic acid, or hydrated minerals—but these do not match the 3.2-μm feature (Fig. 1B). The laboratory spectra of five ammonium salts we investigated are shown in Fig. 2. Ammonium formate, ammonium sulfate, or ammonium citrate all reproduce the 3.1- and 3.3-μm absorption bands observed on the comet. For ammonium carbamate and ammonium chloride, the corresponding bands are shifted to longer wavelengths or have different spectral shapes (Fig. 2).

The similarity of band shapes and positions leads us to conclude that NH_4^+ in ammonium salts is the main species responsible for the

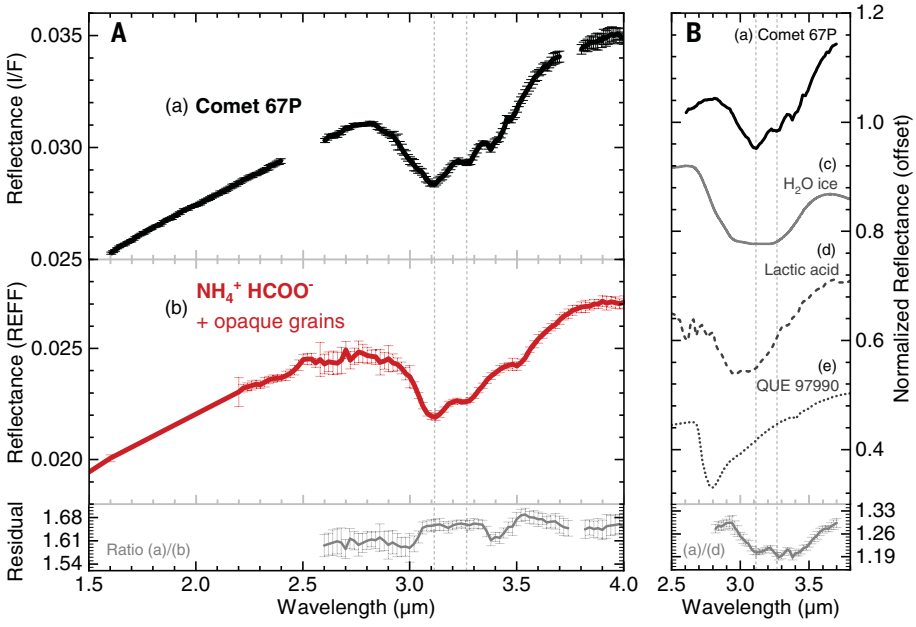


Fig. 1. Comparison of NH_4^+ $HCOO^-$ spectrum with the average spectrum of comet 67P. (**A**) The average reflectance spectrum of comet 67P in the 3.0-μm region [(a), black line; the vertical scale is given in radiance factor I/F] and the spectrum of a sublimate residue containing ≲17 wt % ammonium formate mixed with ≳83 wt % pyrrhotite grains at 170 to 200 K [(b), red line; the vertical scale is given in reflectance factor REFF]. The same spectra overlain on each other is shown in fig. S5. Gaps in the comet 67P spectrum are due to the instrument's diffraction order sorting filters. Both spectra have the same shape and minima at 3.1 and 3.3 μm (dashed gray vertical lines). (**B**) Reflectance spectra (normalized at 2.5 μm) of other compounds that do not match the comet 67P spectrum: (c) a model spectrum of 1-μm-diameter pure water ice grains (solid line) [Hapke model (*60*), using optical constants at 145 K (*61*), spectrum normalized, scaled by a factor of 0.14, and offset by −0.08]; (d) a measured spectrum of a sublimate residue containing ≲17 wt % lactic acid mixed with pyrrhotite grains at 170 to 200 K (dashed line, offset by −0.35) (*12*); and (e) spectrum of the primitive carbonaceous chondrite meteorite QUE 97990, which is rich in hydrous silicates, measured under 400 to 475 K and high vacuum (dotted line, offset by −0.55) (*62*). Between 2.5 and 2.8 μm, the spectra of sublimate residues are affected by measurement artifacts because of the presence of water vapor in the optical path. Residual spectra, calculated by dividing the comet 67P spectrum (a) by the experimental spectrum (b) or (d), are shown at the bottom of (A) and (B). Error bars indicate the ±1σ uncertainties.

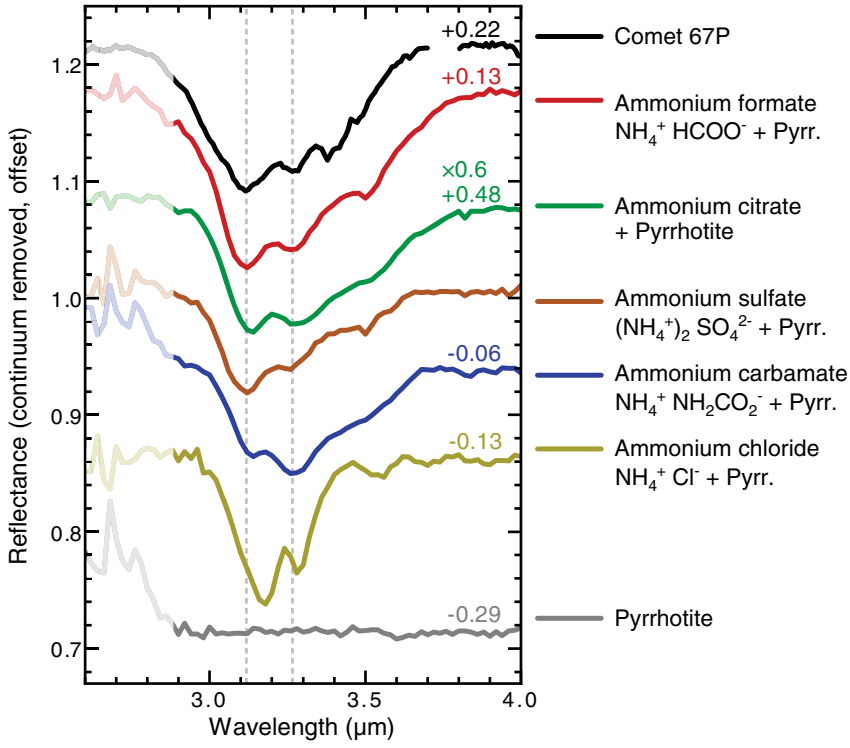


Fig. 2. Reflectance spectra of several ammonium salts. Continuum-removed reflectance spectra of sublimate residues made of ammonium salts (colored lines) mixed with pyrrhotite grains (gray line) measured in high vacuum at 170 to 200 K, compared with the observed average spectrum of comet 67P (black line). Ammonium formate, citrate, and sulfate are the closest matches to the absorption features in comet 67P. We attribute the other absorption features in the comet spectrum at 3.35 to 3.6 μm to C–H stretching modes of organic compounds (*10*). Between 2.6 and 2.8 μm, the laboratory spectra are affected by measurement artifacts because of the presence of water vapor in the optical path. The mass fractions of the salts mixed with pyrrhotite are ≲9 wt % for ammonium sulfate and chloride, ≲17 wt % for ammonium formate and carbamate, and ≲23 wt % for ammonium citrate (the latter has been scaled by a factor of 0.6 for display). These spectra are shown in fig. S6 with uncertainties and before continuum removal.

3.2-μm feature. The counter-ion (in this case, the anion) is not fully constrained. The identification of HCOOH by the Rosetta Orbiter Spectrometer for Ion and Neutral Analysis (ROSINA) mass spectrometer (*16*) and the spectrum of ammonium formate (NH_4^+ $HCOO^-$) in Fig. 2 make it our favored candidate.

Origins of ammonium salts

There are several potential pathways for the synthesis of ammonium salts present at the surface of comet 67P. Ammonia (NH_3) has a high proton affinity, allowing it to transform easily into ammonium (NH_4^+), either in the gas (*17*) or in the solid phase. Ammonium and potential counter-ions (such as $HCOO^-$, CN^-, and OCN^-) may be produced by acid-base reactions of ammonia (NH_3) with the corresponding acids (such as HCOOH, HCN, and HNCO) or by nucleophilic addition of NH_3 with CO_2 or H_2CO, even at cryogenic temperatures in the solid phase (*18*, *19*). These reactions have low activation energies and do not require an external source of photons, electrons, or cosmic rays (*18*). Some ions (such as OCN^- and $HCOO^-$) can be produced at 10 to 14 K, but most of the ions we considered (such as NH_2COO^- and CN^-) are produced at higher temperatures (*19*–*21*). Astronomical observations of interstellar ices have likely identified OCN^- (*22*–*25*) and possibly detected NH_4^+ (*25*–*28*). Ammonium salts can be formed upon sublimation of water ice containing NH_4^+ and counter ions (*19*, *29*). It is possible that the NH_4^+ detected on comet 67P could be inherited from interstellar ices. In that case, the ammonium salts would be produced during further thermal processing of the ices, either in the protoplanetary disk (*25*) or during the sublimation of the ices in the cometary nucleus, through a process similar to the one simulated in our laboratory experiments. The production of ammonium salts by means of a gas phase reaction under astrophysical conditions has not been reported in the literature. Solid-state reactions appear more likely because proton transfer or nucleophilic addition are highly facilitated by a dust surface and a solvent such as ice (*30*, *31*).

Comparison with other small bodies

The 3.2-μm absorption feature of comet 67P shares similarities with the 3-μm features observed on several asteroids, including the position and width of the band from 2.9 to 3.6 μm and the reflectance minimum at 3.1 to 3.2 μm (Fig. 3). However, the bands observed on most of these asteroids are distinct from the one of comet 67P, having a different shape and no secondary minimum at 3.3 μm (Fig. 3). Nevertheless, these spectra are compatible with the presence of ammonium salts, if the spectral differences are due to the environmental conditions at the surface of these small bodies (fig. S7). The spectra of asteroids 24 Themis and 52 Europa are representative of objects found in the asteroid Main Belt (*1*, *32*) and in orbit around Jupiter, such as the irregular moon Himalia (*33*). The dwarf planet 1 Ceres has ammonium-bearing minerals on its surface, with absorption features at 2.72 and 3.06 μm (Fig. 3), mostly in the form of phyllosilicates but with smaller amounts of salts (*34*). Our identification of ammoniated salts on a comet supports the hypothesis that materials on Ceres may have originated from the outer Solar System (*34*).

Volatility of ammonium salts

There is weak evidence for ammonium salts in meteorites, micrometeorites, and interplanetary dust particles (IDPs) (*35*, *36*). Because these salts are more volatile than most refractory material, they might not be preserved during atmospheric entry of small particles and/or during long periods of time under terrestrial environmental conditions (*35*). Ammonium salts contained in grains ejected from cometary nuclei may react and/or sublimate when heated by the Sun and act as distributed sources of gases in comae, the envelopes of gas and dust around cometary nuclei (*37*). This could explain observed increases of NH_3 and HCN when some comets reach short heliocentric distances (<1 au from the Sun), such as comet C/2012 S1 (ISON), which experienced multiple outbursts as it disrupted inside ~0.8 au (*38*, *39*). On comet 67P, the decomposition of ammonium formate (NH_4^+ $HCOO^-$) could produce formamide (NH_2CHO), which has been detected by the ROSINA instrument

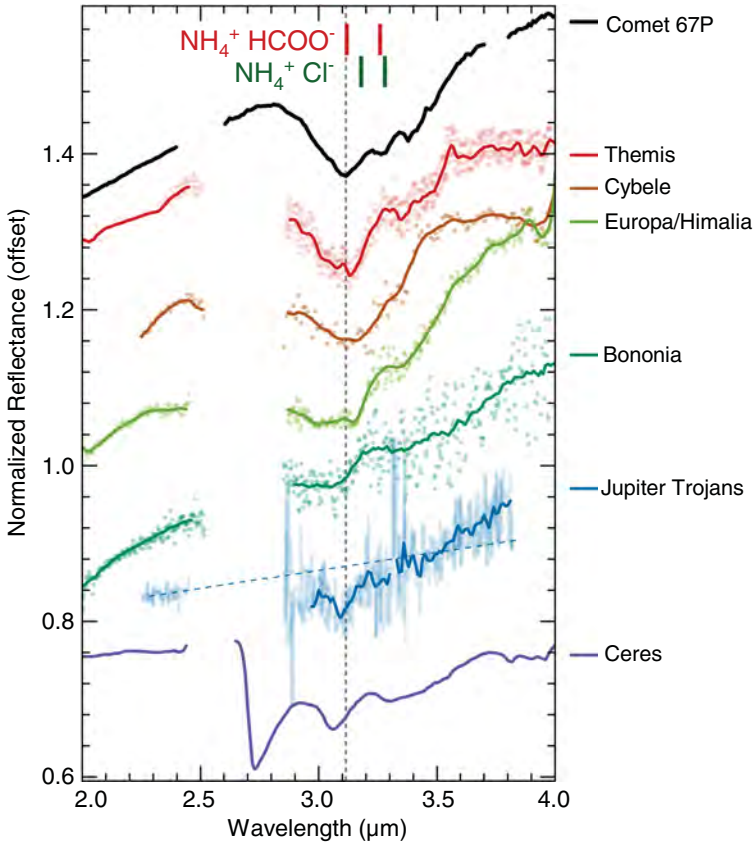


Fig. 3. The spectrum of comet 67P compared with other Solar System bodies. Reflectance spectra normalized at 2.9 μm are of comet 67P (offset by +0.45); the Main Belt asteroids 24 Themis (offset by +0.31) (*63*), 65 Cybele (offset by +0.19) (*64*), 52 Europa (*32*), and 361 Bononia (offset by +0.07) (*32*); the average spectrum of six Jupiter Trojan asteroids [(*65*), their "less red" group] (divided by 3 and offset by +0.49); and the average spectrum of 1 Ceres (scaled by a factor of 0.5, and offset by +0.19) (*34*). The spectrum of Jupiter's irregular moon Himalia is almost indistinguishable from 52 Europa (*33*). For each spectrum, the dots are the observational data (plotted directly for Europa and Bononia, digitized from the literature for the other objects), and the solid lines are running average spectra. The blue dashed line shows the averaged extrapolation of the six Jupiter Trojan's *K*-band spectra (*66*). The gray dashed line shows the position of the band at 3.11 μm on comet 67P spectrum. The red and green vertical marks indicate the positions of the maxima of absorption of ammonium formate and ammonium chloride respectively, shown in Fig. 2. Absorption features around 3.1 to 3.2 μm on some of these bodies are similar to the ammonium salt features on comet 67P. Ceres exhibits different features, which are due to ammoniated phyllosilicates (*34*).

and is compatible with the mass spectrum measured by the Cometary Sampling and Composition (COSAC) experiment on Rosetta's Philae lander (*40*–*42*). The volatility of an ammonium salt strongly depends on the anion; for example, at a pressure of 1 atm, ammonium formate decays at 389 K, ammonium sulfate at 553 K, and ammonium chloride at 611 K (*43*). Under simulated astrophysical conditions (10^{-8} mbar), the sublimation temperatures of ammonium salts are 160 to 180 K for ammonium cyanide (NH_4^+ CN^-) (*44*), 200 to 230 K for ammonium formate (NH_4^+ $HCOO^-$) (*30*, *45*), and 230 to 260 K for ammonium carbamate (NH_4^+ NH_2COO^-) (*46*). Because of these differences of volatility, the composition of ammonium salts observed on comet or asteroid surfaces, and the gases produced by their decomposition, may change with the heliocentric distance.

Nitrogen budget of comet 67P

Rosetta's Cometary Secondary Ion Mass Analyzer (COSIMA) collected coma dust grains 10 to 30 km from comet 67P's nucleus and measured their composition (*47*). Ammoniated salts were not detected, possibly because any semivolatile compounds present in the dust grains would have sublimated during the multiple-day-long pre-analysis storage of the particles at 283 K (*47*). If ammonium salts had been lost from the dust grains analyzed by COSIMA, their measured nitrogen-to-carbon ratio (N/C) would be a lower limit, missing the contribution of the semivolatile nitrogen-bearing salts. COSIMA measured an average N/C of 0.035 ± 0.011, which is similar to the ratio found in the insoluble organic matter extracted from carbonaceous chondrite meteorites and in most micrometeorites and interplanetary dust particles (*47*) but lower than the solar N/C value of 0.29 ± 0.12 (*48*). Similar depletions in nitrogen compared with the Sun have been found in the refractory dust and gas phases of other comets (*49*–*51*).

We propose that ammonium salts may constitute a substantial nitrogen reservoir in comet 67P and possibly other comets and small bodies. The 3.2-μm band observed in the spectrum of comet 67P is 5 to 20% less deep than the band of ammonium formate in our sublimate residues (fig. S6) (*12*). Assuming that the physical parameters that control the light scattering (such as mixing modes and grain sizes) of the sublimate residues are similar to that of the cometary surface, we derived an upper limit of the volumetric abundance of salts in the dark surface material of the comet of ~40 vol %. The dark surface material is a mixture of ~45 wt % organic (~1 g/cm^3) and ~55 wt % mineral (~3.4 g/cm^3) components, estimated from COSIMA measurements (*52*). Taking into account this composition, we derived an upper limit of the mass fraction of ammonium salts mixed with the dust of ~40 wt %, but we cannot determine the surface abundance of ammonium salts on the comet exactly (*12*). If the mass fraction of ammonium formate (NH_4^+ $HCOO^-$) is 5 wt % in the cometary dust, the total atomic nitrogen in the comet is distributed as ~47% N in ammonium salts, ~52% N in refractory organic matter, and ~1% N in volatiles (Fig. 4); the whole comet would then have a N/C ratio of about 0.06 (Fig. 5). If there is a mixture of

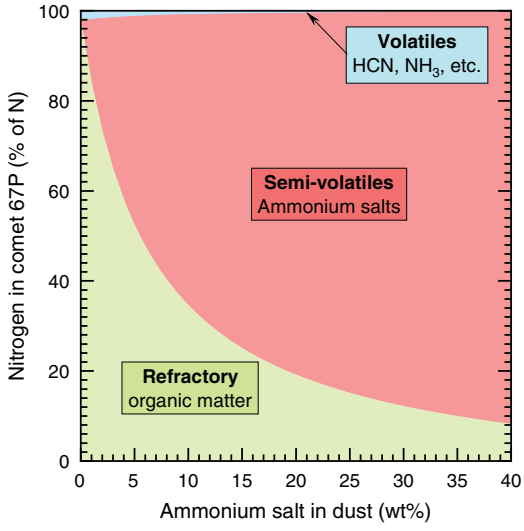


Fig. 4. Reservoirs of nitrogen in comet 67P. Nitrogen in comet 67P is present in the refractory organic matter, in semivolatile ammonium salts, and in volatile molecules. This diagram shows how all the nitrogen atoms in comet 67P are distributed among these three reservoirs, depending on the assumed mass fraction of ammonium salt in the dust (composed of minerals, refractory organic matter, and salts), calculated according to observations from several Rosetta instruments (*12*). If the cometary dust contains more than few percent of ammonium salts, then they form a substantial reservoir of nitrogen in comet 67P.

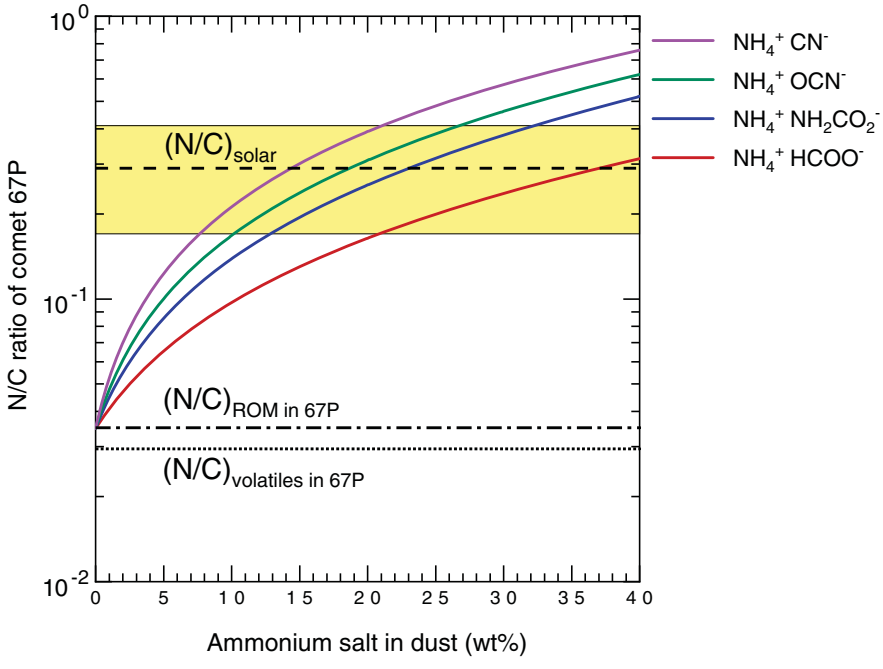


Fig. 5. Nitrogen-to-carbon ratio in comet 67P. The N/C ratio in comet 67P compared with the solar value (dashed line and yellow area indicating the ±1σ uncertainty) (*48*). The colored lines show the contributions of nitrogen in plausible ammonium salts as a function of the mixing with dust (*12*), which are added to the nitrogen in the refractory organic matter (ROM) (*47*) and in volatile species (dashed-dotted and dotted lines, respectively) (*16*). Depending on the counter-ion, the presence of 10 to 30 wt % of ammonium salts in the dust would raise the N/C ratio of comet 67P to be consistent with the solar value.

several ammonium salts, then NH_4^+ would have a range of counter-ions, some of them N-bearing, which would raise the N/C ratio. Shown in Fig. 5 is how the inferred N/C ratio of the comet increases with the assumed concentration of ammonium salts in the dust and depends on the nature of the counter-ions of NH_4^+.

Implications

The identification of ammonium salts on comet 67P shows that this comet, and possibly others, could have a N/C ratio higher than previously thought. If ammonium salts are a substantial repository of nitrogen, assessment of their $^{14}N/^{15}N$ isotopic ratio and comparing it with the proto-solar ratio could inform models of the incorporation and evolution of nitrogen in the early Solar System (*53*). If ammonium salts were also present in sufficient abundance in planetesimals during the early Solar System, they would have provided a solid form of nitrogen closer to the Sun than N_2 and NH_3 ices and therefore available for planetary accretion (*54*). Abundant ammonium salts would have lowered the melting point of water ice in the subsurface of icy bodies (*55*). When mixed in liquid water, ammonium salts are known to participate in potentially prebiotic reactions, such as the formation of pyrimidine and purine nucleobases (*56*), the production of amino acids (*57*), the phosphorylation of nucleosides (*58*), or the formation of sugar molecules (*59*).

REFERENCES AND NOTES

1. A. S. Rivkin *et al*., Astronomical observations of volatiles on asteroids, in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. Arizona Press, 2015), pp. 65–87.
2. A. Coradini *et al*., Virtis: An imaging spectrometer for the Rosetta mission. *Space Sci. Rev.* **128**, 529–559 (2007). doi: 10.1007/s11214-006-9127-5
3. F. Capaccioni *et al*., Cometary science. The organic-rich surface of comet 67P/Churyumov-Gerasimenko as seen by VIRTIS/Rosetta. *Science* **347**, aaa0628 (2015). doi: 10.1126/science.aaa0628; pmid: 25613895
4. G. Filacchione *et al*., The global surface composition of 67P/CG nucleus by Rosetta/VIRTIS. (I) Prelanding mission phase. *Icarus* **274**, 334–349 (2016). doi: 10.1016/j.icarus.2016.02.055
5. M. Ciarniello *et al*., Photometric properties of comet 67P/Churyumov-Gerasimenko from VIRTIS-M onboard Rosetta. *Astron. Astrophys.* **583**, A31 (2015). doi: 10.1051/0004-6361/201526307
6. M. Ciarniello *et al*., The global surface composition of 67P/Churyumov-Gerasimenko nucleus by Rosetta/VIRTIS. II) Diurnal and seasonal variability. *Mon. Not. R. Astron. Soc.* **462**, S443–S458 (2016). doi: 10.1093/mnras/stw3177
7. M. C. De Sanctis *et al*., The diurnal cycle of water ice on comet 67P/Churyumov-Gerasimenko. *Nature* **525**, 500–503 (2015). doi: 10.1038/nature14869; pmid: 26399830
8. G. Filacchione *et al*., Exposed water ice on the nucleus of comet 67P/Churyumov-Gerasimenko. *Nature* **529**, 368–372 (2016). doi: 10.1038/nature16190; pmid: 26760209
9. G. Filacchione *et al*., Comet 67P/CG nucleus composition and comparison to other comets. *Space Sci. Rev.* **215**, 19 (2019). doi: 10.1007/s11214-019-0580-3
10. A. Raponi *et al*., Infrared detection of aliphatic organics on a cometary nucleus. *Nature Astronomy*. (2020). doi: 10.1038/s41550-019-0992-8
11. E. Quirico *et al*., Refractory and semi-volatile organics at the surface of comet 67P/Churyumov-Gerasimenko: Insights from the VIRTIS/Rosetta imaging spectrometer. *Icarus* **272**, 32–47 (2016). doi: 10.1016/j.icarus.2016.02.028
12. Materials and methods are available as supplementary materials.
13. A.-C. Levasseur-Regourd *et al*., Cometary dust. *Space Sci. Rev.* **214**, 64 (2018). doi: 10.1007/s11214-018-0496-3
14. B. Rousseau *et al*., Laboratory simulations of the Vis-NIR spectra of comet 67P using sub-µm sized cosmochemical analogues. *Icarus* **306**, 306–318 (2018). doi: 10.1016/j.icarus.2017.10.015
15. O. Poch *et al*., Sublimation of water ice mixed with silicates and tholins: Evolution of surface texture and reflectance spectra, with implications for comets. *Icarus* **267**, 154–173 (2016). doi: 10.1016/j.icarus.2015.12.017
16. L. Le Roy *et al*., Inventory of the volatiles on comet 67P/Churyumov-Gerasimenko from Rosetta/ROSINA. *Astron. Astrophys.* **583**, A1 (2015). doi: 10.1051/0004-6361/201526450
17. E. Carrasco, I. Tanarro, V. J. Herrero, J. Cernicharo, Proton transfer chains in cold plasmas of H_2 with small amounts of N_2. The prevalence of NH4+. *Phys. Chem. Chem. Phys.* **15**, 1699–1706 (2013). doi: 10.1039/C2CP43438E; pmid: 23247609
18. P. Theulé *et al*., Thermal reactions in interstellar ice: A step towards molecular complexity in the interstellar medium. *Adv. Space Res.* **52**, 1567–1579 (2013). doi: 10.1016/j.asr.2013.06.034
19. O. Gálvez, B. Maté, V. J. Herrero, R. Escribano, Ammonium and formate ions in interstellar ice analogs. *Astrophys. J.* **724**, 539–545 (2010). doi: 10.1088/0004-637X/724/1/539
20. W. A. Schutte *et al*., Weak ice absorption features at 7.24 and 7.41 µm in the spectrum of the obscured young stellar object W 33A. *Astron. Astrophys.* **343**, 966–976 (1999).
21. S. Raunier, T. Chiavassa, F. Marinelli, A. Allouche, J. P. Aycard, Reactivity of HNCO with NH3 at low temperature monitored by FTIR spectroscopy: Formation of $NH_4^+OCN^-$. *Chem. Phys. Lett.* **368**, 594–600 (2003). doi: 10.1016/S0009-2614(02)01919-X
22. K. Demyk *et al*., Laboratory identification of the 4.62µm solid state absorption band in the ISO-SWS spectrum of RAFGL 7009S. *Astron. Astrophys.* **339**, 553–560 (1998).
23. F. A. van Broekhuizen, K. M. Pontoppidan, H. J. Fraser, E. F. van Dishoeck, A 3–5 µm VLT spectroscopic survey of embedded young low mass stars II: Solid OCN-. *Astron. Astrophys.* **441**, 249–260 (2005). doi: 10.1051/0004-6361:20041711
24. K. M. Pontoppidan *et al*., A µm VLT spectroscopic survey of embedded young low mass stars I–Structure of the CO ice. *Astron. Astrophys.* **408**, 981–1007 (2003). doi: 10.1051/0004-6361:20031030
25. A. C. A. Boogert, P. A. Gerakines, D. C. B. Whittet, Observations of the icy universe. *Annu. Rev. Astron. Astrophys.* **53**, 541–581 (2015). doi: 10.1146/annurev-astro-082214-122348
26. W. Schutte, R. Khanna, Origin of the 6.85 µm band near young stellar objects: The ammonium ion (NH4+) revisited. *Astron. Astrophys.* **398**, 1049–1062 (2003). doi: 10.1051/0004-6361:20021705
27. J. V. Keane, A. G. G. M. Tielens, A. C. A. Boogert, W. A. Schutte, D. C. B. Whittet, Ice absorption features in the 5-8 µm region toward embedded protostars. *Astron. Astrophys.* **376**, 254–270 (2001). doi: 10.1051/0004-6361:20010936
28. A. C. A. Boogert *et al*., The c2d Spitzer spectroscopic survey of ices around low-mass young stellar objects. I. H_2O and the 5–8 µm bands. *Astrophys. J.* **678**, 985–1004 (2008). doi: 10.1086/533425
29. B. Maté *et al*., Water-ammonium ices and the elusive 6.85 µm band. *Astrophys. J.* **703**, L178–L182 (2009). doi: 10.1088/0004-637X/703/2/L178
30. J. B. Bergner, K. I. Öberg, M. Rajappan, E. C. Fayolle, Kinetics and mechanisms of the acid-base reaction between NH_3 and HCOOH in interstellar ice analogs. *Astrophys. J.* **829**, 85 (2016). doi: 10.3847/0004-637X/829/2/85
31. A. Potapov, P. Theulé, C. Jäger, T. Henning, Evidence of surface catalytic effect on cosmic dust grain analogs: The ammonia and carbon dioxide surface reaction. *Astrophys. J.* **878**, L20 (2019). doi: 10.3847/2041-8213/ab2538
32. D. Takir, J. P. Emery, Outer Main Belt asteroids: Identification and distribution of four 3-µm spectral groups. *Icarus* **219**, 641–654 (2012). doi: 10.1016/j.icarus.2012.02.022
33. M. E. Brown, A. R. Rhoden, The 3 µm spectrum of Jupiter's irregular satellite himalia. *Astrophys. J.* **793**, L44 (2014). doi: 10.1088/2041-8205/793/2/L44
34. M. C. De Sanctis *et al*., Ceres's global and localized mineralogical composition determined by Dawn's Visible and Infrared Spectrometer (VIR). *Meteorit. Planet. Sci.* **53**, 1844–1865 (2018). doi: 10.1111/maps.13104
35. M. Gounelle, M. E. Zolensky, The Orgueil meteorite: 150 years of history. *Meteorit. Planet. Sci.* **49**, 1769–1794 (2014). doi: 10.1111/maps.12351
36. S. Pizzarello, X. Feng, S. Epstein, J. R. Cronin, Isotopic analyses of nitrogenous compounds from the Murchison meteorite: Ammonia, amines, amino acids, and polar hydrocarbons. *Geochim. Cosmochim. Acta* **58**, 5579–5587 (1994). doi: 10.1016/0016-7037(94)90251-8; pmid: 11539151
37. M. J. Mumma *et al*., paper presented at the 50th American Astronomical Society DPS meeting, Knoxville, TN, 12-26 October 2018; http://adsabs.harvard.edu/abs/2018DPS....5020902M
38. M. A. DiSanti *et al*., En route to destruction: The evolution in composition of ices in comet D/2012 S1 (ISON) between 1.2 and 0.34 AU from the sun as revealed at infrared wavelengths. *Astrophys. J.* **820**, 34 (2016). doi: 10.3847/0004-637X/820/1/34
39. N. Dello Russo *et al*., The compositional evolution of C/2012 S1 (ISON) from ground-based high-resolution infrared spectroscopy as part of a worldwide observing campaign. *Icarus* **266**, 152–172 (2016). doi: 10.1016/j.icarus.2015.11.030
40. K. Altwegg *et al*., Organics in comet 67P–A first comparative analysis of mass spectra from ROSINA–DFMS, COSAC and Ptolemy. *Mon. Not. R. Astron. Soc.* **469** (Suppl_2), S130–S141 (2017). doi: 10.1093/mnras/stx1415
41. F. Goesmann *et al*., Organic compounds on comet 67P/Churyumov-Gerasimenko revealed by COSAC mass spectrometry. *Science* **349**, aab0689 (2015). doi: 10.1126/science.aab0689; pmid: 26228156
42. N. Hänni *et al*., Ammonium salts as a source of small molecules observed with high-resolution electron-impact ionization mass spectrometry. *J. Phys. Chem. A* **123**, 5805–5814 (2019). doi: 10.1021/acs.jpca.9b03534; pmid: 31257892
43. D. R. Lide, Physical constants of inorganic compounds, in *CRC Handbook of Chemistry and Physics* (CRC Press, ed. 90, 2009), pp. 4-46–4-48.
44. G. Danger *et al*., Experimental investigation of aminoacetonitrile formation through the Strecker synthesis in astrophysical-like conditions: Reactivity of methanimine (CH_2NH), ammonia (NH_3), and hydrogen cyanide (HCN). *Astron. Astrophys.* **535**, A47 (2011). doi: 10.1051/0004-6361/201117602
45. V. Vinogradoff, F. Duvernay, G. Danger, P. Theulé, T. Chiavassa, New insight into the formation of hexamethylenetetramine (HMT) in interstellar and cometary ice analogs. *Astron. Astrophys.* **530**, A128 (2011). doi: 10.1051/0004-6361/201116688
46. J. B. Bossa, P. Theulé, F. Duvernay, F. Borget, T. Chiavassa, Carbamic acid and carbamate formation in NH_3:CO_2 ices–UV irradiation versus thermal processes. *Astron. Astrophys.* **492**, 719–724 (2008). doi: 10.1051/0004-6361:200810536
47. N. Fray *et al*., Nitrogen-to-carbon atomic ratio measured by COSIMA in the particles of comet 67P/Churyumov–Gerasimenko. *Mon. Not. R. Astron. Soc.* **469** (Suppl_2), S506–S516 (2017). doi: 10.1093/mnras/stx2002
48. K. Lodders, in *Principles and Perspectives in Cosmochemistry* (Springer, 2010), *Astrophysics and Space Science Proceedings*, pp. 379–417.
49. N. Dello Russo, H. Kawakita, R. J. Vervack Jr., H. A. Weaver, Emerging trends and a comet taxonomy based on the volatile chemistry measured in thirty comets with high-resolution infrared spectroscopy between 1997 and 2013. *Icarus* **278**, 301–332 (2016). doi: 10.1016/j.icarus.2016.05.039
50. M. Rubin *et al*., Molecular nitrogen in comet 67P/Churyumov-Gerasimenko indicates a low formation temperature. *Science* **348**, 232–235 (2015). doi: 10.1126/science.aaa6100; pmid: 25791084
51. S. Wyckoff, S. C. Tegler, L. Engel, Nitrogen abundance in Comet Halley. *Astrophys. J.* **367**, 641 (1991). doi: 10.1086/169659
52. A. Bardyn *et al*., Carbon-rich dust in comet 67P/Churyumov-Gerasimenko measured by COSIMA/Rosetta. *Mon. Not. R. Astron. Soc.* **469** (Suppl_2), S712–S722 (2017). doi: 10.1093/mnras/stx2640
53. E. Füri, B. Marty, Nitrogen isotope variations in the Solar System. *Nat. Geosci.* **8**, 515–522 (2015). doi: 10.1038/ngeo2451
54. K. Lodders, Jupiter formed with more tar than ice. *Astrophys. J.* **611**, 587–597 (2004). doi: 10.1086/421970
55. J. S. Kargel, Ammonia-water volcanism on icy satellites: Phase relations at 1 atmosphere. *Icarus* **100**, 556–574 (1992). doi: 10.1016/0019-1035(92)90118-Q

56. M. P. Callahan *et al*., Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13995–13998 (2011). doi: 10.1073/pnas.1106493108; pmid: 21836052
57. N. R. Lerner, G. W. Cooper, Iminodicarboxylic acids in the Murchison meteorite: Evidence of Strecker reactions. *Geochim. Cosmochim. Acta* **69**, 2901–2906 (2005). doi: 10.1016/j.gca.2004.12.024
58. B. Burcar *et al*., Darwin's warm little pond: A one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angew. Chem. Int. Ed.* **55**, 13249–13253 (2016). doi: 10.1002/anie.201606239; pmid: 27532228
59. P. Dziedzic, A. Bartoszewicz, A. Córdova, Inorganic ammonium salts as catalysts for direct aldol reactions in the presence of water. *Tetrahedron Lett.* **50**, 7242–7245 (2009). doi: 10.1016/j.tetlet.2009.10.014
60. B. Hapke, *Theory of Reflectance and Emittance Spectroscopy* (Cambridge Univ. Press, ed. 2, 2012).
61. B. Schmitt, Near and Mid-IR optical constants of crystalline H2O ice Ih at 140–145K. SSHADE/GhoSST (OSUG Data Center). Dataset/Spectral Data. (2004); doi: 10.26302/SSHADE/EXPERIMENT_BS_20200103_001
62. D. Takir *et al*., Nature and degree of aqueous alteration in CM and CI carbonaceous chondrites. *Meteorit. Planet. Sci.* **48**, 1618–1637 (2013). doi: 10.1111/maps.12171
63. A. S. Rivkin, J. P. Emery, Detection of ice and organics on an asteroidal surface. *Nature* **464**, 1322–1323 (2010). doi: 10.1038/nature09028; pmid: 20428165
64. J. Licandro *et al*., (65) Cybele: Detection of small silicate grains, water-ice, and organics. *Astron. Astrophys.* **525**, A34 (2011). doi: 10.1051/0004-6361/201015339
65. M. E. Brown, The 3-4 μm spectra of Jupiter Trojan asteroids. *Astron. J.* **152**, 159 (2016). doi: 10.3847/0004-6256/152/6/159
66. J. P. Emery, D. M. Burr, D. P. Cruikshank, Near-infrared spectroscopy of Trojan Asteroids: Evidence for two compositional groups. *Astron. J.* **141**, 25 (2011). doi: 10.1088/0004-6256/141/1/25
67. O. Poch, I. Istiqomah, Vis-NIR bidirectional reflection spectra of several ammonium salts mixed with pyrrhotite grains in sublimate residues at 173 K. SSHADE/GhoSST (OSUG Data Center) Dataset/Spectral Data (2018); doi: 10.26302/SSHADE/EXPERIMENT_OP_20191119_001
68. O. Poch, Vis-NIR bidirectional reflection spectra of several ammonium salts mixed with graphite powder at 296 K. SSHADE/GhoSST (OSUG Data Center) Dataset/Spectral Data (2018); doi: 10.26302/SSHADE/EXPERIMENT_OP_20200212_001

ACKNOWLEDGMENTS

O.P., I.I., and E.Q. thank F. Charlot of the Consortium des Moyens Technologiques Communs (CMTC) at the Institut National Polytechnique de Grenoble (INP) for the Scanning Electron Microscopy images of the samples. We acknowledge M. Faure for work on the VIRTIS data and the collaboration of the International Space Science Institute (ISSI) international team number 397 "Comet 67P/Churyumov-Gerasimenko Surface Composition as a Playground for Radiative Transfer Modeling and Laboratory Measurements." O.P. thanks C. Pilorget for insightful comments. **Funding:** O.P. acknowledges a postdoctoral fellowship from the Centre National d'Etudes Spatiales (CNES). I.I. acknowledges a thesis grant from the Lembaga Pengelola Dana Pendidikan (LPDP) Indonesian scholarship. L.M. acknowledges the Deutsche Forschungsgemeinschaft (DFG) grant MO 3007/1-1. D.K. acknowledges DFG-grant KA 3757/2-1. The work of O.P., I.I., E.Q., P.B., B.S., and L.B. was supported by the CNES and the French Agence Nationale de la Recherche (program Classy, ANR-17-CE31-0004). P.B. acknowledges funding from the European Research Council under the SOLARYS grant (ERC-CoG2017-771691). P.H.-B. acknowledges the LabEx Observatoire des Sciences de l'Univers de Grenoble OSUG @ 2020 for funding. S.P. is supported by Université Grenoble Alpes (UGA), Initiatives de Recherche Stratégiques (IRS), and Initiatives d'Excellence UGA (IDEX UGA). The development of the CarboN-IR environmental chamber was supported by the French National Program of Planetology (PNP), the development of the reflectance goniometers was supported by the University of Grenoble Alpes (Initiative de Recherche Stratégique). Université Grenoble Alpes (UGA) and CNES supported the instrumental facilities and activities at IPAG. The following institutions and agencies supported the Rosetta mission: the Italian Space Agency (ASI, Italy), Centre National d'Etudes Spatiales (CNES, France), Deutsches Zentrum für Luft- und Raumfahrt (DLR, Germany), the National Aeronautic and Space Administration (NASA, United States) Rosetta Program and the Science and Technology Facilities Council (United Kingdom). VIRTIS was built by a consortium including Italy, France, and Germany, under the scientific responsibility of the Istituto di Astrofisica e Planetologia Spaziali of INAF, Italy, which also guided scientific operations. The development of the VIRTIS instrument was funded and managed by ASI, with contributions from Observatoire de Meudon, financed by CNES, and from DLR. A.R., M.C., G.F., F.C., M.C.D.S., A.L., E.P., and F.T. acknowledge financial support from the National Institute for Astrophysics (INAF, Italy) and the Italian Space Agency (ASI, Italy) through contract I/024/12/2. Computational resources were provided by INAF-IAPS through the DataWell project. The work of A.P. and N.T. was carried out within the framework of the National Centre of Competence in Research (NCCR) PlanetS supported by the Swiss National Science Foundation. **Author contributions:** O.P. and I.I. carried out the laboratory experiments. O.P. wrote the manuscript with assistance from P.B., E.Q., and B.S., who also contributed to interpreting the results. E.Q., P.H.-B., A.F., and P.B. calculated the nitrogen distribution and N/C ratio. O.P., E.Q., and A.F. produced the figures. A.R., M.C., and G.F. provided the calibrated average reflectance spectrum of comet 67P and wrote parts of the supplementary materials. O.B., S.P., B.R., and L.F. contributed to the experimental work. N.T. and A.P. contributed to the early development of spectroscopic studies of sublimate residues. All co-authors contributed to the preparation of the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** The laboratory reflectance spectra are available in the Grenoble Astrophysics and Planetology Solid Spectroscopy and Thermodynamics (GhoSST) database (*67*, *68*). The average reflectance spectrum of comet 67P measured by VIRTIS is available from (*10*), their 'source data Fig. 1.' The reference spectra of Mars and Lutetia are available from the Planetary Science Archive https://archives.esac.esa.int/psa/#!Table%20View by selecting Mars or 21 Lutetia as the target, VIRTIS (Rosetta) as the instrument, clicking search, then filtering by observation IDs I1_00130974021 for Mars or I1_00237396952 for Lutetia.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/eaaw7462/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 and S2
References (*69*–*89*)

21 January 2019; resubmitted 11 October 2019
Accepted 14 February 2020
10.1126/science.aaw7462

RESEARCH ARTICLE

SPECTROSCOPY

Quantum-nondemolition state detection and spectroscopy of single trapped molecules

Mudit Sinhal*, Ziv Meir*, Kaveh Najafian, Gregor Hegi†, Stefan Willitsch‡

Trapped atoms and ions, which are among the best-controlled quantum systems, find widespread applications in quantum science. For molecules, a similar degree of control is currently lacking owing to their complex energy-level structure. Quantum-logic protocols in which atomic ions serve as probes for molecular ions are a promising route for achieving this level of control, especially for homonuclear species that decouple from blackbody radiation. Here, a quantum-nondemolition protocol on single trapped N_2^+ molecules is demonstrated. The spin-rovibronic state of the molecule is detected with >99% fidelity, and a spectroscopic transition is measured without destroying the quantum state. This method lays the foundations for new approaches to molecular spectroscopy, state-to-state chemistry, and the implementation of molecular qubits.

The impressive advances achieved in the control of ultracold trapped atoms and ions on the quantum level are now increasingly being transferred to molecular systems. Cold, trapped molecules have been created by, for example, binding ultracold atoms via Feshbach resonances (*1*) and photoassociation (*2*, *3*), molecular-beam slowing (*4*), direct laser cooling (*5*, *6*), and sympathetic cooling (*7*, *8*). The trapping of cold molecules enables experiments with long interaction times and thus paves the way for new applications, such as studies of ultracold chemistry (*9*) and precision spectroscopic measurements, which aim at a precise determination of fundamental physical constants (*10*) and their possible time variation (*11*, *12*), as well as tests of fundamental theories that reach beyond the standard model (*13*, *14*).

The complex energy-level structure and the absence of optical cycling transitions in most molecular systems constitute a major challenge for their state preparation, laser cooling, state detection, and coherent manipulation. Molecular ions which are confined in radiofrequency traps and sympathetically cooled by simultaneously trapped atomic ions (*7*, *8*) have proven a promising route for overcoming these obstacles. Recently, their rotational cooling and state preparation have been achieved (*15*–*18*), precision measurements of quantum electrodynamics and fundamental constants have been performed (*10*, *19*), the first studies of dipole-forbidden spectroscopic transitions in the mid-infrared (mid-IR) spectral domain have been reported (*20*), and state- and energy-controlled collisions with cold atoms have been realized (*21*, *22*). However, to reach, for a single molecule, the same exquisite level of control on the quantum level that can be achieved with trapped atoms (*23*), new methodological developments are required. The most promising route for achieving ultimate quantum control of molecular ions in trap experiments is offered by quantum-logic protocols (*24*), in which a co-trapped atomic ion acts as a probe for the quantum state of a single molecular ion (*25*–*27*).

Here, a quantum-logic–based quantum-nondemolition (QND) (*28*–*30*) detection of the spin-rotational-vibrational state of a single molecular nitrogen ion co-trapped with a single atomic calcium ion is demonstrated. N_2^+ is a homonuclear diatomic molecule with no permanent dipole moment, rendering all rotational-vibrational (rovibrational) transitions dipole-forbidden in its electronic ground state (*20*). Therefore, N_2^+ is an ideal test bed for precision spectroscopic studies (*31*), for tests of fundamental physics (*32*), for the realization of mid-IR frequency standards and clocks (*33*), and for implementation of molecular qubits for quantum information and computation applications (*34*, *35*).

The state-detection protocol implemented in this study relies on coherent motional excitation of the Ca^+-N_2^+ two-ion string (*36*–*38*) using an optical dipole force (ODF), which depends on the molecular state and arises from off-resonant dispersive molecule–light interactions. The excited motion was read out on the Ca^+ ion, thus preserving the state of the N_2^+ ion. State-detection fidelities >99% for the ground rovibrational state of N_2^+ were demonstrated, limited only by the chosen bandwidth of the detection cycle. Because the lifetime of the rovibrational levels in N_2^+ is estimated to be on the order of half a year (*20*), the prospects for reaching, and potentially exceeding, the high readout fidelities that are currently achieved with atomic ions (*23*, *39*) are excellent.

The present scheme has immediate applications for marked improvements of the sensitivity and, therefore, precision of spectroscopic measurements on molecular ions. This was demonstrated here by introducing a type of force spectroscopy (*40*) used to study a rovibronic component of the electronic spectrum of a single N_2^+ molecule with a signal-to-noise ratio greatly exceeding that of previous destructive detection schemes for trapped particles. Transition properties such as the line center and the Einstein A coefficient were determined and validated against the results of previous studies, which used conventional spectroscopic methods (*41*–*46*).

Quantum-nondemolition state detection

Our scheme is illustrated in Fig. 1. A detailed description of our experimental apparatus can be found in (*36*). Our setup consists of a molecular-beam machine coupled to a radiofrequency ion trap (Fig. 1A). The experiment started with loading of a small Coulomb crystal of roughly 10 $^{40}Ca^+$ ions into the trap. The atomic ions were Doppler cooled to millikelvin temperatures by scattering photons on the $(4s)^2S_{1/2} \leftrightarrow (4p)^2P_{1/2} \leftrightarrow (3d)^2D_{3/2}$ closed optical cycling transitions (Fig. 1C). A single $^{14}N_2^+$ molecular ion was then loaded into the trap using state-selective resonance-enhanced multiphoton ionization (REMPI) from a pulsed molecular beam of neutral $^{14}N_2$ molecules (*17*, *47*). This ionization scheme preferentially created N_2^+ ions in the ground electronic, vibrational and spin-rotational state, $|X^2\Sigma_g^+, v = 0, N = 0, J = 1/2\rangle$, henceforth referred to as $|\downarrow\rangle_{N_2}$ (Fig. 1D). Here, v denotes the vibrational, N the rotational, and J the total-angular-momentum quantum numbers of the molecule, excluding nuclear spin. The ionized molecule was sympathetically cooled by the Coulomb crystal of atomic ions (*8*) within a few seconds. Then, by lowering the trap depth, Ca^+ ions were successively ejected from the trap while the molecular ion was retained in the trap, until a Ca^+-N_2^+ two-ion crystal was obtained (Fig. 1A, inset) (*36*).

The experimental sequence to detect the spin-rovibrational state of the N_2^+ ion is illustrated in Fig. 1E. First, the in-phase (*48*) motional mode of the Ca^+-N_2^+ crystal (which corresponds to the center-of-mass mode for two ions of the same mass) was cooled to the ground state of the trap, $|0\rangle$, by resolved-sideband cooling ("GSC" in Fig. 1E) on the atomic ion (*36*). The out-of-phase and the radial motional modes were maintained at Doppler-cooling temperature. At the end of the cooling cycle, the Ca^+ ion was optically pumped into its $|S_{1/2}, m = -1/2\rangle$ state, henceforth

Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland.
*These authors contributed equally to this work.
†Present address: RUAG Schweiz AG, Allmendstrasse 86, 3602 Thun, Switzerland.
‡Corresponding author. Email: stefan.willitsch@unibas.ch

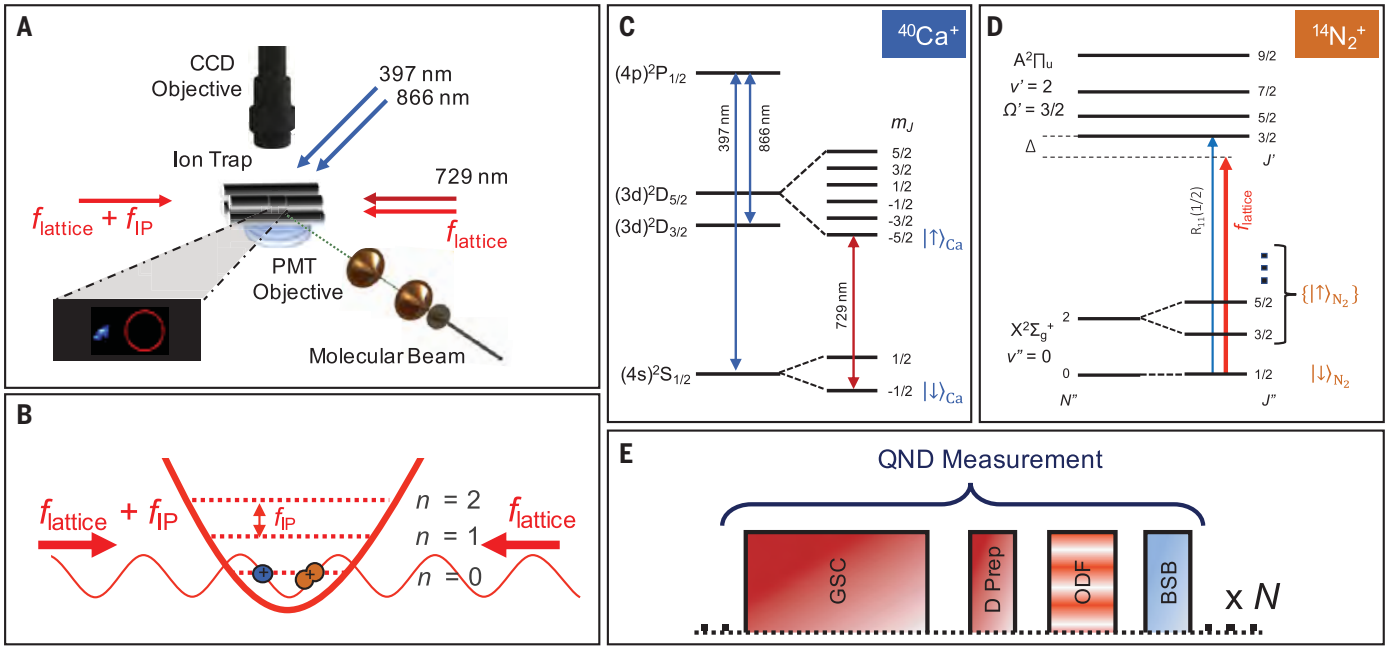


Fig. 1. Experimental scheme. (**A**) Schematic representation of the experimental setup depicting the ion trap and the molecular-beam source. The fluorescence of the atomic ion (inset) was collected by a charge-coupled device (CCD) camera. The red circle marks the position of the nonfluorescing N_2^+ molecular ion in the two-ion crystal. PMT, photomultiplier tube. (**B**) Illustration of a Ca^+-N_2^+ two-ion crystal with its in-phase external motion cooled to the ground state of the ion trap ($n = 0$), which was overlapped with two counter-propagating laser beams forming a running one-dimensional optical lattice. (**C**) Reduced energy-level diagram of $^{40}Ca^+$. The $(4s)^2S_{1/2} \leftrightarrow (4p)^2P_{1/2} \leftrightarrow (3d)^2D_{3/2}$ closed-cycling transitions (blue arrows) were used for Doppler laser cooling and for detection of the states $|\downarrow\rangle_{Ca}$ and $|\uparrow\rangle_{Ca}$. These states were coherently coupled by a narrow-linewidth laser beam (red arrow). (**D**) Reduced energy-level diagram of $^{14}N_2^+$. The lattice laser frequency, $f_{lattice}$ (red arrow), was detuned by Δ from the $R_{11}(1/2)$ transition (blue arrow). The state $|\downarrow\rangle_{N_2}$ was strongly coupled to the lattice, while all other states $\{|\uparrow\rangle_{N_2}\}$ were far detuned and hence did not couple. (**E**) Experimental sequence of a single quantum-nondemolition measurement to be repeated multiple (N) times to increase the fidelity of the determination of the molecular state. See text for details.

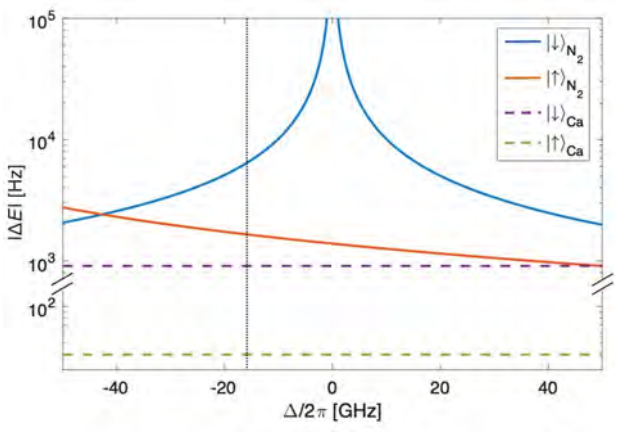


Fig. 2. Ac-Stark shift generated on N_2^+ by the one-dimensional optical lattice. Calculated magnitude of the ac-Stark shift, $|\Delta E|$, experienced by N_2^+, as a function of the laser-frequency detuning, Δ, from the $A^2\Pi_u(v' = 2) \leftarrow X^2\Sigma_g^+ (v'' = 0)$, $R_{11}(1/2)$, spin-rovibronic transition (*52*) at ~787.47 nm (*41*) for a single lattice beam of intensity 2×10^6 W/m^2. The ac-Stark shift experienced by N_2^+ when in the $|\downarrow\rangle_{N_2}$ (not in the $|\downarrow\rangle_{N_2}$) state is given by the blue (red) trace. Ca^+ in the $|\uparrow\rangle_{Ca}(|\downarrow\rangle_{Ca})$ state experienced an ac-Stark shift of 40 Hz (910 Hz) indicated by the dashed green (purple) line. The black dotted line indicates the frequency detuning, $\Delta/2\pi \approx -17$ GHz, of the optical lattice used to generate the optical dipole force, which was used for collecting the data shown in Fig. 3.

referred to as $|\downarrow\rangle_{Ca}$, where m denotes the magnetic quantum number. The Ca^+ ion was then shelved in the metastable $|D_{5/2}, m = -5/2\rangle$ state, henceforth referred to as $|\uparrow\rangle_{Ca}$, using a π-pulse on the narrow $|\uparrow\rangle_{Ca} \leftarrow |\downarrow\rangle_{Ca}$ electric-quadruple transition followed by a D-state purification pulse (*26*) ("D Prep" in Fig. 1E). This pulse allowed us to exclude experiments in which the ion was not successfully shelved to the $|\uparrow\rangle_{Ca}$ state (*49*). Preparing the Ca^+ ion in the $|\uparrow\rangle_{Ca}$ state suppressed the background signal during the state detection and thus enabled an efficient determination of the state of N_2^+ (*36*, *49*) (Fig. 2). For an ideal state preparation, the complete state of the two ions in the trap was given by $|\downarrow\rangle_{N_2}|\uparrow\rangle_{Ca}|0\rangle$.

The detection sequence was continued by applying a state-dependent ODF to excite coherent motion (*36*, *50*), the amplitude of which depends on the rovibrational state of the molecular ion ("ODF" in Fig. 1E). If the molecular ion was in the $|\downarrow\rangle_{N_2}$ state, a coherent motion of amplitude $|\alpha\rangle$ would be excited such that the probability to populate a motional Fock state, $|n\rangle$, is given by $P(n|\alpha) = |\langle n|\alpha\rangle|^2 = e^{-|\alpha|^2}|\alpha|^{2n}n!$ (*51*). If the molecular ion was in any other rotational or vibrational state, henceforth referred to as $\{|\uparrow\rangle_{N_2}\}$, a motional state $\{|\beta\rangle\}$ would be excited, where $\{|\beta|\} \ll |\alpha|$. Here, the curly brackets are a reminder that $\{|\uparrow\rangle_{N_2}\}$ refers to many states, all of which result in vanishingly small amplitudes of the motion, $\{|\beta|\} \ll 1$. Owing to possible experimental imperfections and decoherence—e.g., from coupling of the in-phase mode to the out-of-phase and radial modes, breakdown of the Lamb-Dicke regime because of the finite size of the lattice modulation with respect to the extent of the ion wavefunction at large excitation, imperfect ground-state cooling of the in-phase mode, or possible imperfections in the relative phase stability of the lattice beams—the underlying motional Fock-state distribution, in reality, deviated from the distribution of a perfect coherent state (*49*). However, the knowledge of

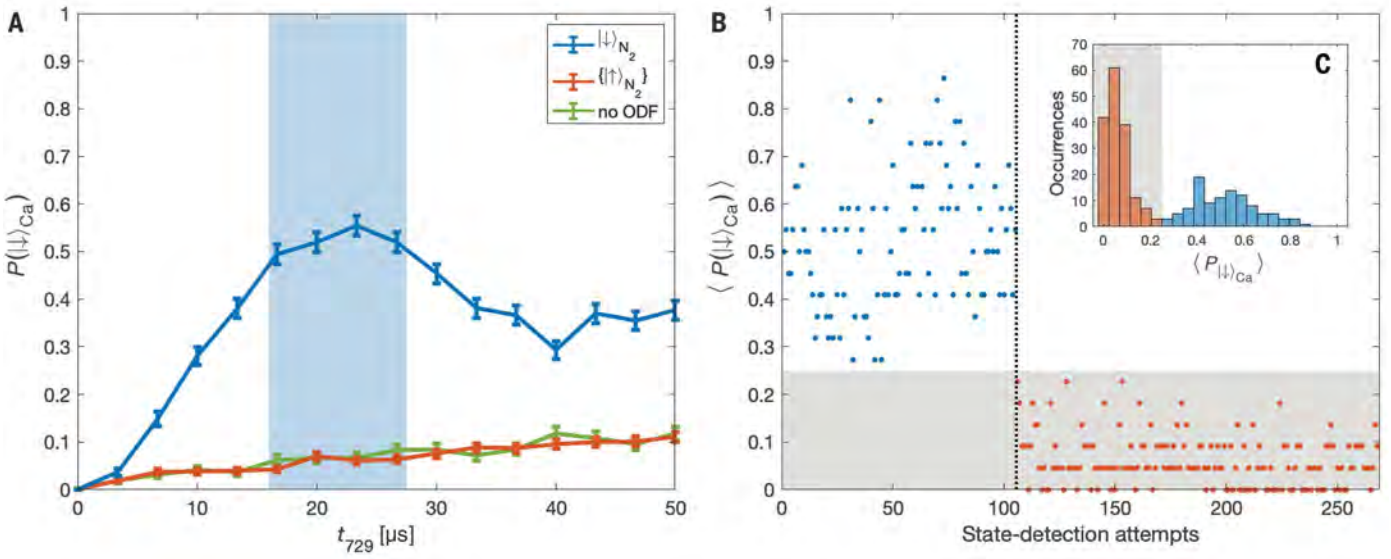


Fig. 3. Quantum-nondemolition state detection of N_2^+. (**A**) Blue-sideband (BSB) Rabi oscillation signal for N_2^+ in the $|\downarrow\rangle_{N_2}$ state (blue) and in one of the $\{|\uparrow\rangle_{N_2}\}$ states (red) as a function of the BSB pulse length t_{729}. Error bars correspond to 1σ binomial errors. The green trace represents a background measurement of the Rabi oscillation signal without ODF beams. The light-blue shaded area indicates the region of BSB pulse lengths for which maximum state-detection contrast was achieved. (**B**) Time trace of state-detection attempts. A single state-detection data point is composed of an average of 22 consecutive BSB-measurement results for pulse lengths indicated by the light-blue shaded area in (A). A threshold of $P(|\downarrow\rangle_{Ca}) = 0.25$ was used to distinguish between N_2^+ in the $|\downarrow\rangle_{N_2}$ or $\{|\uparrow\rangle_{N_2}\}$ states (gray shaded area). The blue (red) dots indicate assignment of $|\downarrow\rangle_{N_2}(\{|\uparrow\rangle_{N_2}\})$ states by the detection scheme. The dotted black line shows the onset of a quantum jump out of the $|\downarrow\rangle_{N_2}$ state to one of the $\{|\uparrow\rangle_{N_2}\}$ states. (**C**) Histogram of state-detection attempts. The gray shaded area indicates separation between $|\downarrow\rangle_{N_2}$ (blue) and $\{|\uparrow\rangle_{N_2}\}$ (red) state-detection assignments.

the exact motional state was not critical for the present molecular-state-detection protocol, as shown below.

The ODF was implemented via a state-dependent ac-Stark shift generated by two counter-propagating laser beams with frequencies $f_{lattice}$, aligned with the crystal axis, which formed a one-dimensional optical lattice (Fig. 1B). By further detuning one of the beams by the frequency of the in-phase motional mode of the two-ion crystal, $f_{IP} \approx 620$ kHz, a running optical lattice was generated, causing a modulation of the amplitude of the ac-Stark shift, which resonantly excited motion of the ion crystal depending on the rotational and vibrational state of the N_2^+ ion.

Figure 2 shows the calculated ac-Stark shift of a single lattice beam as a function of its frequency for N_2^+ in the $|\downarrow\rangle_{N_2}$ state (blue) and the maximum ac-Stark shift experienced by the N_2^+ ion when not in the $|\downarrow\rangle_{N_2}$ state (red) (*49*). The strength of the ac-Stark shift was dependent on the detuning of the lattice laser beam from spectroscopic transitions in the molecule. The peak in the ac-Stark shift of the blue trace corresponds to an on-resonance condition of the $A^2\Pi_u(v' = 2) \leftarrow X^2\Sigma_g^+(v'' = 0)$, $R_{11}(1/2)$, spin-rovibronic transition (*52*) originating from the $|\downarrow\rangle_{N_2}$ state (Fig. 1D), where $''$ ($'$) denotes the lower (upper) level of the transition. By setting the lattice-laser detuning close to this resonance, the N_2^+ ion experienced a much stronger ac-Stark shift leading to a large motional excitation $|\alpha\rangle$ when in the $|\downarrow\rangle_{N_2}$ state, as opposed to the situation where the N_2^+ ion was not in the $|\downarrow\rangle_{N_2}$ state, leading to a much weaker excitation $\{|\beta\rangle\}$. This setup ensured the state selectivity of the present scheme with respect to $|\downarrow\rangle_{N_2}$.

The state-dependent motional excitation (*36*, *37*) mapped the problem of distinguishing between the different internal states $|\downarrow\rangle_{N_2}$ and $\{|\uparrow\rangle_{N_2}\}$ of the molecule to distinguishing between different excited motional states $|\alpha\rangle$ and $\{|\beta\rangle\}$ of the two-ion crystal. The latter was achieved by Rabi sideband thermometry (*50*) on the Ca^+ ion, which shared the motional state with the N_2^+ ion. A blue-sideband (BSB) pulse using a narrow-linewidth laser at 729 nm was used to drive population between $|\uparrow\rangle_{Ca}|n\rangle \rightarrow |\downarrow\rangle_{Ca}|n-1\rangle$ states. It was followed by state-selective fluorescence on Ca^+, which projected the ion either to the $|\uparrow\rangle_{Ca}$ "dark" or the $|\downarrow\rangle_{Ca}$ "bright" state, thereby measuring the success of the BSB pulse ("BSB" in Fig. 1E). The probability of projecting to the "bright" state after the BSB pulse was approximately given by $P(|\downarrow\rangle_{Ca}) = \sum_n P(n)\sin^2(\Omega_n t_{729}/2)$ (*51*) [a more exact form is given in the supplementary materials (*49*)]. Here, t_{729} is the BSB pulse time and $\Omega_n \approx \eta\sqrt{n}\Omega_0$ is the BSB Rabi frequency, where $\eta \approx 0.1$ is the Lamb-Dicke parameter and $\Omega_0 \approx (2\pi)90$kHz is the bare Rabi frequency. The motional Fock-state population distributions were given by $P(n|\alpha)$ or $P(n|\{\beta\})$, depending on the state of N_2^+. Because $0 < P(|\downarrow\rangle_{Ca}|\{\beta\}) < P(|\downarrow\rangle_{Ca}|\alpha) < 1$, the outcome of a single BSB pulse was insufficient to determine the motional state and hence the N_2^+ internal state with high fidelity in a single shot. However, because the internal state of N_2^+ was usually not changed during the measurement, the detection sequence—i.e., cooling of the two-ion string to the motional ground state, preparing Ca^+ in the $|\uparrow\rangle_{Ca}$ state, exciting motion by the ODF, and measuring the result by a BSB pulse—could be repeated until sufficient statistics were obtained to distinguish between different molecular states. The experiment therefore represented a QND measurement (*28*–*30*). Distinguishing between the molecular states $|\downarrow\rangle_{N_2}$ and $|\uparrow\rangle_{N_2}$ is equivalent to distinguishing between two coins, α and β, with biased probabilities to get heads, *h*, in a coin toss given by $0 < p(h|\beta) < p(h|\alpha) < 1$, by repetitively flipping one of the coins *N* times. For $p(h|\alpha) = 0.52$ and $p(h|\beta) = 0.06$, a fidelity of 99.5% can be achieved in the coin (state) determination after $N = 22$ repetitive coin tosses (QND measurements) (*49*).

An experimental demonstration of the presented QND scheme for molecular-state detection is shown in Fig. 3. Here, the ODF beams were turned on for 500 μs with a single lattice-beam intensity of ~2×10^6 W/m^2. The lattice lasers were detuned by $\Delta/2\pi \approx -17$ GHz from

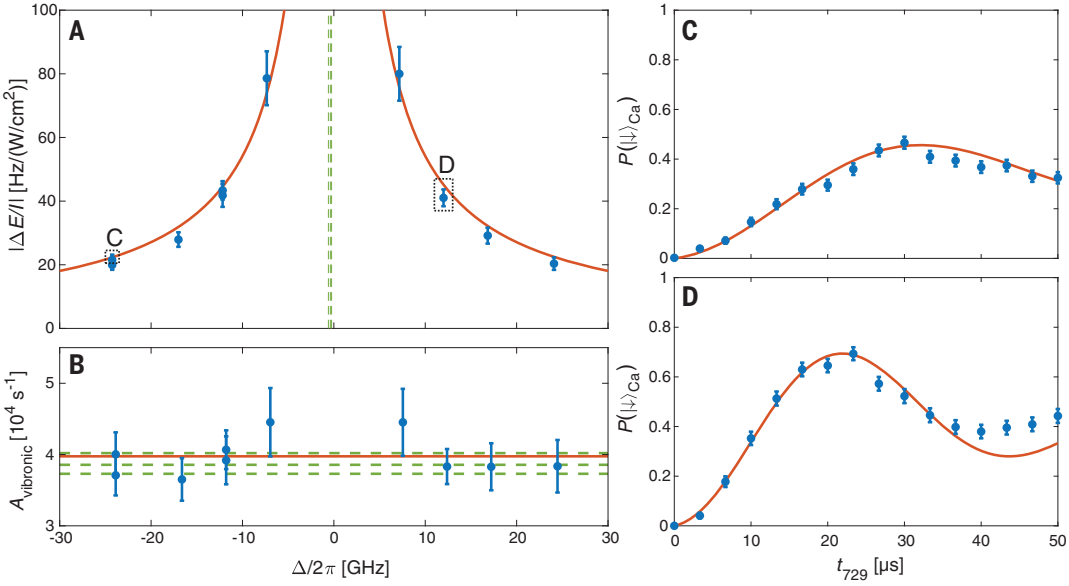


Fig. 4. Nondestructive force spectroscopy on a single N_2^+ molecule. Spectroscopic measurement of the $A^2\Pi_u(v'=2)\leftarrow X^2\Sigma_g^+(v''=0)$, $R_{11}(1/2)$, transition in N_2^+. (**A**) The blue data points represent the amplitude of the ac-Stark shift, ΔE, normalized by the lattice-laser intensity, I, and experienced by the N_2^+ ion as a function of the detuning from resonance, extracted from fits to BSB Rabi oscillation signals [(C) and (D)]. Error bars (1σ) correspond to the uncertainty in the beam intensity and in the extraction of the ODF strength from the BSB signals. The red line is a fit to the experimental data used to determine the line center. The green dashed lines indicate values of the line center reported in the literature (*41*–*43*). (**B**) The blue data points represent values for Einstein *A* coefficients of the $A^2\Pi_u(v'=2)\rightarrow X^2\Sigma_g^+(v''=0)$ vibronic transition in N_2^+ extracted from the measurements in (A). Error bars (1σ) are dominated by the measurement error of the ac-Stark shift in (A). The red line is the mean of all measurements. Different literature values (*44*–*46*) are given by the dashed green lines. (**C** and **D**) Two examples of observed BSB Rabi oscillation signals (blue) for the N_2^+ ion in the $|\downarrow\rangle_{N_2}$ state for lattice-laser detunings, $\Delta/2\pi$, of about −25 GHz and +25 GHz, respectively. Error bars (1σ) correspond to statistical binomial errors. The red line is a fit to determine the ac-Stark shift experienced by the molecule at the respective lattice-laser detunings (*49*).

the $R_{11}(1/2)$ transition (Figs. 1D and 2). The BSB pulse time, t_{729}, was scanned in order to observe Rabi oscillations, shown in Fig. 3A. When the N_2^+ ion was in the $|\downarrow\rangle_{N_2}$ state, a strong Rabi oscillation was observed (blue), in contrast to when the N_2^+ ion was in one of the $\{|\uparrow\rangle_{N_2}\}$ states, where almost no oscillation was observed (red). The residual signal of the $\{|\uparrow\rangle_{N_2}\}$ states is attributed to imperfect ground-state cooling of the two-ion crystal rather than motional excitation by the lattice beams, as can be seen from a comparison to the background signal (green) obtained when the ODF beams were completely turned off.

For the parameters used in the experiment shown in Fig. 3A, the maximum contrast between the $|\downarrow\rangle_{N_2}$ and the $\{|\uparrow\rangle_{N_2}\}$ signals was reached at $t_{729}\approx 20$ μs. For this BSB pulse time, $P(|\downarrow\rangle_{Ca}|\alpha)=0.52$ and $(P(|\downarrow\rangle_{Ca}|\{\beta\})=0.06$, such that 22 QND measurements were sufficient to distinguish between the states at a confidence level of 99.5% (*49*). A lower number of QND measurements would result in a reduced level of confidence for the state determination. Although a higher number of QND measurements would increase the detection fidelity, it would reduce the duty cycle of state determinations. This would imply a smaller number of state-determination attempts before a state-changing event, such as off-resonant photon scattering or inelastic collision with background gas molecules, occurs. Therefore, 22 was chosen as the number of QND measurements per state-detection attempt to optimize the overall fidelity. Such a QND determination of the N_2^+ state is shown in Fig. 3B. The BSB success probability, $P(|\downarrow\rangle_{Ca})$, was determined from the average of the results of 22 BSB pulses with pulse times in the range of 16.7 to 26.7 μs (light-blue shaded area in Fig. 3A). A threshold of $P(|\downarrow\rangle_{Ca})=0.25$ was set to determine whether the molecule was in the $|\downarrow\rangle_{N_2}$ ("bright" molecule) or $\{|\uparrow\rangle_{N_2}\}$ ("dark" molecule) state (blue or red dots in Fig. 3B, respectively). The molecular state was repeatably determined to be "bright" 105 times with zero false detections. Afterward, the molecular state was repeatedly determined to be "dark" 163 times with zero false detections. Using Bayesian inference, the experimentally inferred fidelity was 99.1(9)% and 99.4(6)% for the "bright" and "dark" states, respectively. The sudden change in the state of the molecule from "bright" to "dark" during the experiment was due to a quantum jump that was most likely caused by a state-changing collision with a background gas molecule in our vacuum system at a pressure of 1×10^{-10} mbar. The rate of these state-changing events, together with other types of inelastic processes, such as chemical reactions, is proportional to the partial pressure of background gas molecules in our vacuum system (N_2, H_2O, H_2, and others).

Force spectroscopy

Because the state-detection signal is proportional to the ac-Stark shift experienced by the molecule, it was used to perform measurements of spectroscopic transitions in the molecule. Such a spectroscopic experiment is demonstrated on the $R_{11}(1/2)$ spin-rotational component of the $A^2\Pi_u(v'=2)\leftarrow X^2\Sigma_g^+(v''=0)$ electronic-vibrational transition in N_2^+ (Fig. 4). Rabi oscillations on the BSB transition in Ca^+ resulting from the ODF acting on a molecule in the $|\downarrow\rangle_{N_2}$ state were measured for different detunings of the lattice-laser beams from this resonance. As the resonance was approached, the ac-Stark shift increased as ~$1/\Delta$, leading to a larger excitation, $|\alpha\rangle$, of the ion crystal. The magnitude of the ac-Stark shift was extracted from a fit to the Rabi-oscillation signal (Fig. 4, C and D). The fitting function was experimentally determined by applying a well-defined force on the Ca^+ ion when the N_2^+ ion was in one of the $\{|\uparrow\rangle_{N_2}\}$ states and experienced no force (*49*). The use of an experimentally determined fitting function circumvented the need for characterizing the exact motional state, which may deviate from ideal coherent motion (*49*). For the chosen ODF pulse length of 500 μs, the Rabi signal was sensitive to ac-Stark shifts in the interval from 2.5 to 13 kHz. To extend the dynamic range of our measurement, the lattice-beam powers were scaled to keep the Rabi signal within the experimental sensitivity range.

Figure 4A depicts such a force spectrum of this transition. The experimentally measured ac-Stark shifts were fitted with an $\sim 1/|f-f_0|$ ac-Stark–shift profile to determine the line center, f_0 = 380.7011(2) THz, which agrees well with previous measurements (*41*–*43*) using ensembles of molecules that yielded results in the range f_0 = 380.7007(3) THz (dashed green lines in Fig. 4A). The precision of our

measurement could be enhanced by using smaller detunings, but only at the expense of an increased probability of scattering a photon at the molecule by the lattice beams and thus losing the molecular state. As an example, for the current experimental parameters of ~10 GHz detuning and ~10 kHz ac-Stark shift, an average of 1000 high-fidelity QND state determination cycles (20,000 BSB pulses) can be expected before the molecular state is lost owing to off-resonant scattering (*49*). Decreasing the detuning to 100 MHz would reduce the number of expected QND state determinations to 10 (200 BSB pulses) owing to the $1/\Delta^2$ scaling of the scattering rate (compared with the ac-Stark shift scaling of $1/\Delta$). Nevertheless, these measurements should be possible with efficient and reliable replenishment of molecular ions in the trap. For such close detunings, our method is expected to be sensitive to the hyperfine structure of the transition (*53*), which was not resolved here and has, to our knowledge, not yet been experimentally studied. In this experiment, the absolute accuracy of the wavemeter used to evaluate the lattice-laser frequency was estimated to be better than 50 MHz by repetitive measurements of the $P_{3/2} \leftarrow D_{5/2}$ spectroscopic transition in Ca^+ during the experiment.

The electronic-vibrational (vibronic) part of the Einstein A coefficient, A_{vibronic}, of the $A^2\Pi_u(v'=2) \rightarrow X^2\Sigma_g^+(v''=0)$ transition was extracted from the ac-Stark–shift measurements as shown in Fig. 4B. For each ac-Stark–shift determination, $\Delta E(\Delta)$ (Fig. 4A), a corresponding value $A_{\mathrm{vibronic}}(\Delta)$ was calculated. The mean value of $A_{\mathrm{vibronic}} = 3.98(11) \times 10^4\ \mathrm{s}^{-1}$ is in good agreement with previous results in the range $A_{\mathrm{vibronic}} = 3.87(14) \times 10^4\ \mathrm{s}^{-1}$ (*44*–*46*) (Fig. 4B, dashed green lines). The two data points with the smallest detuning in Fig. 4B seem to slightly deviate from the other points. This effect might be due to the unresolved hyperfine structure of the transition, which becomes nonnegligible at close detunings. Nevertheless, all points were included in the determination of A_{vibronic}.

Outlook

A QND detection of the internal quantum state of a single molecule with >99% fidelity has been demonstrated. The fidelity of the state detection was not limited by the state lifetime and off-resonant scattering, as is the case in atomic ions (*23*, *39*), and hence it can be increased even further, at the expense of a slower data acquisition rate. On the basis of this detection scheme, an approach for measuring spectroscopic line positions and transition strengths in molecules using force spectroscopy has been realized.

The approach presented here can be compared to the pioneering experiment of Wolf *et al.* (*25*), in which a motional qubit was implemented to detect the internal states of polar MgH^+ molecules. Our scheme allows a simpler approach for state detection by coherently exciting motion, which also readily enables the extraction of accurate values for spectroscopic quantities such as transition strengths. Here, a set of tools was developed for the state preparation and quantum manipulation of apolar N_2^+ molecules, which do not couple to the blackbody radiation field. Although immunity to blackbody radiation imposes additional technical challenges, it makes apolar species such as N_2^+ attractive systems, as demonstrated here by the excellent state-detection fidelity.

The present QND scheme should be universally applicable to both polar and apolar molecular ions. It represents a highly sensitive method to repeatedly and nondestructively read out the quantum state of a molecule and thus introduces a molecular counterpart to the state-dependent fluorescence on closed-cycling transitions, which forms the basis of sensitive readout schemes in atomic systems (*23*, *39*). It enables state-selected and coherent experiments with single trapped molecules with duty cycles several orders of magnitude higher than previous destructive state-detection schemes (*15*, *20*). It thus lays the foundations for vast improvements in the sensitivity and, therefore, precision of spectroscopic experiments on molecular ions, as discussed in (*36*). The possibility for efficient, nondestructive state readout also lays the foundation for the application of molecular ions in quantum-information and coherent-control experiments, as are currently being performed with great success using atomic ions (*23*). In this context, the potentially long lifetimes and coherence times of molecular states may offer previously unexplored possibilities for, for example, realizing quantum memories. Additionally, the present scheme also enables studies of cold collisions and chemical reactions between ions and neutrals with state control on the single-molecule level, offering prospects for the exploration of molecular collisions and chemical reaction mechanisms in unprecedented detail. Finally, our approach could be used not only to detect the quantum state of a single molecule but also to prepare a quantum state through a projective measurement down to the Zeeman level (*36*), unlike previous schemes (*15*, *17*), which could prepare only the rovibronic level of the molecule. The present scheme thus also potentially represents a key element in the methodological toolbox of the burgeoning field of molecular quantum technologies and the realization of molecular qubits encoded in the rovibrational spectrum.

REFERENCES AND NOTES

1. K.-K. Ni *et al.*, *Science* **322**, 231–235 (2008).
2. J. M. Sage, S. Sainis, T. Bergeman, D. DeMille, *Phys. Rev. Lett.* **94**, 203001 (2005).
3. L. R. Liu *et al.*, *Science* **360**, 900–903 (2018).
4. S. Y. T. van de Meerakker, H. L. Bethlem, N. Vanhaecke, G. Meijer, *Chem. Rev.* **112**, 4828–4878 (2012).
5. J. F. Barry, D. J. McCarron, E. B. Norrgard, M. H. Steinecker, D. DeMille, *Nature* **512**, 286–289 (2014).
6. L. Anderegg *et al.*, *Nat. Phys.* **14**, 890–893 (2018).
7. K. Mølhave, M. Drewsen, *Phys. Rev. A* **62**, 011401 (2000).
8. S. Willitsch, *Int. Rev. Phys. Chem.* **31**, 175–199 (2012).
9. S. Ospelkaus *et al.*, *Science* **327**, 853–857 (2010).
10. S. Alighanbari, M. G. Hansen, V. I. Korobov, S. Schiller, *Nat. Phys.* **14**, 555–559 (2018).
11. S. Schiller, V. Korobov, *Phys. Rev. A* **71**, 032505 (2005).
12. K. Beloy *et al.*, *Phys. Rev. A* **83**, 062514 (2011).
13. M. S. Safronova *et al.*, *Rev. Mod. Phys.* **90**, 025008 (2018).
14. D. DeMille, J. M. Doyle, A. O. Sushkov, *Science* **357**, 990–994 (2017).
15. P. F. Staanum, K. Højbjerre, P. S. Skyt, A. K. Hansen, M. Drewsen, *Nat. Phys.* **6**, 271–274 (2010).
16. T. Schneider, B. Roth, H. Duncker, I. Ernsting, S. Schiller, *Nat. Phys.* **6**, 275–278 (2010).
17. X. Tong, A. H. Winney, S. Willitsch, *Phys. Rev. Lett.* **105**, 143001 (2010).
18. C.-Y. Lien *et al.*, *Nat. Commun.* **5**, 4783 (2014).
19. J. Biesheuvel *et al.*, *Nat. Commun.* **7**, 10385 (2016).
20. M. Germann, X. Tong, S. Willitsch, *Nat. Phys.* **10**, 820–824 (2014).
21. T. Sikorsky, Z. Meir, R. Ben-Shlomi, N. Akerman, R. Ozeri, *Nat. Commun.* **9**, 920 (2018).
22. A. D. Dörfler *et al.*, *Nat. Commun.* **10**, 5429 (2019).
23. T. P. Harty *et al.*, *Phys. Rev. Lett.* **113**, 220501 (2014).
24. P. O. Schmidt *et al.*, *Science* **309**, 749–752 (2005).
25. F. Wolf *et al.*, *Nature* **530**, 457–460 (2016).
26. C. W. Chou *et al.*, *Nature* **545**, 203–207 (2017).
27. C. W. Chou *et al.*, arXiv:1911.12808 [physics.atom-ph] (28 November 2019).
28. V. B. Braginsky, Y. I. Vorontsov, K. S. Thorne, *Science* **209**, 547–557 (1980).
29. V. B. Braginsky, F. Y. Khalili, *Rev. Mod. Phys.* **68**, 1–11 (1996).
30. D. B. Hume, T. Rosenband, D. J. Wineland, *Phys. Rev. Lett.* **99**, 120502 (2007).
31. M. Kajita, *Phys. Rev. A* **92**, 043423 (2015).
32. M. Kajita, G. Gopakumar, M. Abe, M. Hada, M. Keller, *Phys. Rev. A* **89**, 032509 (2014).
33. S. Schiller, D. Bakalov, V. I. Korobov, *Phys. Rev. Lett.* **113**, 023004 (2014).
34. H. J. Kimble, *Nature* **453**, 1023–1030 (2008).
35. S. Wehner, D. Elkouss, R. Hanson, *Science* **362**, eaam9288 (2018).
36. Z. Meir, G. Hegi, K. Najafian, M. Sinhal, S. Willitsch, *Faraday Discuss.* **217**, 561–583 (2019).
37. D. B. Hume *et al.*, *Phys. Rev. Lett.* **107**, 243902 (2011).
38. J. C. Koelemeij, B. Roth, S. Schiller, *Phys. Rev. A* **76**, 023413 (2007).
39. J. E. Christensen *et al.*, arXiv:1907.13331 [quant-ph] (31 July 2019).
40. M. J. Biercuk, H. Uys, J. W. Britton, A. P. VanDevender, J. J. Bollinger, *Nat. Nanotechnol.* **5**, 646–650 (2010).
41. Y.-D. Wu *et al.*, *Chin. J. Chem. Phys.* **20**, 285–290 (2007).
42. I. H. Bachir, H. Bolvin, C. Demuynck, J. Destombes, A. Zellagui, *J. Mol. Spectrosc.* **166**, 88–96 (1994).
43. K. Harada, T. Wada, T. Tanaka, *J. Mol. Spectrosc.* **163**, 436–442 (1994).
44. D. C. Cartwright, *J. Chem. Phys.* **58**, 178–185 (1973).
45. S. R. Langhoff, C. W. Bauschlicher Jr., H. Partridge, *J. Chem. Phys.* **87**, 4716–4721 (1987).
46. F. R. Gilmore, R. R. Laher, P. J. Espy, *J. Phys. Chem. Ref. Data* **21**, 1005–1107 (1992).
47. A. Gardner, T. Softley, M. Keller, *Sci. Rep.* **9**, 506 (2019).
48. G. Morigi, H. Walther, *Eur. Phys. J. D* **13**, 261–269 (2001).
49. Materials and methods are available as supplementary materials.
50. D. M. Meekhof, C. Monroe, B. E. King, W. M. Itano, D. J. Wineland, *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
51. D. Leibfried, R. Blatt, C. Monroe, D. Wineland, *Rev. Mod. Phys.* **75**, 281–324 (2003).
52. G. Herzberg, *Molecular Spectra and Molecular Structure, Volume I, Spectra of Diatomic Molecules* (Krieger, 1991).
53. P. J. Bruna, F. Grein, *J. Mol. Spectrosc.* **250**, 75–85 (2008).
54. M. Sinhal, Z. Meir, K. Najafian, G. Hegi, S. Willitsch, Raw Data for "Quantum non-demolition state detection and spectroscopy of single trapped molecules," Zenodo repository (2019); https://dx.doi.org/10.5281/zenodo.3532942.

ACKNOWLEDGMENTS

We thank A. Johnson, P. Knöpfel, G. Martin, and G. Holderied for technical support. **Funding:** Funding was provided by the Swiss National Science Foundation as part of the National Centre of Competence in Research, Quantum Science and Technology (NCCR-QSIT), grant CRSII5_183579, and by the University of Basel. **Author contributions:** M.S. and Z.M. carried out the experiments and analyzed the data. K.N. and G.H. contributed to developing the experimental methodology and setup. S.W. conceived and supervised the project. All authors contributed to the writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The raw data supporting the findings of this study are available at the Zenodo repository (*54*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1213/suppl/DC1
Supplementary Text
Fig. S1
References (*55*–*62*)

25 October 2019; accepted 31 January 2020
10.1126/science.aaz9837

FRUSTRATED MAGNETISM

Realization of the kagome spin ice state in a frustrated intermetallic compound

Kan Zhao[1]*†, Hao Deng[2]*, Hua Chen[3]*, Kate A. Ross[3], Vaclav Petříček[4], Gerrit Günther[5], Margarita Russina[5], Vladimir Hutanu[2], Philipp Gegenwart[1]†

Spin ices are exotic phases of matter characterized by frustrated spins obeying local "ice rules," in analogy with the electric dipoles in water ice. In two dimensions, one can similarly define ice rules for in-plane Ising-like spins arranged on a kagome lattice. These ice rules require each triangle plaquette to have a single monopole and can lead to different types of orders and excitations. Using experimental and theoretical approaches including magnetometry, thermodynamic measurements, neutron scattering, and Monte Carlo simulations, we establish HoAgGe as a crystalline (i.e., nonartificial) system that realizes the kagome spin ice state. The system features a variety of partially and fully ordered states and a sequence of field-induced phases at low temperatures, all consistent with the kagome ice rule.

Frustration in spin systems can result in the formation of exotic phases of matter (*1*). One example is the pyrochlore spin ice, in which four nearest-neighbor Ising-like spins sitting at the vertices of a tetrahedron are forced by the exchange and dipolar interactions to obey the "ice rule": two spins pointing into and the other two pointing out of the tetrahedron. Such a local constraint can lead to a macroscopic number of degenerate ground states or an extensive ground state entropy (*2*–*8*).

In two dimensions (2D), ice rules can be similarly defined for in-plane Ising-like classical spins residing on the kagome lattice (*9*–*11*), which require two-in-one-out or one-in-two-out local arrangements of the spins on its triangles. By viewing each spin effectively as a magnetic dipole formed by two opposite magnetic charges or monopoles, the ice rule leaves either a positive or a negative monopole (Q_m = ±1) at each triangle and gives a ground state entropy of ~$0.501k_B$ per spin, where k_B is the Boltzmann constant. However, a $\sqrt{3} \times \sqrt{3}$ ground state can be selected by further-neighbor exchange couplings or the long-range dipolar interaction (*9*–*11*). Consequently, kagome spin ices show a characteristic multistage ordering behavior under changing temperature.

Experimentally, kagome spin ices have only been realized in artificial spin ice systems formed by nanorods of ferromagnets organized into honeycomb networks (*12*–*18*). However, the large magnetic energy scales and system sizes make it challenging to explore the rich phase diagram of spin ices in the thermodynamic limit (*17*, *18*). Alternatively, kagome ice behavior has been reported in pyrochlore spin ices such as $Dy_2Ti_2O_7$ and $Ho_2Ti_2O_7$ under magnetic field along the [111] direction (*19*–*21*). At the right strength, such a magnetic field can align the Ising spins on the triangular layers of the pyrochlore structure; because the field does not break the kagome ice rule, the in-plane components of the spins on the kagome layers can satisfy the rule. However, this is true only in a narrow range of field strength (<1 T) because of the weak exchange or dipolar interactions in such systems. Most recently, a magnetic charge order has been suggested in the tripod kagome compound $Dy_3Mg_2Sb_3O_{14}$ (*22*, *23*), and a dynamic kagome ice has been observed in $Nd_2Zr_2O_7$ under field along the [111] direction (*24*, *25*); however, a long-range spin order does not appear in either case even at the lowest temperature.

Here, we used multiple experimental and theoretical approaches to show that the intermetallic compound HoAgGe is a naturally existing kagome spin ice that exhibits a fully ordered ground state.

[1]Experimentalphysik VI, Center for Electronic Correlations and Magnetism, University of Augsburg, 86159 Augsburg, Germany. [2]Institute of Crystallography, RWTH Aachen University and Jülich Centre for Neutron Science (JCNS) at Heinz Maier-Leibnitz Zentrum (MLZ), D-85747 Garching, Germany. [3]Department of Physics, Colorado State University, Fort Collins, CO 80523, USA. [4]Institute of Physics, Academy of Sciences of the Czech Republic, 18221 Prague, Czech Republic. [5]Helmholtz-Zentrum Berlin für Materialien und Energie, D-14109 Berlin, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: kan.zhao@physik.uni-augsburg.de (K.Z.); philipp.gegenwart@physik.uni-augsburg.de (P.G.)

Structure and magnetometry measurements

HoAgGe is one of the ZrNiAl-type intermetallics with space group P-62m, which is noncentrosymmetric. In particular, Zr sites in the *ab* plane form a distorted kagome lattice (*26*, *27*) (Fig. 1A). The distortion is characterized by opposite rotations of the two types of triangles in the kagome lattice by the same angle (~15.58° in HoAgGe) around the *c* axis. The rotation breaks the spatial inversion symmetry of a single kagome layer, although it does not change the space group of the 3D crystal (*28*). Previous neutron diffraction measurements suggested the presence of noncollinear magnetic structures of HoAgGe (*29*), but the powder samples used in that work yielded limited magnetic peaks that were insufficient to fully determine the magnetic structure, especially in the presence of frustration. Below, we combine neutron diffraction with thermodynamic measurements in single-crystalline HoAgGe to reveal its exotic temperature- and magnetic field–dependent magnetic structures, which we show to be consistent with the kagome ice rule.

Each Ho^{3+} atom in HoAgGe has 10 $4f$ electrons. According to Hund's rules, they should have the ground state of $^5I_8$ with an effective magnetic moment $\mu_{eff} = 10.6\mu_B$, as confirmed by our Curie–Weiss fitting to the anisotropic inverse susceptibilities $\chi^{-1}(T)$ above 100 K (fig. S2A). At lower temperatures, $\chi(T)$ for $\mathbf{H}//b$ under 500 Oe exhibits a relatively sharp peak at 11.6 K (denoted as T_2) and another broad inflection at ~7 K (denoted as T_1), which are more clearly seen in the plot of the temperature derivative of $\chi(T)$ (Fig. 1B). Similar behaviors are also observed for $\mathbf{H}//a$, whereas for $\mathbf{H}//c$, $\chi(T)$ monotonically increases with decreasing temperature (*27*).

Plots of magnetization versus $\mathbf{H}//b$ show a series of plateaus at low temperatures (Fig. 1C). At $T = 5$ K, one can clearly identify three metamagnetic transitions at $H \approx 1$, 2, and 3.5 T. At each transition, the magnetization changes by ~1/3 of the saturated value (M_s) at $H > 4$ T. At

Table 1. Summary of HoAgGe single-crystal neutron data refinement results. Because of the limited number of magnetic and nuclear peaks and the uncertainty in aligning the field exactly with the *b* axis, the final refinement factors under field are usually larger than in the zero-field case yet are still within a reasonable range (mostly smaller than 10%).

Temperature (field)	15 K	10 K	4 K	4 K (1.5 T)	1.8 K (2.5 T)	1.8 K (4 T)
Magnetic space group	P-62m	P-6′m2′	P-6′m2′	Am′m2′	Am′m2′	Am′m2′
Magnetic vector (k, k, 0)		k = 1/3	k = 1/3	k = 1/3	k = 1/3	k = 0
Ho label	Ho1	Ho1– Ho3	Ho1– Ho3	Ho1– Ho6	Ho1–Ho6	Ho1, Ho2
Ordered moment (μ_B)		5.2 (1)	7.5 (1)	7.6 (1)	7.6 (2)	7.4 (3)
Neutron peaks, independent peaks	535, 106	330, 99	971, 217	234, 164	254, 157	220, 137
Magnetic refinement factor (R, wR) (%)	2.95, 3.89	5.20, 6.27	3.38, 3.98	8.61, 10.85	5.90, 7.19	6.52, 8.18

lower temperatures (1.8 K in Fig. 1C), two additional small plateaus with a jump of ~1/6 M_s appear at 0.9 and 3.2 T, respectively, accompanied by a small hysteresis. The $M(H)$ curve for $\mathbf{H}//a$ also shows well-defined plateaus, albeit at different ranges of field (fig. S2B), whereas no plateaus are observed for $\mathbf{H}//c$ (fig. S2E). An H–T phase diagram based on temperature dependence of the peaks in field derivative of $M(H)$ curves (fig. S2, C and D) is constructed in Fig. 1D. Together with the $\chi(T)$ data above, the lack of any clear magnetic transitions for $\mathbf{H}//c$ confirms that the Ho spins in HoAgGe are constrained in the ab plane and have additional in-plane anisotropies, similar to those in the isostructural compounds TmAgGe and TbPtIn (*30*).

Magnetic structures determined from neutron diffraction

To fully determine the nontrivial spin structures of HoAgGe, we performed single-crystal neutron diffraction experiments down to 1.8 K and under $\mathbf{H}//b$ up to 4 T (*31*). Below the high-temperature transition T_2 = 11.6 K, a magnetic peak appears at (1/3, 1/3, 0) (Fig. 2A and fig. S4A), indicating a $\sqrt{3} \times \sqrt{3}$ magnetic unit cell (the green rhombus in Fig. 2B). Below 10 K, most nuclear sites exhibit almost constant intensity but the broad transition at T_1 induces additional magnetic contribution at certain structural diffraction sites, such as (1, 0, 0) (Fig. 2A, inset).

According to neutron data at 10 K (fig. S5A), the magnetic structure belongs to the magnetic space group P-6′m2′ (Table 1), which has three nonequivalent Ho sites labeled by Ho1, Ho2, and Ho3 in Fig. 2B. Six other Ho positions in the magnetic unit cell are obtained from above three Ho sites by threefold rotations around the c axis. Because there are no magnetic contributions at nuclear sites at 10 K, the simplest possibility for (M_{Ho1}, M_{Ho2}, M_{Ho3}) is (M, $-M$, 0), with M determined to be 5.2(1)μ_B (Table 1 and fig. S10B). This corresponds to Ho1 and Ho2 exhibiting ordered moments of the same size but opposite directions and Ho3's moment fluctuating without ordering. Such a partially ordered magnetic structure is shown in Fig. 2C, with the ordered moments forming clockwise or counterclockwise hexagons separated by the unordered moments. The structure thus has a nonzero magnetic toroidal moment defined by $\tau = \frac{1}{V}\int d^3r\,\mathbf{r} \times \mathbf{M}$ (*32*). Similar partially ordered structures have also been observed in the isostructural Kondo lattice CePdAl below 2.7 K with easy c-axis anisotropy (*33*) and in hexagonal UNi_4B below 20 K, with 2/3 of U moments forming in-plane clockwise hexagons (*34*).

Below T_1 ~7 K, Ho3 moments also enter the long-range order, as indicated in the inset of Fig. 2A. Refinement of neutron data at 4 K (figs. S5B and S10C) leads to the magnetic structure shown in Fig. 2E, which also has the P-6′m2′ symmetry, with (M_{Ho1}, M_{Ho2}, M_{Ho3}) = (M, $-M$, $-M$) and M = 7.5(1)μ_B. As illustrated in Fig. 2F, this fully ordered ground state includes alternating clockwise and counterclockwise hexagons of spins, and another 1/3 of hexagons consisting of three pairs of parallel spins. This is exactly the $\sqrt{3} \times \sqrt{3}$ ground state of the classical kagome spin ice predicted theoretically before (*35*–*37*).

To confirm that HoAgGe is indeed a kagome spin ice, however, it is necessary to show that the ice rule is established even outside the fully ordered ground state (*9*–*11*). The kagome ice rule requires dominating nearest-neighbor ferromagnetic coupling between coplanar spins with site-dependent Ising-like uniaxial anisotropy (*9*–*11*). Using neutron diffraction under magnetic fields, we show that these requisites are indeed satisfied in HoAgGe. Figure 2D displays the neutron-scattering integrated intensities of

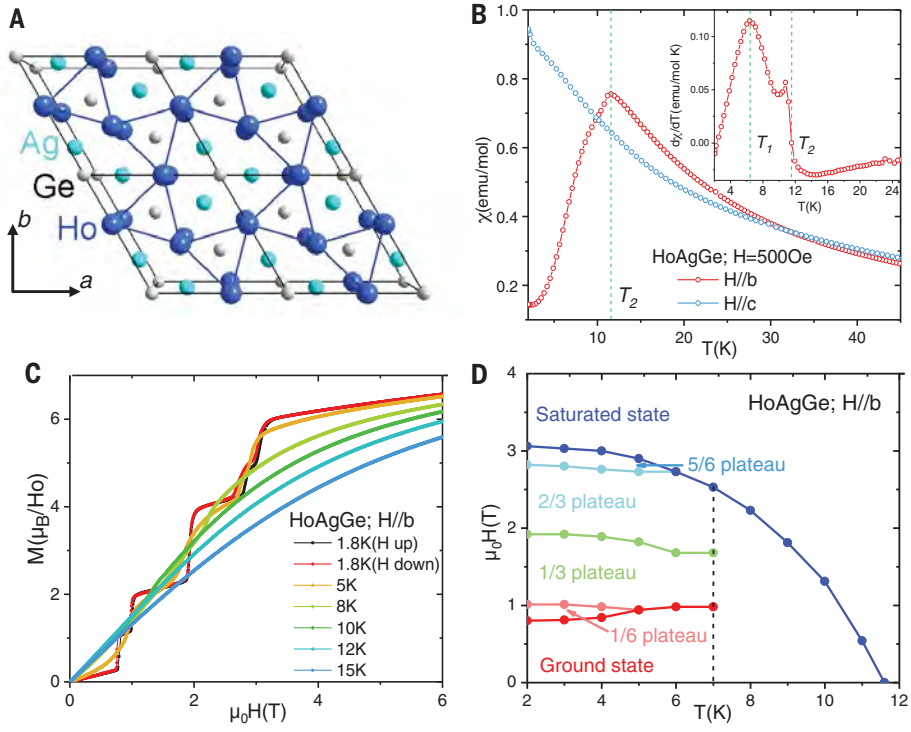


Fig. 1. Crystal structure and magnetic properties of HoAgGe. (**A**) c-Axis projection of the HoAgGe crystal structure, with the definition of a and b directions. (**B**) Low-temperature susceptibility $\chi(T)$ of HoAgGe for both $\mathbf{H}//b$ and $\mathbf{H}//c$ under 500 Oe, with $d\chi(T)/dT$ in the inset. (**C**) Isothermal in-plane ($\mathbf{H}//b$) magnetization for HoAgGe at various temperatures. (**D**) Dependence of the metamagnetic transitions on temperature, with the dotted line indicating T_1 (see text).

Table 2. The four low-energy (<1 meV) CEF modes of Ho^{3+} in HoAgGe.

Irreducible representation	Wave functions
Γ_2	− 0.6186(|7> +|−7>) − 0.1871(|5> +|−5>) − 0.2591(|3> +|−3>) + 0.1234(|1> +|−1>)
Γ_4	− 0.6209(|7> −|−7>) − 0.1961(|5> −|−5>) − 0.2636(|3> −|−3>) − 0.0814(|1> −|−1>)
Γ_3	− 0.0780(|8> −|−8>) − 0.6256(|6> −|−6>) − 0.1512(|4> −|−4>) − 0.2822(|2> −|−2>)
Γ_1	0.0938(|8> +|−8>) + 0.6472(|6> +|−6>) + 0.1865(|4> +|−4>) + 0.1257(|2> +|−2>) − 0.2083|0>

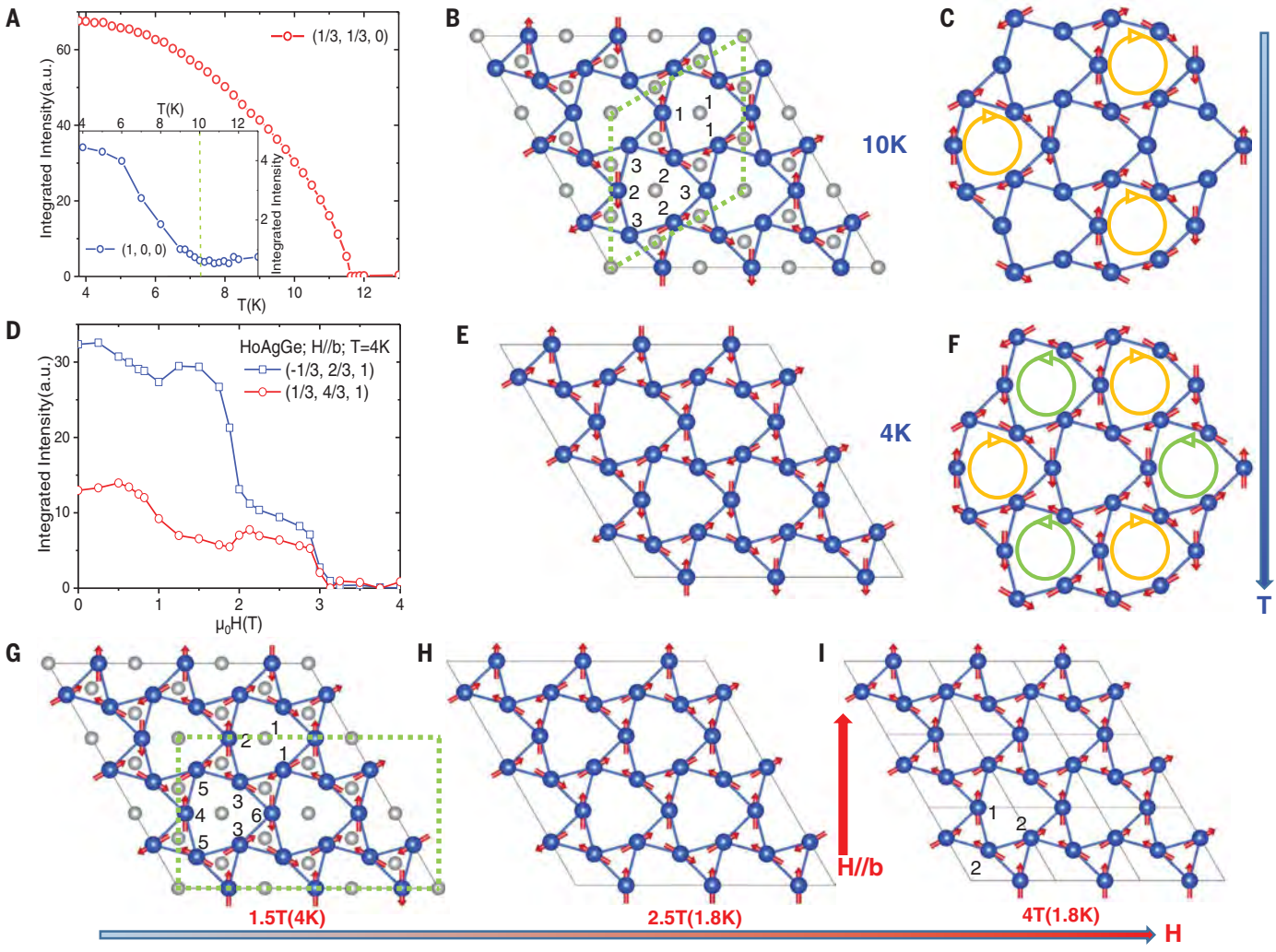


Fig. 2. Magnetic structures of HoAgGe versus temperature and field with *H*//*b*. (**A**) Integrated intensity of the magnetic peak (1/3, 1/3, 0) (fig. S4A) from 13 K down to 3.8 K according to the neutron diffraction, with the integrated intensity of nuclear site (1, 0, 0) as an inset. (**B**) Refined magnetic structures of HoAgGe at 10 K. The magnetic unit cell is indicated by the green rhombus, with the three inequivalent Ho sites Ho1, Ho2, and Ho3 labeled by 1, 2, and 3, respectively, for simplicity. (**C**) Counterclockwise hexagons of spins in the partially ordered structure of HoAgGe at 10 K, with 1/3 spins not participating in the long-range order. (**D**) Integrated intensity of magnetic peak (−1/3, 2/3, 1) (fig. S4B) and (1/3, 4/3, 1) versus field at 4 K. (**E**) Refined magnetic structure of HoAgGe at 4 K. (**F**) Clockwise and counterclockwise hexagons of spins in the magnetic structure of HoAgGe at 4 K, which is exactly the expected $\sqrt{3} \times \sqrt{3}$ ground state of kagome spin ice. (**G**) Refined magnetic structure of HoAgGe at H = 1.5 T and T = 4 K. The refinement was done in the $3 \times \sqrt{3}$ light-green rectangle. The six inequivalent Ho sites are labeled by numbers 1 to 6 for simplicity. (**H**) Refined magnetic structure of HoAgGe at H = 2.5 T and T = 1.8 K. (**I**) Refined magnetic structure of HoAgGe at H = 4 T and T = 1.8 K, with the two inequivalent Ho sites labeled by 1 and 2. The field direction is marked by the red arrow for (G) to (I).

the magnetic peaks at (−1/3, 2/3, 1) (fig. S4B) and (1/3, 4/3, 1) versus the strength of the magnetic field along the *b* axis at 4 K. Overall, the intensity decreases with increasing field and disappears at $H > 3.2$ T, with sudden changes at the metamagnetic transitions depicted in Fig. 1, C and D, suggesting the shrinking of the magnetic unit cell in field. To obtain further information, we refined the magnetic structures at the three major $M(H)$ plateaus from the neutron scattering. The magnetic field breaks the threefold rotational symmetry and turns the ground state magnetic space group P-6′m2′ into Am′m2′, with the nine Ho moments in the $\sqrt{3} \times \sqrt{3}$ unit cell forming six nonequivalent groups (Fig. 2G).

Figure 2, G to I, shows the magnetic structures at the three major plateaus obtained from the neutron data taken at 1.8 K and H = 1.5, 2.5, and 4 T along the *b* axis (also see Table 1 for the refinement factors). One first notices that all of them can be obtained from the ground state by reversing certain Ho spins, with negligible rotation from their local Ising axis (*31*). This is strong evidence for the Ising-like anisotropy of the Ho moments, with the local easy axes defined by a perpendicular mirror plane through each atom. The Ising-like anisotropy is further confirmed by our crystalline electric field (CEF) calculations below. Moreover, in all three structures, the spins are always reversed in such a way that the one-in-two-out or two-in-one-out ice rule is satisfied but the total magnetic moment along *b* increases with increasing field. At H = 4 T, the magnetic unit cell becomes identical to the structural unit cell (*14*, *15*, *18*) and has the largest possible net moment allowed by the ice rule. This is further corroborated by the identical magnetization jump of $1.7\mu_B$/Ho at the three metamagnetic transitions at 1.8 K (fig. S2B). Assuming the magnetic structures in Fig. 2, G to I, this jump can be translated to

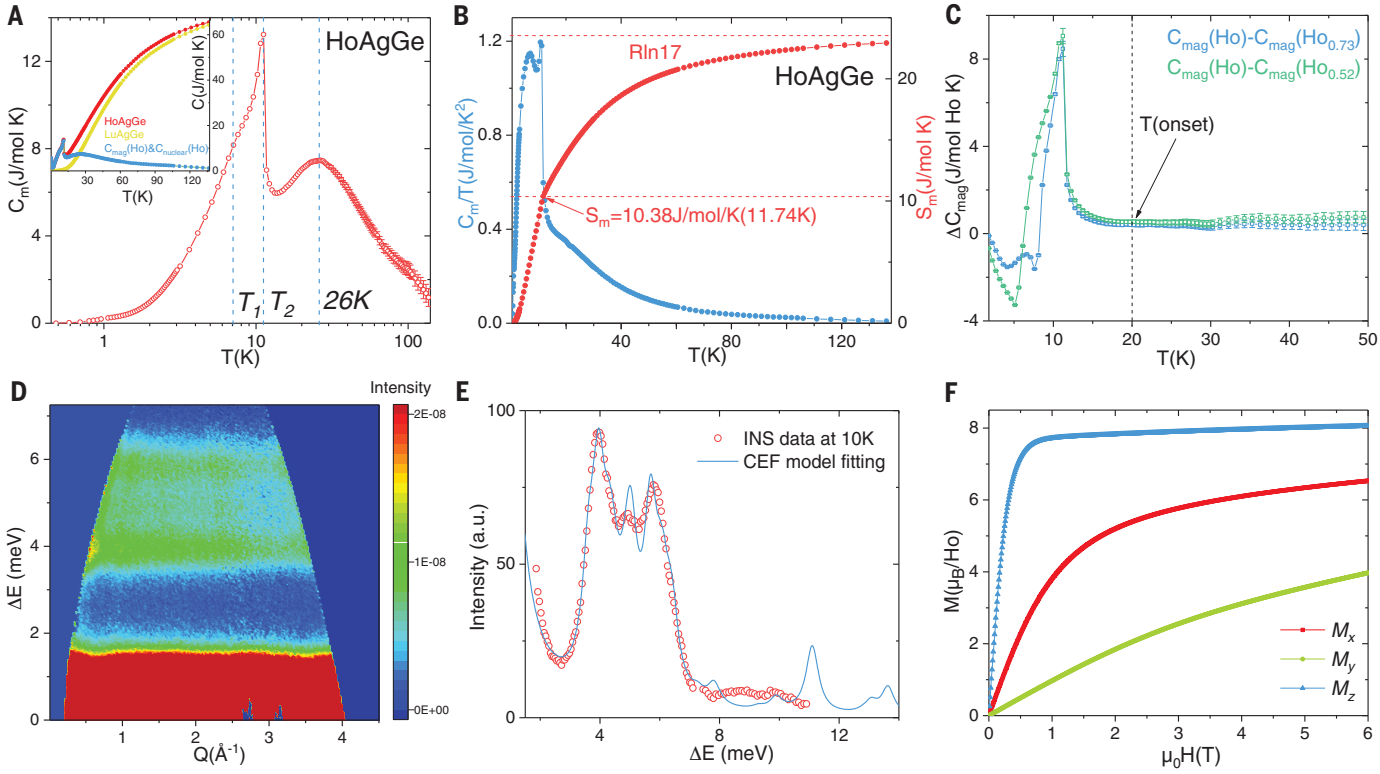


Fig. 3. Magnetic specific heat and INS results of HoAgGe. (**A**) Magnetic contribution to the specific heat C_m of HoAgGe with the dotted lines indicating T_1, T_2, and a broad peak at 26 K (see text). Note that the error bars below 30 K are smaller than the symbol sizes. (Inset) Specific heat of HoAgGe, LuAgGe, and their difference. The latter is defined as the sum of the magnetic and the nuclear contributions to the specific heat of HoAgGe. (**B**) C_m/T data and the corresponding magnetic entropy S_m, which approaches the theoretical value of Rln17 above 100 K. (**C**) Difference between the magnetic specific heat of HoAgGe and that of $Lu_{1-x}Ho_xAgGe$ (x = 0.52 and 0.73) after normalization (see text). (**D**) INS spectra of HoAgGe at 10 K with incident neutron wavelength 3 Å. (**E**) Constant-Q cuts (1.4 < Q < 2.2 Å^{-1}) showing the results of the CEF fitting to neutron-scattering data. (**F**) Isothermal magnetization calculated for CEF-fitting parameters at 1.5 K for three quantization axes.

an ordered moment size of $M = (9/2) \times 1.7\mu_B = 7.65\mu_B$, roughly consistent with that determined from neutron data at zero field [7.5(1)μ_B at 4 K and 5.2(1)μ_B at 10 K]. These results indicate that the Ho moments at low temperatures are constrained by the kagome ice rule. The metamagnetic transitions result from the competition between the external magnetic field and the weaker, further than nearest-neighbor couplings that do not affect the ice rule. For a detailed analysis of the three magnetic structures, see (*31*).

Specific heat and magnetic entropy

Having established the existence of the kagome ice rule in HoAgGe at low temperatures, we then proceeded to examine the thermodynamic behaviors of kagome spin ice. To this end, we isolated the magnetic contribution to the specific heat C_{mag} by subtracting the contributions from nuclei, lattice vibrations, and itinerant electrons (*31*). Figure 3A shows the C_{mag} thus obtained from 136 K down to 0.48 K. In addition to the two peaks at T_1 and T_2, another broad peak appears at 26 K (discussed further below).

Figure 3B shows the magnetic entropy $S_m(T)$ obtained by integrating $C_{mag}(T)/T$ from (nominally) T = 0 K. At high temperatures (>100 K), S_m approaches Rln17, consistent with the $^5I_8$ state of an isolated Ho^{3+} and close to that of the structurally similar intermetallic compounds $HoNiGe_3$ (*38*) and $Ho_3Ru_4Al_{12}$ (*39*). For the ideal kagome spin ice, however, S_m should approach Rln2 at high temperatures because of the Ising anisotropy. The temperature dependence of the magnetic entropy of HoAgGe thus must be analyzed together with the CEF splitting of the Ho^{3+} J = 8 multiplet (see below).

Short-range spin ice correlations stemming from the kagome ice rule can lead to a broad peak in specific heat $C_{mag}(T)$ at the temperature scale corresponding to the nearest-neighbor exchange coupling (*10*, *11*). To investigate the origin of the broad peak at 26 K in Fig. 3A, we also investigated $Lu_{1-x}Ho_xAgGe$ (x = 0.52 and 0.73). Because Lu^{3+} is not magnetic, the exchange interaction between Ho moments is suppressed as x decreases, whereas the CEF splitting that can lead to the Schottky anomaly should not change much. As shown in fig. S11, the T_1 and T_2 for the magnetic transitions shift down to 8 and 4 K for $Lu_{0.27}Ho_{0.73}AgGe$, and for $Lu_{0.48}Ho_{0.52}AgGe$, T_2 shifts to 5 K with T_1 < 1.8 K. However, in both cases, the broad anomaly in the C_{mag} curves still appears at ~26 K. We therefore conclude that the broad peak is a Schottky anomaly caused by the CEF splitting of the Ho^{3+} multiplet.

To clearly see the effects of short-range correlations caused by the exchange interaction between Ho moments, we subtracted the normalized $Lu_{1-x}Ho_xAgGe$ magnetic specific heat from that of pure HoAgGe (Fig. 3C). The resulting ΔC_{mag} is almost constant (within the error bar) above 20 K, but increases as T goes below 20 K until reaching a maximum at the transition to the partially ordered state; therefore, short-range spin ice correlations still exist below 20 K. However, the broad peak characteristic of an ideal kagome ice model (*10*, *11*) is absent, which will be discussed further below.

Inelastic neutron scattering and CEF analysis

To determine to what extent the Ho spins can be approximately viewed as Ising, we next discuss the CEF effects. According to the local orthorhombic symmetry (point group C_{2v}) of Ho sites in HoAgGe, CEF splits the 17-fold multiplet of a non-Kramers Ho^{3+} ion into 17 singlets. To directly probe the CEF splitting, we conducted

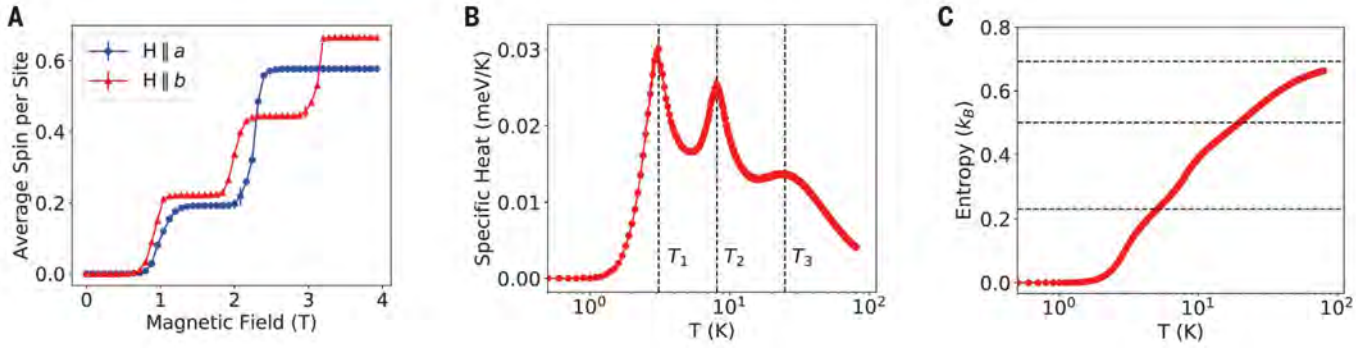


Fig. 4. Monte Carlo simulations of the 2D classical spin model for HoAgGe. (**A**) *M*(*H*) curves at 1 K for **H** along the *a* and *b* axes. (**B**) Temperature dependence of the specific heat per spin. (**C**) Magnetic entropy per spin calculated from the specific heat. The three horizontal dashed lines correspond to ln2 ≈ 0.693 (paramagnetic Ising), 0.501 (short-range ice order), and $\frac{1}{3}$ln2≈0.231 (toroidal order), respectively. An 18 × 18 cell was used for the calculation.

inelastic neutron scattering (INS) experiments of HoAgGe crystals using the time-of-flight spectrometer NEAT at Helmholtz Zentrum Berlin (*31*, *40*). To minimize the influence of internal fields caused by magnetic exchange interactions in the presence of long-range order (*41*), we chose to conduct the measurements at 10 K, very close to T_2, with incident neutron wavelengths of 2.4 and 3 Å. Clear CEF modes, which are independent of momentum transfer *Q*, appear between 4 and 6 meV (Fig. 3D). The same modes are also observed in the INS spectra at 50 K in fig. S14B, indicating that the internal field is already very weak at 10 K. Additional CEF modes with broad features appear between 8 and 11 meV in fig. S14A. With an incident neutron wavelength of 5 Å (3.27 meV) at 15 K, a continuum feature appears in the quasi-elastic scattering plane (ΔE < 0.2 meV) in fig. S14C, indicating diffuse scattering, which is consistent with the observation of short-range spin correlations below 20 K in Fig. 3C.

Using the INS data at 10 K and the magnetic specific heat result >20 K in Fig. 3A, we did a combined fitting (Fig. 3E) to obtain the CEF Hamiltonian. The nine CEF parameters from the fitting are listed in table S3. As shown in Table 2, among the 17 CEF levels, the lowest four with energies <1 meV should be the major ones contributing to the kagome ice behavior at low temperatures. The other 13 CEF modes are listed in table S4. The four low-energy CEF modes indeed have Ising-type anisotropy, as shown in Fig. 3F. Under a magnetic field along the local Ising axis, the Ho moment steeply increases to 7.7μ_B at 1 T (8.1μ_B at 6 T), which is much larger than 6.5μ_B and 4.0μ_B for fields along the two perpendicular directions at 6 T. Because the CEF Hamiltonian does not include the effect of exchange coupling between Ho moments, which is of similar size to the separation between the lower CEF levels, it may not fully account for the anisotropy of the moments, especially at low temperatures.

Classical Monte Carlo simulations

Based on the experimental evidence presented above, we propose a classical spin model consisting of Ising-like in-plane spins on the 2D distorted kagome lattice of the [0001] plane of HoAgGe. The spins are coupled to one another through exchange couplings and long-range dipolar interactions and to external magnetic fields through Zeeman coupling. The comprehensive *M*(*H*) data and magnetic structures from neutron scattering allow us to extract the exchange couplings up to the fourth nearest neighbor, with implicit summation over periodic images along the *c* axis (*31*). The exchange couplings are found to be dominant over the dipolar interaction, quite different from the pyrochlore spin ices $Dy_2Ti_2O_7$ and $Ho_2Ti_2O_7$, as well as $Dy_3Mg_2Sb_3O_{14}$ (*22*, *23*), where they are comparable. This is likely a result of the relatively strong RKKY-type interaction between Ho moments (*27*). As a check, we calculated the *M*(*H*) curves for **H** along the *b* and *a* axes at *T* = 1 K through Monte Carlo simulations (Fig. 4A), which were consistent with the experimental results in Fig. 1C and fig. S2B. We did not consider the effects of the Dzyaloshinskii–Moriya interaction (*31*).

Monte Carlo simulations of the classical spin model on an 18 × 18 lattice with periodic boundary conditions give three peaks in the specific heat versus temperature plot, as in the experiment (Fig. 4B). The positions of the two peaks of *C*(*T*) at lower temperatures agree well with the experimental data in Fig. 3B. The broad peak at the highest temperature (denoted as T_3) is due to the gradual development of ice rule correlations for Ising spins. T_3 is mainly determined by the nearest-neighbor exchange coupling, which, however, is not fixed by the experimental data. Because Ho^{3+} in the real material is not Ising like at *T* > 20 K owing to a strong population of higher CEF levels, such a peak does not have to be present in the experiment (*31*). The ground state and the partially ordered state can both be reproduced by the Monte Carlo simulations (fig. S16). We thus believe that the classical spin model captures the main characteristics of the magnetism of HoAgGe at low temperatures.

The temperature dependence of the magnetic entropy in Fig. 4C confirms that the peaks of *C*(*T*) at T_3 and T_1 in Fig. 4B correspond to the formation of spin ice correlations and the fully ordered ground state, respectively, similar to what was predicted for dipolar kagome ice (*10*, *11*). However, the peak at T_2 in Fig. 4B does not correspond to the transition into the "magnetic charge order" in dipolar kagome ice, which has an entropy of 0.108k_B per spin (*10*, *11*). In fact, the magnetic charge order is destabilized by the further-neighbor exchange couplings in our model. By contrast, the Monte Carlo simulations of the short-range kagome ice, with ferromagnetic nearest-neighbor coupling and antiferromagnetic second–nearest-neighbor coupling (*9*) indeed give an intermediate state similar to that in Fig. 2C through first-order transitions. Simple counting gives an entropy of $\frac{k_B}{3}$ln2≈0.231 k_B per spin for the partially ordered state shown in Fig. 2B. However, it has been shown more recently that the state shown in Fig. 2C (and the ground state as well) has a Z_6-order parameter, similar to a six-state clock model, and the transition into it should be of the Kosterlitz–Thouless (KT) type (*37*).

Our model (as well as the physical system) is different from both dipolar and short-range kagome ice cases because of the coexistence of the further-neighbor exchange couplings and the long-range dipolar interaction. The precise nature of the two low-temperature transitions in our 2D model may only be elucidated through comprehensive finite-size scaling analysis, which deserves a separate study. It also remains an open question how the transitions are influenced by the long-range tail of the RKKY interaction. In reality, however, the KT transition may not be very likely in the 3D HoAgGe, especially considering the strong exchange coupling between neighboring kagome

planes (*31*). Another piece of experimental evidence is the critical exponent $\beta = 0.321(3)$ of the (1/3, 1/3, 0) peak intensity near T_2 (fig. S4A), which indicates the 3D Ising nature of the magnetic order (*42*).

Discussion

Although the Monte Carlo simulations of the classical spin model described above are in partial agreement with our experiments, they do not explain the experimental value of the magnetic entropy $S_m = 10.38\ J \cdot mol^{-1}K^{-1} \approx 1.248R$ at T_2, which is very different from the $0.231R$ given by the model. Qualitatively, the discrepancy should be a result of the thermal population of multiple low-lying CEF levels of Ho^{3+}; however, this leads to the question of why the classical Ising Hamiltonian is applicable. In pyrochlore systems such as $Dy_2Ti_2O_7$, the classical Ising behavior is a consequence of the dominance of CEF splitting over exchange and dipolar interactions. This leads to an effective pseudospin-1/2 Hamiltonian with only the pseudospin components along the local easy axes present (*43*). These pseudospin components are thus good quantum numbers, justifying the use of the classical Ising Hamiltonian.

In HoAgGe, metallicity simultaneously suppresses the CEF splitting of Ho^{3+} ions and enhances the exchange coupling between them, making the two energy scales comparable at least for the low-lying CEF levels. Thus, the large ($J = 8$) Ho^{3+} moments in HoAgGe at moderately low temperatures, when multiple CEF levels are occupied, are closer to semiclassical spins with strong single-ion anisotropy. Such a semiclassical model can still be mapped to an Ising model at the expense of introducing further-neighbor exchange interactions (*44*), which serves as an explanation for the apparent validity of the classical Ising Hamiltonian for HoAgGe. A complete understanding of the entropy data awaits a full quantum mechanical description of the system. It is worth noting that the deviation from an ideal spin-1/2 system can also lead to stronger quantum fluctuations, as in the cases of $Tb_2Ti_2O_7$ (*45*) and $Tb_2Sn_2O_7$ (*46*).

The metallic nature of HoAgGe not only makes it a high-temperature (compared with pyrochlore spin ices) kagome ice, but may also lead to exotic phenomena such as the interaction between electric currents and the magnetic monopoles or the toroidal moments, the relationship between the noncollinear ordering and the anomalous Hall effect (*47*–*50*), and metallic magnetoelectric effects caused by broken inversion symmetry (*51*). Our results suggest that ZrNiAl-type intermetallic compounds are a prototypical family of kagome spin systems that may host other exotic phases beyond the classical spin liquid (*52*, *53*) and deserve further investigation.

REFERENCES AND NOTES

1. L. Balents, *Nature* **464**, 199–208 (2010).
2. M. J. Harris, S. T. Bramwell, D. F. McMorrow, T. Zeiske, K. W. Godfrey, *Phys. Rev. Lett.* **79**, 2554–2557 (1997).
3. A. P. Ramirez, A. Hayashi, R. J. Cava, R. Siddharthan, B. S. Shastry, *Nature* **399**, 333–335 (1999).
4. S. T. Bramwell, M. J. P. Gingras, *Science* **294**, 1495–1501 (2001).
5. C. Castelnovo, R. Moessner, S. L. Sondhi, *Nature* **451**, 42–45 (2008).
6. L. Pauling, *J. Am. Chem. Soc.* **57**, 2680–2684 (1935).
7. D. J. P. Morris *et al*., *Science* **326**, 411–414 (2009).
8. T. Fennell *et al*., *Science* **326**, 415–417 (2009).
9. A. S. Wills, R. Ballou, C. Lacroix, *Phys. Rev. B* **66**, 144407 (2002).
10. G. Möller, R. Moessner, *Phys. Rev. B* **80**, 140409 (2009).
11. G.-W. Chern, P. Mellado, O. Tchernyshyov, *Phys. Rev. Lett.* **106**, 207202 (2011).
12. R. F. Wang *et al*., *Nature* **439**, 303–306 (2006).
13. Y. Qi, T. Brintlinger, J. Cumings, *Phys. Rev. B* **77**, 094418 (2008).
14. S. Ladak, D. Read, G. K. Perkins, L. F. Cohen, W. Branford, *Nat. Phys.* **6**, 359–363 (2010).
15. E. Mengotti *et al*., *Nat. Phys.* **7**, 68–74 (2011).
16. S. Zhang *et al*., *Nature* **500**, 553–557 (2013).
17. C. Nisoli, R. Moessner, P. Schiffer, *Rev. Mod. Phys.* **85**, 1473–1490 (2013).
18. J. C. Gartside *et al*., *Nat. Nanotechnol.* **13**, 53–58 (2018).
19. K. Matsuhira, Z. Hiroi, T. Tayama, S. Takagi, T. Sakakibara, *J. Phys. Condens. Matter* **14**, L559–L565 (2002).
20. Y. Tabata *et al*., *Phys. Rev. Lett.* **97**, 257205 (2006).
21. T. Fennell, S. T. Bramwell, D. F. McMorrow, P. Manuel, A. R. Wildes, *Nat. Phys.* **3**, 566–572 (2007).
22. Z. L. Dun *et al*., *Phys. Rev. Lett.* **116**, 157201 (2016).
23. J. A. M. Paddison *et al*., *Nat. Commun.* **7**, 13842 (2016).
24. E. Lhotel *et al*., *Phys. Rev. Lett.* **115**, 197202 (2015).
25. E. Lhotel *et al*., *Nat. Commun.* **9**, 3786 (2018).
26. R. Gibson, R. Pöttgen, R. K. Kremer, A. Simon, K. R. A. Ziebeck, *J. Alloys Compd.* **239**, 34–40 (1996).
27. E. Morosan, S. L. Bud'ko, P. C. Canfield, M. S. Torikachvili, A. H. Lacerda, *J. Magn. Magn. Mater.* **277**, 298–321 (2004).
28. H. T. Stokes, D. M. Hatch, *J. Appl. Cryst.* **38**, 237–238 (2005).
29. S. Baran *et al*., *J. Alloys Compd.* **281**, 92–98 (1998).
30. E. Morosan, S. L. Bud'ko, P. C. Canfield, *Phys. Rev. B Condens. Matter Mater. Phys.* **71**, 014445 (2005).
31. Materials and methods are available as supplementary materials.
32. V. M. Dubovik, V. V. Tugushev, *Phys. Rep.* **187**, 145–202 (1990).
33. A. Dönni *et al*., *J. Phys. Condens. Matter* **8**, 11213–11229 (1998).
34. S. A. M. Mentink *et al*., *Phys. Rev. Lett.* **73**, 1031–1034 (1994).
35. T. Takagi, M. Mekata, *J. Phys. Soc. Jpn.* **62**, 3943–3953 (1993).
36. M. Wolf, K. D. Schotte, *J. Phys. Math. Gen.* **21**, 2195–2209 (1988).
37. G.-W. Chern, O. Tchernyshyov, *Phil. Trans. R. Soc. A* **370**, 5718–5737 (2012).
38. E. D. Mun, S. L. Bud'ko, H. Ko, G. J. Miller, P. C. Canfield, *J. Magn. Magn. Mater.* **322**, 3527–3543 (2010).
39. D. I. Gorbunov *et al*., *Phys. Rev. B* **97**, 184412 (2018).
40. M. Russina *et al*., *Physica B* **551**, 506–511 (2018).
41. P. Javorský, M. Diviš, H. Sugawara, H. Sato, H. Mutka, *Phys. Rev. B* **65**, 014404 (2001).
42. A. Scheie *et al*., *Phys. Rev. B* **93**, 180407 (2016).
43. J. G. Rau, M. J. P. Gingras, *Phys. Rev. B* **92**, 144417 (2015).
44. Y.-Z. Chou, Y.-J. Kao, *Phys. Rev. B* **82**, 132403 (2010).
45. I. Mirebeau *et al*., *Phys. Rev. Lett.* **94**, 246402 (2005).
46. M. J. P. Gingras *et al*., *Phys. Rev. B* **62**, 6496–6511 (2000).
47. R. Shindou, N. Nagaosa, *Phys. Rev. Lett.* **87**, 116801 (2001).
48. Y. Taguchi, Y. Oohara, H. Yoshizawa, N. Nagaosa, Y. Tokura, *Science* **291**, 2573–2576 (2001).
49. H. Chen, Q. Niu, A. H. MacDonald, *Phys. Rev. Lett.* **112**, 017205 (2014).
50. S. Nakatsuji, N. Kiyohara, T. Higo, *Nature* **527**, 212–215 (2015).
51. H. Saito *et al*., *J. Phys. Soc. Jpn.* **87**, 033702 (2018).
52. J. Carrasquilla, Z. Hao, R. G. Melko, *Nat. Commun.* **6**, 7421 (2015).
53. Y. Huang, M. Hermele, *Phys. Rev. B* **95**, 075130 (2017).
54. K. Zhao, H. Deng, H. Chen, K. A. Ross, V. Petříček, G. Günther, M. Russina, V. Hutanu, P. Gegenwart, Data for: Realization of the kagome spin ice state in a frustrated intermetallic compound, Zenodo (2020); https://doi.org/10.5281/zenodo.3623256.

ACKNOWLEDGEMENTS

We thank Y. Tokiwa, S. Nakatsuji, O. Tchernyshyov, J. Xu, J. Song, V. Peçanha-Antonio, and M. Meven for helpful discussions; A. Jesche for providing access to the MPMS; and T. Delazzer for help with the CEF calculations. K.A.R. acknowledges J. Rau and M. Gingras for CEF discussions. We also thank K. Kiefer and colleagues from Helmholz Zentrum Berlin for providing access and assistance using the VM4 magnet at MLZ. The technical assistance of W. Luberstetter in setting up the Oxford Instruments magnet on POLI is appreciated. **Funding:** The work in Augsburg was supported by the German Science Foundation through SPP1666 (project no. 220179758) and TRR80 (project no. 107745057). The instrument POLI at Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany, was operated by RWTH Aachen University in cooperation with FZ Jülich (Jülich Aachen Research Alliance JARA). Inelastic neutron experiments were conducted at the time-of-flight spectrometer NEAT operated by Helmholtz Zentrum Berlin. This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado–Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado–Boulder and Colorado State University. The work in Prague was supported by the Czech Science Foundation through project no. 18-10504S. **Author contributions:** K.Z. and P.G. proposed the experiments; K.Z. synthesized single crystals and measured magnetic properties and specific heat; H.D. and V.H. conducted single-crystal elastic neutron scattering; V.P., H.D., and K.Z. refined magnetic structures; G.G. and M.R. conducted inelastic neutron scattering; H.C. provided theoretical analysis and MC simulation; K.A.R. performed CEF calculations; K.Z., H.C., and P.G. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data presented in this paper can be found on Zenodo (*54*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1218/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S4
References (*55*–*65*)

26 November 2018; accepted 14 February 2020
10.1126/science.aaw1666

STRUCTURAL BIOLOGY

Structure of CD20 in complex with the therapeutic monoclonal antibody rituximab

Lionel Rougé[1], Nancy Chiang[2], Micah Steffek[3], Christine Kugel[4], Tristan I. Croll[5], Christine Tam[4], Alberto Estevez[1], Christopher P. Arthur[1], Christopher M. Koth[1], Claudio Ciferri[1], Edward Kraft[4], Jian Payandeh[1,2]*, Gerald Nakamura[2]*, James T. Koerber[2]*, Alexis Rohou[1]*

Cluster of differentiation 20 (CD20) is a B cell membrane protein that is targeted by monoclonal antibodies for the treatment of malignancies and autoimmune disorders but whose structure and function are unknown. Rituximab (RTX) has been in clinical use for two decades, but how it activates complement to kill B cells remains poorly understood. We obtained a structure of CD20 in complex with RTX, revealing CD20 as a compact double-barrel dimer bound by two RTX antigen-binding fragments (Fabs), each of which engages a composite epitope and an extensive homotypic Fab:Fab interface. Our data suggest that RTX cross-links CD20 into circular assemblies and lead to a structural model for complement recruitment. Our results further highlight the potential relevance of homotypic Fab:Fab interactions in targeting oligomeric cell-surface markers.

The integral membrane protein cluster of differentiation 20 (CD20) is a B cell–specific marker and a clinically validated therapeutic target for B cell malignancies and autoimmune conditions (*1*). It is ubiquitously expressed on circulating B cells (*2*) and is predicted to have four transmembrane helices (TMs) with two extracellular loops, ECL1 and ECL2, the second of which is much longer and contains a disulfide bond (fig. S1A). These topological features are conserved among a group of membrane proteins called MS4A (membrane-spanning 4-domain family, subfamily A), which includes 18 proteins identified through similarities between their amino acid sequences but whose biological functions are mostly unknown (*3*). Aside from structures of short ECL2 peptide segments from CD20 (*4*–*6*), there exists no high-resolution structural data on any MS4A family member beyond predictions of secondary structure and membrane topology. Although CD20 is the best-studied member of the family, even its oligomeric state is poorly understood: Available evidence suggests that it associates into homo-oligomers and complexes with other proteins (*7*–*9*). In addition, the function of CD20 remains an area of active debate. Early work suggested that CD20 functions as an ion channel because overexpression and knockout of CD20 can increase or decrease Ca^{2+} conductance in B cells, respectively (*7*, *10*). However, more recent work showed that $CD20^+$ B cells lacking the B cell receptor are unable to initiate calcium signaling, suggesting that CD20 indirectly regulates calcium release downstream from the B cell receptor (*11*).

CD20-targeted therapies revolutionized the treatment of B cell malignancies and autoimmune disorders, starting with the monoclonal antibody (mAb) rituximab (RTX; Rituxan), which was the first approved therapeutic mAb for cancer and continues to be the benchmark for second- and third-generation mAbs (*1*, *12*). Another anti-CD20 mAb, ocrelizumab (OCR; Ocrevus), is now used in the treatment of multiple sclerosis (*13*). Although all anti-CD20 mAbs act by depleting B cells, they use at least four distinct mechanisms (*14*): direct cell death, Fc receptor effector functions through antibody-dependent cellular cytotoxicity and phagocytosis, and complement-dependent cytotoxicity (CDC). Each therapeutic antibody varies in its ability to trigger each pathway and there is no molecular-level understanding of why this is the case, but these distinct functional effects have been useful in categorizing anti-CD20 mAbs into either type I or type II (*1*, *14*). Rituximab is the prototypical type I mAb, characterized by high CDC activity and the ability to cluster CD20 into lipid rafts (*12*, *15*). Other type I mAbs include OCR and ofatumumab [OFA; (*16*)]. Type II mAbs such as obinutuzumab (OBZ; Gazyva) and tositumomab (Bexxar) exhibit low CDC activity and lack the ability to localize CD20 into lipid rafts but induce higher levels of direct cell death (*17*).

These broad categorizations do not explain how CD20:mAb binding and mAb features lead to different modes of action. One hint at a possible molecular underpinning for these differences is that twice as many type I mAbs bind the surfaces of $CD20^+$ cells as type II mAbs (*18*), suggesting that CD20:mAb binding stoichiometry plays a role, though it is not clear how such strict stoichiometry might arise. Adding to this mystery, some type II and type I mAbs (e.g., OBZ and RTX) have overlapping epitopes centered around the ^{170}Ala-Asn-Pro-Ser-Glu^{174} (^{170}ANPSE^{174}) motif of ECL2, whereas at least one type I mAb (OFA) has a completely separate, nonoverlapping epitope involving ECL1 and another part of ECL2 (*14*, *19*). Thus, the location of the epitope on CD20 cannot be the sole determinant of mAb stoichiometry or of therapeutic mode of action. How do antibodies with virtually identical epitope sequences centered on ^{170}ANPSE^{174} bind with different stoichiometries and trigger notably different responses? One hint comes from epitope fine-mapping studies showing that residue Asn^{176} of ECL2 is involved in OBZ binding, but not RTX binding, and that OBZ binds to ECL2 peptides (but not $CD20^+$ cells) with higher affinity than RTX (*5*); another hint comes from x-ray crystallographic structures of peptide-bound antigen-binding fragments (Fabs) of RTX, OCR, and OBZ, in which RTX and OCR approach the ECL2 epitope at a similar angle, tilted approximately 70° away from the angle at which OBZ approaches the same peptide (*4*–*6*). However, in the absence of a structure of full-length CD20 or of a MS4A homolog, it is difficult to speculate how such differences in binding geometry might affect overall binding stoichiometry and dictate therapeutic mode of action.

Results

CD20 forms dimers bound by two RTX Fabs

To facilitate biophysical and structural analyses, we produced human CD20 recombinantly in insect cells and optimized the construct for increased expression (fig. S1A). After solubilization and purification in the mild detergent glyco-diosgenin (GDN), CD20 was found to be further stabilized by cholesterol hemisuccinate (CHS; fig. S2). Analyzed in GDN-CHS buffer, size exclusion chromatography with multi-angle light scattering (SEC-MALS) indicated that purified CD20 forms stable complexes in 2:2 stoichiometry with RTX Fabs and 2:1 stoichiometry with OBZ Fabs (fig. S1, C to E, and table S1). Fab binding to purified CD20 was subsequently evaluated using biolayer interferometry (BLI; Fig. 1A) and surface plasmon resonance (SPR; table S2), revealing sensorgrams consistent with a two-state 2:2 binding of RTX [dissociation constant (K_D) = 21.4 nM, SPR], and with 2:1 binding of OBZ Fab (K_D = 58.8 nM, SPR). Imaging the resulting CD20:Fab complexes by negative-stain electron microscopy (nsEM) showed that each CD20 particle is bound by either two RTX Fabs or a single OBZ Fab (Fig. 1B). Cryogenic electron microscopy (cryo-EM) imaging of the RTX:CD20 complex allowed us to determine its structure to a resolution of 3.3 Å (fig. S3), resulting in a near-complete atomic model of the complex (Fig. 1C).

[1]Department of Structural Biology, Genentech Inc., South San Francisco, CA 94080, USA. [2]Department of Antibody Engineering, Genentech Inc., South San Francisco, CA 94080, USA. [3]Department of Biochemical and Cellular Pharmacology, Genentech Inc., South San Francisco, CA 94080, USA. [4]Department of Biomolecular Resources, Genentech Inc., South San Francisco, CA 94080, USA. [5]Cambridge Institute for Medical Research, University of Cambridge, Keith Peters Building, Cambridge CB2 0XY, UK. *Corresponding author. Email: rohou.alexis@gene.com (A.R.); koerber.james@gene.com (J.T.K.); nakamura.gerald@gene.com (G.N.); payandeh.jian@gene.com (J.P.)

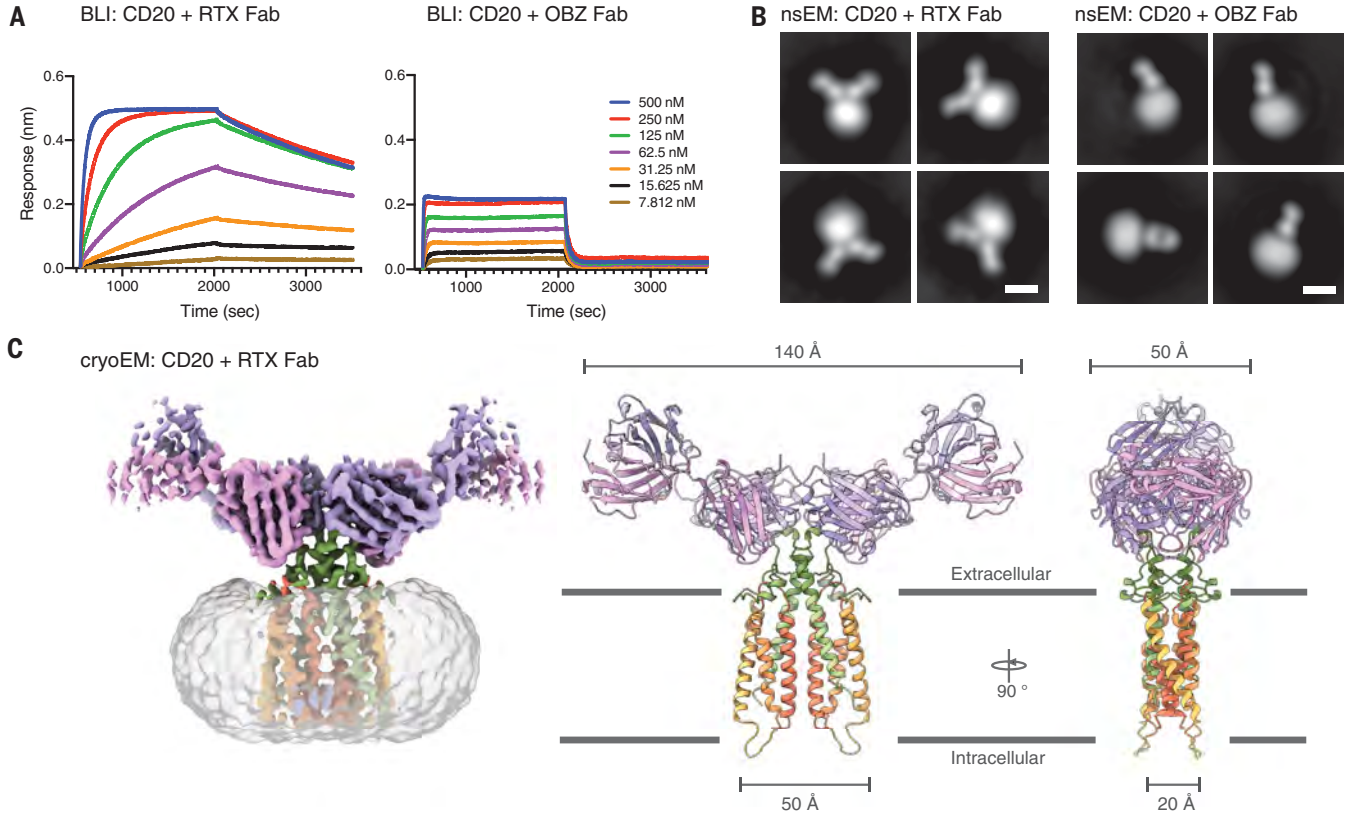


Fig. 1. Characterization of CD20:Fab complexes and cryo-EM structure of CD20:RTX Fab. (**A**) BLI traces. A serial dilution of RTX (left) or OBZ (right) Fabs was flowed in for the first 1500 s of the experiments, followed by a dissociation step. (**B**) nsEM of expressed, solubilized, and purified CD20 in complex with RTX (left) or OBZ (right) Fab. Scale bars, 50 Å. (**C**) Cryo-EM reconstruction of the CD20:RTX Fab complex at a resolution of 3.3 Å. An isosurface rendering, with the GDN micelle rendered in transparent gray, the RTX Fab heavy chain in purple, and the light chain in pink is shown on the left. Two orthogonal side views (along the plane of the membrane) of a ribbon rendering of the structure are shown on the right.

The CD20 fold

The RTX:CD20 complex is Y shaped, with the two RTX Fabs poised on the extracellular face of CD20. CD20 conforms to the predicted four-TM arrangement, with its four TMs arranged antiparallel and clockwise when viewed from the extracellular side (Fig. 2, top view). The core of each CD20 monomer presents a compacted rectangular fold measuring ~25 Å by ~20 Å by ~50 Å (Fig. 2). Despite low overall sequence identity across the MS4A family (~30%), the structure of CD20 reveals that key structural elements are likely shared by all members (fig. S4A). Most notable is a constellation of highly conserved, small residues that allow for the close interhelical packing observed in CD20 and are found along TM1 (Gly^{53}, Gly^{60}, Gly^{67}), TM2 (Gly^{90}, Ser^{97}, Gly^{98}), and TM3 (Ser^{123}, Gly^{130}) (fig. S4A). A set of larger residues contributed by TM2 (Tyr^{86}, Tyr^{94}) and TM4 (Leu^{194}, Met^{197}, Ala^{201}, Gln^{204}) form the bulk of the tightly packed TM-helical core (Fig. 2B) and are conserved as similarly bulky residues in the MS4A family (fig. S4A). Notwithstanding some conformational heterogeneity observed at the intracellular end of TM1, the close interdigitation of highly conserved residues over ~30 Å creates a tightly sealed four-TM bundle within the CD20 monomer that is inconsistent with the formation of a transmembrane permeation pathway.

In contrast to the conserved transmembrane core, the extracellular loops of the MS4A family are extremely diverse in sequence (fig. S4A). In CD20, the first extracellular loop (ECL1) is short and largely shielded by ECL2, leaving only Ile^{76} and Tyr^{77} exposed (Fig. 2A). The first half of the approximately 35-residue-long ECL2 is an amphipathic sequence that may partially partition into the membrane region and surrounds the perimeter of the four-helix bundle until it kinks into a single-turn α helix (Asn^{153} to Arg^{156}), which marks the start of ECL2's extracellular segment (Fig. 2A). Following the α helix, residues His^{158} to Ile^{164} form a circumflex-shaped cap above ECL1, and Ile^{164} and Tyr^{165} appear to plug a cavity at the center of the square CD20 fold. The remaining part of ECL2 is stapled by the landmark disulfide bond between Cys^{167} and Cys^{183}, which is located on an N-terminal extension of TM4 (Fig. 2, side view). This region of ECL2, akin to a turret, is by far the most solvent-accessible region of CD20 and contains the known peptide epitope (^{170}ANPSE^{174}) for most anti-CD20 antibodies (*14*).

A search of the Protein Data Bank for structures similar to CD20 identified claudin-3 and CD81 as the nearest matches (*20*, *21*). Although both present similar topologies, structural superposition with CD20 demonstrates poor overall correspondence (fig. S4B). For example, the transmembrane cavity and cholesterol-coordinating acidic residue present in CD81 (*20*) are absent in CD20, and although claudins appear to share a similar core four-TM packing with CD20, they display a different topology and distinct oligomeric assembly interfaces. We conclude that the three-dimensional structure of CD20 represents a distinct membrane protein fold.

CD20 is a dimer

Previous studies have suggested that CD20 exists as a dimer or a tetramer (*7*–*9*). Viewed from the extracellular side, CD20 forms a dimeric double-barrel assembly of approximate dimensions 20 Å by 50 Å (Fig. 2, top view). Two square CD20 subunits abut each other to form a four-helix antiparallel transmembrane coiled coil involving the upper halves of TM1 (Leu^{61}, Phe^{62}, Ala^{65}, Leu^{69}) and TM4 (Leu^{189}, Ile^{193}, Val^{196}, Phe^{200}) from each protomer (Fig. 2D).

The residues at this intersubunit interface are involved in hydrophobic and van der Waals interactions. In the intracellular half of the transmembrane domain, dimerization is mediated by homotypic contacts between symmetry-related TM1 residues (Thr^{51}, Ala^{54}, Val^{55}, Met^{58}). The close, complementary packing and extensive hydrophobic surface contact at the CD20 dimer interface does not support the existence of an interprotomer transmembrane conduction pathway. Additionally, structure-based sequence alignment suggests that a CD20-like dimer interface may be shared across the MS4A family (fig. S4A).

The dimeric assembly of CD20 is reinforced by extensive contacts between the extracellular α-helical extension of TM4 and the solvent-exposed region of ECL2. Here, numerous hydrophobic interactions and some polar contacts (Ser^{179} to Ser$^{179'}$; Gln^{181} to the backbone amide of Tyr$^{161'}$) contribute to the interface (Fig. 2C). In total, the CD20 dimer buries 1656 Å^2 of surface area and has a shape complementarity score of 0.53, comparable to many established dimeric integral membrane proteins [e.g., (*22*, *23*)]. We further examined Tyr^{182}, located near the symmetry axis in CD20, and found that mutation to cysteine (Tyr^{182}Cys) resulted in purification of a covalent dimeric species in the absence of RTX (fig. S1F), consistent with the dimeric assembly observed in our structure. Overall, we conclude that CD20 forms a tight dimeric assembly that places ECL2 loops in close proximity to each other and presents its main epitope (^{170}ANPSE^{174}) in closely associated pairs, less than 20 Å apart.

ECL2 of CD20 is simultaneously recognized by two RTX Fabs

Previous studies have established that the principal epitope of RTX is centered on the ECL2 sequence ^{170}ANPSE^{174} (*14*). In our structure, this core epitope is simultaneously bound by two Fabs, which we denote RTX and RTX′ (Fig. 3). The RTX Fab sits atop the CD20 protomer, engaging ECL2 at a shallow angle (~22°) relative to the membrane plane (Fig. 3, side view), likely precluding engagement of CD20 by a single immunoglobulin G (IgG). RTX engulfs the core epitope through numerous van der Waals packing contacts as well as multiple polar interactions, including the hydrogen-bond pairs HC.Ser^{58}–Pro^{169} (backbone), HC.His^{35}–Asn^{171}, and HC.Asn^{33}–Ser^{173} (backbone and side chain), where HC indicates heavy chain. The second (RTX′) Fab extends its heavy-chain variable loop 3 (H3) across the dimer interface to present HC.Tyr$^{97'}$, which interacts with Glu^{174} of CD20 (Fig. 3A). At the apex of ECL2, Ser^{173} organizes an extended network of hydrogen bonds spanning from HC.Asn^{33}, through Glu^{174}, to HC.Tyr$^{97'}$ (Fig. 3B). This key interaction network is clamped by both Fabs: Ser^{173} is stabilized by HC.Tyr^{52}, Glu^{174} is sandwiched by HC.Trp^{100b} and HC.Gly$^{100'}$, and HC.Tyr^{97} is stabilized by HC.Trp^{100b}.

The CD20:RTX complex reveals a distinct secondary epitope

Our structure reveals a second CD20 epitope formed by ECL1 and ECL2 and contacted by complementarity-determining region (CDR) loop L1 of RTX (Fig. 3D). Completely distinct from the classic ECL2 turret epitope ^{170}ANPSE^{174}, this secondary epitope is recognized primarily by light-chain (LC) residues LC.Ser^{28}, LC.Ser^{29}, and LC.Ser^{31}. Residues LC.Ser^{28} and LC.Ser^{29} are positioned to make van der Waals contacts with Ile^{76} of ECL1 and Pro^{160} of ECL2, respectively. The side chain of LC.Ser^{31} is situated atop ECL2's circumflex cap and interfaces with both CD20 protomers: It stabilizes Tyr^{161} in a CH_2-arene-CH_2 sandwich also involving Pro^{160} while its hydroxyl moiety makes van der Waals contacts with Pro$^{178'}$, which caps the TM4 α-helical extension of the CD20′ protomer. Earlier crystallographic studies of the primary ECL2 turret epitope in complex with RTX had measured a buried surface area of only 440 Å^2 (*4*), but these L1-ECL1/2 interactions increase the contact surface area by almost 50%, to ~640 Å^2 (Fig. 4B), suggesting that this secondary epitope likely contributes substantially to RTX's affinity for CD20.

RTX Fabs are engaged in homotypic contacts

The close proximity (~20 Å) of the two primary epitopes displayed by the CD20 dimer results in the RTX Fabs accommodating each other along a homotypic interface between their heavy chains (Fig. 3). CDR loop 3 (H3) dominates the Fab:Fab interface, engaging with its symmetry mate (H3′) and with the H1′ and H2′ loops. Residue HC.Tyr^{97}, which is germline-encoded by means of the D gene segment (fig. S6A), seems essential to the Fab-Fab interaction: Its Cβ and Cγ atoms make close van der Waals contacts with their symmetry mates across the dimer axis while its aromatic ring stabilizes the backbone of the H3 loop and its hydroxyl hydrogen-bonds with Glu$^{174'}$ of the contralateral CD20′ protomer (Fig. 3, A and B). Two key additional Fab:Fab interactions are mediated by HC.Ser$^{31'}$, whose backbone and side chain directly engage HC.Gly^{99} while its side chain contacts LC.Tyr^{49} and HC.Tyr^{98} (Fig. 3C). Overall, this Fab:Fab homotypic interface buries 375 Å^2 of solvent-exposed area (Fig. 4B).

Our structure has thus unveiled a composite CD20-Fab′ epitope with three components:

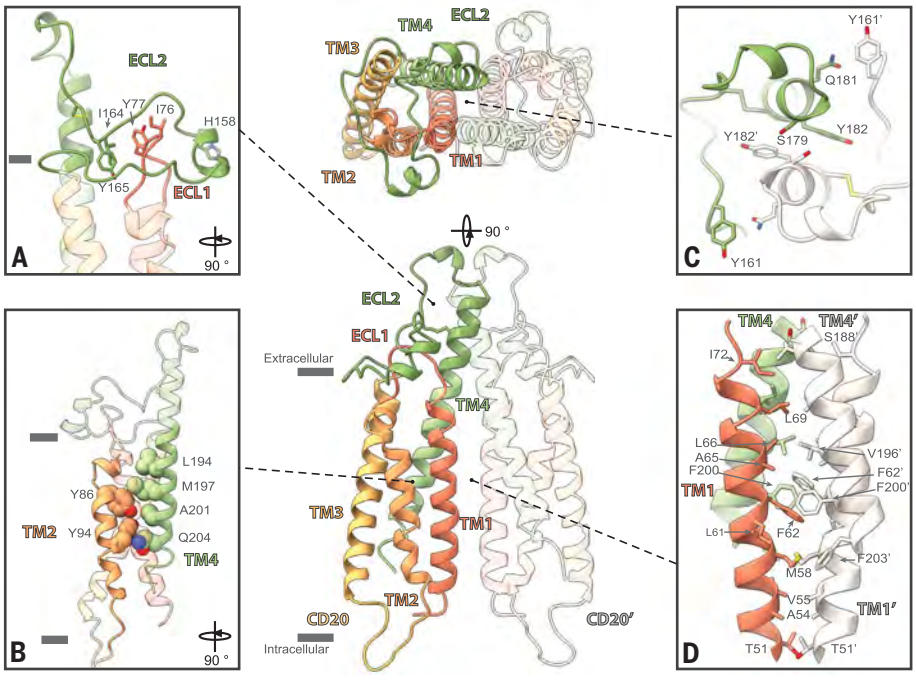


Fig. 2. The CD20 dimer is a compact double–square-barrel structure. Ribbon diagrams of the CD20 structure, with RTX omitted and one of the CD20 protomers transparent for clarity. (**A**) The short loop ECL1 (red), between TM1 and TM2, is almost entirely surrounded by the first half of ECL2 (green). (**B**) The core of each protomer is marked by a number of highly conserved, small (mostly glycine) residues (not shown) and a complementary set of bulkier residues shown here in space-filling representation. (**C** and **D**) The extensive dimeric interface of CD20 involves the extracellular domain (C) as well as TM1 and TM4 (D). In (A), (B), and the center diagram, the gray bars indicate the boundaries of the membrane region. Single-letter abbreviations for 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. Oxygen, nitrogen, and sulfur atoms are colored in red, blue, and yellow, respectively.

the primary epitope on ECL2 (^{170}ANPSE^{174}), a secondary ECL1-ECL2 epitope, and a direct homotypic Fab:Fab interface. The total composite buried surface area of ~1000 Å^2 is comparable to traditional Fab:protein complexes (*24*), and our observations rationalize how RTX achieves nanomolar affinity for full-length CD20 (table S2) and CD20$^+$ cells (*5*) despite low affinity to the ECL2 epitope peptide (table S3).

RTX's target recognition and CDC activity require the full composite epitope

To ascertain the functional relevance of previously unobserved structural features of the complete CD20 epitope, we introduced targeted mutations into RTX (Fig. 4C). We used charge reversals and bulky side chains expected to disrupt these molecular interfaces and measured their effects on CDC (Fig. 4E) and on IgG binding to purified CD20 (SPR; table S2), ECL2 epitope peptide (SPR; table S3), and CD20$^+$ cells (flow cytometry; Fig. 4D). For comparison, we included OBZ, which is known to bind cells at levels ~50% lower than RTX and whose Fab can only bind purified CD20 with 1:2 stoichiometry (Fig. 1).

RTX variants LC.Ser^{28}Asp and LC.Ser^{31}Asp were generated to probe the importance of the secondary epitope. Mutation LC.Ser^{28}Asp resulted in reduced CD20 affinity (~20 fold), cellular binding (~25%), and CDC activity [median inhibitory concentration (IC_{50}) > 10 times higher] relative to wild-type RTX. LC.Ser^{31}Asp had a notably stronger phenotype, which resulted in a ~100-fold reduction in Fab:CD20 binding and ~50% reduction in IgG:cell binding and nearly completely abolished CDC activity. These data substantiate the relevance of this secondary epitope in RTX function.

We next evaluated the role of the germline-encoded HC.Tyr^{97}, because it appears central to complex formation (see above; Fig. 3, A and B). The HC.Tyr^{97}Ser mutation, which we predict may destabilize H3, effectively abolishes target engagement and CDC activity, whereas mutation HC.Tyr^{97}Phe, which removes only the terminal hydroxyl but maintains the aromaticity of the side chain, reduced affinity to CD20 (~15 fold) as well as cellular binding (~25%) and CDC activity (IC_{50} > 10 times higher), suggesting that RTX function is enhanced by the polar interaction between HC.Tyr^{97} and Glu$^{174'}$.

To assess the importance of the homotypic Fab:Fab interface, we introduced mutations at HC.Ser^{31} and HC.Gly^{99}, two positions that are reciprocally involved in interactions at the periphery of the complex. The HC.Ser^{31}Glu mutant had reduced Fab:CD20 affinity (~100 fold), reduced IgG:cell binding, and reduced CDC activity (>100 fold). The effect of HC.Gly^{99}Lys was even more marked, with CDC completely abolished. Because these residues are not involved in any interactions with CD20, we conclude that homotypic Fab:Fab interactions potentiate target engagement, cell binding, and CDC activity of RTX.

In summary, we have discovered several RTX mutants that, despite maintaining cell-binding activity comparable to that of OBZ and largely unaffected binding to the primary ECL2 epitope, are incapable of eliciting CDC. This confirms that the secondary ECL1/2 epitope and Fab:Fab interface contribute to the distinctive binding properties and high CDC activity of RTX.

Full-length RTX cross-links CD20 dimers into higher-order assemblies

In the RTX Fab:CD20 complex, the distance between the C termini of the Fab heavy chains (HC.Pro^{213}) is greater than 120 Å, inconsistent with binding of both Fabs from a single IgG to a CD20 dimer (Fig. 1C). This suggests that two RTX antibodies engage the dimer, each contributing one of its Fab domains. To test this, we formed complexes in vitro between full-length IgG and CD20 and examined their structural arrangement. We found that these complexes are stable (table S2), and nsEM showed that most CD20 particles were bound by two well-resolved Fabs in a similar geometry to that seen in the CD20:RTX Fab structure (fig. S7G). This establishes that 2:2 complex formation is not exclusive to Fab fragments and occurs readily in the context of full-length RTX.

Unlike OBZ or the RTX Fab, RTX IgG cross-links CD20 into cyclical superstructures of 2-to-2 or 3-to-3 IgGs and CD20s, with approximate diameters of 250 and 300 Å, respectively (Fig. 5A). These closed-ring assemblies feature CD20 dimers and Fc domains splayed outward, linked by pairs of Fab arms that position the Fab-Fc hinges of RTX on a circle of approximate diameter 150 Å (Fig. 5, A and B). Because of the marked similarity with our three-dimensional structure of the CD20:RTX Fab complex, we were able to generate a model of these rings

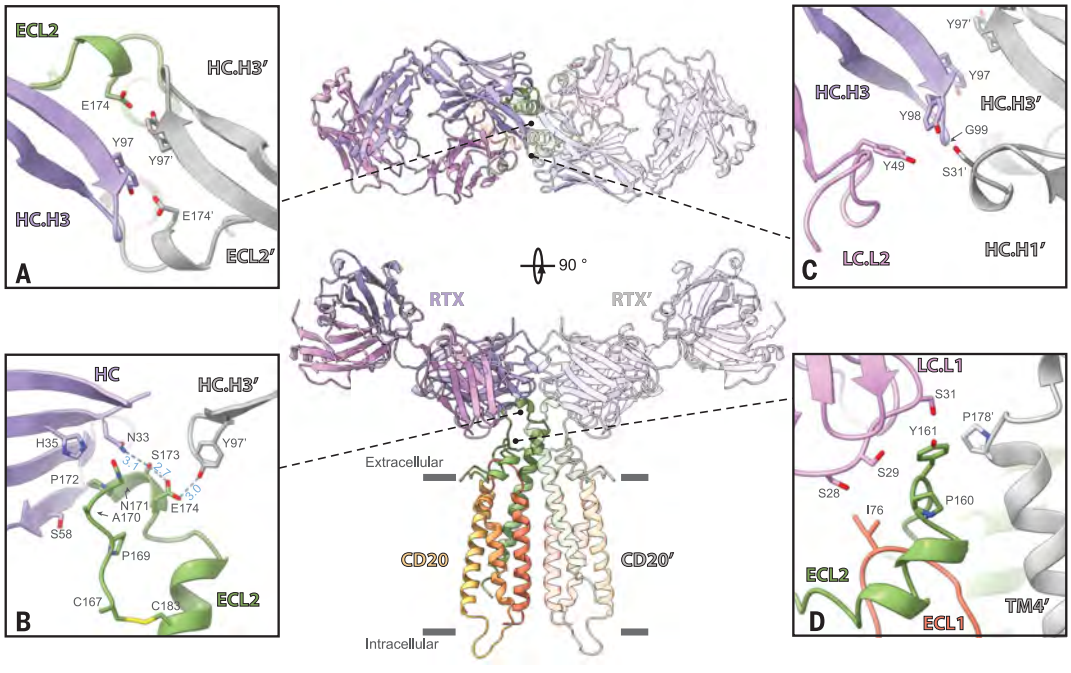


Fig. 3. Key molecular interactions between CD20 and RTX. Ribbon diagrams of the CD20:RTX Fab structure, with key amino acid side chains involved in CD20:RTX or RTX:RTX shown in stick representation. In the center diagram, the gray bars indicate the boundaries of the membrane region. (**A**) Top view of the center of the complex, where HC.Tyr^{97} mediates Fab:Fab and Fab:CD20′ contacts. (**B**) The canonical RTX epitope ^{170}ANPSE^{174}, in addition to being recognized by RTX's heavy chain (left, purple), is also involved in a hydrogen-bond network with Tyr$^{97'}$ from the distal RTX Fab (right, gray). (**C**) Additional Fab:Fab contacts between heavy-chain loops H3 and H1 and light-chain loop L2. (**D**) A secondary epitope consisting of ECL1 and ECL2 is contacted by RTX's LC loop 1.

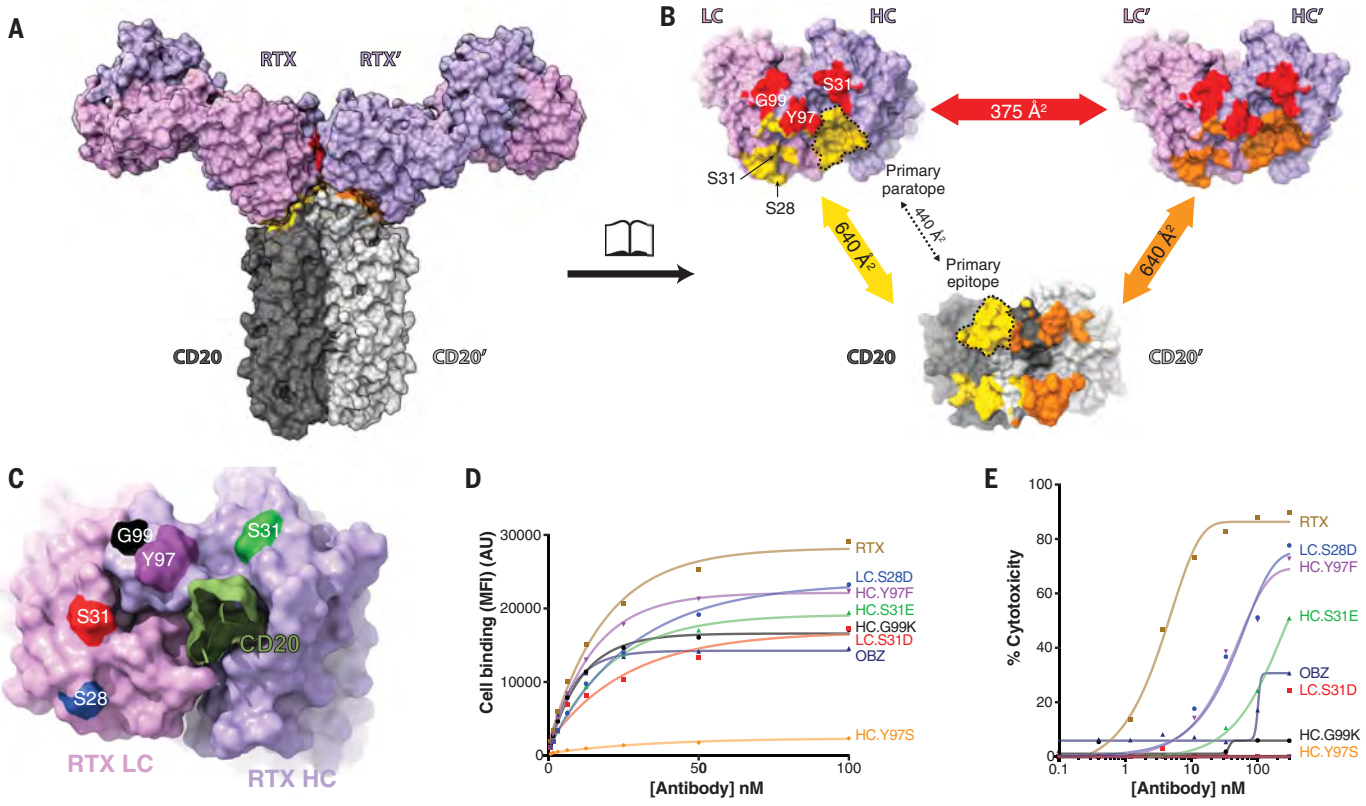


Fig. 4. Multiple CD20:RTX and RTX:RTX interactions enable cell binding and CDC. (**A**) Surface representation of the CD20:RTX Fab complex. The surfaces buried by complex formation are colored in yellow (CD20:RTX), orange (CD20′:RTX′), or red (RTX:RTX′). (**B**) Open-book representation of the same surfaces, with surface area measurements for each buried surface indicated. Residues mutated as part of this study are labeled at the top left and are not involved in the primary paratope. The primary epitope (dashed line) only accounts for less than half (440 out of 1015 Å^2) of the total Fab binding area. (**C**) Surface representation of the CDR face of RTX Fab, with the ECL2 turret epitope shown (green), as well as the positions of the point mutations under study. (**D** and **E**) Cell binding (D) and CDC (E) of RTX mutants are plotted as a function of antibody concentration and compared with wild-type RTX and OBZ. MFI, mean fluorescence intensity; AU, arbitrary units.

as present on the nsEM grids (Fig. 5, B and C). To understand how these assemblies might relate to RTX function, we endeavored to build a model of a CD20:RTX IgG assembly as it might occur on a cell. This was achieved by rotating each CD20:Fab complex 90° while keeping the ends of the Fab domains in close proximity to each other (Fig. 5D, left). This modeled assembly exhibits precisely the dimensions that would be required for the three Fc domains to "fold in" (Fig. 5D, middle) and potentially nucleate assembly of a six-membered Fc platform such as those observed in structures of the complement component C1 in complex with Fc (*25*). The resulting model of a CD20:RTX:C1 complex (Fig. 5D, right) provides a structural hypothesis for how Fab:Fab and Fab:CD20 interactions may lay the molecular foundations that promote tight CD20 clustering and complement recruitment, the hallmarks of RTX.

Discussion

CD20 is a clinically validated target for the treatment of lymphomas and autoimmune diseases, but its structure and function have remained unknown. In contrast to the prevailing view that CD20 is a tetramer (*14*), our structural studies establish CD20 as a compact dimeric double-barrel assembly, with a protein fold that is distinct from any previously determined structure. Electrostatic surface calculations confirm that the transmembrane helices of CD20 are packed predominantly through hydrophobic and van der Waals complementary interactions. Previous reports have suggested the possibility that CD20 may form a plasma membrane ion channel, but our analyses reveal no plausible ion permeation pathway through the monomeric CD20 protomer or along the dimeric packing interface. We conclude that CD20 and other MS4A family members are unlikely to directly function as ion channels.

The dimeric organization of CD20 finally provides a molecular-level explanation for the perplexing observation that twice as many type I as type II mAbs bind CD20$^+$ cells (*18*). Our EM and biophysical studies using purified components establish that each CD20 dimer is bound by two type I RTX Fabs but only one type II OBZ Fab. The two RTX Fabs are brought in close proximity owing to CD20's compact symmetrical dimeric arrangement, resulting in an extensive homotypic Fab:Fab interface, which necessitates a shallow angle of approach of the Fabs. This orientation avoids steric clashes between the two RTX Fabs, whereas OBZ's steeper angle of approach (*5*) would be expected to sterically preclude another Fab from binding.

Though it has long been known that RTX promotes CD20 clustering on the cell surface, our observation of circular RTX:CD20 assemblies with a diameter similar to that required for Fc hexamer formation (Fig. 5, A and B) raises the possibility that RTX-induced cell-surface CD20 clusters may in fact be well-ordered assemblies specifically predisposed to recruit complement (as opposed to loose groupings on the cell surface). Assemblies of this kind could be particularly efficient at complement recruitment by virtue of their biasing of Fc domain positions and orientations toward the formation of the hexameric Fc platforms necessary for complement recruitment (*25*). In

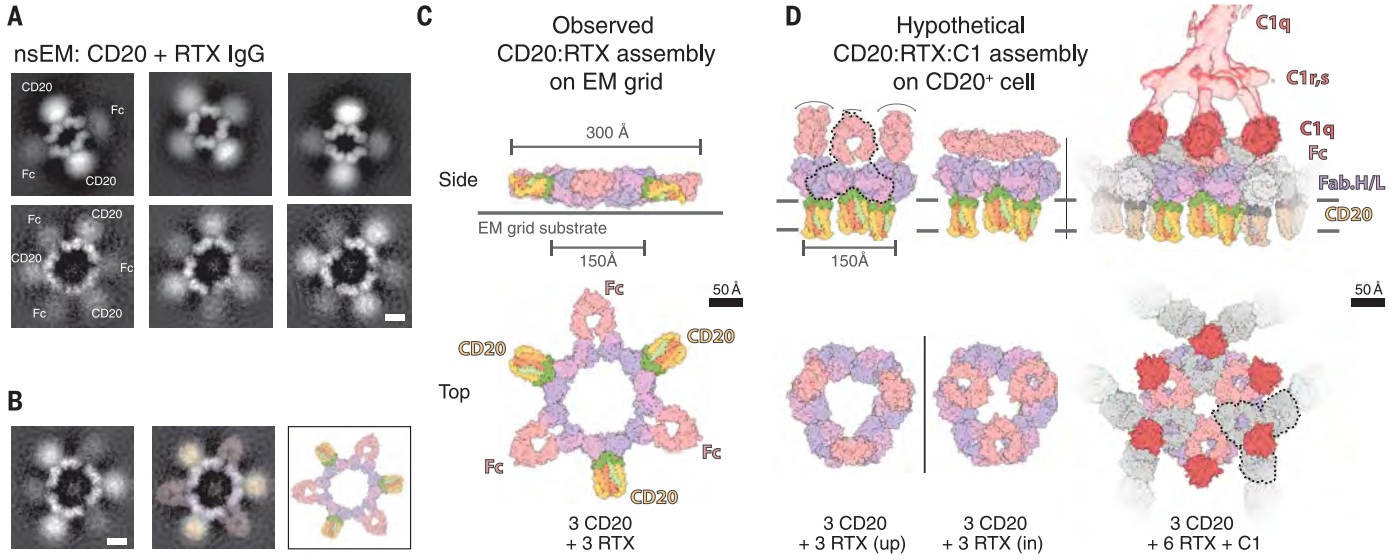


Fig. 5. RTX cross-links CD20 into circular superassemblies. (**A**) Average nsEM images of CD20 incubated with full-length RTX show cyclical higher-order structures of involving 2-to-2 (top row; diameter of 250 Å) or 3-to-3 (bottom row; diameter of 300 Å) CD20-to-RTX complexes. The RTX Fc domains appear disordered, presumably because of IgG hinge flexibility. Scale bar, 50 Å. (**B** and **C**) Interpretation of an nsEM class average of a 3-to-3 assembly. Scale bar in (B), 50 Å. (**D**) Proposed model for CD20:RTX supercomplex formation and complement recruitment. During nsEM experiments, the IgGs and solubilized CD20s are coplanar (C). Modeling these higher-order assemblies as they might occur at the surface of CD20$^+$ cells requires rotating the CD20:Fab complexes 90° [(D), left]. Given the flexibility provided by the IgG hinges, it is then possible to position Fc domains (pink) in a common plane [(D), middle]. The addition of three further Fc domains possibly contributed by neighboring CD20:IgG assemblies (gray) would complete the Fc hexamer formation and enable recruitment of C1q [(D), right]. Dashed lines outline IgG molecules. The following models were used: structure from present work (RTX Fab:CD20 complex), EMDB-4232 (EM map of C1:Fc complex), and Protein Data Bank (PDB) 6FCZ (Fc domains and C1q head domains) (*25*).

a simplistic model, we speculate that dimeric RTX:CD20 building blocks (Fig. 1C) can assemble into a 3-to-3 closed-ring configuration that acts as a nucleating scaffold for IgG hexamer formation to ultimately recruit C1q (Fig. 5D). One unknown or caveat in this model is that it requires the recruitment of three additional RTX IgG molecules that are not involved in the initial 3-to-3 ring but are needed to achieve Fc hexamerization. Once formed, it seems plausible that 3-to-3 ring assemblies could serve to potentiate Fc-Fc interactions by intermingling with various other RTX:CD20 superstructures (Fig. 5A) or assembly intermediates that are likely found and enriched along the cell surface, ultimately leading to efficient Fc-hexamer formation and C1q engagement. Although further experiments are needed to ascertain the precise dynamics and geometrical arrangements of RTX:CD20 complexes on the cell membrane when IgG hexamers are formed, our proposed model for C1q recruitment by RTX (Fig. 5D) provides an initial molecular-level hypothesis for why type I mAbs elicit potent CDC. This speculative structure-based model also suggests that CDC functionality may be shared more generally by antibodies that bind oligomeric cell-surface targets and leave at least one epitope unencumbered and available for further mAb binding.

In the case of RTX, the simultaneous binding of both CD20 subunits is made possible by an intricate geometrical arrangement involving Fab:CD20 contacts at a secondary epitope and a large Fab:Fab interaction surface. We note that all of the RTX residues involved in this homotypic interaction are germline encoded in mice (fig. S6A). This observation suggests that the homotypic Fab:Fab interaction was inherent to the progenitor RTX B cell before somatic hypermutation. These residues are also highly conserved among the RTX-like type I mAbs (fig. S6B), most of which are also mouse-derived. Similar Fab:Fab contacts mediate crystal packing of the isolated RTX Fab-ECL2 peptide complex (*4*) (fig. S5), which, taken together with our findings, indicates that Fab-Fab homotypic interactions are energetically favorable and an essential feature of the RTX-like type I mAbs. This raises the question of whether RTX Fabs may exist as preformed dimers before CD20 binding. We evaluated this possibility but found only weak Fab:Fab interactions at extremely high concentrations (>100 μM; data not shown), suggesting that the homotypic Fab:Fab interactions are nucleated by CD20 binding. We are aware of two other examples of Fab:Fab homotypic interaction, and in each case, these interfaces are central to antibody function: neutralization of a malaria parasite (*26*, *27*) and cross-linking–independent activation of TRAIL-R2 (*28*). Given the functional relevance of homotypic Fab:Fab interfaces in these three exemplar cases, we propose that these observations can be exploited in the discovery and optimization of next-generation therapeutic antibodies.

REFERENCES AND NOTES

1. M. J. E. Marshall, R. J. Stopforth, M. S. Cragg, *Front. Immunol.* **8**, 1245 (2017).
2. P. Stashenko, L. M. Nadler, R. Hardy, S. F. Schlossman, *J. Immunol.* **125**, 1678–1685 (1980).
3. L. Eon Kuek, M. Leffler, G. A. Mackay, M. D. Hulett, *Immunol. Cell Biol.* **94**, 11–23 (2016).
4. J. Du *et al.*, *J. Biol. Chem.* **282**, 15073–15080 (2007).
5. G. Niederfellner *et al.*, *Blood* **118**, 358–367 (2011).
6. J. Du *et al.*, *Mol. Immunol.* **45**, 2861–2868 (2008).
7. J. K. Bubien, L. J. Zhou, P. D. Bell, R. A. Frizzell, T. F. Tedder, *J. Cell Biol.* **121**, 1121–1132 (1993).
8. M. J. Polyak, J. P. Deans, *Blood* **99**, 3256–3262 (2002).
9. M. J. Polyak, H. Li, N. Shariat, J. P. Deans, *J. Biol. Chem.* **283**, 18545–18552 (2008).
10. T. W. Kuijpers *et al.*, *J. Clin. Invest.* **120**, 214–222 (2010).
11. C. A. Walshe *et al.*, *J. Biol. Chem.* **283**, 16971–16984 (2008).
12. D. G. Maloney *et al.*, *Blood* **90**, 2188–2195 (1997).
13. X. Montalban *et al.*, *N. Engl. J. Med.* **376**, 209–220 (2017).
14. C. Klein *et al.*, *MAbs* **5**, 22–33 (2013).
15. J. P. Deans, S. M. Robbins, M. J. Polyak, J. A. Savage, *J. Biol. Chem.* **273**, 344–348 (1998).
16. W. G. Wierda *et al.*, *J. Clin. Oncol.* **28**, 1749–1755 (2010).
17. E. Mössner *et al.*, *Blood* **115**, 4393–4402 (2010).
18. H. T. C. Chan *et al.*, *Cancer Res.* **63**, 5480–5489 (2003).
19. J. Du, H. Yang, Y. Guo, J. Ding, *Mol. Immunol.* **46**, 2419–2423 (2009).
20. B. Zimmerman *et al.*, *Cell* **167**, 1041–1051.e11 (2016).
21. S. Nakamura *et al.*, *Nat. Commun.* **10**, 816 (2019).
22. R. Dutzler, E. B. Campbell, M. Cadene, B. T. Chait, R. MacKinnon, *Nature* **415**, 287–294 (2002).
23. F. Li, J. Liu, Y. Zheng, R. M. Garavito, S. Ferguson-Miller, *Science* **347**, 555–558 (2015).

24. T. Ramaraj, T. Angel, E. A. Dratz, A. J. Jesaitis, B. Mumey, *Biochim. Biophys. Acta.* **1824**, 520–532 (2012).
25. D. Ugurlar *et al*., *Science* **359**, 794–797 (2018).
26. K. Imkeller *et al*., *Science* **360**, 1358–1362 (2018).
27. D. Oyen *et al*., *Sci. Adv.* **4**, eaau8529 (2018).
28. T. Tamada *et al*., *Sci. Rep.* **5**, 17936 (2015).

ACKNOWLEDGMENTS

Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH R01-GM129325 and P41-GM103311. Thanks to the entire BioMolecular Resources department at Genentech for support throughout the project. Thanks to the Peptide Synthesis group in the department of Early Discovery Biochemistry. We thank the Research Material Group for antibody expression as well as the Wu lab in Protein Chemistry and the Antibody Production and Automation Technologies group for antibody purification. **Funding:** T.I.C. was supported by a Principal Research Fellowship from the Wellcome Trust to R. Read (grant number 209407/Z/17/Z) and gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. All other authors are employees of Genentech, Inc. **Author contributions:** Project conceptualization: C.M.K., A.R.; CD20 construct design: C.M.K., L.R.; CD20 protein production: C.K., C.T., E.K.; Electron microscopy: A.E., C.P.A., C.C.; EM image processing: L.R., A.R.; Model building and refinement: A.R., T.I.C.; Structure analysis: A.R., J.P., J.T.K., L.R.; Antibody protein production, flow cytometry, and CDC assays: N.C., G.N.; Antibody mutant design and selection: A.R., J.T.K., G.N., L.R.; Manuscript writing: A.R., J.P., J.T.K.; SPR: M.S.; CD20 purification, EM sample preparation, BLI, differential scanning fluorimetry, and SEC-MALS: L.R. J.P., G.N., J.T.K., and A.R. are co–senior authors. A.R. supervised the project. **Competing interests:** All authors except T.I.C. are employees of Genentech, Inc. **Data and materials availability:** Materials will be made available upon request and material transfer agreement with Genentech. Accession numbers of the CD20:RTX Fab model and map are Protein Data Bank (PDB) 6VJA and EMD-21212, respectively.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1224/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S4
References (*29*–*40*)

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

29 October 2019; accepted 6 February 2020
Published online 20 February 2020
10.1126/science.aaz9356

REPORT

FUNCTIONAL AMYLOIDS

Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in *Drosophila*

Ruben Hervas[1], Michael J. Rau[2], Younshim Park[1,3], Wenjuan Zhang[4], Alexey G. Murzin[4], James A. J. Fitzpatrick[2,5,6], Sjors H. W. Scheres[4], Kausik Si[1,3]*

How long-lived memories withstand molecular turnover is a fundamental question. Aggregates of a prion-like RNA-binding protein, cytoplasmic polyadenylation element–binding (CPEB) protein, is a putative substrate of long-lasting memories. We isolated aggregated *Drosophila* CPEB, Orb2, from adult heads and determined its activity and atomic structure, at 2.6-angstrom resolution, using cryo–electron microscopy. Orb2 formed ~75-nanometer-long threefold-symmetric amyloid filaments. Filament formation transformed Orb2 from a translation repressor to an activator and "seed" for further translationally active aggregation. The 31–amino acid protofilament core adopted a cross-β unit with a single hydrophilic hairpin stabilized through interdigitated glutamine packing. Unlike the hydrophobic core of pathogenic amyloids, the hydrophilic core of Orb2 filaments suggests how some neuronal amyloids could be a stable yet regulatable substrate of memory.

A memory, once formed, is maintained in a physical state that does not require continuous presence of the original experience. The substrate of memory has been extensively studied at the neuronal network level (*1*). However, how the underlying biochemical changes, induced by initial experience, withstand degradation to maintain altered network properties—and hence, memory—remains unclear (*2*). One proposed solution is a self-renewing protein system that can recruit and modify newly made constituents as older ones are degraded (*3*). A prion-like assembly has the ability to promote the conformational conversion of other existing or newly formed monomers, creating a self-sustaining protein "conformational memory" that outlives its individual constituents (*4*). A prion-like protein with a defined role in memory in different species is the mRNA-binding cytoplasmic polyadenylation element–binding (CPEB) family of proteins (*5*–*14*). In *Drosophila melanogaster*, inhibition of Orb2 aggregation interferes with memory consolidation (*6*, *7*, *10*, *15*), facilitation of Orb2 aggregation lowers the threshold for long-term memory formation (*9*), and inactivation of Orb2 after memory formation interferes with memory recall (*9*). These observations suggest that the aggregated state of Orb2 creates a protein conformational memory.

Many prion-like proteins, prevalent among RNA-binding protein (*16*), form amyloids, which are filamentous protein aggregates composed of cross-β cores (*17*, *18*). However, in the nervous system, amyloids are commonly associated with disease (*19*). Recently, new classes of protein assemblies formed by prion-like proteins have been reported, such as the nonamyloid filaments or the gel-like state (*20*, *21*), physical states that are believed to be more amenable to regulation. These observations raise a fundamental question: What is the physical nature of the Orb2 aggregate in the brain that acts as a biochemical substrate of memory?

To characterize the function and structure of endogenous Orb2 aggregates, we purified Orb2 protein from ~3 million 3- to 7-day-old adult *D. melanogaster* heads (Fig. 1A and fig. S1A). There was no qualitative difference between Orb2 purified from synaptosomes or total head extracts, and we used whole-head Orb2 in subsequent experiments. Purified Orb2 contained monomers as well as heat- and SDS-resistant aggregates (Fig. 1B and fig. S1B). Quantitative mass spectrometry [liquid chromatography–tandem mass spectrometry (LC-MS)/MS] revealed that 97.5% of the purified protein is Orb2 (fig. S1C). Unlike in the adult nervous system, Orb2 protein is monomeric in the early embryo (*22*). Orb2 purified from 0- to 2-hour-old embryos (fig. S1C) remained monomeric, suggesting that the purification does not induce Orb2 aggregation (Fig. 1, D and E). Size-exclusion chromatography revealed monomeric (75 to 160 kDa), oligomeric (160 to 670 kDa), and aggregated (>1 MDa) Orb2 in the adult head (Fig. 1C). Monomers were not visible, and oligomers appeared as heterogeneous globules (Fig. 1F and figs. S1D and S2A), which is reminiscent of oligomers of other amyloid-forming proteins (*23*). The aggregates appeared as unbranched helical filaments with an average length of ~750 Å, width of ~100 Å, and helical crossover distance of ~550 Å (Fig. 1F and figs. S1D and S2B). JJJ2, a yeast DnaJ-domain protein, enhances Orb2 aggregation and facilitates long-term memory formation in *Drosophila* (*9*). JJJ2, but not its close relative JJJ3, converted Orb2 oligomers into filaments, suggesting that oligomers are on-pathway precursors for filaments (fig. S3).

Incubation of Orb2 monomers with traces of Orb2 filaments induced monomer aggregation in a time-dependent manner (Fig. 2, A and B), and seeded-aggregates themselves acted as a seed (Fig. 2C). Orb2 monomer alone did not aggregate in the same time period (fig. S4A), and Orb2 filaments did not cause aggregation of monomeric heterologous prion-like RNA-binding protein RBM3 (fig. S4B) (*24*). Both endogenous and seeded Orb2 filaments bound Thioflavin T, were recognized by the amyloid-specific antibody OC, and were resistant to protease treatment (fig. S4, C to E).

Orb2 binds to the 3′ untranslated regions (UTRs) of mRNAs (*25*, *26*); monomeric Orb2, which interacts with CG13928, decreases translation, whereas aggregated Orb2, which interacts with CG4612, enhances translation (Fig. 3A) (*22*). Tequila, a protease, is one of the mRNA targets of Orb2 and required for long-term memory (*26*, *27*). All purified Orb2 species bound to the 3′UTR of Tequila mRNA, and mutations in the Orb2-binding site UUUUGU to GCUUGU reduced binding of all species (fig. S5A). Gold-labeled wild type, but not mutant (M2P) mRNA, colocalized with the oligomers and filaments (Fig. 3B and fig. S5, B to D). In addition to mRNA, when incubated with recombinant binding partners (fig. S6A), Orb2 filaments bound only to GC4612. N-terminal–deleted CG4612, CG4612Δ, and monomer binding-partner CG13928 did not bind to Orb2 filaments (Fig. 3C and fig. S6, B to D). When incubated with wild-type Tequila 3′UTR mRNA and CG4612 protein, the Orb2 filaments were decorated with both mRNA and CG4612 protein (Fig. 3D and fig. S6E). In an Orb2-dependent translation assay, the addition of the Orb2 monomers reduced translation, whereas oligomers and filaments, both endogenous and seeded, enhanced translation (Fig. 3E and fig. S6F). Thus, Orb2 filament purified from fly retains aspects of its physical and functional properties.

We used cryo–electron microscopy (cryo-EM) to determine the atomic structure of Orb2 filaments, which were manually picked and used for helical reconstruction in RELION (*28*) (Fig. 4A, fig. S7, and table S1). The final reconstruction, at an overall resolution of 2.6 Å (figs. S8 and S9), showed that the Orb2

[1]Stowers Institute for Medical Research, Kansas City, MO 64110, USA. [2]Washington University Center for Cellular Imaging, Washington University School of Medicine, St. Louis, MO 63110, USA. [3]Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS 66160, USA. [4]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK. [5]Departments of Neuroscience and Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110, USA. [6]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA.
*Corresponding author. Email: ksi@stowers.org

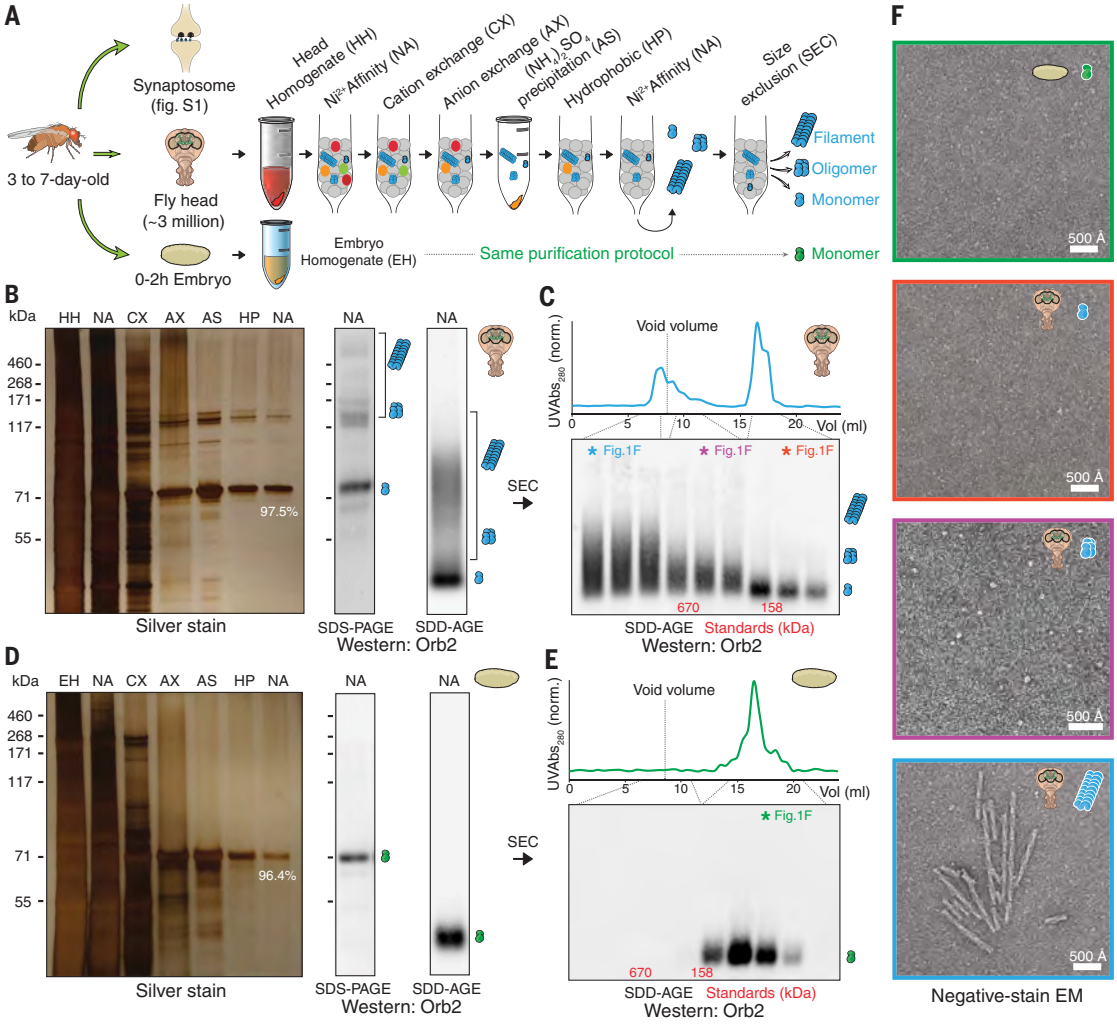


Fig. 1. Purification of different Orb2 species from adult fly head. (**A**) Schematic representation of Orb2 purification method from fly head and embryo. (**B**) Silver stain analysis after each purification step (left) and Western blotting after the final purification step (right) by means of SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and semidenaturing detergent–agarose gel electrophoresis (SDD-AGE). The percent indicates relative Orb2 purity determined by mass spectrometry. The positions of monomer, oligomer, and filaments are indicated schematically. (**C**) Western blotting of purified head Orb2 after size-exclusion chromatography. The colored asterisk fractions were visualized under negative-stain EM in (F). (**D**) Silver staining and Western blotting of Orb2 protein purified from 0- to 2-hour-old embryos. (**E**) Western blotting of purified embryonic Orb2 after size-exclusion chromatography. The colored asterisk fraction was visualized under negative-stain EM in (F). (**F**) Negative-stain electron micrographs of purified Orb2 embryonic monomer (green box), head monomer (red box), head oligomer (magenta box), and head filament (blue box).

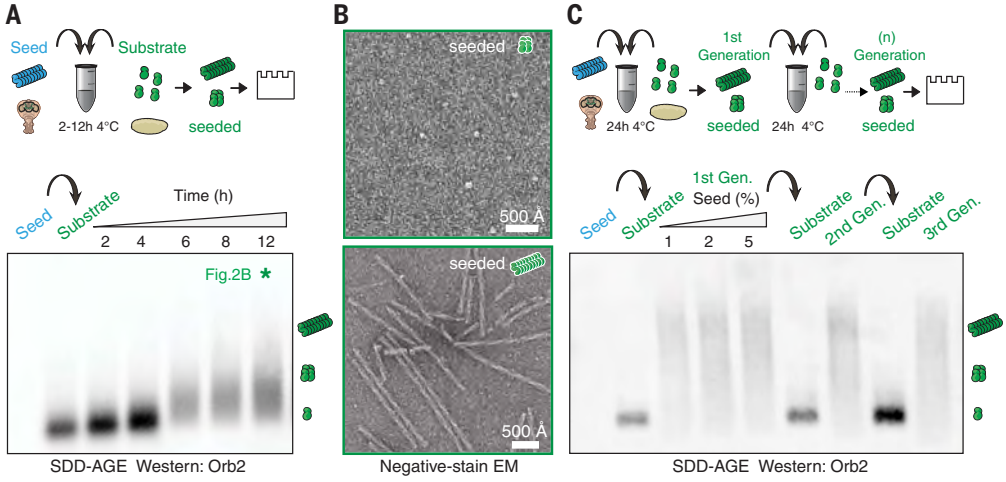


Fig. 2. Endogenous Orb2 filament can seed further filament formation. (**A**) Time course of monomeric Orb2 aggregation seeded by filaments obtained from adult fly head. The seed (~1 ng) is undetectable in western. Within 6 hours, most monomers are aggregated. (**B**) Negative-stain electron micrographs of seeded oligomers and filaments. (**C**) Propagation of seeded aggregation through successive rounds. In the first round, different seed (1, 2, and 5 ng) to monomer (100 ng) ratios were used. In subsequent rounds, only 1 ng of seed was used.

filament is composed of three identical protofilaments related by C3 symmetry with separated β-strands along the helical axis (Fig. 4, B and C). The filament core has a triangular cross section, with a side of ~80 Å and a 15-Å diameter channel along the helical axis (Fig. 4D). Successive rungs of three hairpin-shaped protofilaments form extended β-sheets along the helical axis, which is typical of the cross-β amyloid structure (*17*), with a twist of −1.55° and a rise of 4.75 Å. Surrounding the stable fibril core, lower-resolution densities are visible that

likely correspond to the RNA-recognition motif and protein interaction domain of Orb2 (Fig. 4C, gray density). We could identify densities that correspond to peptide-group oxygen atoms, ordered solvent molecules in each interface, and alternative conformations of histidine side chains (fig. S10, A to C). Consequently, we could build and stereochemically refine an atomic model de novo. The resulting model of the amyloid core comprised 31 amino acid residues from the glutamine-rich prion-like domain of Orb2 (Fig. 4, D and E). Orb2 protein has two isoforms in the fly brain (*10*) that share the prion-like domain but differ in their N-terminal extension (Fig. 4F). The low-abundant Orb2A seeds aggregation of prevalent Orb2B, which can subsequently seed Orb2B protein (*10*). Of the protein in our sample, 97.5% is Orb2B (fig. S1C), and the protofilament core structure extends from residues 176 to 206 of Orb2B (Fig. 4, D and E).

The protofilament core adopts a simple hairpin-like fold, composed of two β-strands, β1 (residues 176 to 186) and β2 (residues 197 to 206), with a wide turn (residues 187 to 196) in between (Fig. 4G). Perpendicular to the helical axis, a difference in height of 6 Å between the highest point in the turn and the lowest point in the tip of the β2 allows an ordered hydrogen-bonding network of the β-stranded regions (Fig. 4H). The cross-β packing of the two β-strands is made of a tight interdigitation of glutamine residues, Q179, Q181, Q183, and Q185 from β1 and Q200, Q202, and Q204 from β2 (Fig. 4I). A single leucine, L198, separates this block of seven glutamines from a cluster of four more interior glutamine residues in the hairpin turn: Q187, Q190, Q193, and Q196 (Fig. 4E). Next to H189, there is an

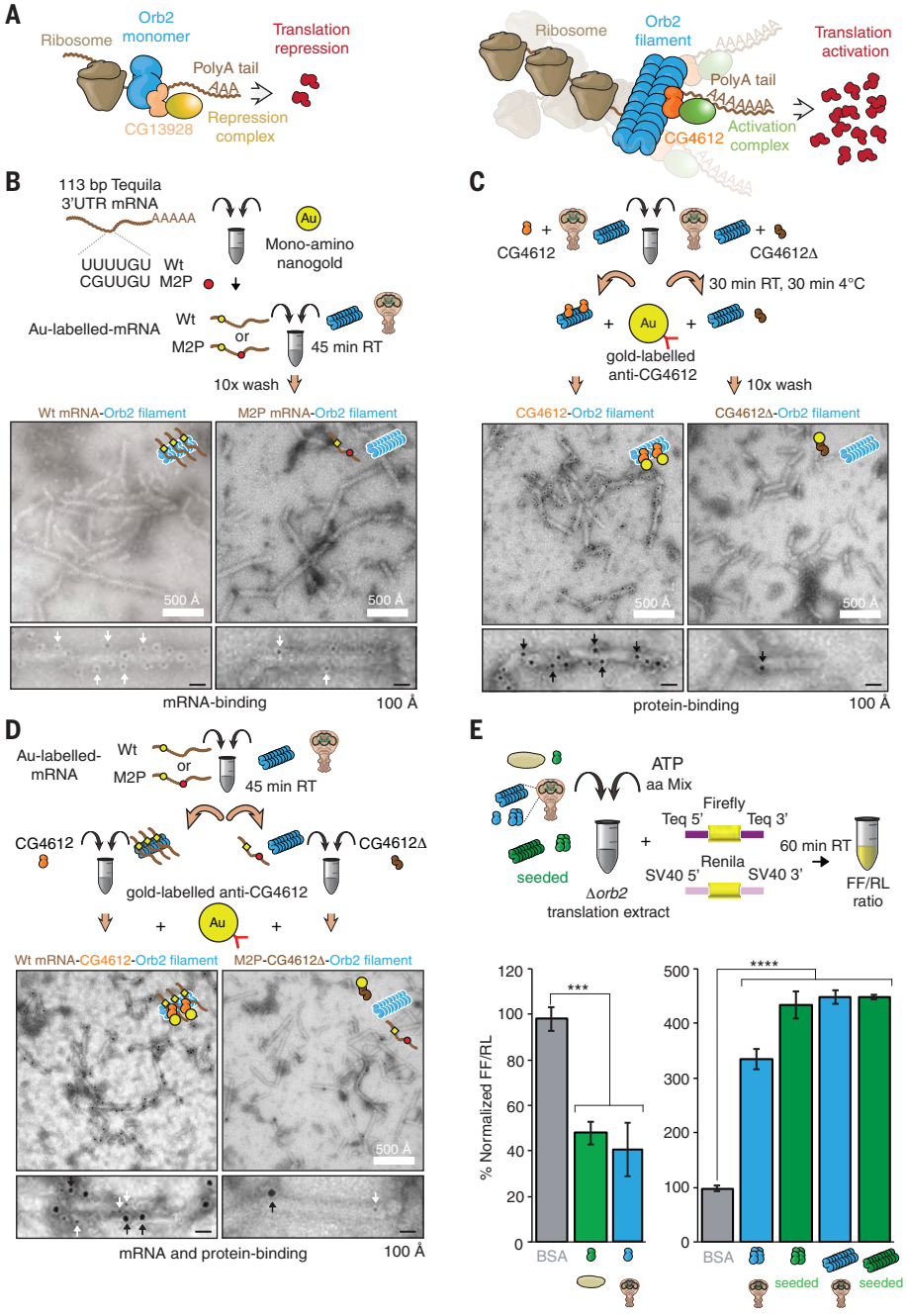


Fig. 3. Biochemical activity of Orb2 monomer, oligomer, and filament isolated from adult head. (**A**) Schematic representation of the biochemical activity of Orb2. CG13928 binds to Orb2 monomer and recruit translation repression complex (indicated in yellow). CG4612 binds to aggregated Orb2 and recruit translation promoting complex (indicated in green). (**B**) Negative-stain EM of nanogold-labeled 3′UTR of Tequila mRNA bound to Orb2 filaments. Black dots indicated with white arrows are nanogold particle (~2 nm) attached to target mRNA. (**C**) Immuno-EM of CG4612 bound to Orb2 filaments. Black dots indicated with black arrows are gold particles (~6 nm) attached to CG4612 protein. (**D**) EM of Orb2 filaments bound to nanogold-labeled 3′UTR of Tequila mRNA and CG4612 protein. The larger gold particle (~6 nm, black arrows) represents CG4612 protein, and the smaller gold particle (~2 nm, white arrows) represents the mRNA. (**E**) Translation of Orb2-target mRNA in presence of different Orb2 species, obtained from embryo, adult fly head, or seeding reaction. Purified Orb2 monomer represses translation, while Orb2 oligomer and filament enhance translation. Bovine serum albumin (BSA) was used as a control. ***P = 0.0003, ****P < 0.0001; Student's t test; two tailed. Data are expressed as mean ± SEM.

additional nonproteinaceous density that runs contiguously across the rungs (Fig. 4D, red arrows, and fig. S10D). This density is compatible with linear chains of aliphatic hydrocarbons and polyamines, which may have copurified with Orb2. The three equivalent protofilament packing interfaces are made of eight polar residues: five glutamines and three histidines. These polar residues form an extensive network of hydrogen bonds on each side and across the interface (fig. S11). Orb2 mutations, known to disrupt Orb2 aggregation (*10*), were mapped on the polar interface between protofilaments and the cross-β packing that stabilizes the protofilament (fig. S12A).

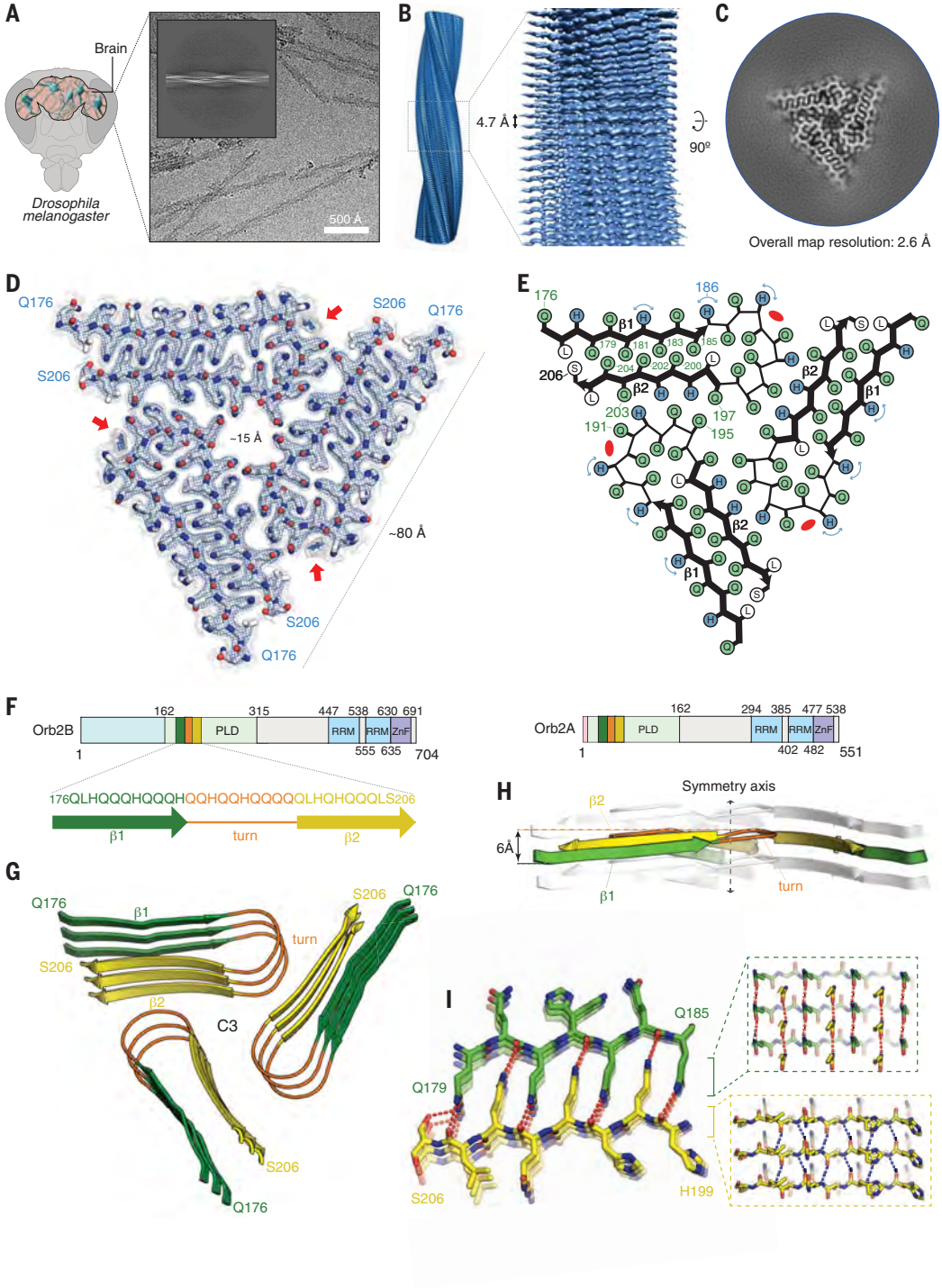


Fig. 4. Atomic structure of Orb2 filaments isolated from head. (**A**) Cryo–electron micrograph obtained at a defocus of −1.8 μm showing individual Orb2 filaments. (Inset) Representative reference-free two-dimensional class average showing an entire helical crossover. (**B**) Side view of the three-dimensional cryo-EM reconstruction illustrating the clear separation of the β-strands. (**C**) Cryo-EM reconstruction of neuronal Orb2 filament at 2.6 Å. (**D**) Sharpened, high-resolution cryo-EM maps with the corresponding atomic model overlaid. Unsharpened, 5-Å low-pass–filtered maps are shown as gray outlines. The weaker densities that border side chains of H182, H186, and H189 correspond to the alternative conformation that those residues adopt. The nonproteinaceous density that runs contiguously across the rungs is highlighted with red arrows. Density maps are shown at a contour level of 2.2σ (blue) and 1.4σ (gray). (**E**) Schematic view of the filament core showing the complementary glutamine packing within the protofilament and the hydrophilic interfaces between protofilaments. (**F**) Primary structure of Orb2A and Orb2B isoforms in fly brain. The prion-like domains are labeled PLD (residues 162 to 315 for Orb2B and residues 9 to 162 for Orb2A). The RNA-recognition motifs are labeled RRM, and the zinc-finger motifs are labeled ZnF. The N-terminal amino acids preceding the prion-like domain of Orb2A (nine residues) and Orb2B (162 residues) are represented in light red and blue, respectively. (Bottom) Schematic depicting the sequence of the Orb2 filament core, with the observed two β-strands colored in green and yellow as well as the connecting turn in orange. (**G**) Rendered view of the secondary structure elements in the Orb2 fold, depicted as three successive rungs. (**H**) Same representation as in (G), but in a view perpendicular to the helical axis, revealing the changes in height within a single molecule. (**I**) Close-up view of the interdigitated glutamine residues. The hydrogen bonds between main chains are shown in blue. The hydrogen bonds involving side chains are shown in red. [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.]

Moreover, removal of 54 amino acids from the glutamine-rich region specifically interfered with long-term, but not short-term, memory (fig. S12B) (*6*).

Several structures of functional amyloids, produced in vitro from truncated protein, have been solved (*29*, *30*). By contrast, Orb2 filament represents a biochemically active full-length functional amyloid extracted from the endogenous source (fig. S13). The threefold symmetry of the ordered core of Orb2 resembles that of β-amyloid$_{1-40}$ filaments seeded from Alzheimer's disease brain tissue (*31*) or assembled in vitro (*32*). However, unlike the hydrophobic β-amyloid$_{1-40}$ core, Orb2 forms a hydrophilic core stabilized by interdigitated glutamines. When Orb2 filaments were assembled from recombinant protein, filaments were longer, morphologically distinct, and less biochemically active and had negligible seeding capacity compared with endogenous ones under our experimental conditions (fig. S14), suggesting that the same prion-like protein can adopt distinct structures in vitro and in vivo. Other proteins with glutamine-rich sequences are known to produce amyloids (*33*). The interdigitated cross-β structure observed in Orb2 filaments could be extended on both sides of a parallel β-sheet made of only glutamine residues, which would allow for the formation of stable, multilayered cross-β structures from long polyglutamine sequences, such as in pathological glutamine expansions.

There are a number of ways that "molecular memories" can be created: increase in the amount of a protein, where the kinetics of decay to the basal state would be the duration of memory; a stable protein or protein state, the half-life of which would inform the duration of memory (*34*, *35*); or a feed-forward molecular circuit, whose activity could be reciprocally perpetuated across time. A self-sustaining amyloid formed by a single polypeptide uses all three mechanisms. The amyloid fold itself is not "memory," but the amyloid fold reorganizes the rest of the Orb2 protein to create a persistent alteration in the synthesis of specific synaptic proteins. Memory is the altered synaptic state that results from the change in functional state of the Orb2 (fig. S13). However, amyloid formation is generally assumed to be irreversible in physiological conditions; then, how can memory be dynamic? Or is it possible that some memories are irreversible and appear lost merely because of an inability to retrieve them? First, amyloids are not necessarily irreversible and could exist in a dynamic equilibrium with the available monomer (*36*). Second, the Orb2 amyloid core is based on a hydrophilic, glutamine- and histidine-rich fold, and the protonation state of histidine residues could influence Orb2 amyloid stability. Lowering pH destabilized Orb2 filaments (fig. S15), suggesting that functional amyloid could be amenable to modification or even dissolution. Our findings question the assumption that amyloid formation in the brain is always an unintended consequence that leads to dysfunction. We postulate that the brain fosters a cellular environment that is permissive to the formation of an amyloid-like state of certain proteins in order to meet the diversity of functional requirements imposed on it.

REFERENCES AND NOTES

1. S. A. Josselyn, S. Tonegawa, *Science* **367**, eaaw4325 (2020).
2. G. Lynch, M. Baudry, *Science* **224**, 1057–1063 (1984).
3. F. Crick, *Nature* **312**, 101 (1984).
4. J. Shorter, S. Lindquist, *Nat. Rev. Genet.* **6**, 435–450 (2005).
5. L. Fioriti *et al*., *Neuron* **86**, 1433–1448 (2015).
6. K. Keleman, S. Krüttner, M. Alenius, B. J. Dickson, *Nat. Neurosci.* **10**, 1587–1593 (2007).
7. S. Krüttner *et al*., *Neuron* **76**, 383–395 (2012).
8. S. Krüttner *et al*., *Cell Rep.* **11**, 1953–1965 (2015).
9. L. Li *et al*., *Curr. Biol.* **26**, 3143–3156 (2016).
10. A. Majumdar *et al*., *Cell* **148**, 515–529 (2012).
11. B. L. Raveendra *et al*., *Nat. Struct. Mol. Biol.* **20**, 495–501 (2013).
12. J. S. Stephan *et al*., *Cell Rep.* **11**, 1772–1785 (2015).
13. K. Si, Y. B. Choi, E. White-Grindley, A. Majumdar, E. R. Kandel, *Cell* **140**, 421–435 (2010).
14. K. Si, S. Lindquist, E. R. Kandel, *Cell* **115**, 879–891 (2003).
15. R. Hervás *et al*., *PLOS Biol.* **14**, e1002361 (2016).
16. V. Iglesias *et al*., *Front. Physiol.* **10**, 314 (2019).
17. R. Nelson *et al*., *Nature* **435**, 773–778 (2005).
18. R. Tycko, *Neuron* **86**, 632–645 (2015).
19. F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **75**, 333–366 (2006).
20. H. Wu, M. Fuxreiter, *Cell* **165**, 1055–1066 (2016).
21. M. P. Hughes *et al*., *Science* **359**, 698–701 (2018).
22. M. R. Khan *et al*., *Cell* **163**, 1468–1483 (2015).
23. H. A. Lashuel, D. Hartley, B. M. Petre, T. Walz, P. T. Lansbury Jr., *Nature* **418**, 291 (2002).
24. D. Peretti *et al*., *Nature* **518**, 236–239 (2015).
25. B. K. Stepien *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7030–E7038 (2016).
26. T. Mastushita-Sakai, E. White-Grindley, J. Samuelson, C. Seidel, K. Si, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11987–11992 (2010).
27. G. Didelot *et al*., *Science* **313**, 851–853 (2006).
28. J. Zivanov *et al*., *eLife* **7**, e42166 (2018).
29. C. Wasmer *et al*., *Science* **319**, 1523–1526 (2008).
30. M. Mompeán *et al*., *Cell* **173**, 1244–1253.e10 (2018).
31. J. X. Lu *et al*., *Cell* **154**, 1257–1268 (2013).
32. A. K. Paravastu, R. D. Leapman, W. M. Yau, R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18349–18354 (2008).
33. E. Scherzinger *et al*., *Cell* **90**, 549–558 (1997).
34. S. Heo *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3827–E3836 (2018).
35. R. Y. Tsien, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12456–12461 (2013).
36. W. Qiang, K. Kelley, R. Tycko, *J. Am. Chem. Soc.* **135**, 6860–6871 (2013).

ACKNOWLEDGMENTS

We thank D. Laurents, J. Oroz, W. Redwine, K. Patton, R. Halfmann, and S. Garcia Alcantara for comments; M. Miller for illustrations; C. Zhang, P. Leal, A. Machen, and A. Rodriguez Gama for experimental assistance; A. Saraf for assistance in mass spectrometry; K. Xi, F. Guo, and T. Parmely for support with electron microscopy; and G. Murshudov and R. Warshamanage for help with REFMAC. This work is dedicated to the memory of Mark T. Fisher, Ph.D. (University of Kansas Medical Center). **Funding:** This work was supported by the UK Medical Research Council (MC_UP_A025_1013, to S.H.W.S.) and Stowers Institute for Medical Research (to K.S). **Author contributions:** Conceptualization, K.S. and R.H.; investigation, R.H., M.J.R., Y.P., W.Z., A.G.M., and S.H.W.S.; resources, K.S, J.A.J.F., and S.H.W.S.; software, S.H.W.S.; supervision, K.S.; writing, original draft, K.S. and R.H.; writing–review and editing, all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Cryo-EM density map for Orb2 has been deposited in the Electron Microscopy Data Bank (EMDB) under accession no. EMD-21316. Refined atomic model has been deposited in the Protein Data Bank (PDB) under accession no. 6VPS. The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) by way of the MassIVE repository with the dataset identifier PXD016266. Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at www.stowers.org/research/publications/LIBPB-1486.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1230/suppl/DC1
Materials and Methods
Figs. S1 to S15
Table S1
References (*37*–*53*)

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

25 November 2019; accepted 18 February 2020
10.1126/science.aba3526

GLACIAL CYCLES

Persistent influence of obliquity on ice age terminations since the Middle Pleistocene transition

Petra Bajo[1,2,3], Russell N. Drysdale[1,4]*, Jon D. Woodhead[5], John C. Hellstrom[5], David Hodell[6], Patrizia Ferretti[7,8], Antje H. L. Voelker[9,10], Giovanni Zanchetta[11], Teresa Rodrigues[9,10], Eric Wolff[6], Jonathan Tyler[12], Silvia Frisia[13], Christoph Spötl[14], Anthony E. Fallick[15]

Radiometric dating of glacial terminations over the past 640,000 years suggests pacing by Earth's climatic precession, with each glacial-interglacial period spanning four or five cycles of ~20,000 years. However, the lack of firm age estimates for older Pleistocene terminations confounds attempts to test the persistence of precession forcing. We combine an Italian speleothem record anchored by a uranium-lead chronology with North Atlantic ocean data to show that the first two deglaciations of the so-called 100,000-year world are separated by two obliquity cycles, with each termination starting at the same high phase of obliquity, but at opposing phases of precession. An assessment of 11 radiometrically dated terminations spanning the past million years suggests that obliquity exerted a persistent influence on not only their initiation but also their duration.

A major challenge of testing the orbital (Milankovitch) theory of the ice ages is the uncertainty associated with the chronology of marine records. Orbital solutions are very accurate over the Pleistocene (*1*), but the age profile of deep-ocean sediments, where much of the evidence for global ice volume changes is preserved, often has large errors. Astronomical tuning of ocean records renders any test of the Milankovitch hypothesis invalid because of circular logic. Testing theories of orbital forcing ultimately requires ocean sediment records firmly anchored in absolute time.

A poorly understood feature of Pleistocene glacial-interglacial (G-IG) cycles is the change in the period of terminations—the relatively rapid switches from glacial to interglacial climate—during the Middle Pleistocene transition (MPT) 1.25 to 0.7 million years ago (Ma) (*2*–*7*). Evidence from ocean sediments shows that most terminations occurred every ~40,000 years (40 kyr) prior to the MPT but averaged ~100 kyr in the post-MPT interval (*5*). Although the precise mechanisms for this switch remain unclear (*4*–*6*), recent studies highlight the critical interval of marine isotope stages (MIS) 24–22, when major changes in ocean circulation and ice sheet dynamics occurred (*7*, *8*). This interval includes a "failed termination" at the MIS 24–23 transition, the residual ice from which probably contributed to the steplike increase in global ice volume observed over the subsequent MIS 22 glacial (the "900-ka event") (*5*, *9*). Accordingly, the interval bounded by the MIS 26–25 and 22–21 transitions—terminations XII and X (TXII and TX), respectively—is often erroneously considered to be the first "100-kyr cycle" (*7*).

The transition to the "100-kyr world" occurred without considerable shifts in astronomical parameters (*4*, *7*, *8*), implying that internal forcing changed the way the Earth system responded to orbital variations. The ~40-kyr period for pre-MPT G-IG cycles (*5*, *10*) suggests pacing by changes in Earth's axial tilt, or obliquity (*1*), which affects the degree of seasonality in a given year. At high obliquity, the polar latitudes in both hemispheres receive more summer insolation, potentially inducing substantial ice sheet ablation (*11*). The dominance of a ~100-kyr periodicity for post-MPT terminations has been linked to forcing by changes in Earth's eccentricity (*1*, *12*), but each ~100-kyr interval is more likely a cluster of climatic precession (herein, precession) and/or obliquity (*8*, *13*, *14*) cycles whose sum averages to ~100 kyr when viewed over the long term. This is supported by an Asian monsoon speleothem record spanning all terminations over the past 640 kyr (*15*), which shows a spacing of four or five precession cycles. Precisely what happened, in terms of forcing, between the MPT and TVII [~635 thousand years ago (ka)] remains unclear, yet the answer may assist in our understanding of the MPT itself.

Studies focusing on G-IG cycles that traverse the MPT (*7*, *8*, *13*) have relied on stacked records of deep-ocean benthic oxygen isotope ($\delta^{18}O$) changes (*5*, *13*), which are driven primarily by variations in global ice volume (*10*) but which also record a prominent deep-ocean temperature component (*7*). Given the inability to directly date marine sediments beyond the limits of radiocarbon dating, and given the phase uncertainties between the benthic ice volume–proxy record and astronomical (or other) tuning targets, precisely datable archives are required. We independently determined the age of terminations across the MPT by tying the radiometric chronology from a speleothem $\delta^{18}O$ time series to North Atlantic ocean sediment records. We then compared our results with astronomical and insolation parameters (*1*, *9*, *13*, *16*) for terminations since 640 ka (*15*, *17*).

Our speleothem record comes from Corchia Cave (Alpi Apuane, Italy) (*18*–*20*) and spans the interval ~970 to ~810 ka, encompassing two complete terminations (TXII and TX) and one uncompleted termination (*7*, *8*). A composite $\delta^{18}O$ time series derived from four stalagmites (CC8, CC30, CC119, and CC122) and a subaqueous speleothem (CD3) (Fig. 1) was anchored in absolute time using the U-Pb method (*18*, *20*–*22*) (figs. S1 and S2 and table S1). Almost the entire record is replicated, and concordance between both the individual stalagmite age models (fig. S3A) and the overlapping stable-isotope profiles (fig. S4A) allows all U-Pb ages to be placed onto a common depth scale to produce a composite age-depth model (*18*, *20*) (fig. S3B). After accounting for all sources of random and correlated uncertainties (*20*), the average model-age precision over the whole record is <7 kyr (95% confidence interval) (fig. S3C).

The climate at Corchia Cave has strong teleconnections with circulation changes in the North Atlantic (*19*, *23*), from where well-resolved marine records of glacial terminations have emerged (*24*, *25*). Previous studies have shown that Corchia speleothem $\delta^{18}O$ tracks changes in sea surface temperature (SST) recorded off the Iberian margin (*19*, *26*) through the effect of SST on moisture advection to, and ultimately rainfall amount above, the cave site. However, during terminations, the link between regional SST and speleothem $\delta^{18}O$ is overridden by large decreases in the $\delta^{18}O$ of surface ocean water ($\delta^{18}O_{sw}$) caused by collapse of continental ice sheets (*18*). This flux of low $\delta^{18}O_{sw}$ values introduces a "source effect" that is captured in rainfall $\delta^{18}O$ at the cave, then recorded in its speleothems (*27*). Similar to

[1]School of Geography, University of Melbourne, Carlton, Victoria 3053, Australia. [2]Australian Nuclear Science and Technology Organisation, Lucas Heights, New South Wales 2234, Australia. [3]Croatian Geological Survey, 10000 Zagreb, Croatia. [4]Laboratoire EDYTEM–UMR5204, Université de Savoie Mont Blanc, 73376 Le Bourget du Lac, France. [5]School of Earth Sciences, University of Melbourne, Parkville, Victoria 3010, Australia. [6]Godwin Laboratory for Palaeoclimate Research, Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK. [7]Istituto per la Dinamica dei Processi Ambientali, Consiglio Nazionale delle Ricerche (IDPA-CNR), Venice 30172, Italy. [8]Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari, Venice 30172, Italy. [9]Instituto Português do Mar e da Atmosfera (IPMA), Divisão de Geologia e Georecursos Marinhos, 1495-165 Alges, Portugal. [10]Centre of Marine Sciences (CCMAR), University of the Algarve, 8005-139 Faro, Portugal. [11]Department of Earth Sciences, University of Pisa, Pisa 56100, Italy. [12]Department of Earth Sciences, University of Adelaide, North Terrace, South Australia 5005, Australia. [13]School of Environmental and Life Sciences, University of Newcastle, Callaghan, New South Wales 2308, Australia. [14]Institute of Geology, University of Innsbruck, 6020 Innsbruck, Austria. [15]Scottish Universities Environmental Research Centre, East Kilbride G75 0QF, Scotland, UK.
*Corresponding author. Email: rnd@unimelb.edu.au

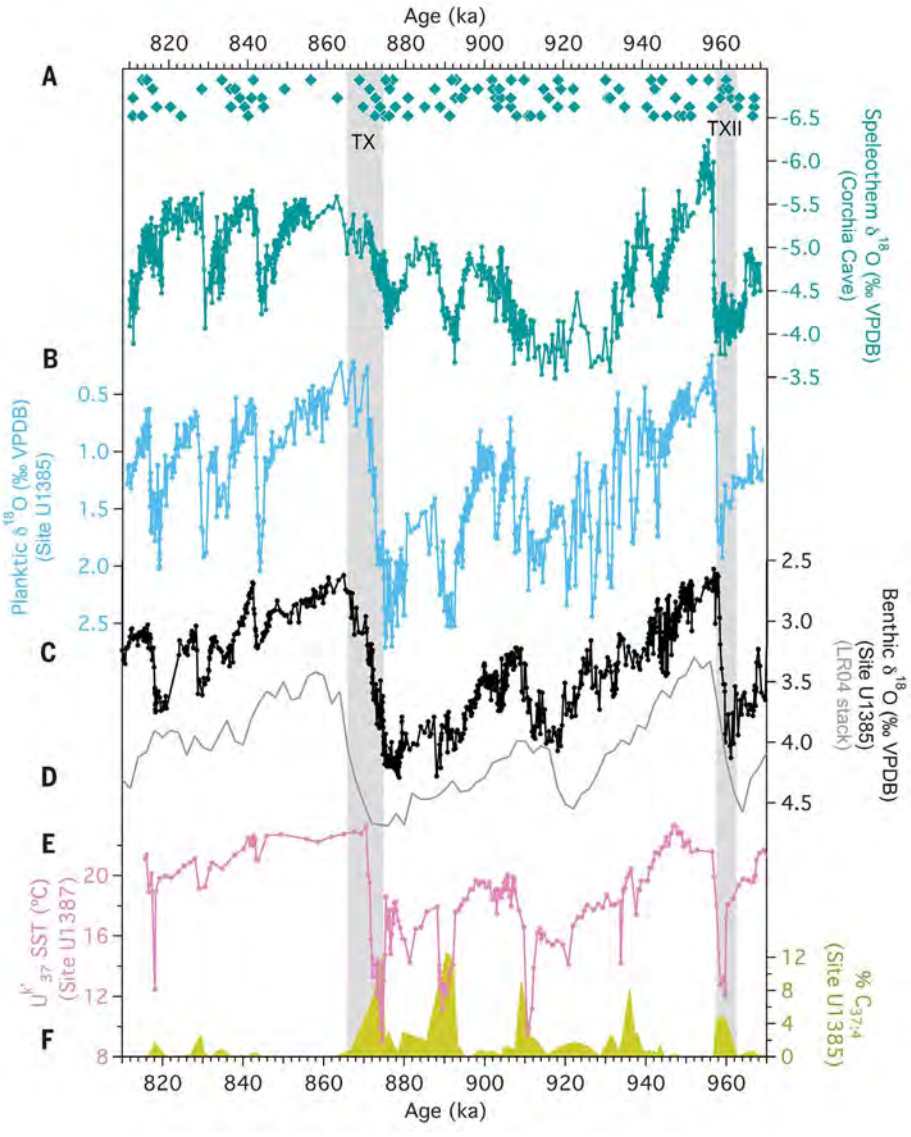


Fig. 1. North Atlantic ocean core records from IODP sites U1385 and U1387 for the period 970 to 810 ka, tuned to the Corchia Cave speleothem $\delta^{18}O$ stack. (**A**) Corchia Cave speleothem $\delta^{18}O$ stack with U-Pb ages shown (diamonds). VPDB, Vienna Pee Dee belemnite. (**B** and **C**) Planktic (blue, *Globigerina bulloides*) and benthic $\delta^{18}O$ (black, *Cibicidoides wuellerstorfi*) from site U1385. (**D**) The LR04 benthic $\delta^{18}O$ stack (gray) of global ice volume and deep-water temperature changes (*5*). (**E**) Alkenone $U^{k'}_{37}$ sea-surface temperatures from site U1387. (**F**) Percent concentration of the $C_{37:4}$ alkenone from site U1385 (*31*). The time series from sites U1385 and U1387 are plotted on the Corchia Cave speleothem U-Pb chronology (table S1 and figs. S1 to S4) using the tuning procedure outlined in the methods (*18*) and graphically presented in figs. S4B and S5. The LR04 stack is plotted on its original published age model (*5*). The two vertical gray bars highlight the position of terminations X (TX) and XII (TXII).

speleothems, the $\delta^{18}O$ of planktic foraminifera from the Iberian margin and the western Mediterranean Sea is also sensitive to changes in both SST and $\delta^{18}O_{sw}$. SST dominates the signal, except during times of large meltwater incursions, such as terminations (*23*, *24*, *27*–*29*), making the planktic $\delta^{18}O$ a robust tuning target for synchronizing the cave and ocean records (*18*). Accordingly, we tied our speleothem chronology to a newly produced, high-resolution ocean sediment record from North Atlantic Integrated Ocean Drilling Program (IODP) site U1385 (*30*) by synchronizing the planktic $\delta^{18}O$ to the Corchia $\delta^{18}O$ time series (Fig. 1, figs. S4B to S6, and table S2) (*18*). Previous cores from this drilling site (*23*, *24*) register the commencement of terminations as large decreases in benthic $\delta^{18}O$. The consistency of this pattern can be evaluated by comparing the phasing of these decreases with changes in planktic $\delta^{18}O$ and the tetraunsaturated alkenone ($C_{37:4}$) meltwater proxy from the same core (*31*), together with changes in SST at IODP site U1387, nearby in the Gulf of Cádiz (fig. S5).

The multiproxy ocean data show that the commencement of large, near-monotonic benthic $\delta^{18}O$ decreases for both terminations is approximately synchronous with rapid SST cooling and increased percent concentration of $C_{37:4}$ (Fig. 2) caused by meltwater from ice sheet collapse reaching the Iberian margin. These terminal stadial events provide unequivocal evidence for the onset of the two terminations, as is the case with younger terminations recorded at the Iberian margin (*23*, *29*). The larger percent $C_{37:4}$ value witnessed during TX relative to TXII is consistent with the concurrent planktic $\delta^{18}O$ decrease and SST cooling at the beginning of the termination, suggesting release of a larger meltwater volume (Fig. 2). This caused a prominent decoupling between SST and planktic $\delta^{18}O$, similar to that observed during TII (*23*, *27*, *29*).

Applying the Corchia chronology to both ocean records allows the onset of TXII and TX to be dated with a precision of ~0.5%, with TXII starting at 960.1 ± 4.7 ka, and TX at 875.4 ± 4.7 ka (Fig. 2). The corresponding LR04 benthic stack (*5*) onset ages for TXII and TX suggest an intervening interval of 92 kyr (Fig. 1D). Our newly generated chronology yields a somewhat shorter interval of ~85 ± 7 kyr (Fig. 2), constituting the first radiometric evidence that the period between TXII and TX represents a single G-IG spanning about two obliquity and four precession cycles. The chronology also reveals that both terminations started at similar phases of high obliquity, whereas the corresponding precession phases were almost diametrically opposed (Fig. 2). Furthermore, the two terminations were completed at different rates (Fig. 1C). At TXII, ice sheet collapse was initiated when obliquity and precession approached maximum values

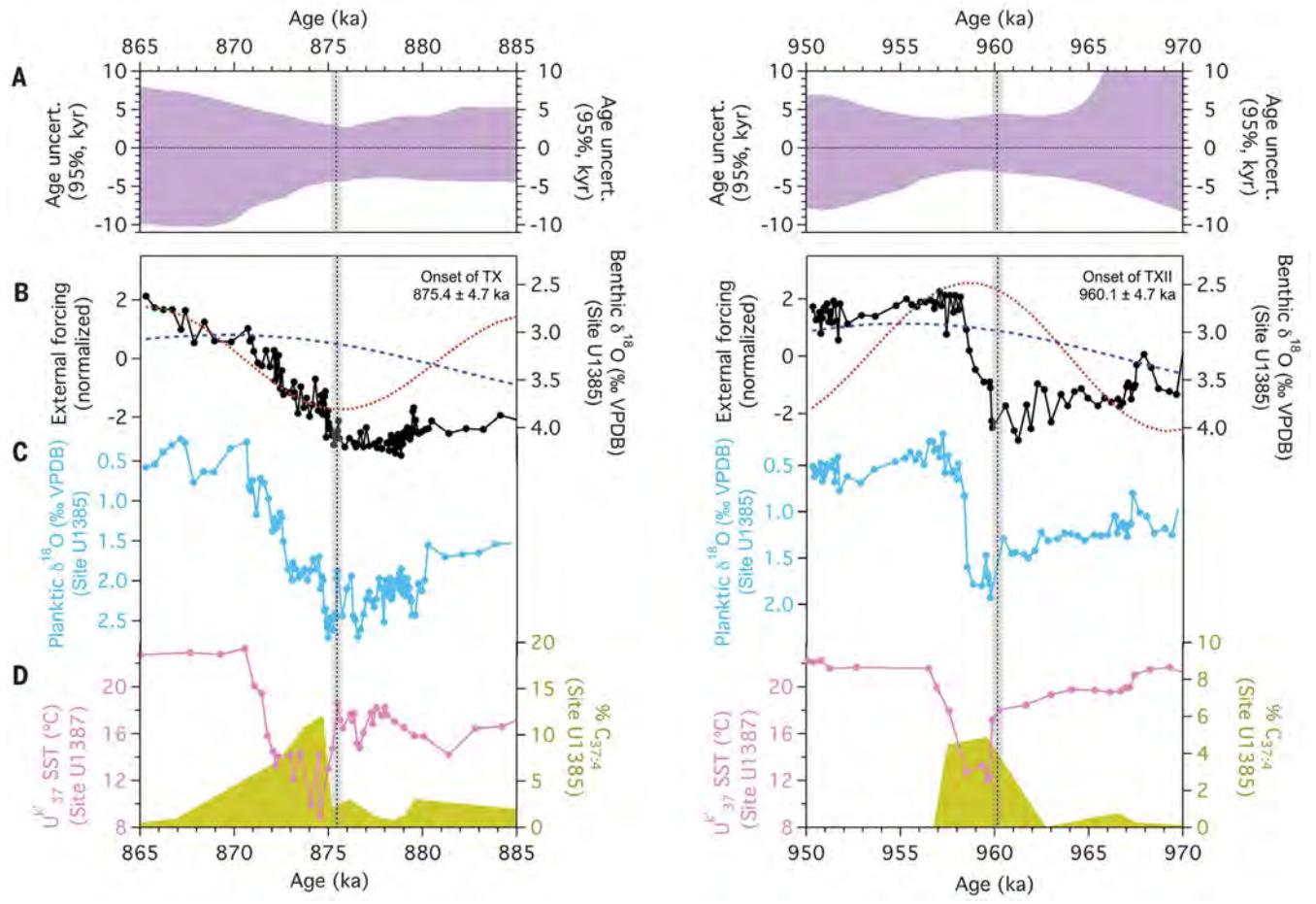


Fig. 2. Detail of ocean changes through terminations X (left) and XII (right). (**A**) Time series of variations in age uncertainty for the Corchia Cave U-Pb chronology underpinning the ocean records from IODP sites U1385 and U1387. (**B** and **C**) Benthic (*C. wuellerstorfi*) and planktic $\delta^{18}O$ (*G. bulloides*) representing deep-ocean and surface-ocean temperature and sea water $\delta^{18}O$ changes at IODP site U1385. Normalized obliquity (blue dashed curve) and climatic precession (brown dotted curve) are also shown (*1*). The precession index of (*1*) is multiplied by −1 so that peak scores—referred to here as "maximum precession"—represent points when perihelion aligns with the Northern Hemisphere summer solstice. (**D**) Alkenone $U^{k'}_{37}$ sea-surface temperatures (SST) from IODP site U1387 and percent concentration of the $C_{37:4}$ alkenone from site U1385 (*31*), a proxy for freshwater incursions at the coring site. The black vertical dashed line and gray vertical bands mark the position of the commencement of each termination (age and 95% uncertainty envelope, respectively). Termination ages are shown; the ± values incorporate uncertainties from both the age modeling and a conservative estimate of the component attributable to the synchronization procedure (*18*).

(*18*), resulting in strong Northern Hemisphere (NH) summer insolation and a very rapid termination, whereas the more prolonged TX started at near-minimum precession but was completed at near-maximum obliquity and precession (Fig. 2). These observations suggest that insolation changes more closely associated with obliquity than with precession initiated the two terminations, while summer insolation status at initiation controlled termination duration.

Next, we explore whether these relationships hold for TVII to TI, for which previous assessments favor precession over obliquity (*15*, *17*). Estimates for their timing can be determined using a principle similar to our approach for TXII and TX. The precisely dated Chinese speleothems, to which the younger terminations are anchored, register perturbations to the Asian monsoon at the onset of a terminal stadial event (*15*, *17*), enabling the start of each termination to be tied to a radiometric chronology (table S3 and fig. S7) (*18*). Our analysis of all 11 radiometrically constrained terminations shows that the phasing of precession and obliquity at the start of TXII and TX falls within the range of values for post-MPT terminations (Fig. 3A, right panel). However, there is a clear obliquity phase lead of at least ~30° (table S4) (*18*) for 8 of the 11 terminations (Fig. 3, B and C). Seven terminations began when integrated summer energy >275 W/m^2 at 65°N (predominantly obliquity-driven) (*16*) was above average (Fig. 3D), whereas NH summer insolation intensity at 65°N (predominantly precession-driven) was below average in eight cases (Fig. 3E). A similar finding emerges for the termination midpoints, the classical metric for quantifying termination pacing (*5*): These midpoints are, overall, positioned at closer proximity to maximum integrated summer energy values (Fig. 3D) than to maximum NH summer insolation intensity values (Fig. 3E). We also find that the interval between each termination midpoint is a multiple of both precession (23 ± 2 kyr) and obliquity (~41 ± 7 kyr) periods (table S4) (*18*). Finally, terminations never commence in a precession cycle that does not align with the rising limb or peak of an obliquity cycle (Fig. 3, B and C). In light of this evidence, a predominance of precession over obliquity seems unlikely in the pacing of post-MPT terminations (*15*, *17*). Obliquity has clearly played an equal, if not greater, role in their timing.

We also determined the age at the end of each termination to calculate the time it took

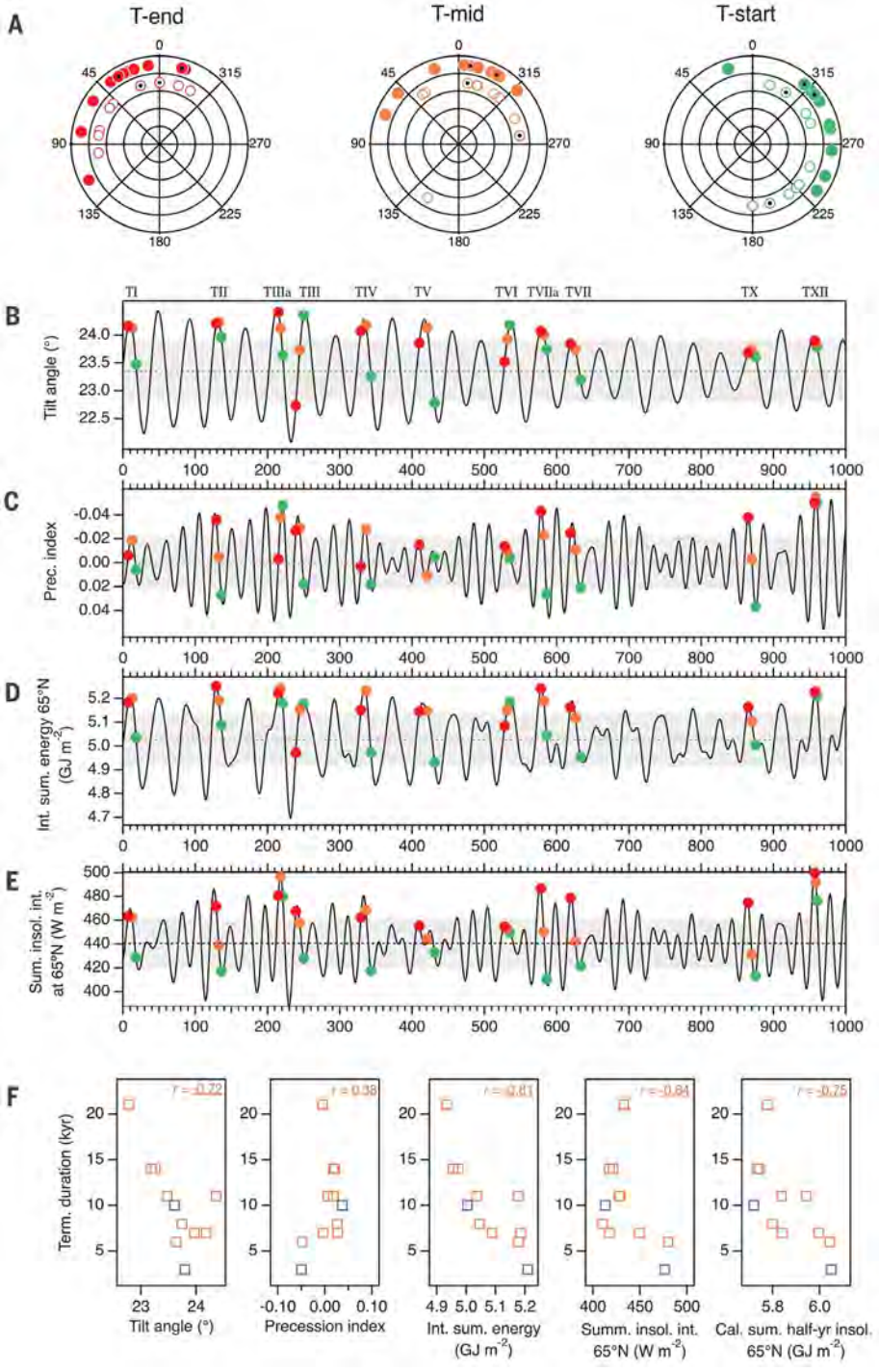


Fig. 3. Radiometric-based timing and duration of 11 terminations (TI to TVII, TX, and TXII) compared with astronomical and insolation parameters. (**A**) Polar plots showing phasing in degrees for both obliquity (solid circles) and precession (open circles) at the start (green), midpoint (orange), and end (red) of each termination. The black dots in each series highlight the phasings for TX and TXII. (**B** to **E**) Phasing between the timing of the start (green), midpoint (orange), and end (red) of each termination and (B) obliquity (*1*), (C) climatic precession (*1*), (D) integrated summer energy (>275 W/m^2) at 65°N (*16*), and (E) insolation intensity for July at 65°N (*1*). (**F**) Scatterplots and Pearson *r* correlation coefficients for the duration of 11 terminations (*18*) (tables S3 and S4) versus orbital and insolation metrics at the start of each termination [left to right: obliquity (*1*), climatic precession (*1*), integrated summer energy (>275 W/m^2) at 65°N (*16*), July insolation intensity at 65°N (*1*), and caloric summer half-year energy at 65°N (*9*)]. The blue squares highlight the data for TX and TXII. Underlined *r* values are statistically significant ($P < 0.05$; degrees of freedom = 9).

for each one to run to completion (*18*). We find that duration is significantly correlated with caloric summer half-year energy (approximately equal contributions from obliquity and precession), integrated summer energy, and NH summer insolation intensity (all at 65°N) (Fig. 3F) at the commencement of a termination; the correlation with the precession index is much weaker but remains significant for tilt. This reinforces the strong role of obliquity in post-MPT terminations.

The radiometrically constrained ensemble of 11 terminations allows us to evaluate the findings of a recent study implicating a combination of obliquity and precession in controlling termination timing over the past 1 million years (*13*). In this study, an approximate age of each termination midpoint was estimated using the rate of change in benthic $\delta^{18}O$ on the basis of a stack of benthic $\delta^{18}O$ records tied to a depth-derived (rather than orbitally tuned) age model (*13*, *14*). This study suggested that the ~100-kyr G-IG spacing consists of clusters of two (80-kyr) or three (~120-kyr) tilt cycles (*13*, *14*), with the interval between each termination controlled by obliquity, but the exact timing within a given cycle occurring when Earth is at perihelion during the NH summer solstice (i.e., maximum precession) (*13*). Our results show that the spacing of termination midpoints is consistent with obliquity forcing (Figs. 3B and 4A), and that the midpoints are most consistently aligned

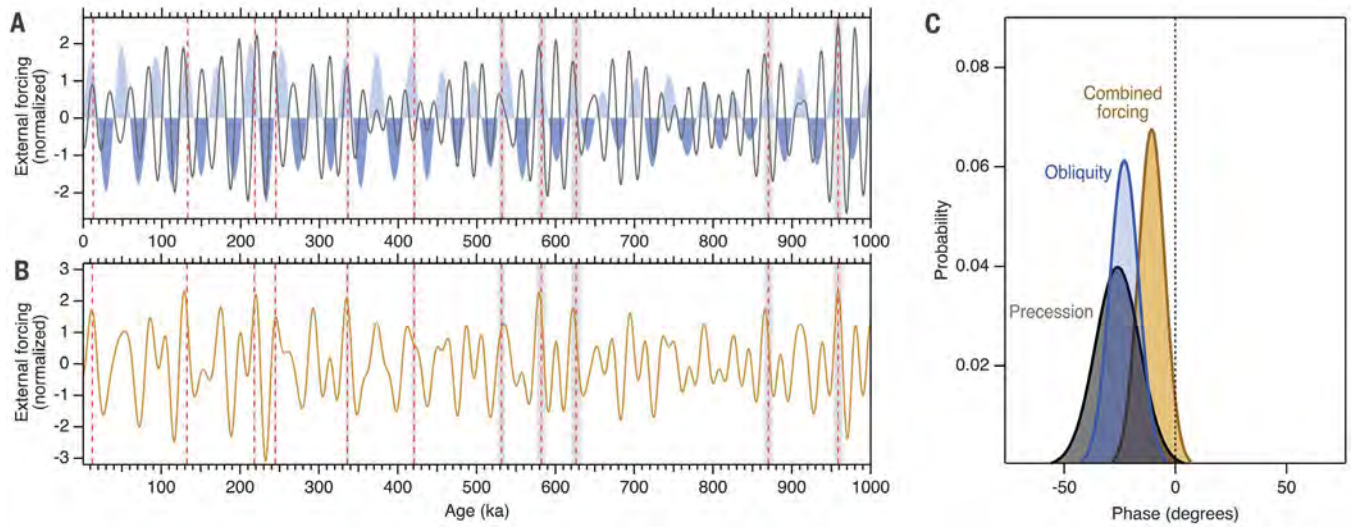


Fig. 4. Comparison of the timing of 11 termination midpoints and normalized orbital and insolation metrics. (**A**) Termination timing (red vertical dashed lines) versus obliquity (light and dark blue shading) and climatic precession (dark gray curve) (*1*). Precession is multiplied by −1, as in Fig. 2. Gray vertical bands are the 95% uncertainties of the midpoint-age estimates, which for the younger terminations (*18*) are small compared with the line thickness. (**B**) Termination timing, as in (A), versus an insolation forcing metric that combines both obliquity and climatic precession variability (*13*, *18*). (**C**) Phase probability distributions for climatic precession, obliquity, and the combined precession-obliquity insolation forcing metric of (*13*) [see also (*18*)]. Each distribution is an error-weighted phase mean and uncertainty based on the phase and uncertainty of 11 individual terminations at their midpoint age (fig. S8, A, C, and E). Individual phase uncertainties were derived using the 95% uncertainties of the midpoint ages (*18*). The vertical zero line represents the phase maximum for each parameter. A negative phase represents termination ages that precede the maximum phase of the orbital parameter.

with peaks in an insolation forcing metric (almost identical to caloric summer half-year insolation at 65°N), which integrates approximately equal amounts of obliquity and precession (Fig. 4, A to C) (*13*, *18*).

Newly determined radiometric ages for TXII and TX coupled with a reassessment of well-dated younger terminations (TVII to TI) suggest that obliquity pacing of G-IG cycles continued beyond the 40-kyr world. A termination onset was more likely to occur at a higher phase of obliquity than precession. Once ice sheet collapse was initiated, insolation changes driven by both precession and obliquity propelled the climate toward full interglacial conditions, but at a rate dependent upon the prevailing levels of predominantly obliquity-controlled summer energy. The results presented here suggest that the term "100-kyr world" is both inaccurate and misleading, and that its usage should probably be discontinued.

REFERENCES AND NOTES

1. A. Berger, M. F. Loutre, *Quat. Sci. Rev.* **10**, 297–317 (1991).
2. J. D. Hays, J. Imbrie, N. J. Shackleton, *Science* **194**, 1121–1132 (1976).
3. N. G. Pisias, T. C. Moore Jr., *Earth Planet. Sci. Lett.* **52**, 450–458 (1981).
4. P. U. Clark *et al*., *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
5. L. E. Lisiecki, M. E. Raymo, *Paleoceanogr. Paleoclimatol.* **20**, PA1003 (2005).
6. M. J. Head, P. L. Gibbard, *Quat. Int.* **389**, 7–46 (2015).
7. H. Elderfield *et al*., *Science* **337**, 704–709 (2012).
8. L. D. Pena, S. L. Goldstein, *Science* **345**, 318–322 (2014).
9. P. C. Tzedakis, M. Crucifix, T. Mitsui, E. W. Wolff, *Nature* **542**, 427–432 (2017).
10. N. J. Shackleton, N. D. Opdyke, *Mem. Geol. Soc. Am.* **145**, 449–464 (1976).
11. P. Huybers, C. Wunsch, *Nature* **434**, 491–494 (2005).
12. L. E. Lisiecki, *Nat. Geosci.* **3**, 349–352 (2010).
13. P. Huybers, *Nature* **480**, 229–232 (2011).
14. P. Huybers, *Quat. Sci. Rev.* **26**, 37–55 (2007).
15. H. Cheng *et al*., *Nature* **534**, 640–646 (2016).
16. P. Huybers, *Science* **313**, 508–511 (2006).
17. H. Cheng *et al*., *Science* **326**, 248–252 (2009).
18. See supplementary materials.
19. R. N. Drysdale *et al*., *Earth Planet. Sci. Lett.* **227**, 215–229 (2004).
20. P. Bajo, R. Drysdale, J. Woodhead, J. Hellstrom, G. Zanchetta, *Quat. Geochronol.* **14**, 5–17 (2012).
21. J. Woodhead *et al*., *Quat. Geochronol.* **1**, 208–221 (2006).
22. J. Woodhead *et al*., *Quat. Geochronol.* **14**, 105–113 (2012).
23. P. C. Tzedakis *et al*., *Nat. Commun.* **9**, 4235 (2018).
24. B. Martrat *et al*., *Science* **317**, 502–507 (2007).
25. I. Hernández-Almeida, F. J. Sierro, I. Cacho, J. A. Flores, *Paleoceanogr. Paleoclimatol.* **27**, PA3214 (2012).
26. R. N. Drysdale *et al*., *Science* **325**, 1527–1531 (2009).
27. G. Marino *et al*., *Nature* **522**, 197–201 (2015).
28. B. Martrat *et al*., *Science* **306**, 1762–1765 (2004).
29. B. Martrat, P. Jimenez-Amat, R. Zahn, J. O. Grimalt, *Quat. Sci. Rev.* **99**, 122–134 (2014).
30. D. Hodell *et al*., *Global Planet. Change* **133**, 49–64 (2015).
31. T. Rodrigues *et al*., *Quat. Sci. Rev.* **172**, 118–130 (2017).

ACKNOWLEDGMENTS

Funding: We acknowledge financial support from Australian Research Council Discovery Project grants 0664621 (to J.D.W.), 110102185 (to R.N.D., J.D.W., J.C.H., E.W., A.E.F., and S.F.), and 160102969 (to R.N.D., J.D.W., G.Z., E.W., and P.F.). We thank the Gruppo Speleologico Lucchese and the Federazione Speleologica Toscana for logistic and funding support. The SUERC contribution to this study falls within the framework of the Scottish Alliance for Geoscience, Environment and Society (SAGES). P.B. was the recipient of a University of Melbourne International Postgraduate Research Scholarship and Postgraduate Writing-Up Award supported by the Albert Shimmins Fund. J.C.H. was the recipient of an Australian Research Council Future Fellowship (FT130100801). P.F. acknowledges support from the European Union through a Marie-Curie Reintegration grant (PERG-GA-2010-272134 - MILLEVARIABILI). D.H. acknowledges support from the UK Natural Environmental Research Council. E.W. is supported by a Royal Society Professorship. A.H.L.V. and T.R. received financial support from Fundação para a Ciência e a Tecnologia (FCT, Portugal) projects MOWCADYN (PTDC/MAR-PRO/3761/2012), WarmWorld (PTDC/CTA-GEO/29897/2017), and CCMAR (UID/Multi/04326/2019). Samples from the marine sites were provided by the Integrated Ocean Drilling Program to D.H., A.H.L.V., and T.R. **Author contributions:** R.N.D., J.D.W., J.C.H., G.Z., and P.B. initiated the study. P.B. and J.D.W. performed the U-Pb analyses; P.B. and J.C.H., the $^{234}U/^{238}U$ analyses; P.B., J.C.H., and J.T., the age-depth modeling; and P.B., R.N.D., C.S., and A.E.F., the stable isotope measurements. D.H., A.H.L.V., and T.R. provided the marine core data. P.B. and R.N.D. performed the speleothem-ocean synchronization, which was scrutinized by D.H., A.H.L.V., T.R., P.F., and E.W. R.N.D. performed the analysis of the termination data for TI to TVII. S.F. provided the petrographic interpretations. All the authors contributed to the interpretation of the results. R.N.D. and P.B. wrote the manuscript, with all authors contributing to reviewing and editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data produced and used in this study are available from the World Data Center PANGAEA online repository at www.pangaea.de. The computer code for the finite growth rate depth-age model is available upon request from J.C.H. (j.hellstrom@unimelb.edu.au) and will be published in full in a future publication.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1235/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S4
References (*32*–*67*)

20 November 2018; accepted 11 February 2020
10.1126/science.aaw1114

STRUCTURAL BIOLOGY

Structure of V-ATPase from the mammalian brain

Yazan M. Abbas[1], Di Wu[2], Stephanie A. Bueler[1], Carol V. Robinson[2], John L. Rubinstein[1,3,4]*

In neurons, the loading of neurotransmitters into synaptic vesicles uses energy from proton-pumping vesicular- or vacuolar-type adenosine triphosphatases (V-ATPases). These membrane protein complexes possess numerous subunit isoforms, which complicates their analysis. We isolated homogeneous rat brain V-ATPase through its interaction with SidK, a *Legionella pneumophila* effector protein. Cryo–electron microscopy allowed the construction of an atomic model, defining the enzyme's ATP:proton ratio as 3:10 and revealing a homolog of yeast subunit f in the membrane region, which we tentatively identify as RNAseK. The c ring encloses the transmembrane anchors for cleaved ATP6AP1/Ac45 and ATP6AP2/PRR, the latter of which is the (pro)renin receptor that, in other contexts, is involved in both Wnt signaling and the renin-angiotensin system that regulates blood pressure. This structure shows how ATP6AP1/Ac45 and ATP6AP2/PRR enable assembly of the enzyme's catalytic and membrane regions.

Vesicular- or vacuolar-type adenosine triphosphatases (V-ATPases) are ATP-hydrolysis–driven proton pumps that are essential for acidification of endosomes, lysosomes, and the trans Golgi network, as well as for acid secretion by osteoclasts, kidney intercalated cells, and some tumor cells (*1*, *2*). ATP hydrolysis in the V-ATPase catalytic V_1 region drives rotation of a central rotor subcomplex and leads to proton translocation through the membrane-embedded V_O region. V-ATPase activity is regulated by reversible separation of the V_1 and V_O regions, which inhibits ATP hydrolysis in the isolated V_1 complex and makes the V_O complex impermeable to protons (*3*, *4*). In neurons, V-ATPase activity energizes synaptic vesicle membranes, allowing transporters to load the vesicles with neurotransmitters (Fig. 1A) (*1*, *2*). Fusion of synaptic vesicles with the presynaptic membrane requires separation of the V_1 and V_O regions, but it is not known how these events are coordinated (*5*). The regulated release of neurotransmitters from synaptic vesicles into the synaptic cleft allows signal propagation from the axon terminal of a presynaptic neuron to the dendrite of a postsynaptic neuron. After the release of neurotransmitters, subsequent endocytosis and regeneration of synaptic vesicles occurs via clathrin-independent and clathrin-mediated routes (*6*), with the formation of clathrin-coated vesicles temporarily blocking V-ATPase activity (*7*). In the *Saccharomyces cerevisiae* enzyme, which is the most thoroughly characterized enzyme to date, the V_1 region contains subunits A_3, B_3, C, D, E_3, F, G_3, and H, and the V_O region contains subunits a, c_8, c′, c″, d, e, f, and Voa1p (*8*, *9*). The V_1 regions of mammalian V-ATPases contain the same subunits as the yeast enzyme, whereas mammalian V_O regions are thought to be composed of subunits a, c_x, c″, d, and e as well as ATP6AP1, also known as Ac45, and ATP6AP2, also known as the (pro)renin receptor (*10*, *11*). ATP6AP2/PRR is involved in several signaling pathways (*12*), including the renin-angiotensin system for regulating blood pressure and electrolyte balance (*10*, *11*, *13*) and Wnt signaling in stem cells and embryo development (*14*). The precise arrangement of subunits in the mammalian V_O region remains unclear. Further, mammals have multiple isoforms of some subunits in both V_1 and V_O that are expressed in a tissue-dependent and cellular-compartment–dependent way. These include two isoforms of subunit B, two of C, two of E, three of G, four of a, two of d, and two of e (*15*, *16*). Mass spectrometry of purified rat-brain synaptic vesicles detected V-ATPase subunits A, B1, B2, C1, D, E1, F, G1, G2, a1, a4, c, d1, ATP6AP1/Ac45, and ATP6AP2/PRR (*17*).

To isolate V-ATPase for structural analysis, we developed a purification strategy based on the high-affinity interaction of the enzyme with the *Legionella pneumophila* effector protein SidK (*18*, *19*). Although procedures capable of obtaining highly purified synaptic vesicles have been described previously (*17*, *20*), these procedures enhance purity at the cost of yield, which complicates or precludes structure determination. Instead, we used a two-step differential centrifugation procedure to collect rat synaptic vesicles and clathrin-coated vesicles along with other cell membranes. Membranes were solubilized with detergent and SidK (residues 1 to 278) fused to a C-terminal 3×FLAG tag was used with M2-agarose to purify V-ATPase. With this approach, a single ~2-g rat brain provided ~200 μg of highly purified V-ATPase (Fig. 1B). Mass spectrometry of tryptic fragments (fig. S1A and tables S1 and S2) identified the bands on the SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gel as V-ATPase subunits A, B2, C1, D, E1, F, G2, a1, c, and d1. A low-intensity band on the gel corresponding to subunit G1 was also detected, which, from gel densitometry, appears to constitute 17 ± 3% of subunit G in the preparation (±SD, three independent purifications). Subunit G1 may be a component of some V-ATPase complexes in synaptic vesicles (*17*). Alternatively, subunit G1 may be from copurifying lysosomal V-ATPase (*21*). The glycoprotein ATP6AP1/Ac45 and the small and hydrophobic proteins ATP6AP2/PRR and subunit e2, which do not stain clearly with Coomassie, were detected after cleavage with trypsin or chymotrypsin (Fig. 1B, fig. S1A, and table S3). The subunit isoforms identified are all consistent with the V-ATPase from synaptic vesicles (*17*). This homogeneity could be because of the synaptic vesicle V-ATPase being the predominant form of the enzyme in the brain or because of V-ATPases from other cellular compartments in the brain having the same isoform composition as the synaptic vesicle enzyme. RNAseK, a hydrophobic protein recently found to associate with mammalian V-ATPase (*22*), was also detected after cleavage with trypsin or chymotrypsin (Fig. 1B, fig. S1A, and table S3). Subunit H, which dissociates from bovine brain V-ATPase when treated with oxidizing agents (*23*), was not detected by mass spectrometry, despite the absence of these reagents from the preparation. Subunit H is part of the mammalian V-ATPase (*7*, *17*) and is needed for full enzyme activity (*24*). Loss of subunit H during purification of mammalian V-ATPase is markedly different from the subunit's behavior in *S. cerevisiae* (*25*). This difference is noteworthy because the protein's physiological role of mechanically blocking ATP hydrolysis in the isolated V_1 region likely requires a strong attachment to the enzyme (*26*, *27*). Although tryptic peptides from subunits B1, C2, a2, a3, and a4 were also detected (table S4), integrated peak intensities for these peptides in extracted ion chromatograms were two to three orders of magnitude lower than for peptides from subunits B2, C1, and a1, which indicates that their abundance is negligible for structural studies.

Native mass spectrometry further demonstrated the homogeneity of the enzyme preparation (Fig. 1C, red, and fig. S1, B and C). Spectra show the V_1 region, presumably because of dissociation of the complex during analysis. The V_1 region has a native mass of 683369 ± 144 Da, consistent with a subunit composition of A_3, $B2_3$, C1, D, $E1_3$, F, $G2_3$, and $SidK_3$ (Fig. 1C, fig. S1B, and table S5). Fragmentation of the V_1 region using a higher-energy collisional dissociation (HCD) voltage of 50 to 250 V confirmed the isoform composition of the V_1 region as B2, C1, and E1 (fig. S1B). Specifically, spectra for subunit G showed a native mass of 13578 ± 1 Da (Fig. 1C, right), consistent with subunit G2 (13578 Da) but not G1 (13621 Da), both of which are N-terminally acetylated (fig. S1C). Lower abundance peaks were also seen for the V_1 region missing subunit C1 (Fig. 1C, green) and

[1]Molecular Medicine Program, The Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada. [2]Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, UK. [3]Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada. [4]Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada.
*Corresponding author. Email: john.rubinstein@utoronto.ca

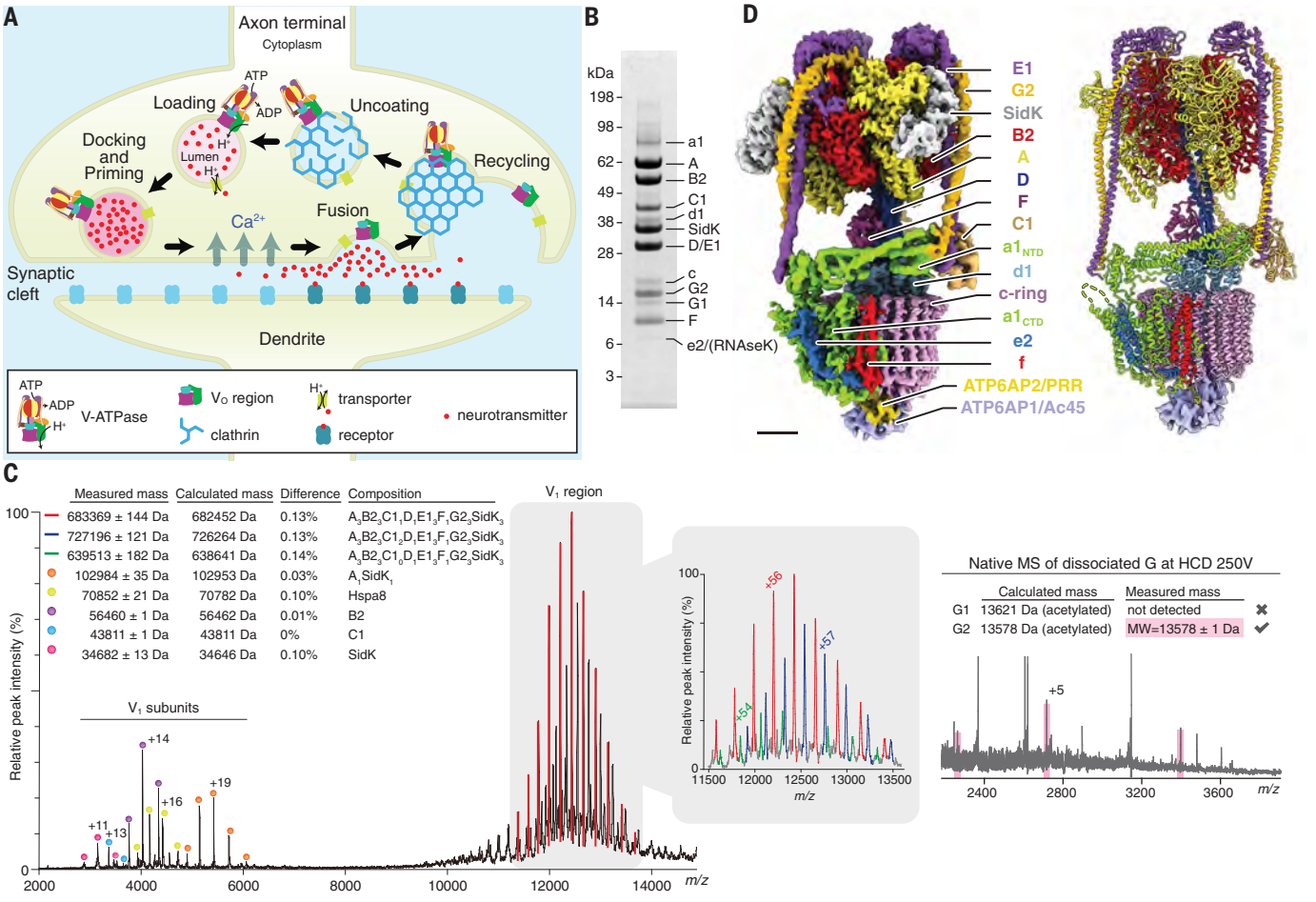


Fig. 1. Overall structure of brain V-ATPase. (**A**) Cycle of synaptic vesicle loading, docking and priming, fusion, and recycling. (**B**) SDS-PAGE of rat brain V-ATPase isolated with 3×FLAG $SidK_{1-278}$ and gel filtration chromatography. (**C**) Native mass spectrometry (MS) of V_1 region (left) and native mass spectrometry of dissociated subunit G (right) at a higher-energy collisional dissociation (HCD) voltage of 250 V. The charge state for one peak per subunit is indicated. The table shows the measured mass for each peak (± SD of fit) and the calculated mass depending on subunit composition (table S5). The difference between calculated and measured masses is indicated. *m/z*, mass/charge ratio. (**D**) Composite cryo-EM map (left) and atomic model (right) of brain V-ATPase in rotational state 1. Scale bar, 25 Å.

even with two copies of C1 (Fig. 1C, blue), the latter possibly attributable to the loss of subunit H. The enzyme preparation had an ATP hydrolysis rate that decreased during hydrolysis, consistent with the complex lacking subunit H (*24*, *26*) (fig. S2). Further, the preparation did not demonstrate full coupling of ATPase activity in the V_1 region to proton transport through the V_O region: Bafilomycin, which blocks the V_O region, only partially inhibited ATPase activity. This lack of coupling could be due to free V_1 complexes that remain active because subunit H is not present or to the detergent-solubilized V-ATPase disassembling during ATP hydrolysis, as seen with the *Manduca sexta* enzyme (*28*). The extent of SidK inhibition of ATP hydrolysis activity could not be quantified because only the SidK-bound form of the enzyme was available, although previous work has shown that SidK inhibits *S. cerevisiae* V-ATPase by ~40% (*19*).

Cryo–electron microscopy (cryo-EM) of the preparation yielded three three-dimensional (3D) maps, corresponding to ~120°-rotations of the rotor subcomplex between states (*25*). These rotational states of the enzyme had overall resolutions of 3.9, 4.0, and 3.9 Å (figs. S3 and S4). Focused refinement of rotational state 1 was able to improve the resolution to 3.8 Å for the membrane region and 3.6 Å for the catalytic region of the complex. Together, these maps allowed the construction of an atomic model for most of the complex, with a few components—including parts of subunits E1 and G2, the soluble N-terminal domain of subunit a1, subunit C1, and several luminal loops in the membrane region—modeled as backbone with truncated side chains (Fig. 1D, table S6, and fig. S5). Similar to the yeast V_O structure (*8*, *29*), no density was apparent for the loop between residues 667 and 712 in subunit a1. Owing to the averaging that occurs during cryo-EM image analysis, the minor population of complexes in the preparation possessing subunit G1 rather than G2 could produce a map that shows a weighted average of both isoforms. Alternatively, images of complexes containing subunit G1 may be excluded during 2D- and 3D-image classification if they do not average coherently with the major population of complexes to produce high-resolution map features. Where the map shows high-resolution features for subunit G, it accommodates subunit G2 better than G1 (fig. S6), which is consistent with most of the V-ATPase complexes containing subunit G2. Interpolating between the three rotational states produced a movie (movie S1) that shows the conformational changes in the enzyme that couple ATP hydrolysis in the V_1 region to proton pumping through the V_O region. These changes illustrate the flexibility of the enzyme, particularly in the peripheral stalks, subunit C1, and N-terminal domain of subunit a1.

Previous high-resolution insight into the structure of the eukaryotic V_1 region has been limited to cryo-EM of the intact yeast V-ATPase at ~7-Å resolution (*19*, *25*, *29*) and a 6.2- to 6.7-Å resolution crystal structure of the autoinhibited

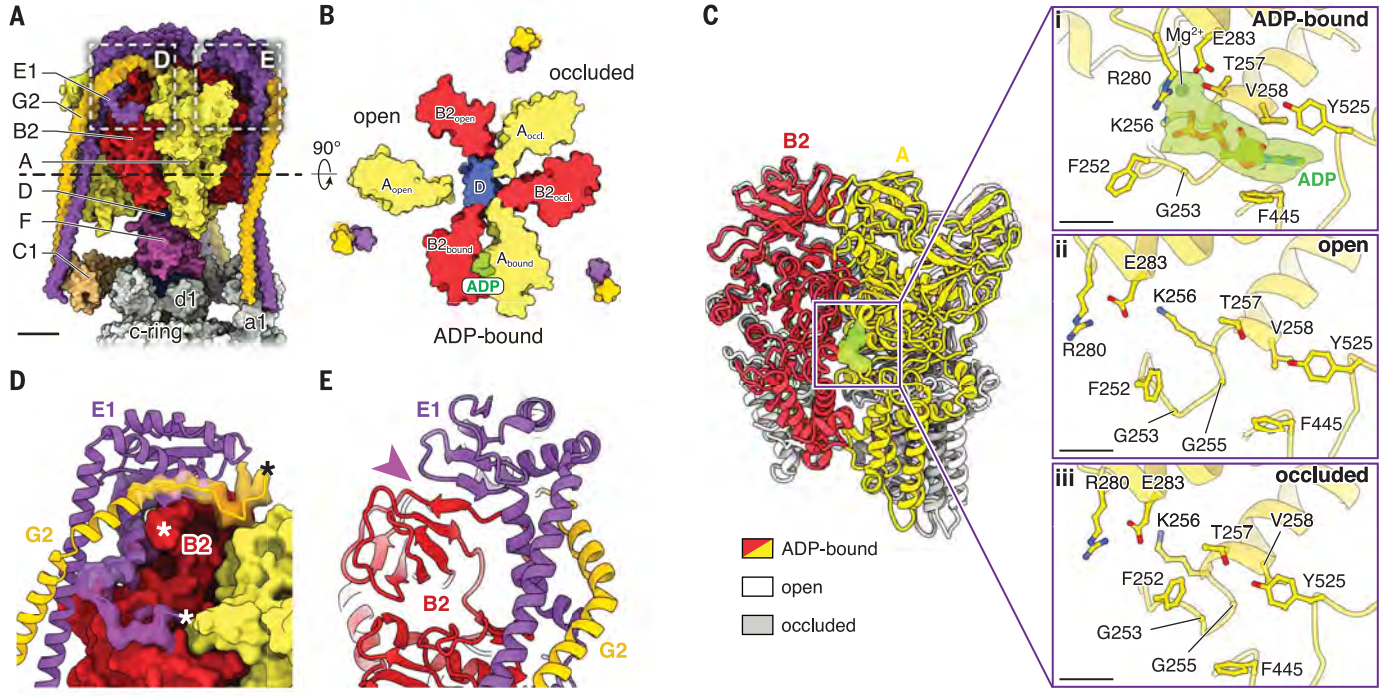


Fig. 2. Structure of the V_1 region. (**A**) Surface representation. Scale bar, 25 Å. (**B**) Cross-section through the V_1 region, viewed from V_1 toward V_O. ADP is shown in green. occl, occluded. (**C**) Superposition of the catalytic AB pairs (left) and close-ups of the conformations of the nucleotide binding sites (right). Scale bars, 5 Å. (**D**) Close-up of the interaction of subunits B2, E1, and G2 from (A). N and C termini of models are indicated by asterisks. (**E**) Close-up of the continuous β sheet formed by E1 and B2 from (A) (purple arrowhead). 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.

yeast V_1(-C) complex (*27*). These maps only allowed visualization of α helices in the structure. The current structure of the V_1 region (Fig. 2A) enables comparison with atomic models of prokaryotic V/A-ATPase catalytic regions (*30*–*32*) as well as numerous atomic models of ATP synthase F_1 regions (*33*). ATP synthases possess three catalytic and three noncatalytic nucleotide binding sites, found at the interfaces between subunits α (corresponding to V-ATPase subunit B) and β (corresponding to subunit A) (*33*). In contrast, and consistent with structures of the bacterial V/A-ATPase (*31*, *34*), the mammalian V_1 region lacks noncatalytic nucleotide binding sites and shows only a single-bound nucleotide in one of the catalytic sites (Fig. 2, B and Ci). This nucleotide could be modeled in the density map as adenosine 5′-diphosphate (ADP), consistent with biochemical analysis of the *M. sexta* V-ATPase (*28*). The three pairs of catalytic subunits in F- and V-type ATPases interchange between three different conformations, originally described for the F_1-ATPase as ATP-bound, ADP-bound, and empty (*33*). In the present structure, each catalytic AB pair similarly adopts one of three conformations (Fig. 2, B and C). Comparison of the conformation of catalytic AB pairs with crystal structures of a prokaryotic V_1/A_1 region suggests that the ADP-bound site is in a posthydrolysis state (Fig. 2, B and Ci) (*34*) and that the site in an open conformation (clockwise from the ADP-bound site when viewed from V_1 toward V_O) is in a state with high affinity for ATP (*31*) (Fig. 2, B and Cii). The conformation of the third pair of AB subunits does not appear to correspond to previous structures, but its nucleotide-binding pocket is occluded (Fig. 2, B and Ciii), suggesting a low affinity for nucleotide (*31*). Therefore, the enzyme imaged here appears to be poised to bind ATP. Three copies of $SidK_{1\text{-}278}$ are bound to the three A subunits (fig. S7). The structural consequences of SidK binding to the mammalian V_1 region are not known, but SidK perturbs the conformation of the yeast V_1 region only subtly (*19*). Although lower-resolution structures of the yeast enzyme have suggested that subunit G does not contact the rest of the V_1 region (*25*, *27*), the present structure shows that it does participate in linking the EG heterodimers of the peripheral stalks to the catalytic A_3B_3 subcomplex in the mammalian enzyme (Fig. 2D). This connection relies on residues from the N terminus of subunit B2 and C-terminal residues of subunits E1 and G2 (Fig. 2D, asterisks) and includes a structure where β strands from both subunit B2 and E1 form a single β sheet (Fig. 2E, purple arrowhead), also seen in a recent cryo-EM map of a prokaryotic V/A-ATPase (*32*). Numerous mutations in subunits A, B1, B2, and E1 are linked to disease and can be mapped onto the structure (fig. S8).

The V_O region contains subunits a1, d1, e2, f, ATP6AP1/Ac45, ATP6AP2/PRR, and the c ring (Fig. 3A). Previous high-resolution structures of *S. cerevisiae* V_O regions were determined from auto-inhibited complexes separated from their V_1 regions (*8*, *9*, *29*), whereas the structure presented here is within the context of an assembled V-ATPase. Consequently, the soluble N-terminal domain of subunit a1 is found in its non-inhibitory conformation and does not make contact with subunit d1 (Fig. 3B, green). The membrane-embedded C-terminal domain of subunit a1 includes eight transmembrane α helices (Fig. 3C) and creates the two offset half-channels that allow proton translocation (*8*). The subunit starts with a v-shaped insertion into the lipid bilayer formed by two short α helices that do not entirely cross the membrane (α1 and α2), followed by four transmembrane α helices (α3 to α6) and two highly tilted transmembrane α helices (α7 and α8) that create the surface that contacts the c ring. This fold closely follows yeast Vph1p and Stv1p (*8*, *29*), with root mean square deviations between C_α atoms of 1 and 0.8 Å, respectively (fig. S9A). The major difference between these structures is the insertion of a structured linker with α helices connecting transmembrane α3 and α4 in subunit a1 (Fig. 3C and fig. S9B). The c ring

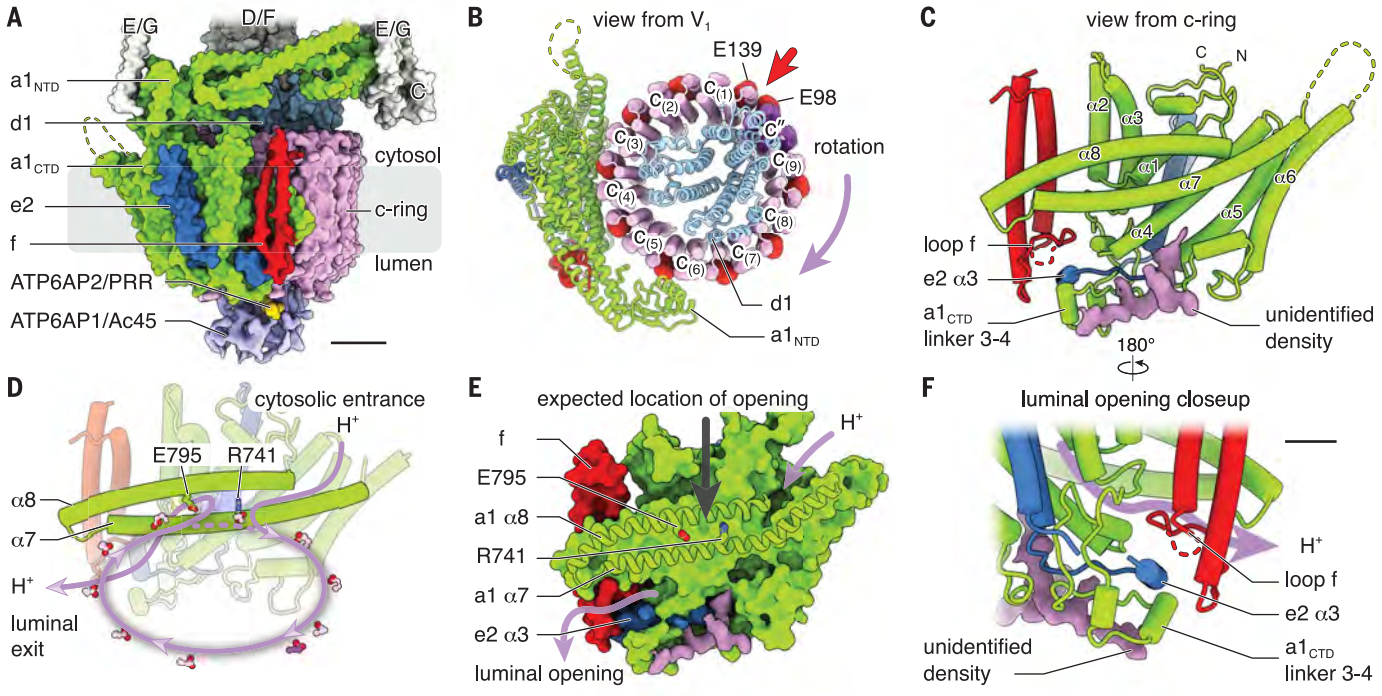


Fig. 3. Structure of the V_O region. (**A**) Surface representation with cryo-EM density for subunits f, ATP6AP1/Ac45, and ATP6AP2/PRR. Scale bar, 25 Å; NTD, N-terminal domain; CTD, C-terminal domain. (**B**) Viewed from V_1 with conserved proton-carrying Glu residues as red spheres. The direction of ATP-hydrolysis–driven rotation of the ring is indicated. Red arrow indicates the symmetry-breaking Glu residue of subunit c″. (**C**) Cartoon representation viewed parallel to the plane of the lipid bilayer. (**D**) Proton path through the V_O region. (**E**) Surface representation of $a1_{CTD}$ viewed parallel to the plane of the lipid bilayer. The gray arrow indicates the expected location of an opening between α7 and α8 leading toward the luminal half-channel. (**F**) Close-up view of the luminal terminus of the luminal half-channel, showing the interaction of subunits f, e2, a1, and the unidentified density. Scale bar, 10 Å.

contains nine copies of subunit c and one of subunit c″ (Fig. 3B, pink and purple). As expected from the lack of a homologous gene in mammals, the ring does not contain the subunit c′ that is found in yeast. Conserved Glu residues from protomers of the c ring are each capable of carrying a single proton during proton translocation (Fig. 3B, red spheres). As with the yeast V_O region, subunit c″ breaks the pattern of conserved proton-carrying Glu residues on alternating outer α helices of the c ring (Fig. 3B, red arrow) (*8*). The presence of ten protomers in the c ring sets the ATP:proton ratio at three ATP molecules hydrolyzed for every ten protons translocated. With a Gibbs free energy for ATP hydrolysis of ~−60 kJ/mol (*35*), this ratio limits the transmembrane proton motive force established by the brain V-ATPase, $\Delta\tilde{\mu}_{H^+}$, to ~18 kJ/mol, which is equivalent to ~3 pH units or ~180 mV (*35*). Similarity between the mammalian and yeast V-ATPase structures suggests that they share the same mechanism for rotation-driven proton pumping (*8*) (Fig. 3D). In this mechanism, rotation of the c ring drives abstraction of a proton from the cytoplasmic half-channel to neutralize the charge on a conserved c-ring Glu residue as it enters the lipid bilayer. Rotation brings a protonated Glu residue close to Arg^{741} of subunit a1, with formation of a salt bridge between the Arg and Glu causing release of the proton into the luminal half-channel that begins near Glu^{795} of subunit a1. Although structures of ATP synthases (*32*, *36*–*38*) have resolved the gap between the two tilted α helices of subunit a where protons must pass, this opening cannot be seen here (Fig. 3E, gray arrow) or in other V_O region structures (*8*, *9*, *29*). This difference is either because of limited resolution or because the opening only forms transiently in V-ATPases that, unlike ATP synthases, pump protons against a proton motive force. The surface of subunit a1 that contacts the c ring has the pattern of a positive charge at the conserved Arg residue and negative charges at the midmembrane termini of the two half channels (*39*) that has now been seen in all other rotary ATPases (fig. S9C).

Subunit e2 is adjacent to transmembrane α helices 3 and 4 of subunit a1 and consists of an α-helical hairpin with a C-terminal tail, similar to the corresponding yeast protein Vma9p (*8*) (Fig. 3, A and C, blue, and fig. S10). However, unlike yeast Vma9p, subunit e2 has an extended C-terminal sequence that terminates in a third short α helix (α3) encircled by the linker between α3 and α4 of subunit a1 and the luminal loop of subunit f (Fig. 3C and fig. S10, blue arrow). The function of subunit e2 is unknown (*40*), but deletion of Vma9p causes the Vma^- V-ATPase deficiency phenotype in yeast, where cells can grow on medium buffered to pH 5.5 but not pH 7.5, are sensitive to extracellular calcium and a variety of heavy metals, and cannot grow on medium with typical concentrations of nonfermentable carbon sources (*41*). Recent analysis of both the vacuolar and Golgi forms of the *S. cerevisiae* V-ATPase identified the hypothetical protein YPR170W-B as a transmembrane α-helical hairpin subunit, named subunit f (*8*, *29*). YPR170W-B is highly conserved in fungi, but deletion of the gene did not cause the Vma^- phenotype (*8*). The structure of the brain V-ATPase revealed density for a similar transmembrane α-helical hairpin in the V_O region in a position corresponding to the yeast subunit f (Fig. 3, red). Bioinformatic analysis suggests RNAseK, a conserved metazoan protein (*42*), as a homolog of *S. cerevisiae* subunit f, sharing 32% sequence identity and 52% sequence similarity (fig. S11A). Immunoprecipitation of RNAseK previously revealed that it is associated with V-ATPase (*22*) and mass spectrometry done in this study shows that RNAseK is present in this enzyme preparation (tables S1 to S3). Consistent with a role

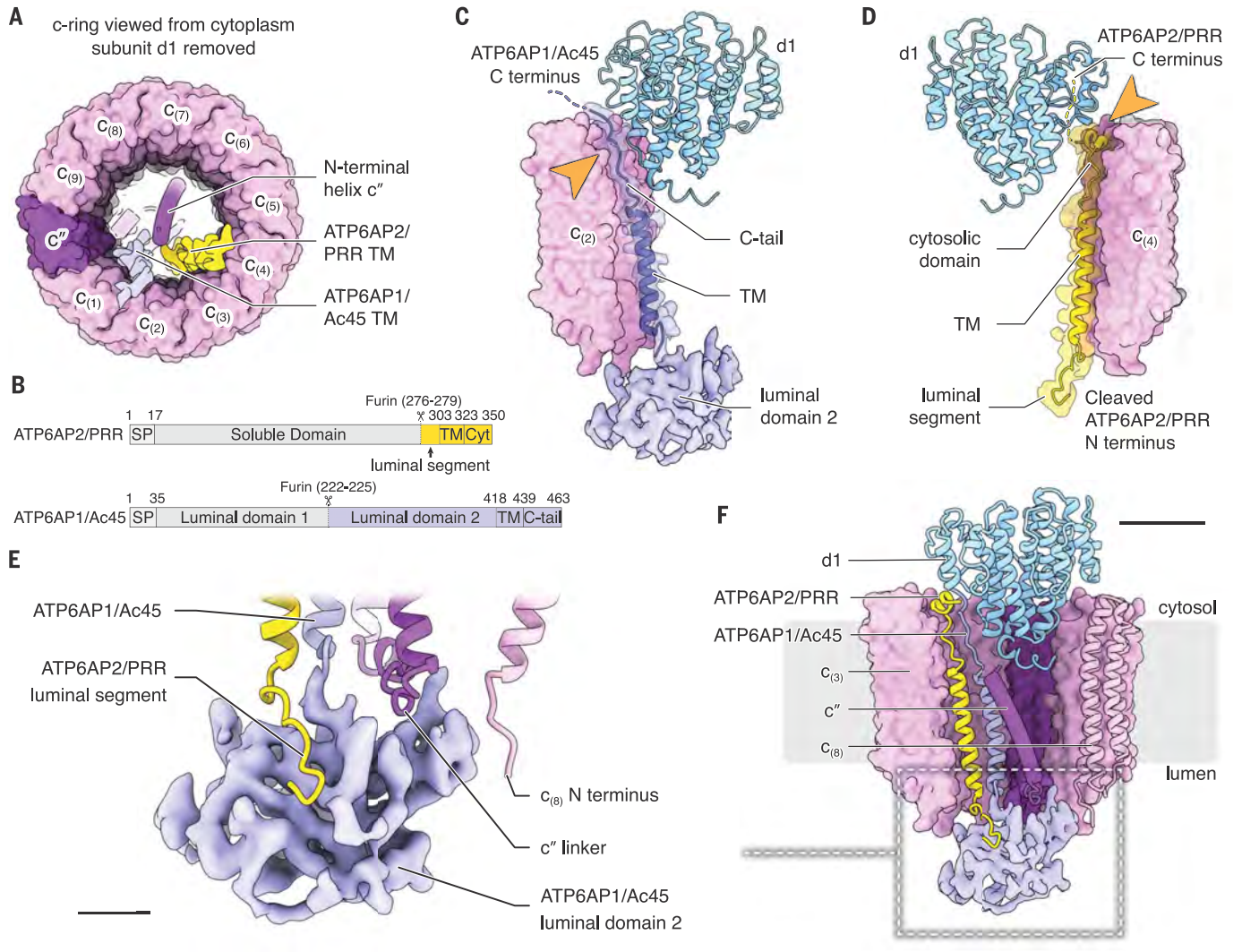


Fig. 4. Interaction of subunits within the membrane-embedded rotor subcomplex. (**A**) Subunits ATP6AP1/Ac45, ATP6AP2/PRR, and c″ possess transmembrane α helices in the center of the c ring. (**B**) ATP6AP2/PRR (top) and ATP6AP1/Ac45 (bottom) possess luminal domains that are absent from the structure. SP, signal peptide; TM, transmembrane; Cyt, cytosolic. (**C** and **D**) Subunit d1 interacts with the C termini of ATP6AP1/Ac45 (C) and ATP6AP2/PRR (D). Arrowheads indicate regions of ATP6AP1/Ac45 and ATP6AP2/PRR that contribute to the binding surface for subunit d1. (**E**) Subunits c″, $c_{(8)}$, and ATP6AP2/PRR interact with the second luminal domain of ATP6AP1/Ac45. Scale bar, 10 Å. (**F**) Subunits c″, ATP6AP1/Ac45, ATP6AP2/PRR, and d1 create a network of interactions that connect the vesicle lumen and the cytoplasm. Scale bar, 25 Å.

in V-ATPase activity, RNAseK is necessary for viral endocytosis and replication (*22*, *43*). The cryo-EM density for subunit f, although lacking the necessary resolution to unambiguously identify the protein as RNAseK, is consistent with the expected size of RNAseK and accommodates most of its bulky side chains (fig. S5B). Together, this evidence tentatively identifies RNAseK as subunit f in mammalian V-ATPases. Whether or not RNAseK functions as a ribonuclease within the V-ATPase is not known, but we note that the in vitro assays that established this protein and its homologs as ribonucleases (*42*, *44*) were performed in the absence of detergents or other membrane mimetics, despite clear transmembrane α helices in hydropathy analysis of the protein's amino acid sequence (fig S11B). An additional unidentified protein-like density previously found in both the vacuolar and Golgi forms of the *S. cerevisiae* V-ATPase (*9*, *29*) is also present in the mammalian complex (Fig. 3, C to F, purple). All four of the membrane-embedded components from this part of the enzyme interact at the opening of the luminal proton half-channel, with the linker between α3 and α4 of subunit a1 acting as a scaffold for packing the luminal loop connecting the transmembrane α helices of subunit f, the C-terminal sequence of subunit e2, and part of the unknown component (Fig. 3F). These four components are all elongated relative to their yeast counterparts to accommodate their interaction (figs. S9 to S11).

Unlike in the yeast V-ATPase, three rather than two transmembrane α helices are located within the lumen of the c ring (Fig. 4A). The centermost α helix corresponds to the N-terminal α helix of subunit c″, which is also seen in the yeast V-ATPase (*8*, *9*, *29*). The other two transmembrane α helices are from ATP6AP1/Ac45 and ATP6AP2/PRR (Fig. 4A and fig. S5C). ATP6AP1/Ac45 mutations in humans can lead to immunodeficiency, cognitive impairment, liver dysfunction, and abnormal protein glycosylation (*45*), whereas mutations in ATP6AP2/PRR can result in neurodegeneration as well as X-linked parkinsonism and epilepsy (*12*). Both proteins are frequently mutated in granular cell cancers (*46*). Immature ATP6AP1/Ac45 contains two soluble N-terminal domains on the luminal side of the membrane, separated by a furin cleavage site (*47*) (Fig. 4B, bottom). In the map, the soluble region of APT6AP1/

Ac45 appears as a single low-resolution domain approximately 30 Å by 20 Å by 30 Å in size (Fig. 4C, light blue), which is consistent with only the second luminal domain being in the complex. Further, although mass spectrometry detected peptides from uncleaved ATP6AP1/Ac45, it suggested that the cleaved form is predominantly present (fig. S12, A and C). The C-terminal transmembrane α helix of ATP6AP1/Ac45 is homologous with that of the yeast V-ATPase subunit Voa1p (*45*) and is found in an equivalent position inside the c ring (*9*). This transmembrane α helix stretches approximately halfway across the thickness of the lipid bilayer, terminating as a C-terminal tail that interacts with two of the subunits in the c ring and subunit d1 before entering the cytoplasm (Fig. 4C, light blue). ATP6AP2/PRR is known to associate with mammalian V-ATPase (*10*, *11*) and is essential for biogenesis of active V-ATPase (*48*). The strong density for its transmembrane α helix in the map suggests that every complex includes the subunit (fig. S5C). Consistent with an emerging role for V-ATPases in cell signaling (*49*), the presence of ATP6AP2/PRR in the complex has tied V-ATPases to the renin-angiotensin system for regulation of blood pressure and electrolyte balance (*10*, *11*, *13*), Wnt signaling (*14*), and other pathways (*12*). The gene for ATP6AP2/PRR encodes an N-terminal extracellular or luminal soluble domain and a transmembrane anchor (Fig. 4B, top). The soluble domain increases the angiotensin I–generating activity of renin and can function in both a membrane-bound form and, when released by proteolysis, a soluble form (*50*, *51*). In the structure, the renin-activating domain of ATP6AP2/PRR is missing (Fig. 4D, yellow). The transmembrane anchor consists of a long α helix and a short α-helical turn, connected by an extended linker with N- and C-terminal tails. The location of ATP6AP2/PRR's transmembrane region, captured within the stable c ring, dictates that, after its incorporation into the c ring, the protein must remain associated with the V-ATPase. Mass spectrometry detected some intact ATP6AP2/PRR in the preparation, but, consistent with the cryo-EM density, the cleaved transmembrane region alone was much more abundant (fig. S12, B and D). In contrast to the renin-angiotensin system, Wnt signaling relies on ATP6AP2/PRR remaining membrane-anchored, with interaction of proteins with the extracellular or luminal part of ATP6AP2/PRR leading to signaling in the cytoplasm (*14*). The absence of the soluble domain of ATP6AP2/PRR from the structure suggests that ATP6AP2/PRR's function in Wnt signaling involves either a subpopulation of intact ATP6AP2/PRR molecules that are not associated with V-ATPase or a different population of V-ATPase or V_O complexes that contain intact ATP6AP2/PRR.

The structure shows numerous interactions between ATP6AP1/Ac45, ATP6AP2/PRR, subunits d1 and c″, and multiple c-subunits (Fig. 4, E and F). Because these proteins are all part of the rotor subcomplex, the interactions persist during rotation and are also found in rotational states 2 and 3 (fig. S13). In the lumen, the soluble domain of ATP6AP1/Ac45 interacts with the N-terminal tail of cleaved ATP6AP2/PRR, the linker that connects subunit c″ to its N-terminal α helix in the middle of the c ring, and the N terminus of subunit $c_{(8)}$ (Fig. 4E). Near the cytoplasmic surface of the V_O region, the C-terminal tail of ATP6AP1/Ac45 and the short C-terminal α helix of ATP6AP2/PRR, including 12 of the 19 residues in its intracellular domain (*14*), are sandwiched between subunits of the c ring and subunit d1 (Fig. 4, C and D, arrow). V_O complexes assemble in the endoplasmic reticulum (ER), with the incorporation of subunit d allowing ER release (*52*, *53*) and subsequent binding of the V_1 region (*53*). Free subunit d in the cytoplasm (*54*) does not bind the V_1 complexes that are preassembled there (*55*), which suggests that the conformation of subunit d changes upon incorporation into V_O, making it competent to interact with subunit D of V_1 (Fig. 1D). Indeed, subunit d adopts a more open conformation when it engages subunit D of the central rotor in the intact V-ATPase and a more closed conformation when V_1 is detached and subunits d and D are disconnected (fig. S14). The C-terminal tails of ATP6AP1/Ac45 and ATP6AP2/PRR produce part of the surface onto which subunit d1 assembles (Fig. 4, C, D, and F). This structure explains why mutations in ATP6AP1/Ac45 and ATP6AP2/PRR are associated with related disease phenotypes (*12*, *45*, *46*). It also explains why ATP6AP1/Ac45 and ATP6AP2/PRR appear to work together to allow subunit d binding, ER release, and subsequent V_1 assembly onto V_O in the mammalian V-ATPase (*9*, *56*). ATP6AP1/Ac45 and ATP6AP2/PRR could also be involved in the reverse process, with conformational changes within them altering the conformation of subunit d1, disrupting its interaction with subunit D, and triggering separation of the V_1 and V_O regions.

REFERENCES AND NOTES

1. M. Forgac, *Nat. Rev. Mol. Cell Biol.* **8**, 917–929 (2007).
2. P. M. Kane, *Microbiol. Mol. Biol. Rev.* **70**, 177–191 (2006).
3. P. M. Kane, *J. Biol. Chem.* **270**, 17025–17032 (1995).
4. J. P. Sumner *et al*., *J. Biol. Chem.* **270**, 5649–5653 (1995).
5. A. Bodzęta, M. Kahms, J. Klingauf, *Cell Rep.* **20**, 1348–1359 (2017).
6. S. Watanabe *et al*., *Nature* **515**, 228–233 (2014).
7. Z. Farsi *et al*., *eLife* **7**, e32569 (2018).
8. M. T. Mazhab-Jafari *et al*., *Nature* **539**, 118–122 (2016).
9. S.-H. Roh *et al*., *Mol. Cell* **69**, 993–1004.e3 (2018).
10. N. L'Huillier, M. G. F. Sharp, D. R. Dunbar, J. J. Mullins, in *The Local Cardiac Renin-Angiotensin Aldosterone System*, E. D. Frohlich, R. N. Re, Eds. (Springer, 2006), pp. 17–34.
11. C. Burcklé, M. Bader, *Hypertension* **48**, 549–551 (2006).
12. A. Ichihara, M. S. Yatabe, *Nat. Rev. Nephrol.* **15**, 693–712 (2019).
13. J. Ludwig *et al*., *J. Biol. Chem.* **273**, 10939–10947 (1998).
14. C.-M. Cruciat *et al*., *Science* **327**, 459–463 (2010).
15. M. Toei, R. Saum, M. Forgac, *Biochemistry* **49**, 4715–4723 (2010).
16. K. G. Blake-Palmer, Y. Su, A. N. Smith, F. E. Karet, *Gene* **393**, 94–100 (2007).
17. S. Takamori *et al*., *Cell* **127**, 831–846 (2006).
18. L. Xu *et al*., *PLOS Pathog.* **6**, e1000822 (2010).
19. J. Zhao *et al*., *PLOS Pathog.* **13**, e1006394 (2017).
20. S. Ahmed, M. Holt, D. Riedel, R. Jahn, *Nat. Protoc.* **8**, 998–1009 (2013).
21. Y. Murata *et al*., *J. Biol. Chem.* **277**, 36296–36303 (2002).
22. J. M. Perreira *et al*., *Cell Rep.* **12**, 850–863 (2015).
23. S. Wilkens, T. Inoue, M. Forgac, *J. Biol. Chem.* **279**, 41942–41949 (2004).
24. Z. Zhou *et al*., *J. Biol. Chem.* **274**, 15913–15919 (1999).
25. J. Zhao, S. Benlekbir, J. L. Rubinstein, *Nature* **521**, 241–245 (2015).
26. K. J. Parra, K. L. Keenan, P. M. Kane, *J. Biol. Chem.* **275**, 21761–21767 (2000).
27. R. A. Oot, P. M. Kane, E. A. Berry, S. Wilkens, *EMBO J.* **35**, 1694–1706 (2016).
28. M. Huss, H. Wieczorek, *FEBS Lett.* **581**, 5566–5572 (2007).
29. T. Vasanthakumar *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7272–7277 (2019).
30. Y. Nagamatsu, K. Takeda, T. Kuranaga, N. Numoto, K. Miki, *J. Mol. Biol.* **425**, 2699–2708 (2013).
31. S. Arai *et al*., *Nature* **493**, 703–707 (2013).
32. L. Zhou, L. A. Sazanov, *Science* **365**, eaaw9144 (2019).
33. J. P. Abrahams, A. G. Leslie, R. Lutter, J. E. Walker, *Nature* **370**, 621–628 (1994).
34. K. Suzuki *et al*., *Nat. Commun.* **7**, 13235 (2016).
35. D. G. Nicholls, S. J. Ferguson, *Bioenergetics* (Academic Press, ed. 4, 2013).
36. A. Hahn, J. Vonck, D. J. Mills, T. Meier, W. Kühlbrandt, *Science* **360**, eaat4318 (2018).
37. H. Guo, T. Suzuki, J. L. Rubinstein, *eLife* **8**, e43128 (2019).
38. B. J. Murphy *et al*., *Science* **364**, eaaw9128 (2019).
39. H. Guo, S. A. Bueler, J. L. Rubinstein, *Science* **358**, 936–940 (2017).
40. S. A. Bueler, J. L. Rubinstein, *Biochemistry* **54**, 853–858 (2015).
41. M. Sambade, P. M. Kane, *J. Biol. Chem.* **279**, 17361–17365 (2004).
42. T. N. Rampias, D. C. Sideris, E. G. Fragoulis, *Nucleic Acids Res.* **31**, 3092–3100 (2003).
43. B. A. Hackett *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7797–7802 (2015).
44. M.-A. I. Economopoulou, E. G. Fragoulis, D. C. Sideris, *Nucleic Acids Res.* **35**, 6389–6398 (2007).
45. E. J. R. Jansen *et al*., *Nat. Commun.* **7**, 11600 (2016).
46. F. Pareja *et al*., *Nat. Commun.* **9**, 3533 (2018).
47. E. Louagie *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12319–12324 (2008).
48. K. Kinouchi *et al*., *Circ. Res.* **107**, 30–34 (2010).
49. R. Zoncu *et al*., *Science* **334**, 678–683 (2011).
50. G. Nguyen *et al*., *J. Clin. Invest.* **109**, 1417–1427 (2002).
51. C. Cousin *et al*., *Hypertension* **53**, 1077–1082 (2009).
52. M. Ryan, L. A. Graham, T. H. Stevens, *Mol. Biol. Cell* **19**, 5131–5142 (2008).
53. L. A. Graham, K. J. Hill, T. H. Stevens, *J. Cell Biol.* **142**, 39–49 (1998).
54. C. Bauerle, M. N. Ho, M. A. Lindorfer, T. H. Stevens, *J. Biol. Chem.* **268**, 12749–12757 (1993).
55. R. D. Doherty, P. M. Kane, *J. Biol. Chem.* **268**, 16845–16851 (1993).
56. M. C. Guida *et al*., *Mol. Biol. Cell* **29**, 2156–2164 (2018).

ACKNOWLEDGMENTS

We thank S. Benlekbir for assistance with cryo-EM data collection; T. Vasanthakumar for advice on ATPase assays and assistance with Fig. 1A; and S. Angers, P. Brzezinsky, and members of the Rubinstein laboratory for discussions. **Funding:** This work was supported by the Canadian Institutes of Health Research Grant PJT166152 (J.L.R.), grant 69551-ENABLE from the European Research Council (C.V.R.), and Wellcome Trust Investigator Award 104633/Z/14/Z (C.V.R.). Y.M.A. was supported by a Restracomp fellowship and a Canadian Institutes of Health Research Postdoctoral fellowship. J.L.R. was supported by the Canada Research Chairs program. Cryo-EM data was collected at the Toronto High-Resolution High-Throughput cryo-EM facility, supported by the Canada Foundation for Innovation and Ontario Research Fund. **Author contributions:** J.L.R. conceived the project and experimental approach and supervised the research. S.A.B. made the SidK expression vector and developed an initial protein purification strategy. Y.M.A. developed the final protein purification procedure, prepared cryo-EM specimens, collected cryo-EM data, calculated the cryo-EM maps, built and refined the atomic models, and performed biochemical assays. D.W. performed

proteomics and native mass spectrometry experiments and C.V.R. supervised the mass spectrometry. J.L.R. and Y.M.A. wrote the manuscript and prepared the figures with input from the other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Cryo-EM maps are deposited in the Electron Microscopy Data Bank under accession numbers EMD-21317 to 21319 for maps of the intact complex and 21345 to 213453 for maps from focused refinement. Atomic models are deposited in the Protein Data Bank under accession numbers 6VQ6, 6VQ7, and 6VQ8 for composite models and 6VQ9, 6VQA, 6VQB, 6VQC, 6VQG, 6VQH, 6VQI, 6VQJ, and 6VQK for models built into maps from focused refinement. The SidK-3×FLAG expression vector is available from J.L.R.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1240/suppl/DC1
Materials and Methods
Figs. S1 to S14
Tables S1 to S6
References (*57*–*78*)
Movie S1

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

28 August 2019; accepted 12 February 2020
10.1126/science.aaz2924

ORGANIC CHEMISTRY

Enantioselective remote C–H activation directed by a chiral cation

Georgi R. Genov*, James L. Douthwaite*, Antti S. K. Lahdenperä, David C. Gibson, Robert J. Phipps†

Chiral cations have been used extensively as organocatalysts, but their application to rendering transition metal–catalyzed processes enantioselective remains rare. This is despite the success of the analogous charge-inverted strategy in which cationic metal complexes are paired with chiral anions. We report here a strategy to render a common bipyridine ligand anionic and pair its iridium complexes with a chiral cation derived from quinine. We have applied these ion-paired complexes to long-range asymmetric induction in the desymmetrization of the geminal diaryl motif, located on a carbon or phosphorus center, by enantioselective C–H borylation. In principle, numerous common classes of ligand could likewise be amenable to this approach.

Ion-pairing has been put to extensive use as a key design feature in the field of asymmetric catalysis (*1*). In the 1980s, pioneering studies on enantioselective phase-transfer catalysis paired a chiral cation with a reactive anionic intermediate in the enantiodetermining transition state (*2*), with cinchona alkaloid-derived cations dominating as effective and readily accessible scaffolds (*3*). The numerous subsequent developments in this area have had enormous impact in the field of asymmetric organocatalysis, encapsulating such important transformations as Michael and aldol additions, as well as Mannich, fluorination, alkylation, and oxidative cyclization reactions, to name but a few (*4*–*7*) (Fig. 1A, left). Over the past decade, the inverse strategy of using a chiral anion to associate with a cationic reaction intermediate has also proven extremely successful (*1*, *8*, *9*). This latter strategy has been effective not only in an organocatalytic context (*10*, *11*) but also in powerful combination with transition metal catalysts (*12*–*14*), cleverly capitalizing on the relatively common occurrence of cationic transition metal complexes in catalytic cycles. By contrast, it is far rarer to encounter anionic transition metal complexes as key intermediates. As such, the charge-inverted approach of pairing a chiral cation with an anionic transition metal catalyst has only been demonstrated in a handful of pioneering cases, notably asymmetric oxidation reactions involving anionic diphosphatobisperoxotungstate (*15*) and peroxomolybdate (*16*) complexes as catalysts (Fig. 1A, center) (*17*–*21*). Owing to this scarcity of anionic metal complexes in the most commonly used processes, the broader potential of uniting chiral cations with the versatile reactivity of transition metals has remained underexplored, despite the obvious potential presented by several privileged classes of chiral cation. Given the success of these motifs as chiral controllers in asymmetric organocatalysis (vide supra), a general strategy to integrate them with transition metal catalysis would likely have broad impact in the field of asymmetric catalysis.

In an important advance, which compellingly demonstrates this potential, Ooi and co-workers incorporated a chiral cation covalently into the structure of a phosphine ligand, resulting in highly stereocontrolled formation of contiguous all-carbon quaternary stereocenters under palladium catalysis (Fig. 1A, right) (*22*, *23*). At the outset of this project, we envisioned a potentially more generally applicable approach whereby an anionic handle is incorporated into a common ligand scaffold, providing the key point of interaction with the chiral cation (Fig. 1B). Judicious placement of this anionic group would be crucial to success—not close enough to the metal center to disrupt reactivity but not so far that the chiral environment imparted by the cation would be ineffective. Various chiral cations could be introduced in the final step by simple ion exchange, allowing for rapid catalyst optimization. In pioneering work, Ooi and co-workers previously demonstrated the productive combination of cationic ligands with chiral anions, as demonstrated effectively in enantioselective allylic alkylation (*24*, *25*). We envisaged that, in principle, a wide variety of privileged ligand scaffolds for transition metal catalysis could be rendered anionic, creating exciting opportunities to explore the use of chiral cations as chiral controllers in a wealth of powerful transition metal–catalyzed reactions.

In seeking a rigorous and relevant test of the above-described approach, we targeted a transformation that lies at the cutting edge of what is currently possible in enantioselective catalysis. Although enantioselective, desymmetrizing C–H activation of arenes has been extensively explored with palladium (*26*, *27*), rhodium (*28*, *29*), and iridium (*30*, *31*) catalysis, all but a single case functionalize at the arene ortho position (*32*). Only very recently did Yu and co-workers achieve enantioselective desymmetrization through direct arylation at the arene meta position (Fig. 1C) (*33*), taking advantage of an ingenious relay strategy via the ortho position, although relatively high loadings of the chiral norbornene mediator (**CTM**, 20 to 50 mol %) were required. C–H borylation reactions have the useful attribute that the new C–B bond can undergo numerous diverse transformations (*34*, *35*), but so far, enantiocontrol in arene borylation has been realized only in two recent reports, from Shi, Hartwig, and co-workers (*30*) and Xu, Ke, and co-workers (*31*). In both cases, the chiral information is covalently incorporated into the ligand scaffold in the conventional manner and a directing group guides borylation to the ortho position. By contrast, the creation of chirality over long ranges, where the enantiotopic site is far from the new stereocenter, is an outstanding challenge in which catalyst designs that incorporate noncovalent interactions offer numerous opportunities (*36*–*38*).

We recently developed anionic bipyridine ligands that bear a remote sulfonate group to impart control of regioselectivity in iridium-catalyzed C–H borylation via noncovalent interactions with the substrate (*39*–*41*). Throughout these studies, a single ligand scaffold consistently gave the optimal regiocontrol. In one particular study, we attributed the high regioselectivity for borylation at the arene meta position to the existence of a hydrogen bond between the substrate and the sulfonate group of the ligand in the regiodetermining transition state for C–H activation (Fig. 1D) (*40*). We hypothesized that exchange of the achiral tetrabutylammonium counterion of the ligand for a chiral cation might allow enantioselective, desymmetrizing C–H activation in a prochiral substrate (as in Fig. 1E). Herein, we demonstrate that, using this approach, remote, enantioselective C–H borylation can be achieved for formation of chiral-at-carbon and chiral-at-phosphorous compounds, showcasing the thus far unexplored approach of combining a chiral cation with an anionic ligand for a reactive transition metal.

We commenced our studies with symmetrical benzhydrylamide **2a** (Fig. 2A). Numerous ion-paired ligands **L·1**, possessing a variety of chiral cations **1a** to **1i**, could be readily obtained through counterion exchange. The chiral cations were all derived from dihydroquinine (DHQ) with varying *N*-benzyl substitution. At room temperature in tetrahydrofuran (THF) as solvent, low but encouraging levels of enantioselectivity were obtained with 3,5-dimethoxy benzyl and 3,5-di-*tert*-butyl groups [**L·1a** and **L·1b**, 31 and 30% enantiomeric excess (ee)]. We next investigated placing substituted aromatic rings at the 3- and 5-positions

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: rjp71@cam.ac.uk

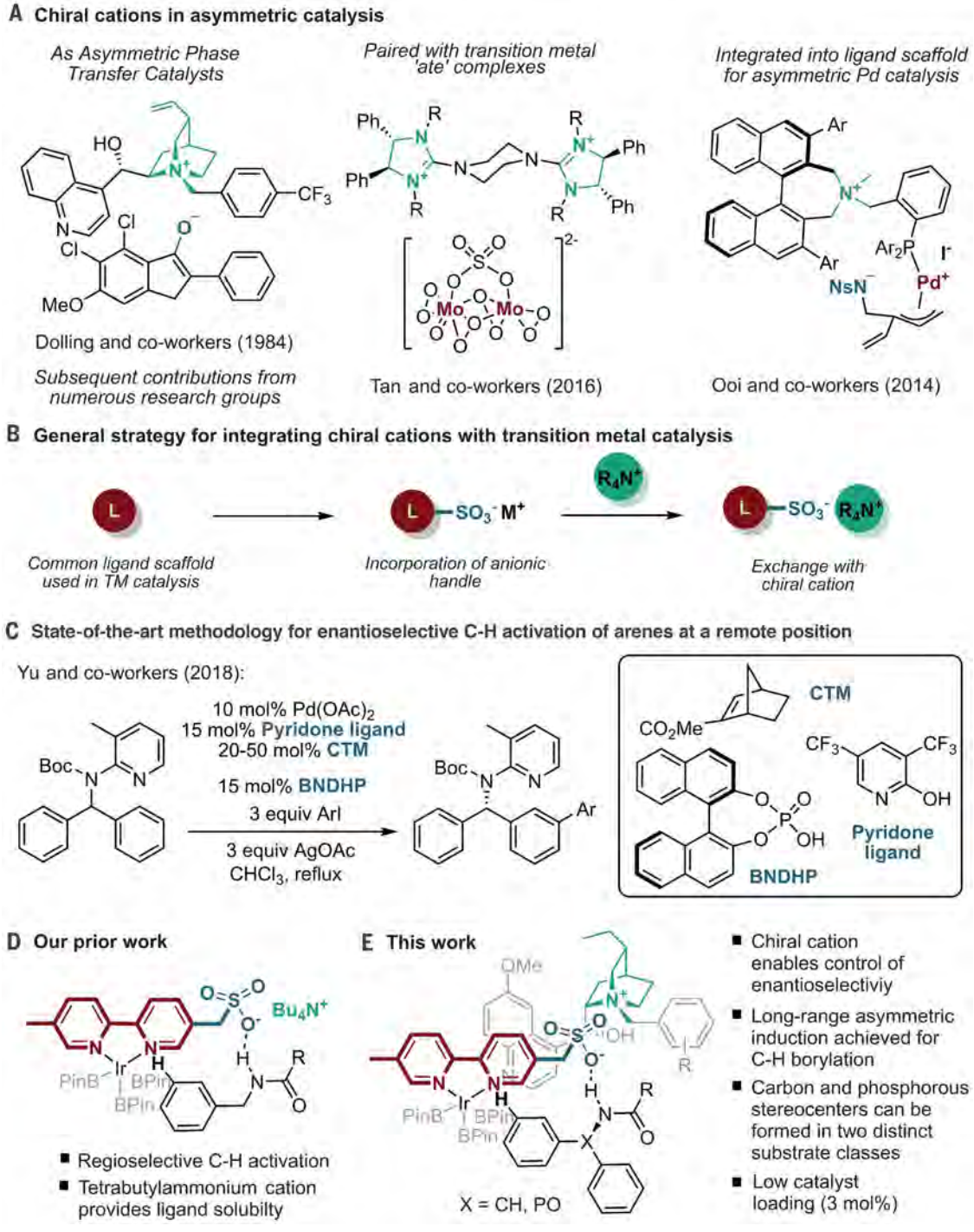


Fig. 1. Strategy for incorporating chiral cations with transition metals. (**A**) Applications of chiral cations in asymmetric catalysis. Me, methyl; Ph, phenyl; R, substituent. (**B**) General strategy for integration of transition metal (TM) catalysis with chiral cations. M, metal. (**C**) State-of-the-art methodology for enantioselective remote C–H activation of arenes. Boc, butyloxycarbonyl; Ac, acetyl; BNDHP, 1,1′-binaphthyl-2,2′-diyl hydrogenphosphate. (**D**) Our prior work controlling regioselectivity in arene borylation. Bu, butyl. (**E**) Summary of this work.

of the quaternizing *N*-benzyl group. Encouragingly, **L·1c** (4-$CF_3C_6H_4$) gave increased enantioselectivity (39% ee) and **L·1d** (3,4,5-$F_3C_6H_2$) resulted in a further improvement (52% ee). Focusing attention on the meta positions of the outer arenes of the teraryl system, we then evaluated a series of substituents (**L·1e** to **L·1i**). Trifluoromethyl (**L·1e**) and methoxy (**L·1f**) substitution again gave increases (both 60% ee), but the biggest gain came from the *tert*-butyl substituted **L·1g** (73% ee). At this point, we investigated aryl groups in these positions to extend the reach even further, but both of these proved detrimental (**L·1h** and **L·1i**). Thus, we shifted our attention to other reaction parameters with **L·1g**. A solvent evaluation identified cyclopentyl methyl ether (CPME) as being optimal, in that the reaction temperature could be reduced to –10°C while high reactivity was maintained, resulting in isolation of **3a** in 72% yield and with 96% ee, following oxidation to the corresponding phenol with H_2O_2 (see inset in Fig. 2A). This derivatization aided separation from any remaining starting material or difunctionalized material. The undesired borylation of monoborylated **3a′** to give a symmetrical diborylated by-product did occur to varying degrees in the reactions, being unavoidable at higher conversions. We thus carried out careful experiments to establish whether kinetic resolution may be occurring in such instances, resulting in possible enhancement of the observed ee of **3a′** at the

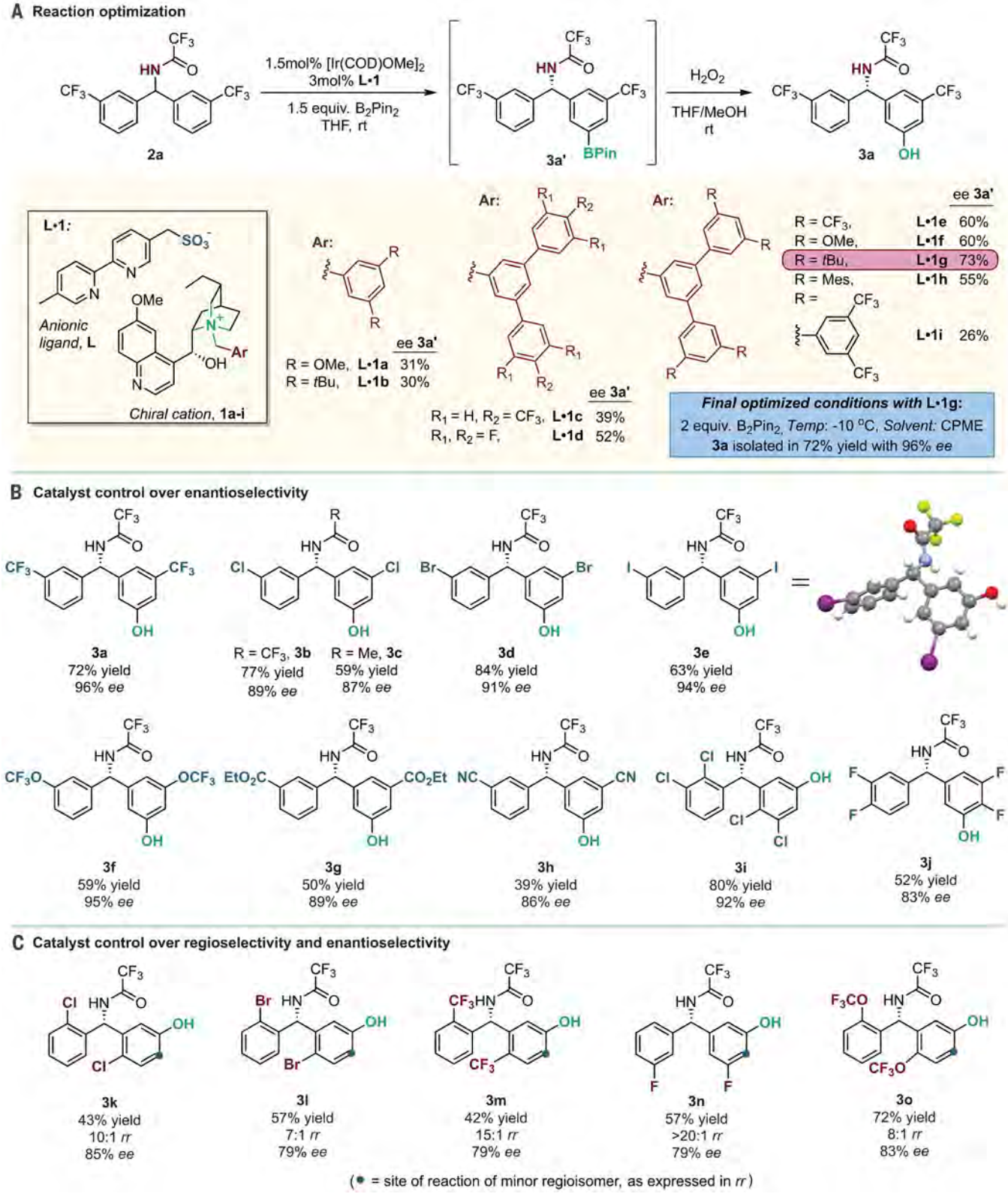


Fig. 2. Enantioselective desymmetrizing C–H borylation of benzhydrylamides. (**A**) Reaction optimization. COD, 1,5-cyclooctadiene; rt, room temperature; *t*Bu, *tert*-butyl. (**B**) Scope of enantioselective borylation using **L-1g** in substrates bearing no regioselectivity challenge. Et, ethyl. (**C**) Examples in which the catalyst is controlling regioselectivity and enantioselectivity. Yield values refer to isolated yields. Regioisomeric ratios were determined from the crude ^{1}H–nuclear magnetic resonance (NMR) spectrum before isolation. Enantiomeric excesses determined by chiral high-performance liquid chromatography (HPLC) or supercritical fluid chromatography (SFC) analysis.

end of the reaction. Evaluating ee of **3a′** at various levels of conversion as well as submitting racemic **3a′** to the enantioselective borylation conditions with **L·1g** showed that there is no appreciable kinetic resolution occurring (figs. S1 and S2). Finally, we evaluated a ligand paired with a Maruoka-type chiral cation which gave racemic product, a variant of **L·1g** in which the quinine hydroxyl group is methylated, which gave a reduced ee of 72%, and a variant of **L·1g** in which the stereochemistry of the quinine hydroxyl group is inverted, which gave only 11% ee (see supplementary materials and table S1 for full optimization details). A survey of N-protecting groups demonstrated that trifluoroacetyl is optimal, although acetyl also performed well (fig. S3).

We proceeded to examine the scope of the reaction in terms of versatile substituents on the substrate aryl rings (Fig. 2B). Postreaction derivatization of the introduced boronic acid pinacol ester (BPin) group facilitated purification, and we used oxidation with hydrogen peroxide to give the corresponding phenols. We were pleased to find that halide substitution was very well tolerated in the enantioselective borylation. Chloro-substituted (**3b**), bromo-substituted (**3d**), and iodo-substituted (**3e**) arenes all delivered excellent levels of enantioselectivity, the latter being of particular note because it would likely be incompatible with palladium catalysis and is a testament to the mild conditions and functional group tolerance of iridium-catalyzed borylation. The *N*-trifluoroacetyl group could be replaced by acetyl with little drop in ee, as demonstrated on substrate **3c**. The absolute stereochemistry of compound **3e** was determined by x-ray crystallographic analysis, and all other compounds were assigned by analogy with this. Further variation of substituents revealed that trifluoromethoxy (**3f**), ester (**3g**), and nitrile (**3h**) were all well accommodated at the 3-position of the substrates. We also examined vicinally dichlorinated (**3i**) and difluorinated (**3j**) substrates, which both worked effectively. Substrates bearing electron-donating substituents exhibited lower reactivity under our conditions—3-methoxy gave no conversion and 3-methyl gave <5% conversion, likely owing to the reaction temperatures being lower than those typically used in C–H borylation. Performing the reactions at room temperature gave conversion, but with moderate enantioselectivity (fig. S4). The substrates examined so far have all presented no regioselectivity challenge, owing to the well-established preference for C–H borylation at the least hindered arene position (*42*). Given that the sulfonated bipyridine ligand scaffold was originally designed for the purpose of controlling regioselectivity in substrates that would typically be nonselective, we were keen to evaluate whether **L·1g** would be able to control both of these important selectivity factors for a substrate that possessed ortho-substituted aromatic rings (Fig. 2C) (*40*). We were concerned that the introduction of ortho substituents may substantially change the preferred substrate conformation, potentially affecting crucial interactions with the chiral cation. Also, it was possible that the complex chiral cation might disrupt the regioselectivity that we had previously observed when using tetrabutylammonium as the cation. However, we were delighted to find that an *ortho*-chloro substrate gave the *meta*-borylated product **3k** with excellent regioselectivity [10:1 regioisomeric ratio (rr)] and only a small reduction in enantioselectivity (85% ee). In contrast to this, the control borylation with standard borylation ligand **dtbpy** resulted in a 1.6:1 ratio of regioisomers (fig. S5). An *ortho*-bromo substrate performed similarly (**3l**), as did an *ortho*-CF_3 (**3m**) and *ortho*-OCF_3 (**3o**). We also examined a *meta*-fluoro substrate, which presents regioselectivity challenges using standard ligands owing to the small size of the fluorine atom (*42*), but with **L·1g**, high regioselectivity was observed (**3n**). In addition, we carried out preliminary experiments with nonsymmetrical substrates to assess the viability of using the reaction in kinetic resolution mode. These showed that it is indeed viable, although

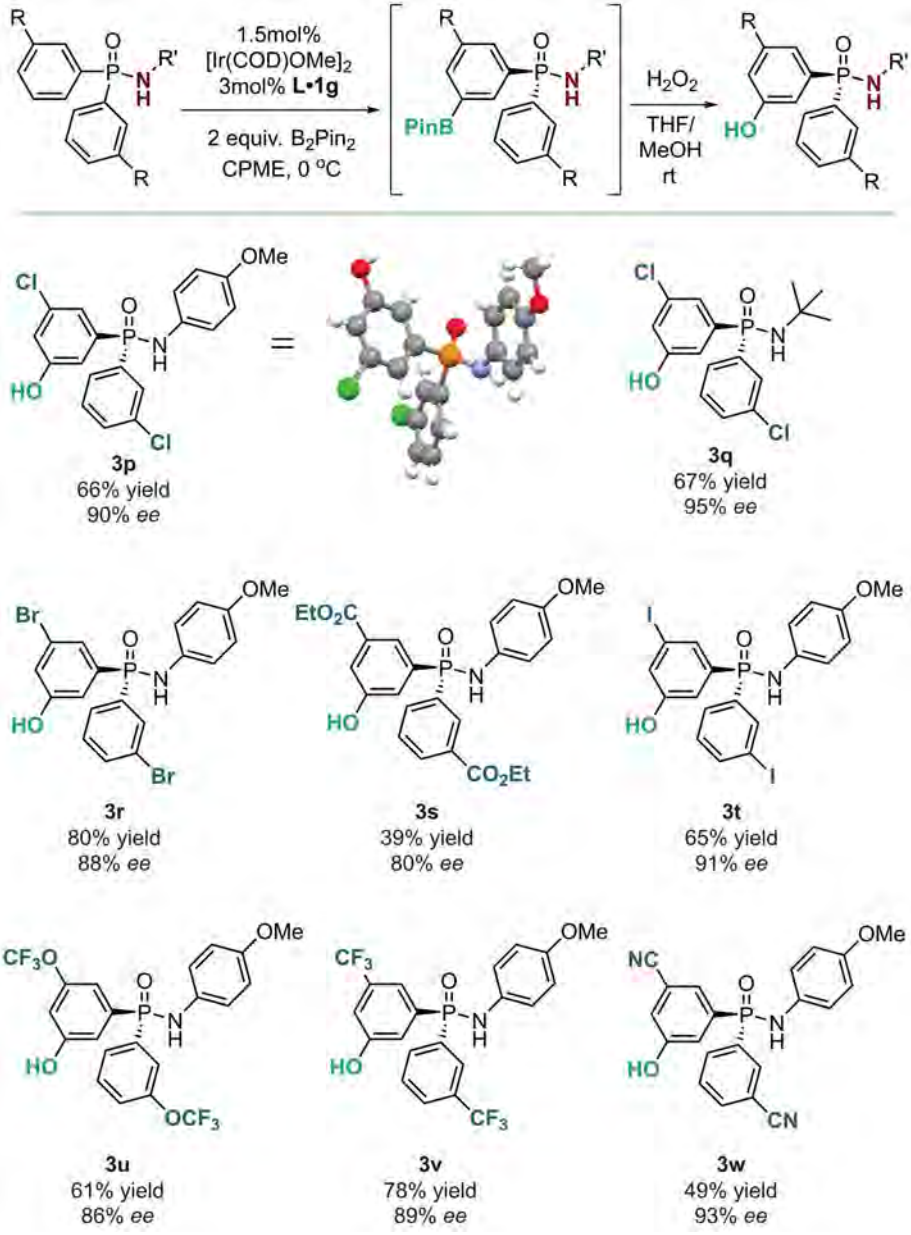


Fig. 3. Substrate scope of the enantioselective C–H borylation of diaryl phosphinamides. Yield values refer to isolated yields.

further investigations and optimization are likely required to enable this to be a general procedure (fig. S6).

At this stage, we envisaged that a compelling demonstration of the potential of this approach would be to successfully apply it to a different class of compound entirely. For this purpose, we identified symmetrical diarylphosphinamides, which contain a prochiral, configurationally stable phosphorus atom at the heart of the compound. We reasoned that such substrates would test our chiral cation–directed C–H borylation strategy in tackling an additional prominent challenge to synthetic chemists—that of how to synthesize P-chiral compounds in a catalytic asymmetric manner (*43*). Although there are several recently reported methods for enantioselective desymmetrizing C–H activation of phosphinamides using chiral Pd and Rh complexes, both result in ortho-functionalized products (*44*, *45*). Given the broad utility of P-chiral compounds in catalysis as well as increasingly in medicinal chemistry, we envisaged that remote desymmetrization would be of substantial practical utility (*46*). We were pleased to observe that a symmetrical phosphinamide, bearing a *para*-methoxy phenyl group on the phosphinamide nitrogen, was borylated to give **3p** with 90% ee using ligand **L·1g**, which had been optimal for the benzhydrylamide substrate class (Fig. 3). X-ray crystallographic analysis of **3p** showed that this product had analogous absolute stereochemistry to that obtained in the amide series, relative to the position of the NH hydrogen-bond donor. Experiments stopped at various conversions demonstrated that secondary kinetic resolution to form diborylated product is not contributing to the observed high enantioselectivity (fig. S7). N-substitution was found not to be limited to aromatic moieties, as demonstrated by *N-tert*-butyl substituted **3q** (95% ee). As in the amide substrate class, a variety of useful functional groups were tolerated on the aromatic ring, encompassing bromide (**3r**), ester (**3s**), iodide (**3t**), trifluoromethoxy (**3u**), trifluoromethyl (**3v**), and nitrile (**3w**). In some cases, yields are modest owing to poor substrate solubility under the reaction conditions (as in **3s**). There are numerous established avenues for the manipulation of the phosphinamide functional group in a stereospecific manner, such as to tertiary phosphine oxides, which have been amply demonstrated elsewhere (*44*). To test whether the catalyst may be able to influence both regioselectivity and enantioselectivity in this substrate class, we tested an ortho-substituted symmetrical phosphinamide but found that both outcomes were poor (fig. S8). We speculate that this may arise owing to the ortho-substituted aromatic ring and bulky nature of the quaternary phosphorous center, relative to the benzhydrylamide, having a conformational impact on the substrate that adversely affects crucial substrate-ligand interactions.

For both classes of compounds demonstrated, the C–H borylation products typically possess three versatile functional groups on the aromatic rings for further elaboration into complex scaffolds, at the heart of which lies the newly formed stereocenter. By virtue of the desymmetrization strategy used, two of these functional groups must necessarily be identical, and we sought to demonstrate that site selectivity between these in the product should be possible in many instances by electronic differentiation arising from introduction of the new substituent. In the first example, we carried out borylation and oxidation of dichloride **2b** to give the phenol **3b** with good yield and high enantioselectivity (Fig. 4A, upper scheme). By carrying out

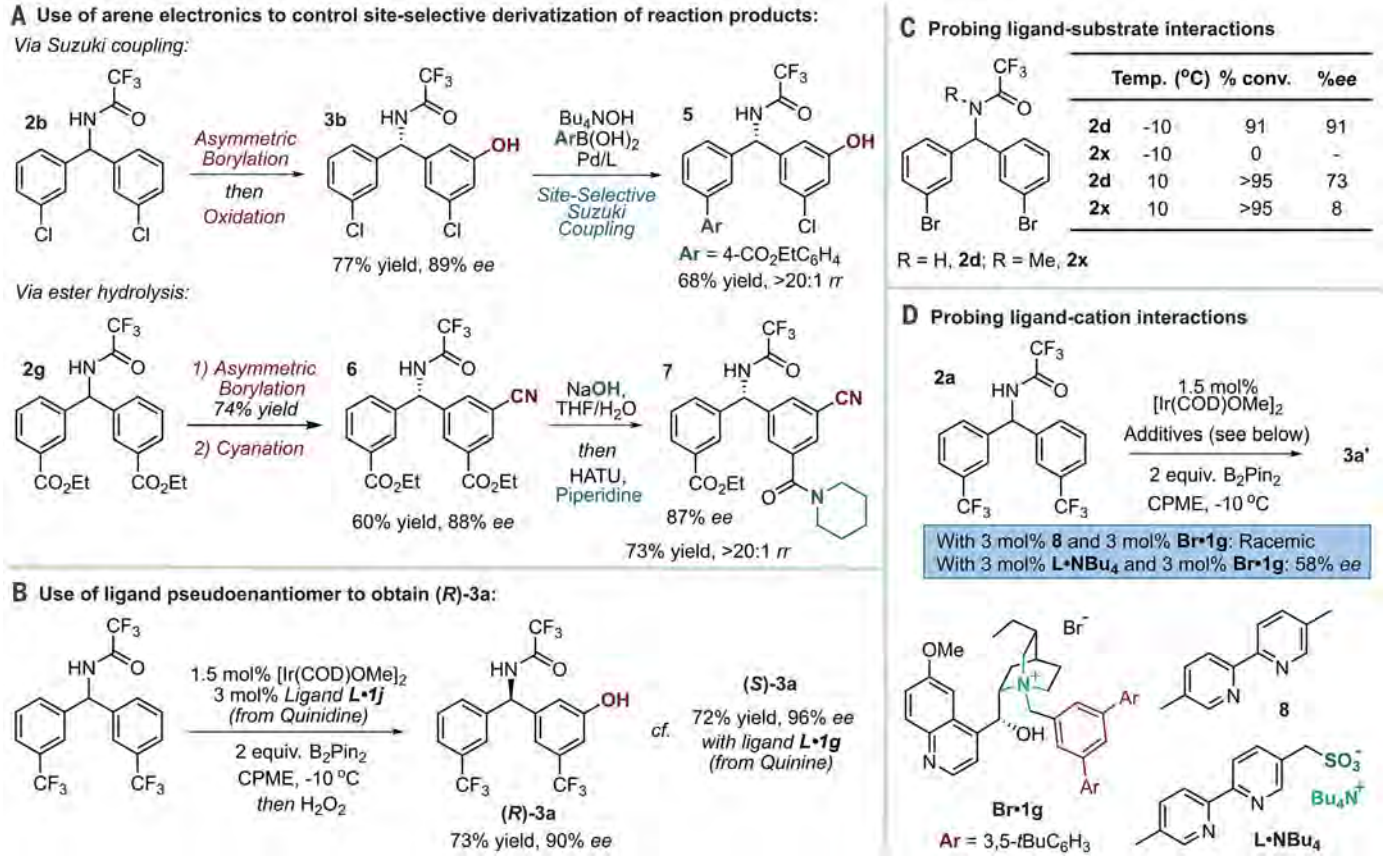


Fig. 4. Product elaboration and further experiments. (**A**) Use of arene electronics to control site-selective derivatization of reaction products. HATU, hexafluorophosphate azabenzotriazole tetramethyl uronium. (**B**) Use of a pseudoenantiomeric chiral cation to form (*R*)-**3a**. (**C**) Control experiments to probe ligand-substrate interactions. % conv., % conversion. (**D**) Control experiments to probe ligand-cation interactions. Yield values refer to isolated yields.

Suzuki-Miyaura coupling on **3b** in the presence of one equivalent of tetrabutylammonium hydroxide, we were able to achieve >20:1 site selectivity for cross-coupling on the nonphenolic aromatic ring. We anticipate that this is a result of the highly electron-rich nature of the in situ–generated phenolate disfavoring oxidative addition to the C–Cl bond on the same ring. In the second case, we carried out borylation of diester **2g** followed by cyanation to obtain **6** (Fig. 4A, lower scheme) (*47*). Careful treatment of **6** with NaOH selectively hydrolyzed the ester on the same ring as the nitrile owing to electronic factors that can be readily rationalized and predicted using substituent Hammett parameters (*48*), giving **7** in >20:1 rr after amide coupling.

To emphasize the practicality of our process, we demonstrated borylation using ligand **L·1j**, possessing a diastereomeric chiral cation derived from quinidine, the pseudoenantiomer of quinine. This proceeded smoothly, giving (*R*)-**3a** with 90% ee (Fig. 4B). Next, we performed experiments to probe the hydrogen-bonding interaction of substrate with ligand. The N-methylated variant (**2x**) of successful substrate **2d** underwent no borylation under the optimized conditions at –10°C, and the temperature had to be raised to 10°C to obtain product, which was found to have only 8% ee (Fig. 4C). This outcome highlights the importance of the hydrogen-bond donor in the substrate for both reactivity and selectivity, in line with our initial hypothesis (Fig. 1E). We also performed an experiment in which ion-paired ligand **L·1g** was replaced with neutral 5,5′-dimethylbipyridine (**8**) together with the optimal chiral cation as its bromide salt (**Br·1g**). The product was racemic, demonstrating the requirement for ligand and chiral cation to be associated to achieve enantioinduction. We also ran a reaction in which ligand **L·NBu$_4$**, bearing achiral tetrabutylammonium as the cation, was used in conjunction with **Br1g**. In this case, 58% ee was obtained in the product, consistent with some degree of counterion exchange occurring between the two, leading to moderate enantioinduction.

We have demonstrated a strategy for pairing privileged chiral cations with an iridium-bipyridine complex, enabled by incorporation of an anionic sulfonate group into the ligand scaffold. In principle, numerous widely used transition metal–catalyzed reactions could be amenable to this approach, as evidenced by the numerous common ligand classes that have been sulfonated for the purpose of engendering water solubility (*49*). We anticipate that wider incorporation of chiral cations into mainstream transition metal catalysis could have broad implications in asymmetric organic synthesis.

REFERENCES AND NOTES

1. K. Brak, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **52**, 534–561 (2013).
2. U. H. Dolling, P. Davis, E. J. J. Grabowski, *J. Am. Chem. Soc.* **106**, 446–447 (1984).
3. H.-G. Park, B.-S. Jeong, in *Cinchona Alkaloids in Synthesis and Catalysis*, C. E. Song, Ed. (Wiley, 2009), pp. 131–169.
4. T. Ooi, K. Maruoka, *Angew. Chem. Int. Ed.* **46**, 4222–4266 (2007).
5. S. Shirakawa, K. Maruoka, *Angew. Chem. Int. Ed.* **52**, 4312–4348 (2013).
6. M. Uyanik, H. Okamoto, T. Yasui, K. Ishihara, *Science* **328**, 1376–1379 (2010).
7. M. T. Oliveira, J.-W. Lee, *ChemCatChem* **9**, 377–384 (2017).
8. R. J. Phipps, G. L. Hamilton, F. D. Toste, *Nat. Chem.* **4**, 603–614 (2012).
9. M. Mahlau, B. List, *Angew. Chem. Int. Ed.* **52**, 518–533 (2013).
10. S. Mayer, B. List, *Angew. Chem. Int. Ed.* **45**, 4193–4195 (2006).
11. G. L. Hamilton, T. Kanai, F. D. Toste, *J. Am. Chem. Soc.* **130**, 14984–14986 (2008).
12. G. L. Hamilton, E. J. Kang, M. Mba, F. D. Toste, *Science* **317**, 496–499 (2007).
13. S. Mukherjee, B. List, *J. Am. Chem. Soc.* **129**, 11336–11337 (2007).
14. S. Liao, B. List, *Angew. Chem. Int. Ed.* **49**, 628–631 (2010).
15. X. Ye *et al.*, *Angew. Chem. Int. Ed.* **55**, 7101–7105 (2016).
16. L. Zong *et al.*, *Nat. Commun.* **7**, 13455 (2016).
17. R. C. D. Brown, J. F. Keily, *Angew. Chem. Int. Ed.* **40**, 4496–4498 (2001).
18. C. Wang, L. Zong, C.-H. Tan, *J. Am. Chem. Soc.* **137**, 10677–10682 (2015).
19. L. Zong, C.-H. Tan, *Acc. Chem. Res.* **50**, 842–856 (2017).
20. S. Paria, H.-J. Lee, K. Maruoka, *ACS Catal.* **9**, 2395–2399 (2019).
21. K. Ohmatsu, T. Ooi, *Top. Curr. Chem.* **377**, 31 (2019).
22. K. Ohmatsu, N. Imagawa, T. Ooi, *Nat. Chem.* **6**, 47–51 (2014).
23. K. Ohmatsu, S. Kawai, N. Imagawa, T. Ooi, *ACS Catal.* **4**, 4304–4306 (2014).
24. K. Ohmatsu, M. Ito, T. Kunieda, T. Ooi, *Nat. Chem.* **4**, 473–477 (2012).
25. K. Ohmatsu, M. Ito, T. Kunieda, T. Ooi, *J. Am. Chem. Soc.* **135**, 590–593 (2013).
26. X.-F. Cheng *et al.*, *J. Am. Chem. Soc.* **135**, 1236–1239 (2013).
27. L. Chu, K.-J. Xiao, J.-Q. Yu, *Science* **346**, 451–455 (2014).
28. B. Ye, N. Cramer, *Science* **338**, 504–506 (2012).
29. B. Ye, N. Cramer, *Acc. Chem. Res.* **48**, 1308–1318 (2015).
30. B. Su, T.-G. Zhou, P.-L. Xu, Z.-J. Shi, J. F. Hartwig, *Angew. Chem. Int. Ed.* **56**, 7205–7208 (2017).
31. X. Zou *et al.*, *J. Am. Chem. Soc.* **141**, 5334–5342 (2019).
32. C. G. Newton, S.-G. Wang, C. C. Oliveira, N. Cramer, *Chem. Rev.* **117**, 8908–8976 (2017).
33. H. Shi, A. N. Herron, Y. Shao, Q. Shao, J.-Q. Yu, *Nature* **558**, 581–585 (2018).
34. J.-Y. Cho, M. K. Tse, D. Holmes, R. E. Maleczka Jr., M. R. Smith III, *Science* **295**, 305–308 (2002).
35. T. Ishiyama *et al.*, *J. Am. Chem. Soc.* **124**, 390–391 (2002).
36. B. Kim *et al.*, *J. Am. Chem. Soc.* **138**, 7939–7945 (2016).
37. A. J. Metrano, S. J. Miller, *Acc. Chem. Res.* **52**, 199–215 (2019).
38. S. Y. Hsieh, Y. Tang, S. Crotti, E. A. Stone, S. J. Miller, *J. Am. Chem. Soc.* **141**, 18624–18629 (2019).
39. H. J. Davis, M. T. Mihai, R. J. Phipps, *J. Am. Chem. Soc.* **138**, 12759–12762 (2016).
40. H. J. Davis, G. R. Genov, R. J. Phipps, *Angew. Chem. Int. Ed.* **56**, 13351–13355 (2017).
41. M. T. Mihai, H. J. Davis, G. R. Genov, R. J. Phipps, *ACS Catal.* **8**, 3764–3769 (2018).
42. J. F. Hartwig, *Chem. Soc. Rev.* **40**, 1992–2002 (2011).
43. M. Dutartre, J. Bayardon, S. Jugé, *Chem. Soc. Rev.* **45**, 5771–5794 (2016).
44. Z.-J. Du *et al.*, *J. Am. Chem. Soc.* **137**, 632–635 (2015).
45. Y. Sun, N. Cramer, *Angew. Chem. Int. Ed.* **56**, 364–367 (2017).
46. D. A. DiRocco *et al.*, *Science* **356**, 426–430 (2017).
47. C. W. Liskey, X. Liao, J. F. Hartwig, *J. Am. Chem. Soc.* **132**, 11389–11391 (2010).
48. C. Hansch, A. Leo, R. W. Taft, *Chem. Rev.* **91**, 165–195 (1991).
49. K. H. Shaughnessy, *Chem. Rev.* **109**, 643–710 (2009).

ACKNOWLEDGMENTS

We are grateful to A. Bond (University of Cambridge) for solving and refining the x-ray crystal structures, the EPSRC UK National Mass Spectrometry Facility at Swansea University, M. Gaunt for useful discussion and use of equipment, and C. Hunter for useful discussion. We also thank A. Turner (AstraZeneca) and P. Seden (Syngenta) for useful discussion. **Funding:** We are grateful to The Royal Society for a University Research Fellowship (R.J.P., UF130004), the EPSRC (EP/N005422/1), the European Research Council under the Horizon 2020 Program (Starting Grant no. 757381), the EPSRC and AstraZeneca for a CASE studentship (J.L.D.), The Emil Aaltonen Foundation for a fellowship (A.S.K.L.), and the EPSR, and Syngenta for a CASE studentship (D.C.G.). **Author contributions:** G.R.G. developed the catalysts and reactions and performed and analyzed experiments. J.L.D. developed the reactions and performed and analyzed experiments. A.S.K.L and D.C.G. performed and analyzed experiments. R.J.P. conceived and supervised the project and wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The supplementary materials contain additional spectral and chromatographic data. Crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC 1959892 and CCDC 1960549.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1246/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 and S2
HPLC and SFC Traces
NMR Spectra
References (*50*–*63*)

8 November 2019; accepted 19 February 2020
10.1126/science.aba1120

WATER RESOURCES

Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation

P. C. D. Milly* and K. A. Dunne

The sensitivity of river discharge to climate-system warming is highly uncertain, and the processes that govern river discharge are poorly understood, which impedes climate-change adaptation. A prominent exemplar is the Colorado River, where meteorological drought and warming are shrinking a water resource that supports more than 1 trillion dollars of economic activity per year. A Monte Carlo simulation with a radiation-aware hydrologic model resolves the longstanding, wide disparity in sensitivity estimates and reveals the controlling physical processes. We estimate that annual mean discharge has been decreasing by 9.3% per degree Celsius of warming because of increased evapotranspiration, mainly driven by snow loss and a consequent decrease in reflection of solar radiation. Projected precipitation increases likely will not suffice to fully counter the robust, thermodynamically induced drying. Thus, an increasing risk of severe water shortages is expected.

The Upper Colorado River Basin (UCRB) supplies water to ~40 million people and supports ~16 million jobs (*1*). Atmospheric warming and recent precipitation deficits have heightened concern about the future (*2*–*6*), but the response of river discharge to warming remains highly uncertain. An implicit assumption in the literature on UCRB hydroclimatic change is that two climatic mean variables—precipitation and temperature—determine runoff (hence, river discharge) response, following constant sensitivities α [percent discharge change per percent precipitation change (dimensionless)] and β [percent discharge change per degree Celsius of warming (% $°C^{-1}$)]. Empirical regression analyses imply large values of β (−13 to −15% $°C^{-1}$) (*4*, *6*–*8*), which is inconsistent with estimates in the range −2 to −9% $°C^{-1}$ obtained from perturbation of temperature inputs (the delta method) to hydrologic model simulations (*2*, *9*, *10*) and from theory (*11*). For α, regression and delta estimates are in much better agreement (*10*). The discrepancy in β, which is seen for rivers around the globe (*11*), translates into great uncertainty in the magnitude of future effects on human livelihood, economic activity, and ecosystem health. The situation is exacerbated by limited process understanding in the presence of hydroclimatic nonstationarity (*12*). The empiricism that is inherent in the regression approach, and even that which is inherent in the estimation of energy-driven evaporative demand in the hydrologic models (*13*), leaves the use of such methods for extrapolation of past observations to the future, under anthropogenic climate change, open to question. Accordingly, we gave special attention to surface net radiation—the ultimate driver of evapotranspiration—and to its modulation by snow-affected surface albedo (*14*) rather than relying on temperature measurements as a surrogate for energy availability. We found a strong influence of snow-affected albedo on radiation balance in the UCRB (Fig. 1) (*15*), which necessitated its consideration in a process-based estimation of β.

Herein, we address the following questions, in turn, by use of a monthly water-balance model grounded in a suite of observations: Does the model reproduce the historical regression-based β? What is the model's delta-based β, and why does it differ from the regression-based value? Can the two values be reconciled? What physical processes control β? How sensitive is our β estimate to the assumptions in our analysis? How much did warming contribute to the historical hydrological drying in the UCRB? What future changes in UCRB discharge can be expected?

In addition to the snow-water equivalent (SWE), albedo, and radiation measurements used to develop the relations in Fig. 1, we used observations of precipitation and temperature

U.S. Geological Survey, Princeton, NJ, USA.
*Corresponding author. Email: cmilly@usgs.gov

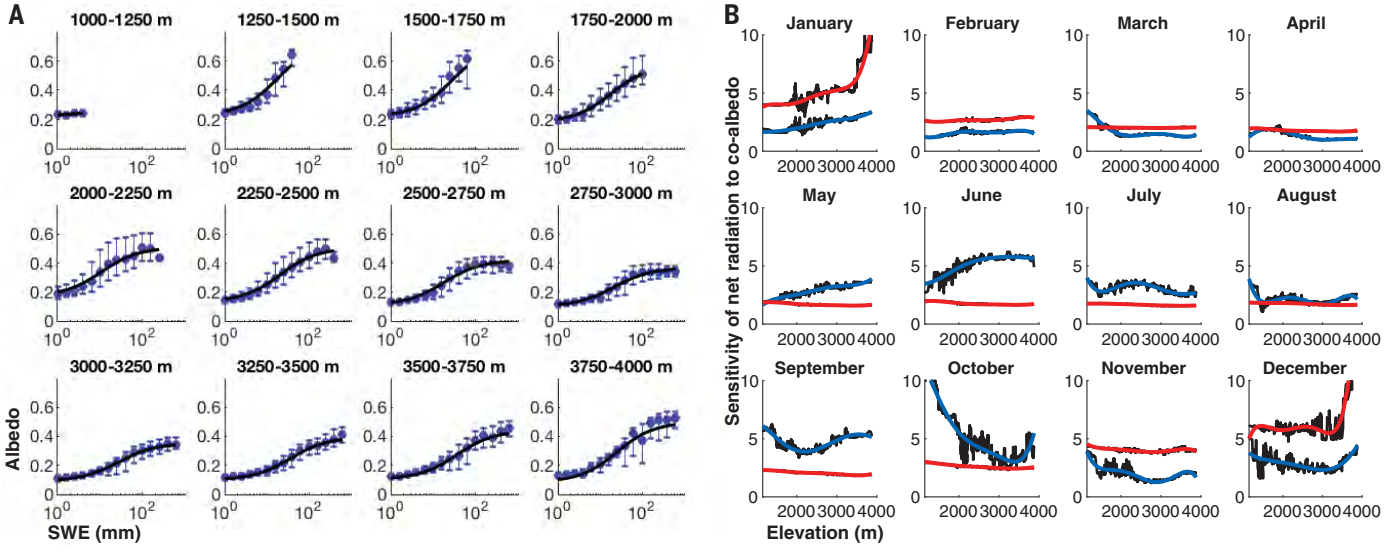


Fig. 1. Observed relations among monthly SWE, surface albedo, and surface net radiation in the UCRB. (**A**) Dependence of surface albedo on SWE (logarithmic scale) for each of 12 elevation ranges. 1st, 2nd, and 3rd quartiles of binned data are shown. Curves are least-squares fits to the unbinned data and are used in the model. (**B**) Inferred dimensionless sensitivity $\frac{\bar{C}}{\bar{R}_n}\frac{dR_n}{dC}$ of net radiation R_n to co-albedo (one minus albedo) C as a function of mean elevation of 960 subareas by month of the year. Blue curves are fitted to smoothed (30-point moving median; black) data from empirical regression estimates. Red curves are analogous fits for theoretical case where a change in absorbed solar radiation causes no radiative feedbacks. Fits to regressions are used in the model, except that fits to no-feedback data are used in a sensitivity experiment.

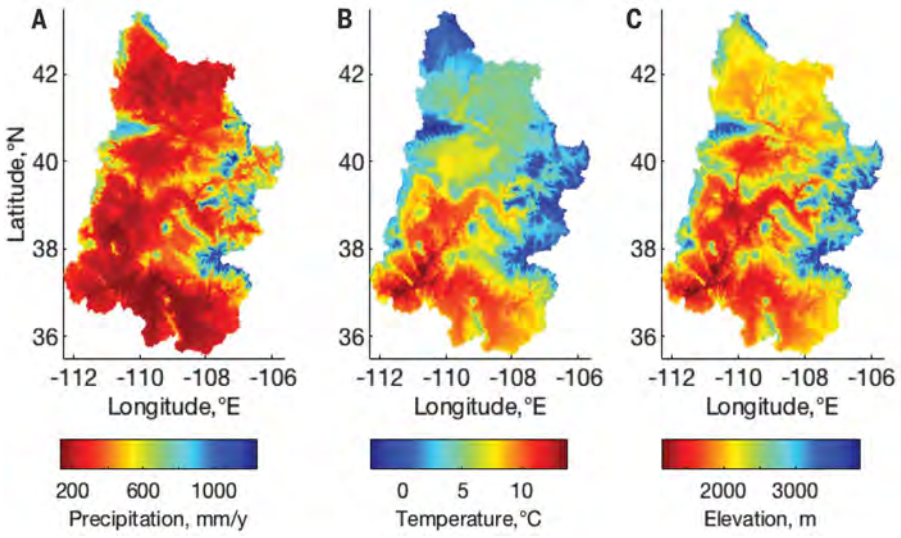


Fig. 2. Spatial distributions of annual climate variables and elevation over the UCRB, as resolved by the model discretization of space into 960 subareas. (**A**) Total precipitation. (**B**) Mean temperature. (**C**) Elevation.

(Fig. 2 and Fig. 3, A and B) as well as discharge (Fig. 3G) to constrain a hydrologic simulation model (*15*), which we used to elucidate the processes that control sensitivity and to reconcile divergent, previously published sensitivity estimates. The model has a monthly time step and divides the 290,000-km^2 UCRB into 960 subareas to capture the strong heterogeneity induced by rugged (2700-m relief) topography (Fig. 2C). Rain-snow partitioning depends on temperature. Evaporative potential is set to the rate of non–water-stressed evapotranspiration under conditions of minimal advection (*16*). Fifteen model parameters were estimated by maximizing goodness of fit to observed discharge (*15*). We measured goodness of fit with respect to mean, linear trend, regression-based sensitivities α and β, and Nash-Sutcliffe coefficient of efficiency. [Including a correction that accounts for temporary subsurface storage of runoff before entering the river (*11*), which has previously been neglected, and using an October to September water year, we found observational regression-based α and β, ± one standard error of estimation, to be 1.98 ± 0.16 and −16.1 ± 2.9% °C^{-1}, respectively. Neglecting the storage correction yields β = −13.1 ± 2.4% °C^{-1}, consistent with earlier analyses.] The sensitivity of our results to the goodness-of-fit criteria is presented in the supplementary materials (*15*).

Of 500,000 trial parameter sets, 171 satisfied the goodness-of-fit criteria (*15*), and these formed a model ensemble for subsequent analyses. As the temperature rose, the ensemble-mean SWE and—hence, following the relations in Fig. 1—albedo decreased, which led to a basin-mean increase of net radiation by 3.0% per century over the study period (Fig. 3, C to E). With an associated increase in evapotranspiration (Fig. 3F), the ensemble mean–modeled annual discharge (Fig. 3G) fell by 20.1% per century, compared with 19.6% per century observed; the square of the correlation coefficient (r^2) between the observed and ensemble mean–modeled annual discharge is 0.82. Within the ensemble, the models' regression-based, storage-corrected sensitivities α and β ranged from α = 1.89 ± 0.16 to 2.08 ± 0.18 (mean, 1.99) and β = −15.4 ± 2.9 to −16.9 ± 3.0% °C^{-1} (mean, −15.9% °C^{-1}), which is consistent with observational estimates.

The ensemble was rerun with the temperature increased by 1°C, and differences from the base simulations were used to estimate sensitivities. The delta-based β ranged from −7.8 to −12.2% °C^{-1} (mean, −9.3% °C^{-1}), which is consistent with higher-magnitude values from previous delta-based analyses (*2*, *10*) and lower than regression-based estimates. Simulations with precipitation perturbed by +1% each month yielded a delta-based α of 2.21 to 2.83 (mean, 2.52). It is not surprising that some previous delta estimates of β were as high as ours nor that some were lower, because models had a variety of features, with varying physical realism, differentially sensitizing their evapotranspiration to temperature, and the mechanisms underlying β generally were neither identified nor constrained with measurements.

We found that the difference between the model's regression- and delta-based sensitivities is explained by confounding variables: seasonal shifts of precipitation that historically accompanied annual anomalies of temperature. During warm water years, precipitation tended to shift from December–April to August–September. Because discharge sensitivity to precipitation is strong in winter (when extra precipitation tends to run off) (*3*) and weak in summer (when extra precipitation

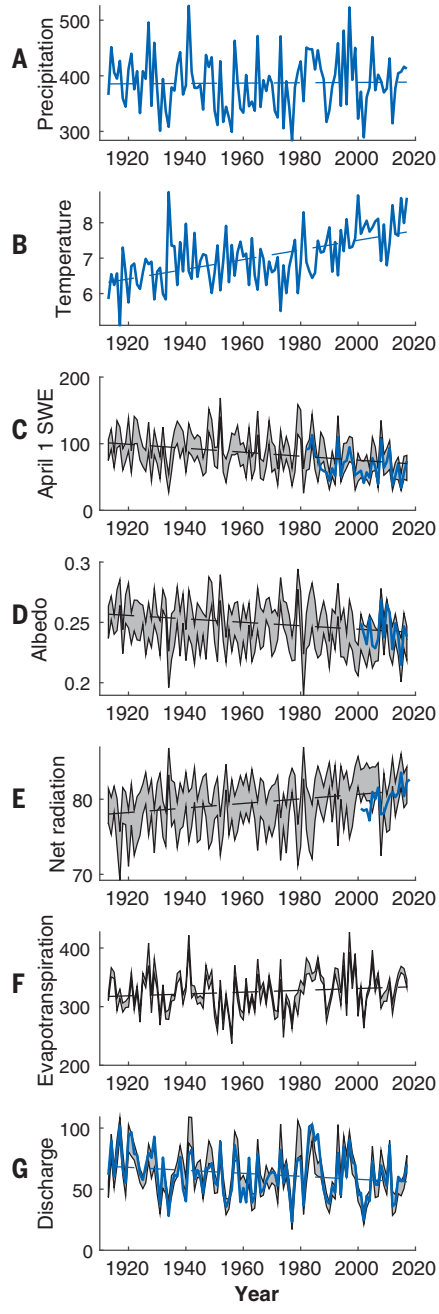


Fig. 3. Water-year time series of basin-mean, annual-mean values. (**A** to **G**) Precipitation (millimeters per year) (A), temperature (degrees Celsius) (B), April 1 SWE (millimeters) (C), surface albedo (D), surface net radiation (watts per square meter) (E), evapotranspiration (millimeters per year) (F), and discharge per unit area (millimeters per year) (G). Blue curves represent estimates from observations, and gray bands represent ensemble range of model outputs. Least-squares linear fits also are shown.

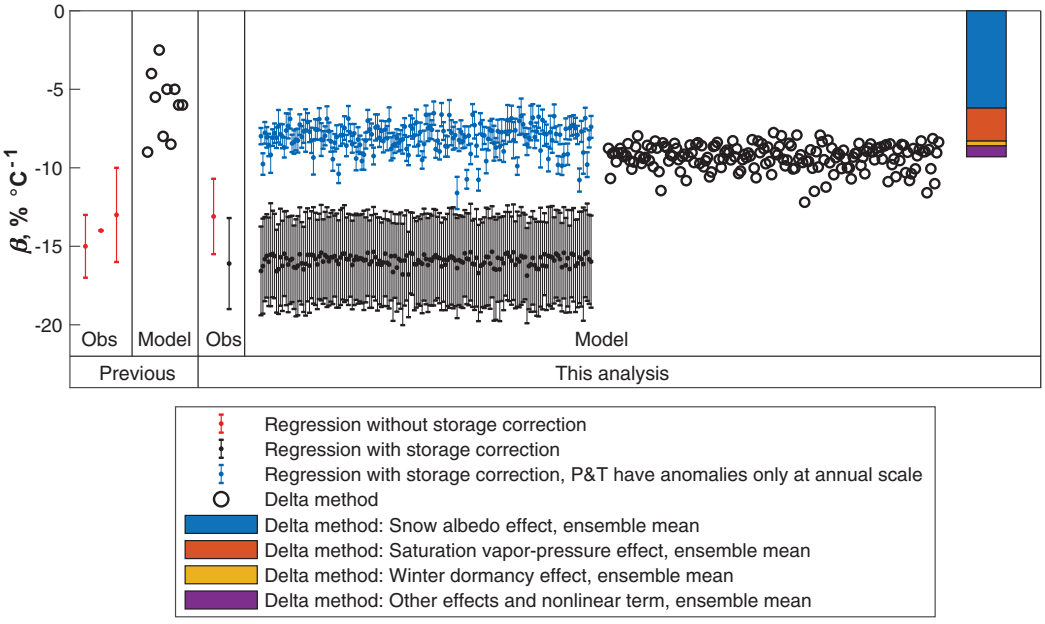


Fig. 4. Summary of estimates of β from previous studies and from this analysis. Left to right: previous observational (Obs) and model analyses (*2*, *4*, *6*–*10*) and results from this analysis. Error bars represent ± one standard error of estimation from the regressions. The multicolored bar shows the contribution of each of the temperature-sensitive mechanisms to the magnitude of β. Excluded as unrealistic from the previous delta analyses are cases in which maximum daily temperature was perturbed whereas minimum was not (*10*). The label "P&T have anomalies only at annual scale" refers to the computations in which the monthly course of precipitation in every year was set to climatology times a factor that preserved the observed annual anomaly.

tends to evaporate), runoff is suppressed during warm years. To remove the confounding variables from our comparison of the delta and regression estimates, we modified the original experiments so that the monthly course of precipitation in every year was set to climatology times a factor that preserved the observed annual anomaly; for temperature, the annual anomalies were applied as additive constants to the climatology. For these experiments, the models' regression-based α and β were α = 2.26 ± 0.03 to 3.00 ± 0.08 (mean, 2.59 ± 0.04) and β = −6.5 ± 0.7 to −11.6 ± 1.0% °C^{-1} (mean, −8.1 ± 0.7% °C^{-1}), which is in reasonable agreement with the delta-based values; the difference between −8.1 ± 0.7 and −9.3 is only marginally significant, and allowances must be made for the simple formulation of the storage correction for the regression estimate (*15*).

The substantial dependence of inferred sensitivities on seasonal distributions of climate perturbations implies that the use of simple annual sensitivity parameters (α and/or β) can severely distort climate-change analyses. This is a shortcoming of both the regression and delta approaches. With regression, the derived sensitivities depend on basin-specific historical intra-annual and interannual variability, including the confounding precipitation-temperature covariance. In the usual delta approach, the perturbations have no seasonal variations, and the roles of precipitation and temperature are decoupled, so delta sensitivities are more readily interpreted. However, the best approach for hydroclimatic projections is to use the delta approach with projected monthly varying climate changes.

To understand the ensemble-mean magnitude −9.3% °C^{-1} of the delta-based β and its potential relevance for ongoing anthropogenic climate change, we consider the physical processes at play. Temperature enters the model in four ways: (i) Because SWE depends on the phase of precipitation and the rate of snow melt, the surface albedo and, hence, the evaporative potential are temperature-dependent. (ii) The maximum fraction of net radiation that is converted to latent heat flux depends on the temperature-dependent slope of the saturation vapor-pressure curve (*11*, *16*). (iii) Evapotranspiration from soil ceases below a critical temperature, simulating winter dormancy of vegetation. (iv) Temperature affects the timing of snow melt and, thus, causes differences in sublimation and evapotranspiration in the model. By disabling these processes one at a time, we found that the contributions from the first three processes were −6.2, −2.1, and −0.3% °C^{-1}, respectively, and other snow-storage effects and nonlinear interactions accounted for the remainder. Figure 4 summarizes the foregoing reconciliation of sensitivity estimates.

We repeated the analysis under the assumption that a change in albedo induces negligible radiation feedbacks (Fig. 1B, red). We found an ensemble mean β of −7.8% °C^{-1}, indicating that our findings are somewhat sensitive to uncertainties in albedo-radiation feedback.

Unaccounted factors in our analysis include externally driven changes in radiation (e.g., from changing atmospheric composition), changes in boundary-layer entrainment (*17*), and stomatal responses to CO_2 fertilization (*18*). The latter two factors tend to decrease the efficiency of the conversion of net radiation to potential evapotranspiration. We found the potential net effect of these factors on β to be negligible (*15*).

Our parameterization of potential evapotranspiration by use of the Priestley-Taylor formulation (*16*), which allowed for no atmospheric aridity feedback caused by actual (nonpotential) evapotranspiration, could be questioned. We therefore repeated our analysis with allowance for this feedback (*15*), finding a negligible difference in results. Another caveat to consider is that our adoption of the Priestley-Taylor formulation, even when we consider the aridity feedback, implicitly assumes that variabilities (in particular, long-term trends) of wind speed and humidity will not affect the value of the Priestley-Taylor α, even though they do, on certain time scales, play a documented role in variabilities of pan evaporation (*19*) and of the American Society of Civil Engineers Standardized Reference Evapotranspiration (*20*).

How much have temperature changes contributed to the period-of-record discharge trend (−19.6 and −20.1% per century observed and modeled, respectively) and the 2000 to 2017 discharge deficit (−15.9 and −17.6% of previous mean observed and modeled, respectively)? If we set temperature every year to its climatology, the model yields a discharge trend of −8.4% per century and a discharge deficit of −8.1%. We conclude that temperature sensitivity accounts for more than half of both drying phenomena, which is consistent with a previous analysis (*21*).

What about the future? To characterize future temperature and precipitation, we used the 8 out of 24 Coupled Model Intercomparison Project Phase 5 (CMIP5) climate models that simulated 1913 to 2017 discharge (area-weighted runoff) within a factor of two of the observed discharge. (The constraints used for climate-model selection were much less stringent than those used for selection of hydrologic model parameter sets because the hydrologic model was driven by historical climate time

series, whereas the modeled climate time series were biased and independent of actual history.) From CMIP5's historical, Representative Concentration Pathway 4.5 (RCP4.5), and RCP8.5 scenarios, we computed month-of-year temperature climatology increases from 1913–2017 to 2036–2065, added them to the observed historical record, and reran the ensemble. Across the set of eight climate models, ensemble-mean discharge decreased 14 to 26% (RCP4.5) and 19 to 31% (RCP8.5).

Could possible future increases in precipitation counteract the temperature-driven drying? When month-of-year temperature increases and precipitation ratios from the climate models were both applied, the ensemble-mean discharge decreased 5 to 24% (RCP4.5); under RCP8.5, changes ranged from an increase of 3% to a decrease of 40%. Thus, it appears unlikely that precipitation changes will be sufficient to fully counter the temperature-induced drying, though they might moderate it.

Many water-stressed regions around the world depend on runoff from seasonally snow-covered mountains, and more than one-sixth of the global population relies on seasonal snow and glaciers for water supply (*22*). It has been well established that snowpack serves as a reservoir that beneficially regulates the timing of water availability (*23*). Our findings imply that snow cover is also a protective shield that limits radiation absorption by, and consequently evaporative losses from, this natural reservoir; incidentally, this explains the observed phenomenon of precipitation as snowfall favoring runoff (*24*). The progressive diminution of this ecosystem service as a result of climate change will have a deleterious effect on water availability in snow-fed regions that are already stressed, including the UCRB.

REFERENCES AND NOTES

1. T. James, A. Evans, E. Madly, C. Kelly, "The economic importance of the Colorado River to the basin region" (L. William Seidman Research Institute, Arizona State Univ., 2014); https://businessforwater.org/wp-content/uploads/2016/12/PTF-Final-121814.pdf.
2. G. J. McCabe, D. W. Wolock, *Geophys. Res. Lett.* **34**, L22708 (2007).
3. C. A. Woodhouse, G. T. Pederson, K. Morino, S. A. McAfee, G. J. McCabe, *Geophys. Res. Lett.* **43**, 2174–2181 (2016).
4. G. J. McCabe, D. M. Wolock, G. T. Pederson, C. A. Woodhouse, S. McAfee, *Earth Interact.* **21**, 1–14 (2017).
5. B. Udall, J. Overpeck, *Water Resour. Res.* **53**, 2404–2418 (2017).
6. J. A. Vano *et al.*, *Bull. Am. Meteorol. Soc.* **95**, 59–78 (2014).
7. R. R. Revelle, P. E. Waggoner, in *Changing Climate: Report of the Carbon Dioxide Assessment Committee* (National Academy Press, 1983), pp. 418–432.
8. K. Nowak, M. Hoerling, B. Rajagopalan, E. Zagona, *J. Clim.* **25**, 4389–4403 (2012).
9. L. L. Nash, P. H. Gleick, "The Colorado River basin and climatic change: The sensitivity of streamflow and water supply variations to temperature and precipitation" (U.S. Environmental Protection Agency, EPA 230-R-93-009, 1993).
10. J. A. Vano, T. Das, D. P. Lettenmaier, *J. Hydrometeorol.* **13**, 932–949 (2012).
11. P. C. D. Milly, J. Kam, K. A. Dunne, *Water Resour. Res.* **54**, 2624–2641 (2018).
12. P. C. D. Milly *et al.*, *Science* **319**, 573–574 (2008).
13. P. C. D. Milly, K. A. Dunne, *J. Am. Water Resour. Assoc.* **53**, 822–838 (2017).
14. F. G. Hall *et al.*, in *Vegetation, Water, Humans and the Climate: A New Perspective on an Interactive System*, P. Kabat *et al.*, Eds. (Springer, 2004), pp. 93–114.
15. Methods and supplementary text can be found in the supplementary materials.
16. C. H. B. Priestley, R. J. Taylor, *Mon. Weather Rev.* **100**, 81–92 (1972).
17. P. A. Dirmeyer, Y. Jin, B. Singh, X. Yan, *J. Hydrometeorol.* **14**, 829–849 (2013).
18. P. C. D. Milly, K. A. Dunne, *Nat. Clim. Chang.* **6**, 946–949 (2016).
19. M. L. Roderick, L. D. Rotstayn, G. D. Farquhar, M. T. Hobbins, *Geophys. Res. Lett.* **34**, L17403 (2007).
20. M. T. Hobbins, *Trans. Am. Soc. Ag. and Biol. Engineers* **59**, 561–576 (2016).
21. M. Xiao, B. Udall, D. P. Lettenmaier, *Water Resour. Res.* **54**, 6739–6756 (2018).
22. T. P. Barnett, J. C. Adam, D. P. Lettenmaier, *Nature* **438**, 303–309 (2005).
23. I. T. Stewart, D. R. Cayan, M. D. Dettinger, *Clim. Change* **62**, 217–232 (2004).
24. W. R. Berghuijs, R. A. Woods, M. Hrachowitz, *Nat. Clim. Chang.* **4**, 583–586 (2014).

ACKNOWLEDGMENTS

This study was facilitated by the Geophysical Fluid Dynamics Laboratory of the National Oceanic and Atmospheric Administration and by several data providers cited in the supplementary materials. The authors gratefully acknowledge colleague reviews by R. Koster and T. Delworth. **Funding:** The authors are supported by the U.S. Geological Survey. **Author contributions:** P.C.D.M. was responsible for the conceptualization and overall direction of the work and wrote the original draft. P.C.D.M. and K.A.D. carried out computations. K.A.D. performed data curation and reviewed the original draft. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** No original data collection was performed. The results of this study are reproducible and extensible by use of the cited data sources and other information in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1252/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Tables S1 to S5
References (*25–32*)
Data S1 to S3

24 August 2019; accepted 4 February 2020
Published online 20 February 2020
10.1126/science.aay9187

IMMUNOLOGY

mRNA destabilization by BTG1 and BTG2 maintains T cell quiescence

Soo Seok Hwang[1]*, Jaechul Lim[1]*, Zhibin Yu[1,2,3], Philip Kong[1], Esen Sefik[1], Hao Xu[1], Christian C. D. Harman[1], Lark Kyun Kim[4], Gap Ryol Lee[5], Hua-Bing Li[1,2,3], Richard A. Flavell[1,6]†

T cells maintain a quiescent state prior to activation. As inappropriate T cell activation can cause disease, T cell quiescence must be preserved. Despite its importance, the mechanisms underlying the "quiescent state" remain elusive. Here, we identify BTG1 and BTG2 (BTG1/2) as factors responsible for T cell quiescence. BTG1/2-deficient T cells show an increased proliferation and spontaneous activation due to a global increase in messenger RNA (mRNA) abundance, which reduces the threshold to activation. BTG1/2 deficiency leads to an increase in polyadenylate tail length, resulting in a greater mRNA half-life. Thus, BTG1/2 promote the deadenylation and degradation of mRNA to secure T cell quiescence. Our study reveals a key mechanism underlying T cell quiescence and suggests that low mRNA abundance is a crucial feature for maintaining quiescence.

Quiescence is a cellular state in which cells undergo reversible cell cycle arrest in response to environmental challenges or lack of stimuli (*1*). For T lymphocytes, quiescence is a ground state where they are poised in the absence of activation signals. In this state, T cells barely proliferate and maintain a low basal metabolic rate, which is required for survival (*2*, *3*). Upon encountering cognate antigen through T cell receptor (TCR)–costimulatory signals, naïve T cells exit quiescence and undergo clonal expansion and differentiation (*4*). Studies have shown that T cell quiescence is actively maintained, rather than a default state (*5*–*7*). In naïve T cells, FOXO1 and KLF2 control cell trafficking by modulating genes such as *Ccr7*, *S1pr1*, and *Sell* (*8*, *9*). Additionally, cytokines and receptors important for T cell quiescence are regulated by FOXO1, FOXP1, and TOB (*7*, *8*, *10*). Metabolic programming by TSC1 also supports T cell quiescence (*11*). Although these factors play crucial roles in T cell quiescence through specific genes or programs, how "minimalism" operates in quiescent T cells remains largely unknown.

To reveal additional factors responsible for orchestrating quiescence in T cells, we searched for genes regulated by super-enhancers [strong histone H3 lysine 27 acetylation (H3K27ac) signals] in quiescent T cells. Our rationale was that super-enhancers are closely associated with cell-type specificity (*12*). We found a super-enhancer linked to *BTG1* and *BTG2* in the top ~1% of hits, which was comparable to the percentage for *KLF2* and *IL7R* (Fig. 1A and fig. S1A). BTG1/2 belong to the B cell translocation gene/Transducer of ERBB2 (BTG/TOB) family (*13*, *14*), which shows antiproliferative activity and is implicated in numerous cellular processes (*14*, *15*). Notably, *BTG1/2* were highly expressed in lymph nodes and white blood cells among the BTG/TOB family, suggesting

[1]Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06510, USA. [2]Shanghai Institute of Immunology, Department of Microbiology and Immunology, Shanghai Jiao Tong University School of Medicine (SJTU-SM), Shanghai 200025, China. [3]Yale Center for ImmunoMetabolism, Shanghai Jiao Tong University School of Medicine (SJTU-SM), Shanghai 200025, China [4]Severance Biomedical Science Institute and BK21 PLUS Project for Medical Sciences, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul 06230, Republic of Korea. [5]Department of Life Science, Sogang University, Seoul 04107, Republic of Korea. [6]Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: richard.flavell@yale.edu

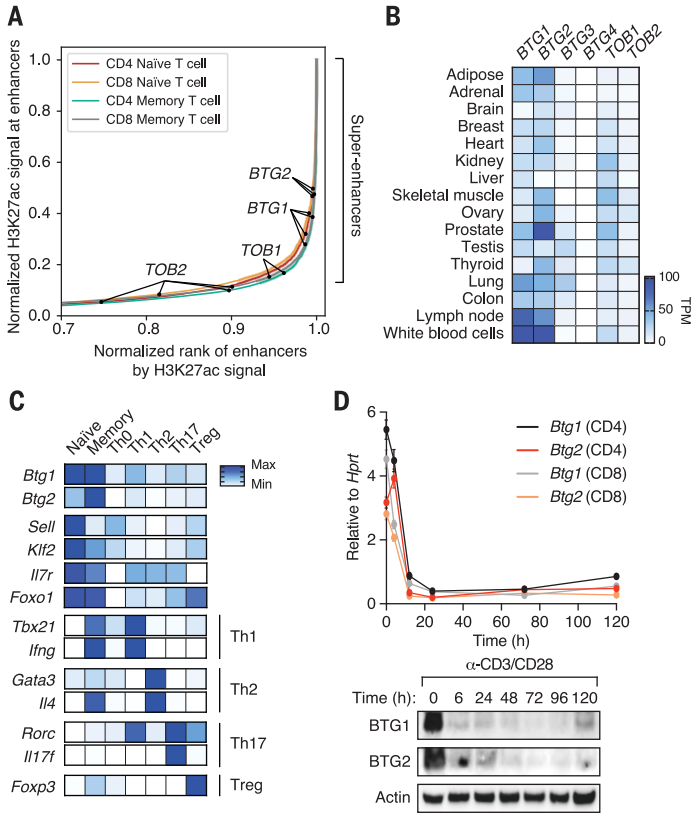


Fig. 1. BTG1 and BTG2 are mainly expressed in quiescent T cells. (**A**) Super-enhancers are ranked by normalized H3K27ac chromatin immunoprecipitation sequencing signals. Super-enhancer signals are adopted from (*12*). (**B**) Tissue-specific expression of BTG1/2 in the immune system. RNA-seq data are adopted from Illumina Body map 2.0 project (PRJEB2445). TPM, transcripts per million. (**C**) Relative expression of *Btg1/2* in naïve and memory CD4 T cells. (**D**) Decrease of *Btg1* and *Btg2* expression upon TCR stimulation. Mouse naïve CD4 and CD8 T cells were cultured with antibodies against CD3 (α-CD3) (5 μg/ml) and α-CD28 (2 μg/ml) for 5 days. The expression of *Btg1/2* was quantified by reverse transcription–quantitative polymerase chain reaction (RT-qPCR) (top) and immunoblotting (bottom), respectively. Error bars represent SEM (n = 2). Each symbol represents the mean of two independent experiments with two technical replicates.

their specific role in the immune system (Fig. 1B). The same pattern was observed in mouse tissues (fig. S1B). Next, we confirmed that *Btg1/2* were mainly expressed in naïve and memory T cells rather than in other T cell subsets (Fig. 1C). This distinct pattern was similar to that of known quiescence markers, such as *Foxo1, Il7r*, and *Klf2*. Like those factors, both mRNA and protein abundances of *Btg1/2* were rapidly reduced upon TCR stimulation and remained low during the activated state (Fig. 1D and fig. S1C). Thus, *Btg1* and *Btg2* are selectively expressed in quiescent T cells.

To understand their roles in vivo, we generated *Btg1* and *Btg2* conditional knockout

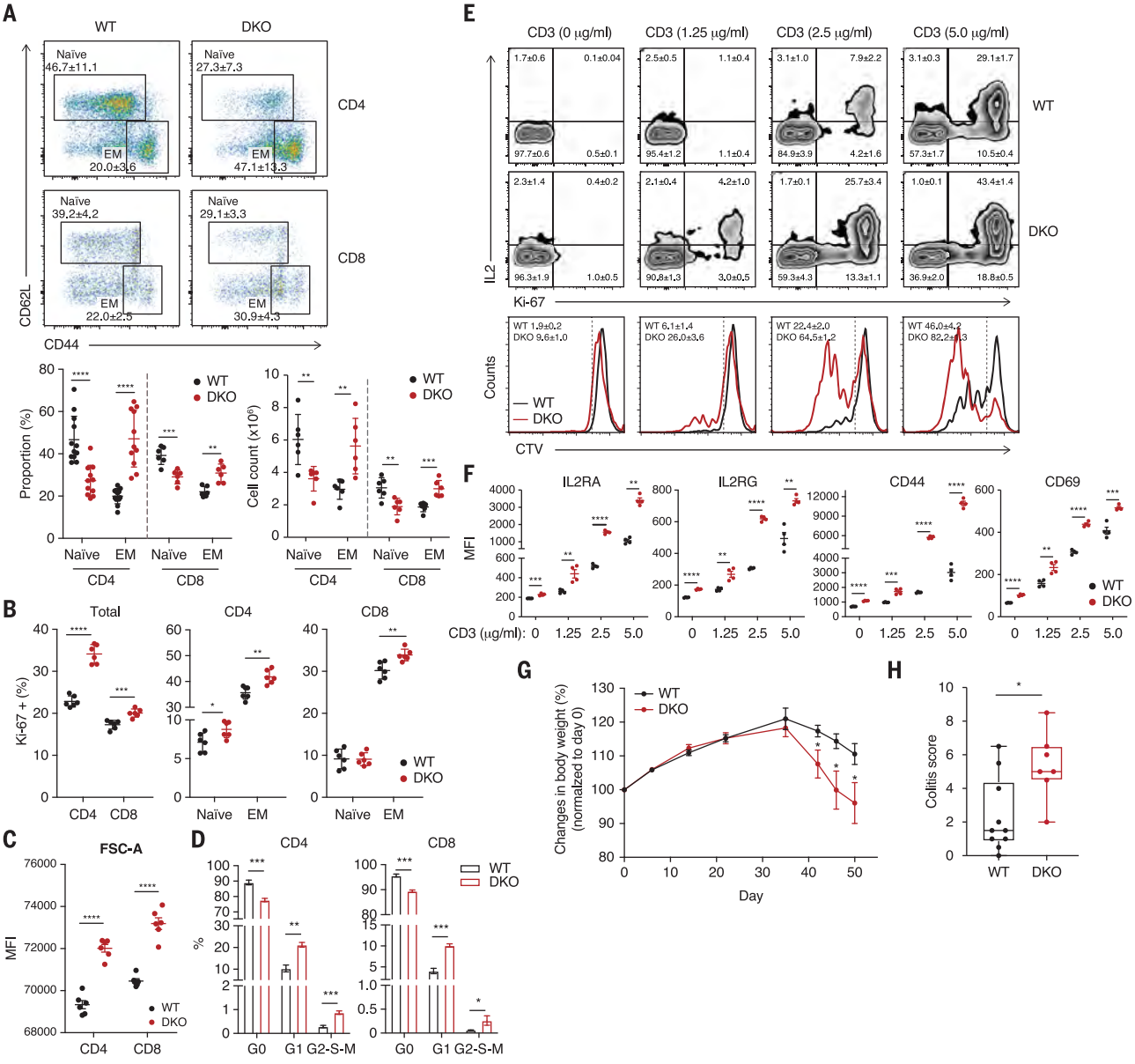


Fig. 2. Loss of *Btg1* and *Btg2* results in T cell quiescence defects. (**A**) Representative dot plots of naïve and effector and memory cells T cells from the spleens of WT and DKO mice (top). Populations are represented as means ± SD. Frequencies and numbers of naïve and effector memory CD4 and CD8 T cells in steady-state mice (n = 6 to 11) (bottom). (**B**) Ki-67 production by CD4 or CD8 T cells was measured from WT and DKO spleens (n = 6 mice per group). (**C**) Size of WT or DKO naïve T cells was measured by flow cytometry (n = 6 mice per group). (**D**) Cell-cycle analyses by 4′,6-diamidino-2-phenylindole (DAPI) and Ki-67 staining from naïve T cells (n = 4). (**E** and **F**) In vitro T cell proliferation assays. Naïve T cells were cultured for 4 days with plate-bound α-CD3 (0, 1.25, 2.5, and 5.0 μg/ml) and α-CD28 (2 μg/ml). (E) Representative dot plots showing the production of IL-2 and Ki-67 in WT and DKO cells (top). Populations are represented as means ± SD. Cell divisions are indicated by labeling with CellTrace Violet (CTV) (bottom). Populations of >1 division are represented as means ± SD. (F) Flow cytometry plots show the cell-surface expression of IL2RA, IL2RG, CD44, and CD69 (n = 4 mice per group). (**G**) Weight-loss curve after adoptive transfer of naïve CD4 T cells to *Rag2* KO mice. Mice were monitored for 7 weeks (WT, n = 10; DKO, n = 8). (**H**) Colitis score was recorded by colonoscopy. A box represents the first and third quartiles, and an internal bar indicates median. Error bars in (A) and (B) represent SD, and in (D) to (G) represent SEM. P values <0.05 were considered significant (*P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001); unpaired Student's t test was used for (B) to (G). Mann–Whitney U-test was used for (H).

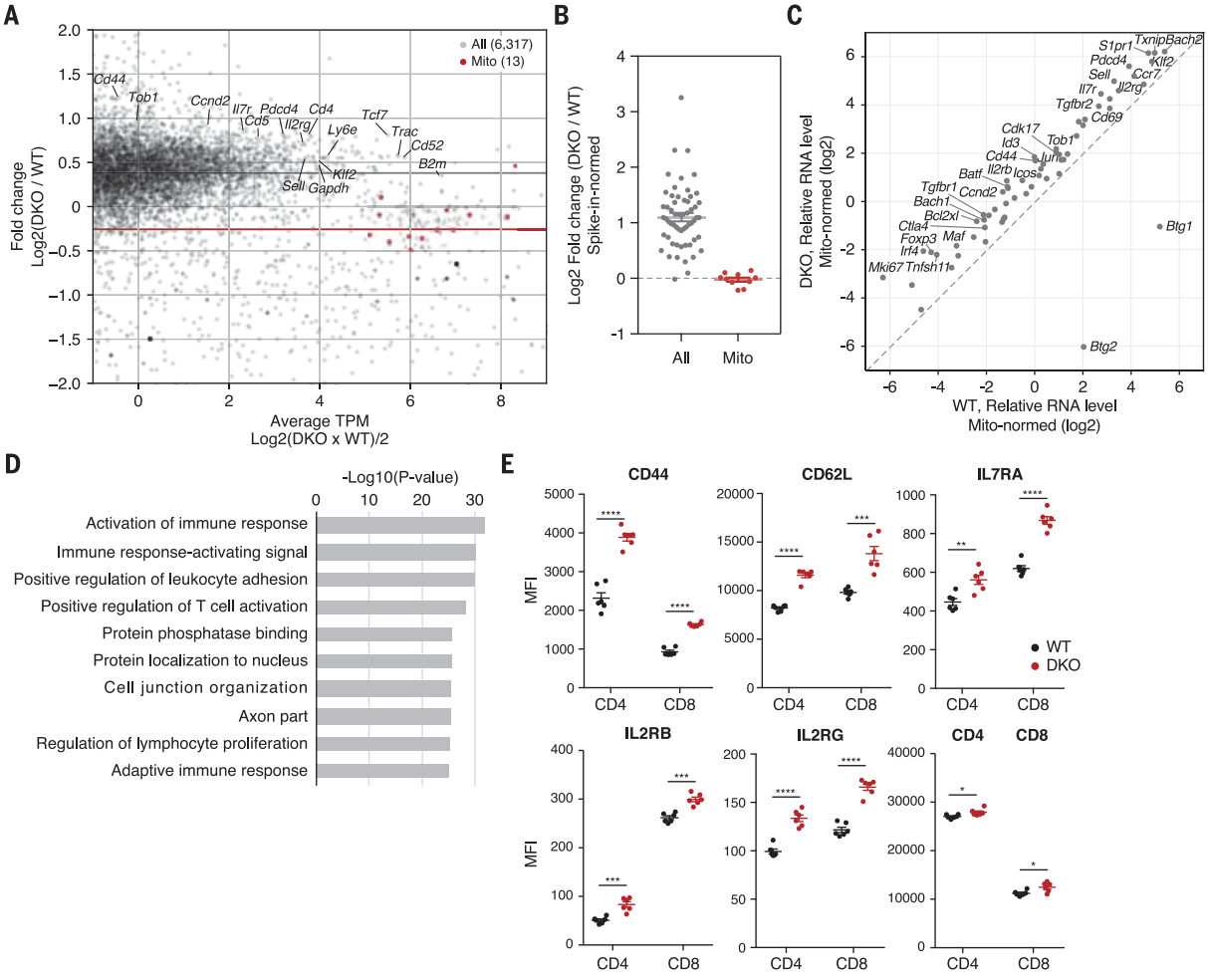


Fig. 3. BTG1 and BTG2 decrease mRNA and protein abundance in naïve T cells. (**A**) Plot of total RNA-seq reads from WT and DKO naïve CD4 T cells (n = 2 per group). Dots represent whole transcripts (gray) and mtRNAs (red), respectively. Gray and red lines indicate median fold changes from whole transcripts and mtRNAs, respectively. (**B**) RT-qPCR validation of WT and DKO naïve CD4 T cells. mRNA (n = 80) and mtRNA (n = 10) transcripts are normalized by spike-in RNA. (**C**) mRNA transcripts known as key T cell genes (n = 73) from WT and DKO naïve CD4 T cells are validated by RT-qPCR. Each value is normalized by the average of mtRNA transcripts. Each group from (B) and (C) was pooled from two or three mice. Each symbol represents the mean of two technical replicates. Two independent experiments showed similar results. (**D**) Differentially regulated gene sets analyzed by gene ontology analysis. The top 10 representative terms that are all positively regulated are shown. (**E**) Protein abundance was measured in WT (black) and DKO (red) naïve T cells (gated on $CD4^+CD62L^{hi}CD44^{lo}$ or $CD8^+CD62L^{hi}CD44^{lo}$) purified by fluorescence-activated cell sorting (n = 6 mice per group). Error bars in (E) denote SEM. P values <0.05 were considered significant (*P < 0.05; **P < 0.01; ***P < 0.001;****P < 0.0001), unpaired Student's t test.

mice using CRISPR-Cas9 technology (*16*). We confirmed correct insertions of two loxP sites in both *Btg1/2* and checked their efficient deletion after crossing with *Cd4*-Cre transgenic mice (fig. S2). T cell development in the thymus and spleen of knockout (KO) mice did not differ markedly from wild-type littermate controls (WT) (fig. S3A). However, *Btg1* and *Btg2* double conditional KO (DKO) mice showed decreased numbers and frequencies of naïve CD4 and CD8 T cells compared to WT in the periphery (Fig. 2A). Single-KO mice did not show any significant changes, suggesting that BTG1/2 are functionally redundant (fig. S3B). BTG1/2 are indeed more closely related to each other than to their family members on the basis of their phylogenetic tree (fig. S4) (*13*, *14*). Consistent with the reduced numbers and proportions of naïve T cells in DKO mice, the amounts of Ki-67 were significantly increased (Fig. 2B). Moreover, there was a significant increase in cell size and a concomitant reduction in the proportion of cells in G_0 phase of G_0 phase (Figs. 2, C and D, and fig. S5), indicating that T cell quiescence is compromised in the absence of *Btg1/2*.

To determine if DKO T cells are more prone to escape from quiescence, we performed in vitro T cell proliferation assays (Fig. 2E and fig. S6). Even under weak TCR stimulation conditions, DKO T cells readily underwent clonal expansion as evidenced by the production of interleukin-2 (IL-2) and Ki-67, and dilution of a cell division–tracking dye (Fig. 2E). By contrast, WT T cells barely proliferated under the same conditions. There were also substantial changes in the amounts of IL2RA, IL2RG, CD44, and CD69, indicating enhanced proliferation and activation (Fig. 2F). DKO naïve T cells could proliferate even in the presence of tonic cytokines (IL-2 and/or IL-7) alone absent any TCR stimulation (fig. S7). Thus, DKO T cells appear to be overly sensitive to activation signals and can easily overcome the activation threshold.

In support of the proclivity of DKO cells to exit quiescence, DKO naïve T cells adoptively transferred to *Rag2*-deficient mice significantly exacerbated colitis (Fig. 2, G and H). DKO naïve T cells underwent more activation and proliferation with accompanying

pathogenic T cell responses in vivo (fig. S8). BTG1/2 appear to play an indispensable role in limiting spontaneous activation of T cells to prevent autoimmune disease. Similarly, upon *Listeria monocytogenes* infection, the number of pathogenic T cells in DKO mice increased more than in WT, which in turn resulted in the rapid clearance of *Listeria* (fig. S9). Thus, BTG1/2 maintain T cell quiescence through controlling proliferation and activation in vivo and in vitro.

To gain insight into the molecular mechanism underlying T cell quiescence, we carried out total RNA sequencing (RNA-seq) on naïve T cells from WT or DKO mice. Intriguingly, we observed a modest but global increase [5610 out of 6317 (88.8%)] in mRNA abundance in DKO T cells (Fig. 3A and table S1). This shift in the entire transcriptome may imply an alteration in the general machinery, such as deadenylases. We validated the global increase by normalizing cytoplasmic mRNA abundance to that of mitochondrial genome–encoded mRNAs (mtRNAs) and spike-in RNAs (materials and methods) (*17*, *18*). Intriguingly, we confirmed a similar increase in mRNA abundance by both methods, whereas mtRNAs remain unaffected (Fig. 3, B and C, and fig. S10). Because the vast majority of transcripts were up-regulated, we next asked which processes were modulated. Consistent with earlier data (Fig. 2), pathways related to activation and proliferation of the immune system were particularly enriched (Fig. 3D and table S2). To check if the increase in mRNAs was reflected in an increased number of functional proteins, we measured mean fluorescence intensity (MFI) by flow cytometry. The proteins that we tested were significantly increased in abundance in both CD4 and CD8 T cells (Fig. 3E). Notably, the protein abundance of stimulatory genes whose mRNAs were relatively profuse in naïve T cells was increased, implying a lowered threshold for activation (fig. S11). Thus, mRNA and protein abundance is up-regulated in the absence of BTG1/2, which likely results in the activation of naïve T cells.

The transcriptome-wide increase in mRNA abundance hinted that BTG1/2 may play a role in mRNA stability control. Because BTG1/2 have been shown to associate with poly(A)-binding protein (PABP) and subunits of CCR4-NOT (CNOT) deadenylase complex in vitro (*19*–*23*), we examined whether BTG1/2 interact with PABP and/or the core deadenylation machinery in naïve T cells. By coimmunoprecipitation and in situ proximity ligation assay, BTG1/2 were shown to be physically associated with PABP and CNOT7 in naïve T cells in an RNA-independent manner (Fig. 4, A and B). Furthermore, deletion mutants of BTG1 that lost the ability to interact with either PABP or CNOT7 failed to inhibit proliferation (fig. S12). Thus, BTG1/2 functions are critically dependent on the interaction with PABP and CNOT7.

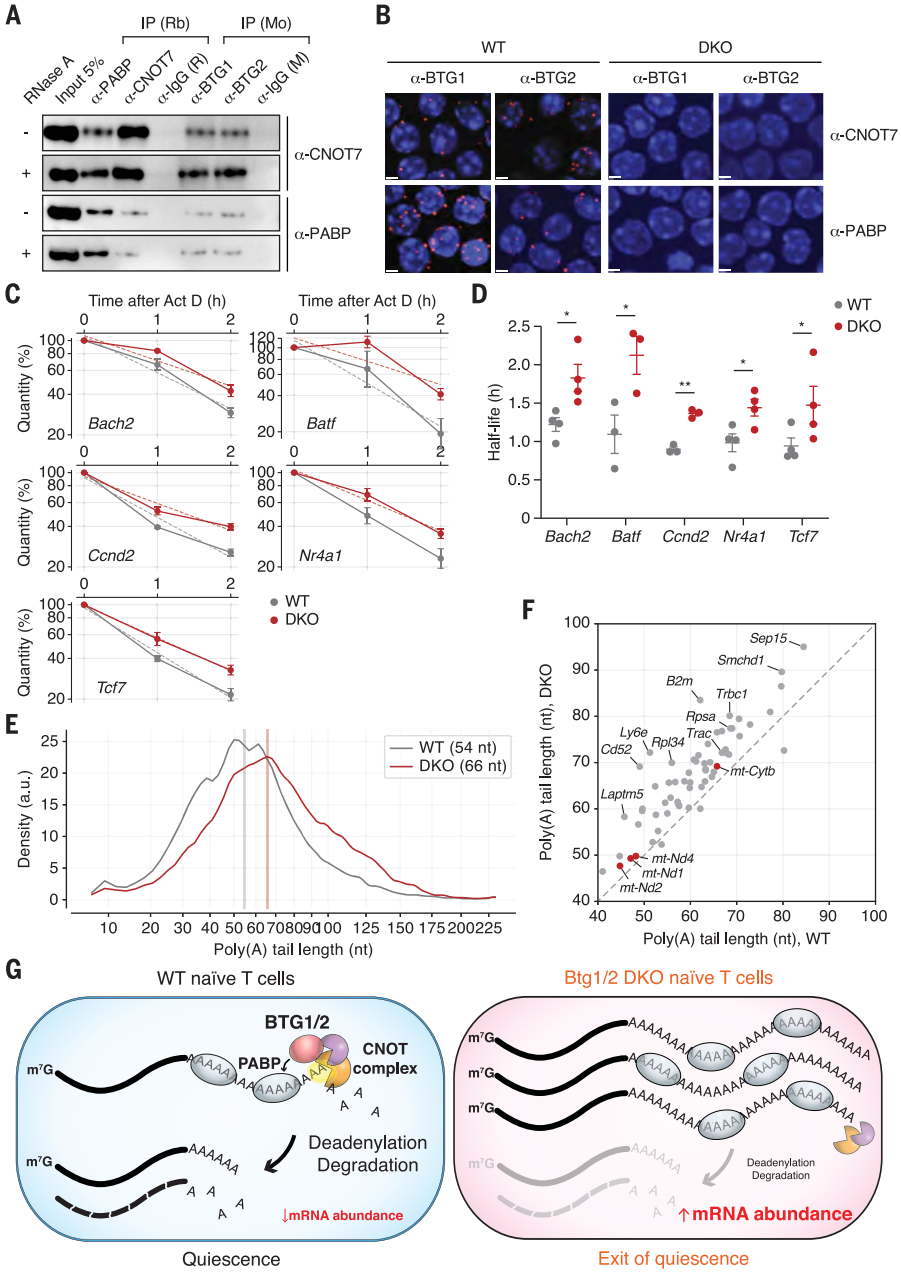


Fig. 4. BTG1 and BTG2 promote mRNA deadenylation and decay in naïve T cells. (**A**) Coimmunoprecipitation with antibodies against PABP, CNOT7, BTG1, BTG2, and control immunoglobulin G (R, rabbit; M, mouse) with or without ribonuclease A (RNase A) treatment. (**B**) Direct interactions between BTG1 or BTG2 with CNOT7 or PABP were detected as red dots in WT naïve T cells, but not in DKO cells. Scale bars indicate 2 μm. (**C**) mRNA decay is delayed in DKO naïve T cells. *Gapdh* was used for normalization. Error bars represent SEM (WT, $n = 3$ mice; DKO, $n = 4$ mice). Each symbol represents the mean of three independent experiments with two technical replicates. (**D**) Half-lives are calculated from (C) by linear fitting of the log-transformed exponential decay function. P values <0.05 were considered significant (*$P < 0.05$; **$P < 0.01$), unpaired Student's t test. (**E**) Global poly(A) length distribution determined by mTAIL-seq from WT and DKO naïve CD4 T cells. The median poly(A) length of each sample is shown in parentheses and marked by vertical lines. (**F**) A scatter plot showing geometric mean poly(A) length. Transcripts supported by >20 mean poly(A)$^+$ tags are shown ($n = 3$). Red dots represent mtRNA transcripts. (**G**) Schematic illustration of how naïve T cells maintain the quiescent state.

Next, to assess whether mRNA turnover is impaired in DKO cells, we measured the decay rates of mRNAs. Notably, decay rates of cytoplasmic mRNAs, but not mtRNAs, were retarded in DKO T cells compared to WT (Figs. 4, C and D; figs. S13 and S14; and table S3). mRNA decay rates in activated T cells, where BTG1/2

are no longer expressed, showed no difference in mRNA degradation between WT and DKO (fig. S15), suggesting that BTG1/2 contribute to lowering mRNA amounts through mRNA turnover in naïve but not activated T cells.

To examine if poly(A) length is indeed elongated in DKO cells, we performed mTAIL-seq (fig. S16A) (*24*). In line with the stabilization of mRNA, the poly(A) tail length of DKO T cells was longer than that of WT (Fig. 4, E and F, and table S4). Again, mtRNA poly(A) tail lengths were nearly unchanged (Fig. 4F), suggesting that BTG1/2-mediated deadenylation was responsible for degrading the bulk of cytoplasmic mRNA in naïve T cells. Notably, the difference in poly(A) tail length in naïve T cells was diminished in activated T cells (fig. S16, B and C, and table S5).

In conclusion, we propose a model to explain how the quiescent state in T cells is maintained (Fig. 4G). BTG1 and BTG2 are highly and specifically expressed in quiescent T cells and promote mRNA deadenylation and degradation. Unlike BTG1/2, deadenylases appear to be ubiquitously expressed at a low level, implying a specialized function of BTG1/2 in immune cells (fig. S17). Because BTG1/2 interact with PABP and the CNOT complex, which bind nonspecifically to mRNAs (*25*), BTG1/2 can direct the down-regulation of mRNAs at a global level. This mechanism is seemingly inefficient in terms of cost, but the availability of presynthesized mRNA provides a benefit to quiescent T cells of a rapid response to activation signals. Given that naïve T cells differentiate into multiple lineages, having such a primed state on a hair trigger would be ultimately beneficial. Upon activation, BTG1/2 quickly disappear, which results in an accumulation of mRNAs and exit from the quiescent state (Figs. 1D and 4G).

Here, we show the functional consequences of BTG1/2-mediated deadenylation in vivo. B cells may also use a similar mechanism to maintain quiescence through BTG1/2, as BTG1/2 are commonly dysregulated in leukemia and lymphoma (*15*, *26*–*28*). Likewise, other cofactors, in addition to BTG1/2, may recruit and enhance mRNA deadenylation and degradation in certain cell types. Thus, the mechanism that we propose may be broadly applicable to other cells and tissues, serving as a general system to secure the quiescent state.

REFERENCES AND NOTES

1. T. H. Cheung, T. A. Rando, *Nat. Rev. Mol. Cell Biol.* **14**, 329–340 (2013).
2. G. J. van der Windt, E. L. Pearce, *Immunol. Rev.* **249**, 27–42 (2012).
3. N. J. MacIver, R. D. Michalek, J. C. Rathmell, *Annu. Rev. Immunol.* **31**, 259–283 (2013).
4. J. Zhu, H. Yamane, W. E. Paul, *Annu. Rev. Immunol.* **28**, 445–489 (2010).
5. J. Sprent, C. D. Surh, *Nat. Immunol.* **12**, 478–484 (2011).
6. W. Ouyang, O. Beckett, R. A. Flavell, M. O. Li, *Immunity* **30**, 358–371 (2009).
7. X. Feng *et al.*, *Nat. Immunol.* **12**, 544–550 (2011).
8. Y. M. Kerdiles *et al.*, *Nat. Immunol.* **10**, 176–184 (2009).
9. W. Ouyang, M. O. Li, *Trends Immunol.* **32**, 26–33 (2011).
10. D. Tzachanis *et al.*, *Nat. Immunol.* **2**, 1174–1182 (2001).
11. K. Yang, G. Neale, D. R. Green, W. He, H. Chi, *Nat. Immunol.* **12**, 888–897 (2011).
12. D. Hnisz *et al.*, *Cell* **155**, 934–947 (2013).
13. F. Mauxion, C. Y. Chen, B. Séraphin, A. B. Shyu, *Trends Biochem. Sci.* **34**, 640–647 (2009).
14. G. S. Winkler, *J. Cell. Physiol.* **222**, 66–72 (2010).
15. L. Yuniati, B. Scheijen, L. T. van der Meer, F. N. van Leeuwen, *J. Cell. Physiol.* **234**, 5379–5389 (2019).
16. J. Henao-Mejia *et al.*, *Cold Spring Harb. Protoc.* **2016**, prot090704 (2016).
17. N. Gadir, L. Haim-Vilmovsky, J. Kraut-Cohen, J. E. Gerst, *RNA* **17**, 1551–1565 (2011).
18. H. Yi *et al.*, *Mol. Cell* **70**, 1081–1088.e5 (2018).
19. J. P. Rouault *et al.*, *J. Biol. Chem.* **273**, 22563–22569 (1998).
20. D. Prévôt *et al.*, *J. Biol. Chem.* **276**, 9640–9648 (2001).
21. F. Mauxion, C. Faux, B. Séraphin, *EMBO J.* **27**, 1039–1048 (2008).
22. R. Doidge, S. Mittal, A. Aslam, G. S. Winkler, *PLOS ONE* **7**, e51331 (2012).
23. B. Stupfler, C. Birck, B. Séraphin, F. Mauxion, *Nat. Commun.* **7**, 10811 (2016).
24. J. Lim, M. Lee, A. Son, H. Chang, V. N. Kim, *Genes Dev.* **30**, 1671–1682 (2016).
25. I. A. Eliseeva, D. N. Lyabin, L. P. Ovchinnikov, *Biochemistry (Mosc.)* **78**, 1377–1391 (2013).
26. R. P. Kuiper *et al.*, *Leukemia* **21**, 1258–1266 (2007).
27. R. D. Morin *et al.*, *Nature* **476**, 298–303 (2011).
28. E. Waanders *et al.*, *PLOS Genet.* **8**, e1002533 (2012).

ACKNOWLEDGMENTS

We are grateful to R. Medzhitov for support and discussions. We thank J. Y. Hwang and W. J. Chae for helpful comments. We also thank all members of the Flavell laboratory for technical and administrative support. We thank M. Ha for the gift of RNA spike-in and J. Choi for illustrating the graphical abstract. **Funding:** This work was supported by the Howard Hughes Medical Institute (R.A.F.). The work of S.S.H. was supported by a Leslie H. Warner Fellowship from Yale Cancer Center (YCC). The work of J.L. was supported by postdoctoral fellowships from the HFSPO (LT000037/2018-L) and Jane Coffin Childs Memorial Fund. G.R.L was supported by National Research Foundation of Korea (NRF-2017R1A2B3008621). H.B.L. was supported by the National Natural Science Foundation of China (91753141) and the start-up fund from the Shanghai Jiao Tong University School of Medicine. **Author contributions:** S.S.H. conceived and initiated the project. S.S.H. and J.L. designed the experiments, analyzed and discussed the results, and wrote the manuscript. S.S.H. performed most of the experiments in this study. J.L. performed total RNA-seq and mTAIL-seq, analyzed the data, and illustrated the figures. Z.Y. and P.K. helped to perform the FACS analysis and in vitro experiments. E.S. and H.X. helped to carry out the colitis experiment. C.C.D.H. helped to analyze our deep RNA-seq data. L.K.K., G.R.L., and H.-B.L. contributed to key discussions. R.A.F. supervised the project, participated in interpreting the results, and edited the manuscript. **Competing interests:** R.A.F is a consultant for GSK and Zai Lab Ltd. All other authors declare no competing financial interests. **Data and materials availability:** Sequenced reads have been deposited in the NCBI Gene Expression Omnibus (GEO) database (accession no. GSE125890).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1255/suppl/DC1
Materials and Methods
Figs. S1 to 17
Tables S1 to S6
References (*29*–*33*)

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

14 February 2019; resubmitted 22 October 2019
Accepted 19 February 2020
10.1126/science.aax0194

CORONAVIRUS

Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation

Daniel Wrapp[1]*, Nianshuang Wang[1]*, Kizzmekia S. Corbett[2], Jory A. Goldsmith[1], Ching-Lin Hsieh[1], Olubukola Abiona[2], Barney S. Graham[2], Jason S. McLellan[1]†

The outbreak of a novel coronavirus (2019-nCoV) represents a pandemic threat that has been declared a public health emergency of international concern. The CoV spike (S) glycoprotein is a key target for vaccines, therapeutic antibodies, and diagnostics. To facilitate medical countermeasure development, we determined a 3.5-angstrom-resolution cryo–electron microscopy structure of the 2019-nCoV S trimer in the prefusion conformation. The predominant state of the trimer has one of the three receptor-binding domains (RBDs) rotated up in a receptor-accessible conformation. We also provide biophysical and structural evidence that the 2019-nCoV S protein binds angiotensin-converting enzyme 2 (ACE2) with higher affinity than does severe acute respiratory syndrome (SARS)-CoV S. Additionally, we tested several published SARS-CoV RBD-specific monoclonal antibodies and found that they do not have appreciable binding to 2019-nCoV S, suggesting that antibody cross-reactivity may be limited between the two RBDs. The structure of 2019-nCoV S should enable the rapid development and evaluation of medical countermeasures to address the ongoing public health crisis.

The novel coronavirus 2019-nCoV has recently emerged as a human pathogen in the city of Wuhan in China's Hubei province, causing fever, severe respiratory illness, and pneumonia—a disease recently named COVID-19 (*1*, *2*). According to the World Health Organization (WHO), as of 16 February 2020, there had been >51,000 confirmed cases globally, leading to at least 1600 deaths. The emerging pathogen was rapidly characterized as a new member of the betacoronavirus genus, closely related to several bat coronaviruses and to severe acute respiratory syndrome coronavirus (SARS-CoV) (*3*, *4*). Compared with SARS-CoV, 2019-nCoV appears to be more readily transmitted from human to human, spreading to multiple continents and leading to the WHO's declaration of a Public Health Emergency of International Concern (PHEIC) on 30 January 2020 (*1*, *5*, *6*).

2019-nCoV makes use of a densely glycosylated spike (S) protein to gain entry into host cells. The S protein is a trimeric class I fusion protein that exists in a metastable prefusion conformation that undergoes a substantial structural rearrangement to fuse the viral membrane with the host cell membrane (*7*, *8*). This process is triggered when the S1 subunit binds to a host cell receptor. Receptor binding destabilizes the prefusion trimer, resulting in shedding of the S1 subunit and transition of the S2 subunit to a stable postfusion conformation (*9*). To engage a host cell receptor, the receptor-binding domain (RBD) of S1 undergoes hinge-like conformational movements that transiently hide or expose the determinants of receptor binding. These two states are referred to as the "down" conformation and the "up" conformation, where down corresponds to the receptor-inaccessible state and up corresponds to the receptor-accessible state, which is thought to be less stable (*10*–*13*). Because of the indispensable function of the S protein, it represents a target for antibody-mediated neutralization, and characterization of the prefusion S structure would provide atomic-level information to guide vaccine design and development.

Based on the first reported genome sequence of 2019-nCoV (*4*), we expressed ectodomain residues 1 to 1208 of 2019-nCoV S, adding two stabilizing proline mutations in the C-terminal S2 fusion machinery using a previous stabilization strategy that proved effective for other betacoronavirus S proteins (*11*, *14*). Figure 1A shows the domain organization of the expression construct, and figure S1 shows the purification process. We obtained ~0.5 mg/liter of the recombinant prefusion-stabilized S ectodomain from FreeStyle 293 cells and purified the protein to homogeneity by affinity chromatography and size-exclusion chromatography (fig. S1). Cryo–electron microscopy (cryo-EM) grids were prepared using this purified, fully glycosylated S protein, and preliminary screening revealed a high particle density with little aggregation near the edges of the holes.

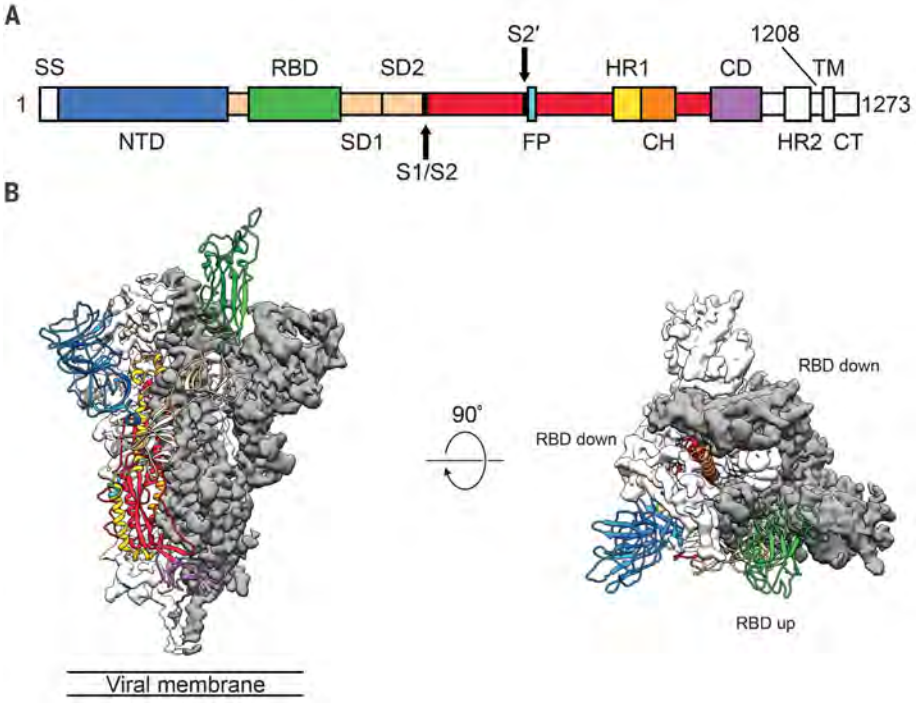


Fig. 1. Structure of 2019-nCoV S in the prefusion conformation. (**A**) Schematic of 2019-nCoV S primary structure colored by domain. Domains that were excluded from the ectodomain expression construct or could not be visualized in the final map are colored white. SS, signal sequence; S2′, S2′ protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Arrows denote protease cleavage sites. (**B**) Side and top views of the prefusion structure of the 2019-nCoV S protein with a single RBD in the up conformation. The two RBD down protomers are shown as cryo-EM density in either white or gray and the RBD up protomer is shown in ribbons colored corresponding to the schematic in (A).

[1]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA. [2]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: jmclellan@austin.utexas.edu

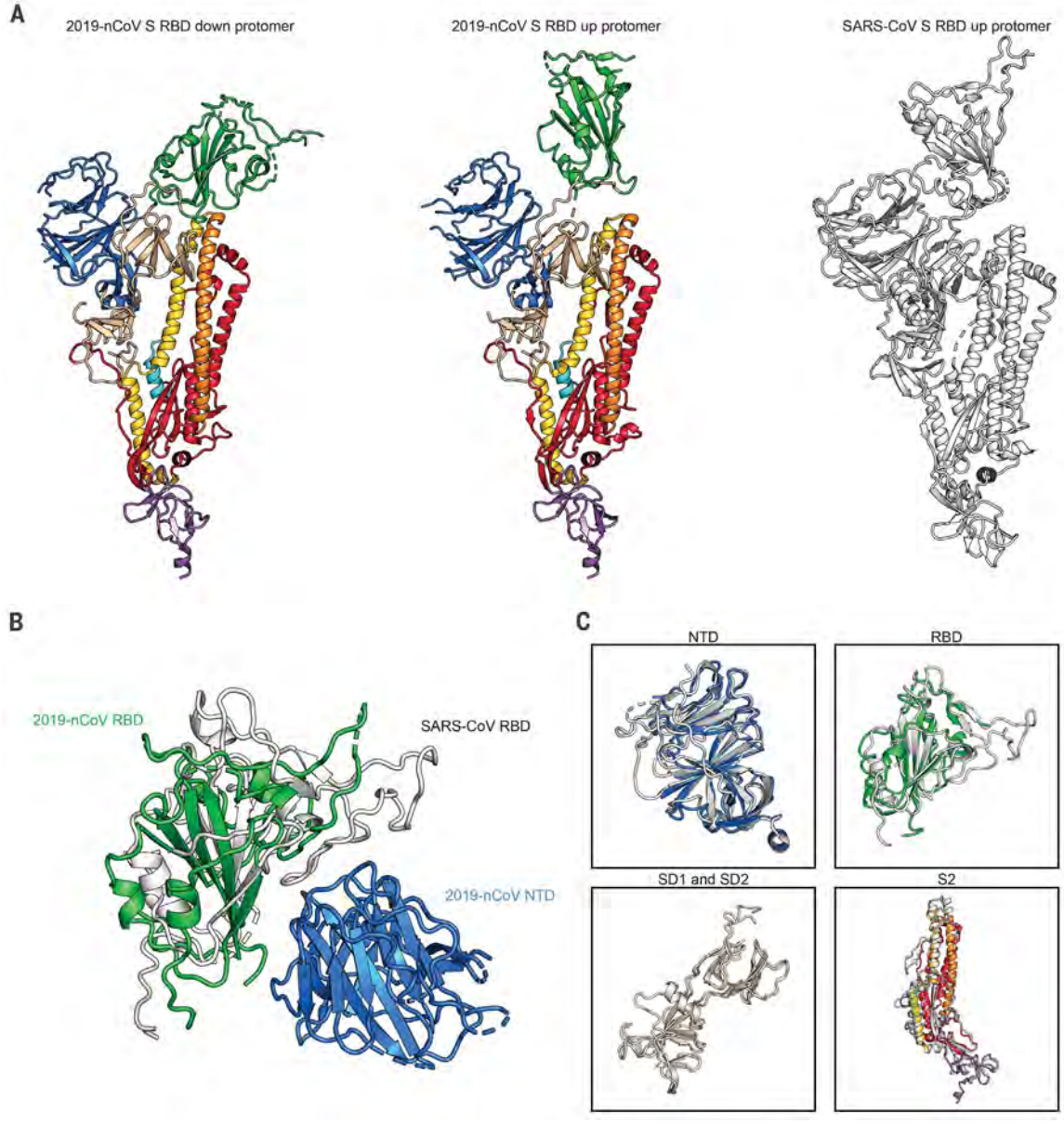
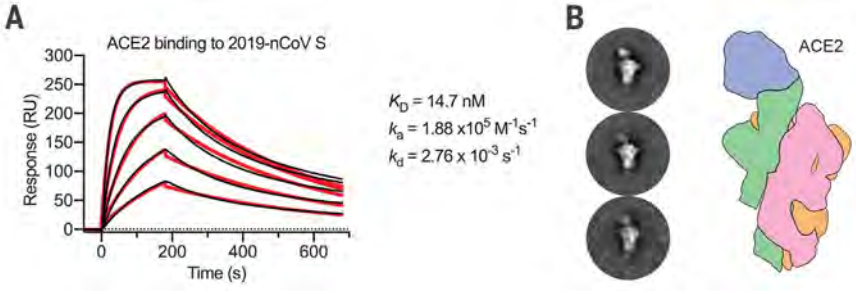


Fig. 2. Structural comparison between 2019-nCoV S and SARS-CoV S. (**A**) Single protomer of 2019-nCoV S with the RBD in the down conformation (left) is shown in ribbons colored according to Fig. 1. A protomer of 2019-nCoV S in the RBD up conformation is shown (center) next to a protomer of SARS-CoV S in the RBD up conformation (right), displayed as ribbons and colored white (PDB ID: 6CRZ). (**B**) RBDs of 2019-nCoV and SARS-CoV aligned based on the position of the adjacent NTD from the neighboring protomer. The 2019-nCoV RBD is colored green and the SARS-CoV RBD is colored white. The 2019-nCoV NTD is colored blue. (**C**) Structural domains from 2019-nCoV S have been aligned to their counterparts from SARS-CoV S as follows: NTD (top left), RBD (top right), SD1 and SD2 (bottom left), and S2 (bottom right).

Fig. 3. 2019-nCoV S binds human ACE2 with high affinity. (**A**) Surface plasmon resonance sensorgram showing the binding kinetics for human ACE2 and immobilized 2019-nCoV S. Data are shown as black lines, and the best fit of the data to a 1:1 binding model is shown in red. (**B**) Negative-stain EM 2D class averages of 2019-nCoV S bound by ACE2. Averages have been rotated so that ACE2 is positioned above the 2019-nCoV S protein with respect to the viral membrane. A diagram depicting the ACE2-bound 2019-nCoV S protein is shown (right) with ACE2 in blue and S protein protomers colored tan, pink, and green.



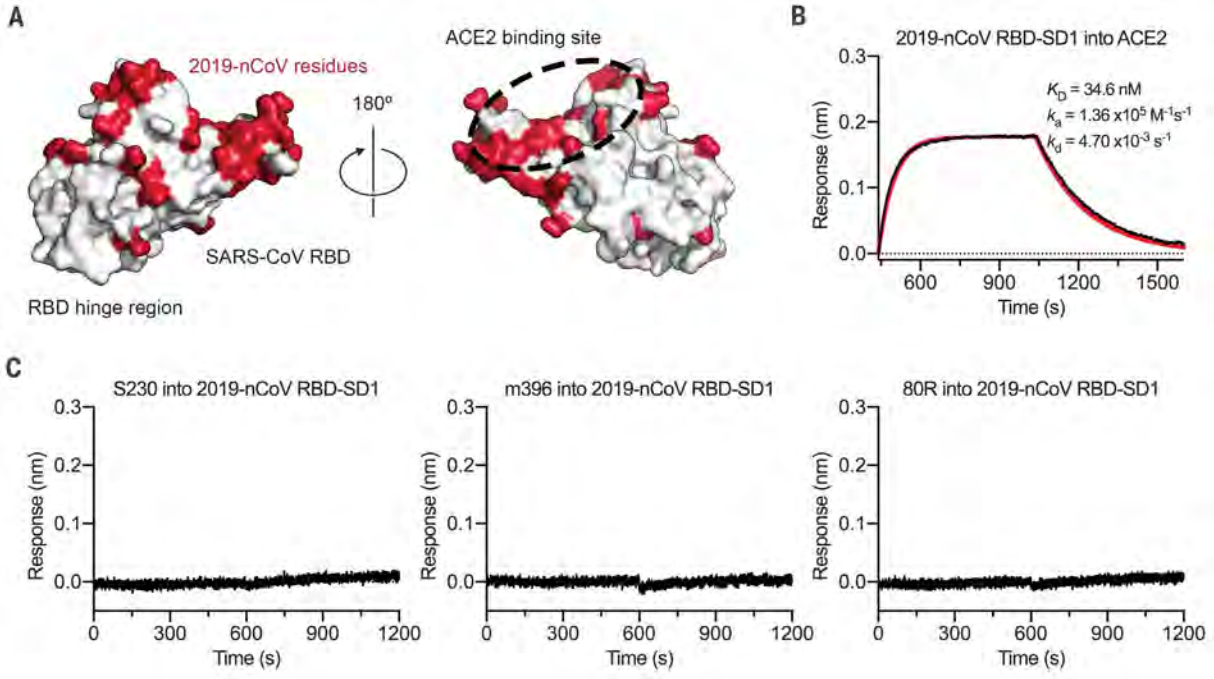


Fig. 4. Antigenicity of the 2019-nCoV RBD. (**A**) SARS-CoV RBD shown as a white molecular surface (PDB ID: 2AJF), with residues that vary in the 2019-nCoV RBD colored red. The ACE2-binding site is outlined with a black dashed line. (**B**) Biolayer interferometry sensorgram showing binding to ACE2 by the 2019-nCoV RBD-SD1. Binding data are shown as a black line, and the best fit of the data to a 1:1 binding model is shown in red. (**C**) Biolayer interferometry to measure cross-reactivity of the SARS-CoV RBD-directed antibodies S230, m396, and 80R. Sensor tips with immobilized antibodies were dipped into wells containing 2019-nCoV RBD-SD1, and the resulting data are shown as a black line.

After collecting and processing 3207 micrograph movies, we obtained a 3.5-Å-resolution three-dimensional (3D) reconstruction of an asymmetrical trimer in which a single RBD was observed in the up conformation. (Fig. 1B, fig. S2, and table S1). Because of the small size of the RBD (~21 kDa), the asymmetry of this conformation was not readily apparent until ab initio 3D reconstruction and classification were performed (Fig. 1B and fig. S3). By using the 3D variability feature in cryoSPARC v2 (*15*), we observed breathing of the S1 subunits as the RBD underwent a hinge-like movement, which likely contributed to the relatively poor local resolution of S1 compared with the more stable S2 subunit (movies S1 and S2). This seemingly stochastic RBD movement has been captured during structural characterization of the closely related betacoronaviruses SARS-CoV and MERS-CoV, as well as the more distantly related alphacoronavirus porcine epidemic diarrhea virus (PEDV) (*10*, *11*, *13*, *16*). The observation of this phenomenon in 2019-nCoV S suggests that it shares the same mechanism of triggering that is thought to be conserved among the Coronaviridae, wherein receptor binding to exposed RBDs leads to an unstable three-RBD up conformation that results in shedding of S1 and refolding of S2 (*11*, *12*).

Because the S2 subunit appeared to be a symmetric trimer, we performed a 3D refinement imposing C3 symmetry, resulting in a 3.2-Å-resolution map with excellent density for the S2 subunit. Using both maps, we built most of the 2019-nCoV S ectodomain, including glycans at 44 of the 66 *N*-linked glycosylation sites per trimer (fig. S4). Our final model spans S residues 27 to 1146, with several flexible loops omitted. Like all previously reported coronavirus S ectodomain structures, the density for 2019-nCoV S begins to fade after the connector domain, reflecting the flexibility of the heptad repeat 2 domain in the prefusion conformation (fig. S4A) (*13*, *16*–*18*).

The overall structure of 2019-nCoV S resembles that of SARS-CoV S, with a root mean square deviation (RMSD) of 3.8 Å over 959 Cα atoms (Fig. 2A). One of the larger differences between these two structures (although still relatively minor) is the position of the RBDs in their respective down conformations. Whereas the SARS-CoV RBD in the down conformation packs tightly against the N-terminal domain (NTD) of the neighboring protomer, the 2019-nCoV RBD in the down conformation is angled closer to the central cavity of the trimer (Fig. 2B). Despite this observed conformational difference, when the individual structural domains of 2019-nCoV S are aligned to their counterparts from SARS-CoV S, they reflect the high degree of structural homology between the two proteins, with the NTDs, RBDs, subdomains 1 and 2 (SD1 and SD2), and S2 subunits yielding individual RMSD values of 2.6 Å, 3.0 Å, 2.7 Å, and 2.0 Å, respectively (Fig. 2C).

2019-nCoV S shares 98% sequence identity with the S protein from the bat coronavirus RaTG13, with the most notable variation arising from an insertion in the S1/S2 protease cleavage site that results in an "RRAR" furin recognition site in 2019-nCoV (*19*) rather than the single arginine in SARS-CoV (fig. S5) (*20*–*23*). Notably, amino acid insertions that create a polybasic furin site in a related position in hemagglutinin proteins are often found in highly virulent avian and human influenza viruses (*24*). In the structure reported here, the S1/S2 junction is in a disordered, solvent-exposed loop. In addition to this insertion of residues in the S1/S2 junction, 29 variant residues exist between 2019-nCoV S and RaTG13 S, with 17 of these positions mapping to the RBD (figs. S5 and S6). We also analyzed the 61 available 2019-nCoV S sequences in the Global Initiative on Sharing All Influenza Data database (https://www.gisaid.org/) and found that there were only nine amino acid substitutions among all deposited sequences. Most of these substitutions are relatively conservative and are not expected to have a substantial effect on the structure or function of the 2019-nCoV S protein (fig. S6).

Recent reports demonstrating that 2019-nCoV S and SARS-CoV S share the same functional host

cell receptor, angiotensin-converting enzyme 2 (ACE2) (*22*, *25*–*27*), prompted us to quantify the kinetics of this interaction by surface plasmon resonance. ACE2 bound to the 2019-nCoV S ectodomain with ~15 nM affinity, which is ~10- to 20-fold higher than ACE2 binding to SARS-CoV S (Fig. 3A and fig. S7) (*14*). We also formed a complex of ACE2 bound to the 2019-nCoV S ectodomain and observed it by negative-stain EM, which showed that it strongly resembled the complex formed between SARS-CoV S and ACE2 that has been observed at high resolution by cryo-EM (Fig. 3B) (*14*, *28*). The high affinity of 2019-nCoV S for human ACE2 may contribute to the apparent ease with which 2019-nCoV can spread from human to human (*1*); however, additional studies are needed to investigate this possibility.

The overall structural homology and shared receptor usage between SARS-CoV S and 2019-nCoV S prompted us to test published SARS-CoV RBD-directed monoclonal antibodies (mAbs) for cross-reactivity to the 2019-nCoV RBD (Fig. 4A). A 2019-nCoV RBD-SD1 fragment (S residues 319 to 591) was recombinantly expressed, and appropriate folding of this construct was validated by measuring ACE2 binding using biolayer interferometry (BLI) (Fig. 4B). Cross-reactivity of the SARS-CoV RBD-directed mAbs S230, m396, and 80R was then evaluated by BLI (*12*, *29*–*31*). Despite the relatively high degree of structural homology between the 2019-nCoV RBD and the SARS-CoV RBD, no binding to the 2019-nCoV RBD could be detected for any of the three mAbs at the concentration tested (1 μM) (Fig. 4C), in contrast to the strong binding that we observed to the SARS-CoV RBD (fig. S8). Although the epitopes of these three antibodies represent a relatively small percentage of the surface area of the 2019-nCoV RBD, the lack of observed binding suggests that SARS-directed mAbs will not necessarily be cross-reactive and that future antibody isolation and therapeutic design efforts will benefit from using 2019-nCoV S proteins as probes.

The rapid global spread of 2019-nCoV, which prompted the PHEIC declaration by WHO, signals the urgent need for coronavirus vaccines and therapeutics. Knowing the atomic-level structure of the 2019-nCoV spike will allow for additional protein-engineering efforts that could improve antigenicity and protein expression for vaccine development. The structural data will also facilitate the evaluation of 2019-nCoV spike mutations that will occur as the virus undergoes genetic drift and help to define whether those residues have surface exposure and map to sites of known antibody epitopes for other coronavirus spike proteins. In addition, the structure provides assurance that the protein produced by this construct is homogeneous and in the prefusion conformation, which should maintain the most neutralization-sensitive epitopes when used as candidate vaccine antigens or B cell probes for isolating neutralizing human mAbs. Furthermore, the atomic-level detail will enable the design and screening of small molecules with fusion-inhibiting potential. This information will support precision vaccine design and the discovery of antiviral therapeutics, accelerating medical countermeasure development.

REFERENCES AND NOTES

1. J. F. Chan *et al.*, *Lancet* **395**, 514–523 (2020).
2. C. Huang *et al.*, *Lancet* **395**, 497–506 (2020).
3. R. Lu *et al.*, *Lancet* S0140-6736(20)30251-8 (2020).
4. F. Wu *et al.*, *Nature* (2020).
5. N. Chen *et al.*, *Lancet* **395**, 507–513 (2020).
6. Q. Li *et al.*, *N. Engl. J. Med.* NEJMoa2001316 (2020).
7. F. Li, *Annu. Rev. Virol.* **3**, 237–261 (2016).
8. B. J. Bosch, R. van der Zee, C. A. de Haan, P. J. Rottier, *J. Virol.* **77**, 8801–8811 (2003).
9. A. C. Walls *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11157–11162 (2017).
10. M. Gui *et al.*, *Cell Res.* **27**, 119–129 (2017).
11. J. Pallesen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7348–E7357 (2017).
12. A. C. Walls *et al.*, *Cell* **176**, 1026–1039.e15 (2019).
13. Y. Yuan *et al.*, *Nat. Commun.* **8**, 15092 (2017).
14. R. N. Kirchdoerfer *et al.*, *Sci. Rep.* **8**, 15701 (2018).
15. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nat. Methods* **14**, 290–296 (2017).
16. D. Wrapp, J. S. McLellan, *J. Virol.* **93**, e00923-19 (2019).
17. A. C. Walls *et al.*, *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
18. R. N. Kirchdoerfer *et al.*, *Nature* **531**, 118–121 (2016).
19. B. Coutard *et al.*, *Antiviral Res.* **176**, 104742 (2020).
20. B. J. Bosch, W. Bartelink, P. J. Rottier, *J. Virol.* **82**, 8887–8890 (2008).
21. I. Glowacka *et al.*, *J. Virol.* **85**, 4122–4134 (2011).
22. W. Li *et al.*, *Nature* **426**, 450–454 (2003).
23. S. Belouzard, V. C. Chu, G. R. Whittaker, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5871–5876 (2009).
24. J. Chen *et al.*, *Cell* **95**, 409–417 (1998).
25. M. Hoffmann *et al.*, The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. bioRxiv 929042 [Preprint]. 31 January 2020. https://doi.org/10.1101/2020.01.31.929042.
26. Y. Wan, J. Shang, R. Graham, R. S. Baric, F. Li, *J. Virol.* JVI.00127-20 (2020).
27. P. Zhou *et al.*, *Nature* (2020).
28. W. Song, M. Gui, X. Wang, Y. Xiang, *PLOS Pathog.* **14**, e1007236 (2018).
29. W. C. Hwang *et al.*, *J. Biol. Chem.* **281**, 34610–34616 (2006).
30. P. Prabakaran *et al.*, *J. Biol. Chem.* **281**, 15829–15836 (2006).
31. X. Tian *et al.*, *bioRxiv* **9**, 382–385 (2020).

ACKNOWLEDGMENTS

We thank J. Ludes-Meyers for assistance with cell transfection, members of the McLellan laboratory for critical reading of the manuscript, and A. Dai from the Sauer Structural Biology Laboratory at the University of Texas at Austin for assistance with microscope alignment. **Funding:** This work was supported in part by a National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases (NIAID) grant awarded to J.S.M. (R01-AI127521) and by intramural funding from NIAID to B.S.G. The Sauer Structural Biology Laboratory is supported by the University of Texas College of Natural Sciences and by award RR160023 from the Cancer Prevention and Research Institute of Texas (CPRIT). **Author contributions:** D.W. collected and processed cryo-EM data. D.W., N.W., and J.S.M. built and refined the atomic model. N.W. designed and cloned all constructs. D.W., N.W., K.S.C., J.A.G., and O.A. expressed and purified proteins. D.W., J.A.G., and C.-L.H. performed binding studies. B.S.G. and J.S.M. supervised experiments. D.W., B.S.G., and J.S.M. wrote the manuscript with input from all authors. **Competing interests:** N.W., K.S.C., B.S.G., and J.S.M. are inventors on U.S. patent application no. 62/412,703 ("Prefusion Coronavirus Spike Proteins and Their Use"), and D.W., N.W., K.S.C., O.A., B.S.G., and J.S.M. are inventors on U.S. patent application no. 62/972,886 ("2019-nCoV Vaccine"). **Data and materials availability:** Atomic coordinates and cryo-EM maps of the reported structure have been deposited in the Protein Data Bank under accession code 6VSB and in the Electron Microscopy Data Bank under accession codes EMD-21374 and EMD-21375. Plasmids are available from B.S.G. under a material transfer agreement with the NIH or from J.S.M. under a material transfer agreement with The University of Texas at Austin.

SUPPLEMENTARY MATERIAL

science.sciencemag.org/content/367/6483/1260/suppl/DC1
Materials and Methods
Figs S1 to S8
Table S1
Movies S1 and S2
References (*32*–*41*)

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

10 February 2020; accepted 17 February 2020
Published online 19 February 2020
10.1126/science.abb2507

CANCER

Rare driver mutations in head and neck squamous cell carcinomas converge on NOTCH signaling

Sampath K. Loganathan[1], Krista Schleicher[1,2], Ahmad Malik[1,2], Rene Quevedo[3,4], Ellen Langille[1,2], Katie Teng[1], Robin H. Oh[1,2], Bhavisha Rathod[1], Ricky Tsai[1], Payman Samavarchi-Tehrani[1], Trevor J. Pugh[3,4,5], Anne-Claude Gingras[1,2], Daniel Schramek[1,2]*

In most human cancers, only a few genes are mutated at high frequencies; most are mutated at low frequencies. The functional consequences of these recurrent but infrequent "long tail" mutations are often unknown. We focused on 484 long tail genes in head and neck squamous cell carcinoma (HNSCC) and used in vivo CRISPR to screen for genes that, upon mutation, trigger tumor development in mice. Of the 15 tumor-suppressor genes identified, *ADAM10* and *AJUBA* suppressed HNSCC in a haploinsufficient manner by promoting NOTCH receptor signaling. *ADAM10* and *AJUBA* mutations or monoallelic loss occur in 28% of human HNSCC cases and are mutually exclusive with NOTCH receptor mutations. Our results show that oncogenic mutations in 67% of human HNSCC cases converge onto the NOTCH signaling pathway, making NOTCH inactivation a hallmark of HNSCC.

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common human cancer, and the 5-year survival rate is <50% (*1*). HNSCC arises in the mucosal lining of the upper aerodigestive tract and is tightly linked to tobacco use, alcohol consumption, and human papillomavirus (HPV) infection. The most common genetic alterations in HNSCC affect *p53* (71%), *FAT1* (23%), *CDKN2A* (22%), *PIK3CA* (18%), *NOTCH1* (17%), and *HRAS* (6%), followed by a "long tail" of hundreds of individually rare mutations, most of which lack biological or clinical validation (*2*, *3*) (fig. S1, A and B).

The long tails of recurrent but rare mutations remain enigmatic in cancers (*4*, *5*). It is difficult to reconcile their apparent positive selection with their low frequency and immense diversity. The long tail might reflect the stochasticity of cancer evolution, where each mutation confers only a small fitness advantage, but several such low-penetrant mutations might cooperate and substantially promote tumor progression (*6*, *7*). Some long tail mutations may be highly penetrant but simply affect genes that are rarely mutated. Several mutations may also converge on the same cellular signaling pathway, which could explain some of the diversity and low frequency, highlighting the importance of a pathway-centric view of the cancer landscape (*8*, *9*). Given that long tail driver mutations can be the genetic basis for exceptional responses to therapy (*10*, *11*), it is crucial to identify clinically relevant mutations and define their oncogenic mechanisms.

To functionally assess HNSCC long tail genes, we developed a CRISPR screen to identify genes that, upon mutation, predispose mice to HNSCC development. We constructed lentiviruses that coexpress a single-guide RNA (sgRNA) and Cre recombinase and used ultrasound-guided in utero microinjections to deliver these lentiviruses to the single-layered surface ectoderm of live mouse embryos (*12*) (Fig. 1A). The surface ectoderm generates several structures, including the skin epithelium and oral mucosa. We used multicolor Lox-Stop-Lox-Confetti mice, which, upon Cre-mediated excision of the Lox-Stop-Lox (LSL) cassette, stochastically switch on one of four fluorescent proteins, and determined the viral titer required to generate thousands of discrete clones within the oral cavity and epidermis (Fig. 1B).

Next, we tested the efficiency of CRISPR/Cas9-mediated in vivo mutagenesis (fig. S2A). We found that LSL-Cas9-GFP;LSL-tdTomato mice transduced with a Cre lentivirus encoding a scrambled sgRNA (LV-sgScr-Cre) coexpressed green fluorescent protein (GFP) and tdTomato. By contrast, mice transduced with sgRNA-Cre lentivirus targeting GFP (LV-sgGFP-Cre) displayed tdTomato$^+$ cells lacking GFP, demonstrating a knockout efficiency of 85 ± 5% (fig. S2B). Next, we targeted the heme biosynthesis gene *Urod*, the loss of which leads to the accumulation of unprocessed, fluorescent porphyrins (*13*), which resulted in bright red fluorescence in the oral mucosa and skin (fig. S2C), demonstrating efficient targeting of an endogenous gene.

To determine whether this approach can reveal genetic interactions, we recapitulated cooperation between oncogenic phosphatidylinositol 3-kinase (PI3K)/Akt signaling and loss of tumor suppressors such as transforming growth factor-β receptor II (*Tgfβr II*) or *p53* (*14*). Cas9-mediated ablation of *Tgfβr II* or *p53* in conditional *Pik3ca*^{H1047R} mice (LSL-*Pik3ca*^{H1047R};LSL-*Cas9*) triggered rapid formation of HNSCC tumors, whereas littermates transduced with scrambled control sgRNAs remained asymptomatic (fig. S2, D and E). Because of the transduction method used, these mice simultaneously developed HNSCC as well as cutaneous SCC (cSCC). The latter tumors are genetically, histologically, and pathologically related to HNSCC (*3*).

Next, we generated a lentiviral sgRNA library to mutate the mouse homologs of 484 recurrent but infrequent human HNSCC long tail genes and control libraries containing 418 nontargeting sgRNAs or 414 randomly picked genes (four sgRNAs per gene) (figs. S1 and S2F and tables S1 and S2) (*2*). To identify tumor-suppressor genes that cooperate with known HNSCC oncogenic driver mutations, we transduced these sgRNA libraries into LSL-Cas9 mice that harbor (i) a conditional *Pik3ca* oncogene (LSL-*Pik3ca*^{H1047R}); (ii) a conditional *HRas* oncogene (LSL-*HRas*^{G12V}); (iii) a conditional, dominant-negative p53 mutation (LSL-*p53*^{R270H}); or (iv) epithelium-specific expression of the HPV16-E6/E7 oncogenes (K14-*HPV16*). Within a year, none of the mice transduced with the control libraries developed tumors, highlighting that *Pik3ca*^{H1047R}, *HRas*^{G12V}, *p53*^{R270H}, and HPV16-E6/E7 on their own are insufficient to initiate SCC. By contrast, all mice transduced with the long tail sgRNA library developed multiple HNSCC and cSCC tumors within weeks (Fig. 1, C and D, and fig. S3, A and B). *Pik3ca*^{H1047R};*Cas9* mice transduced with an sgRNA library targeting 215 breast cancer long tail genes did not develop tumors over a 4-month observation period (fig. S2G and table S2), indicating the existence of strong, HNSCC-specific long tail tumor-suppressor genes.

To identify these tumor-suppressor genes, we determined sgRNA representation in 205 mouse tumors. Most tumors showed strong enrichment for a single sgRNA compared with tumor-adjacent, phenotypically normal epithelium (fig. S3C). Fifteen genes showed enrichment of two or more independent sgRNAs in multiple tumors, with *Adam10*, *Ripk4*, and *Ajuba* being the most prevalent hits, followed by *Notch2* and *Notch3* (Fig. 1E, fig. S3D, and table S3). Although most genes scored in all oncogenic backgrounds, *Ripk4*, for example, did not surface in HPV-mutant mice. For validation of our top hits, we injected *Pik3ca*^{H1047R};*Cas9* and *HRas*^{G12V};*Cas9* mouse embryos with *Adam10*, *Ajuba*, or *Ripk4* sgRNAs that were not present in the initial library. These mice rapidly developed invasive SCC tumors but were more

[1]Centre for Molecular and Systems Biology, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [3]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. [4]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [5]Ontario Institute for Cancer Research, Toronto, Ontario, Canada.
*Corresponding author. Email: schramek@lunenfeld.ca

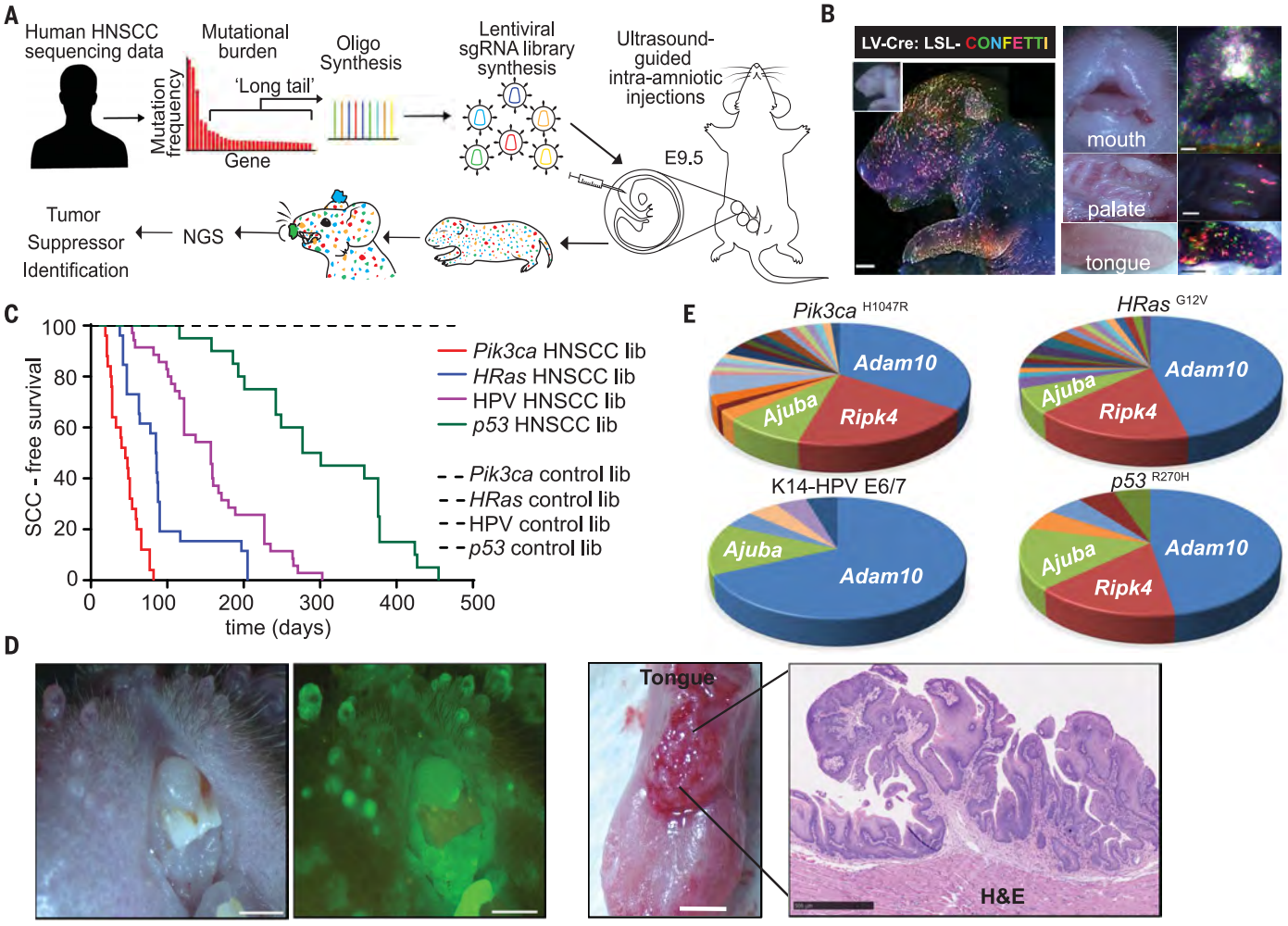


Fig. 1. Direct in vivo CRISPR screen for HNSCC tumor suppressors. (**A**) Experimental workflow for pooled in vivo CRISPR screen. A lentiviral sgRNA library targeting mouse homologs of human HNSCC long tail genes is introduced into the surface epithelium of mice, and tumors are analyzed by next-generation sequencing (NGS). (**B**) Representative images of whole body, oral cavity, tongue, and palate of newborn Cre-reporter LSL-Confetti mice transduced with Cre lentivirus. Scale bars, 500 μm. (**C**) Tumor-free survival of mice transduced with an sgRNA library targeting putative HNSCC genes or with a control sgRNA library ($n > 20$ per group; $P < 0.0001$, log-rank test). (**D**) Representative images of *Pik3ca*^{H1047R};Cas9 mice transduced with the HNSCC sgRNA library showing multiple, progressively growing tumors in the oral cavity and tongue. Scale bars, 500 μm. (**E**) Pie chart showing tumor-suppressor genes with enriched sgRNAs in tumor DNA obtained from the four different HNSCC mouse models.

prone to developing HNSCC than cSCC compared with mice with *p53* or *Tgfβr1I* loss (fig. S4, A to E). Sequencing of the sgRNA target sites in tumor DNA confirmed frame-shift mutations in all tested tumors (fig. S5). Inducible activation of the *Pik3ca*^{H1047R} oncogene and Cas9-mediated mutagenesis of *Adam10* or *Ripk4* in adult mice also led to rapid HNSCC development (fig. S6, C and D). Loss of *Adam10* or *Ripk4* triggered tumor development even in wild-type mice with very long latency (fig. S6, A and B). Thus, *Adam10*, *Ripk4*, and *Ajuba* are strong HNSCC suppressors.

The membrane-anchored metalloproteinase ADAM10 controls diverse cellular processes through regulated intramembrane proteolysis of numerous proteins (*15*–*17*). To confirm its role in suppressing HNSCC, we generated conditional *Adam10*$^{fl/fl}$;*Pik3ca*^{H1047R} compound mutant mice and transduced embryos with lentiviral Cre. Homozygous *Adam10* deletion rapidly induced HNSCC tumors, recapitulating our CRISPR/Cas9 findings (Fig. 2A and fig. S7, A to C). Even heterozygous *Adam10*$^{fl/+}$;*Pik3ca*^{H1047R} mice developed HNSCC, although with longer latency (Fig. 2A). To assess whether tumor development was due to *Adam10* loss of heterozygosity, we used fluorescence-activated cell sorting to isolate tumor cells from *Adam10*-homozygous and -heterozygous *Pik3ca*^{H1047R} tumors and from control sg*p53*;*Pik3ca*^{H1047R} tumors. Transcriptional profiling and Western blot analysis revealed wild-type Adam10 expression in *Adam10*-heterozygous tumors, albeit at a reduced level compared with control tumors (Fig. 2B and fig. S7D), indicating that *Adam10* functions as a haploinsufficient tumor suppressor in HNSCC.

Gene set enrichment analysis (GSEA) of these tumor cell transcriptomes revealed differentially expressed gene sets specifically associated with "Hallmarks of G2M checkpoint" and "E2F targets" as well as "NOTCH signaling" (Fig. 2C and fig. S8A). NOTCH receptors are transmembrane proteins that are proteolytically cleaved upon binding to JAG1/2 or DLL1/3/4 ligands. The cleaved NOTCH intracellular domain (NOTCH^{IC}) enters the nucleus, binds the DNA-binding protein RBPJ, and regulates gene expression. *Adam10*-knockout tumor cells showed reduced expression of NOTCH target genes such as *Hes1* and *Heyl* (fig. S8E), consistent with ligand-dependent activation of NOTCH by ADAM10-mediated cleavage (*14*, *17*).

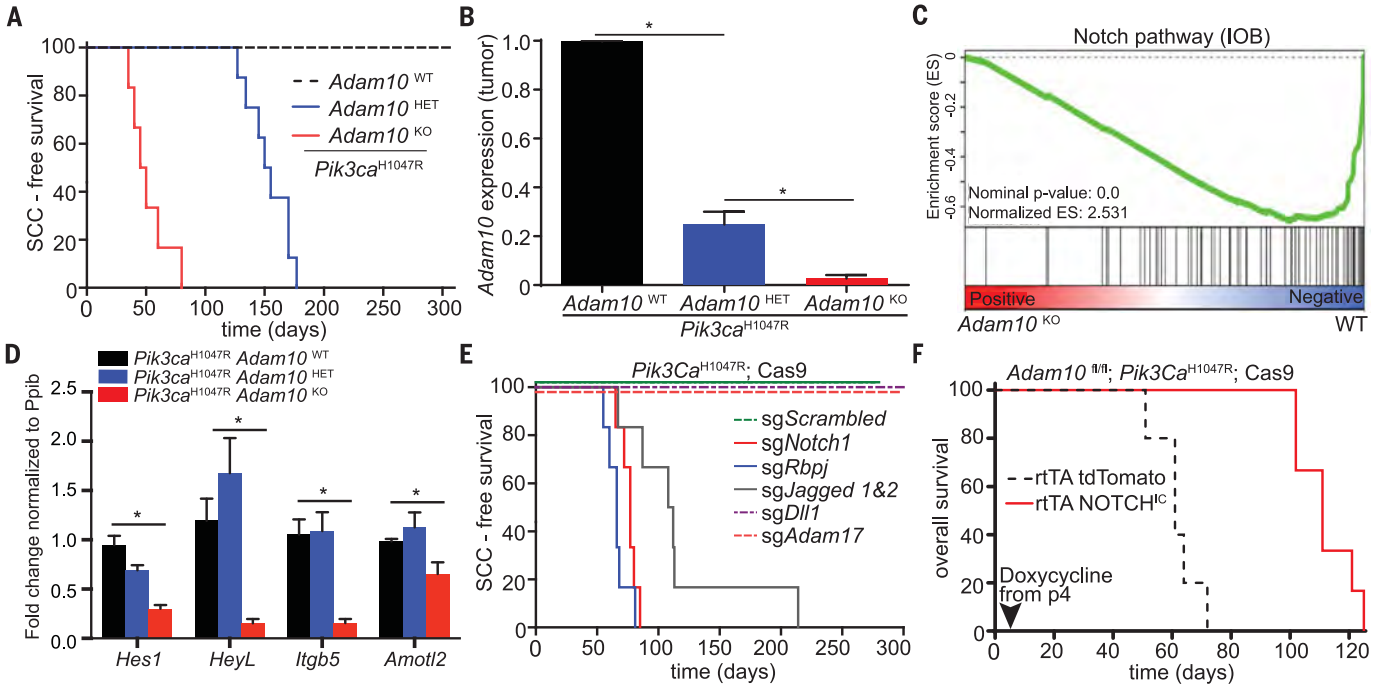


Fig. 2. Adam10 suppresses HNSCC development by regulating NOTCH. (**A**) Tumor-free survival for *Adam10*$^{fl/fl}$, *Adam10*$^{fl/+}$, and *Adam10*$^{+/+}$;*Pik3ca*^{H1047R} mice transduced with Cre lentivirus. (**B**) *Adam10* mRNA expression levels in tumors of the indicated genotype. (**C**) GSEA reveals down-regulation of the NOTCH pathway in *Adam10*$^{fl/fl}$;*Pik3ca*^{H1047R} tumors. (**D**) RT-PCR analysis of NOTCH target gene expression in keratinocytes isolated from postnatal day 4 pups of the indicated genotype. Data are shown as mean ± SEM (n = 3). *P < 0.05. (**E**) Tumor-free survival for *Pik3ca*^{H1047R} mice transduced with lentiviral Cre-sgRNA targeting the indicated NOTCH pathway components. (**F**) NOTCH^{IC} expression rescues loss of *Adam10*. Tumor-free survival of *Adam10*$^{fl/fl}$;*Pik3ca*^{H1047R} mice transduced with Cre lentiviral constructs that allow for doxycycline-inducible expression of NOTCH^{IC}. Inducible expression of tdTomato served as a control.

Down-regulated NOTCH pathway activity was also seen in primary keratinocytes isolated from *Adam10*-deficient *Pik3ca*^{H1047R} mice 4 days after birth, before any overt phenotypic changes could be observed (Fig. 2D and fig. S8, B to D), indicating that ADAM10 directly activates NOTCH.

Given the prominent role of NOTCH signaling in suppressing mouse and human HNSCC development (*18*), we tested whether genetic ablation of canonical NOTCH signaling components in *Pik3ca*^{H1047R} mice phenocopies *Adam10* loss. Indeed, sgRNA/Cas9-mediated mutagenesis of *Notch1*, *Jag1*, and *Jag2* or the obligate transcriptional NOTCH effector *Rbpj*, as well as conditional deletion of *Notch1* and *Notch2*, caused HNSCC development. Targeting the NOTCH ligand *Dll1* or *Adam17*, a metalloproteinase implicated in ligand-independent NOTCH activation that is required for epidermal barrier function (*19*, *20*), did not trigger tumor development (Fig. 2E; fig. S9, A and B; and fig. S5D). Finally, we tested whether restoring NOTCH signaling in *Adam10*-deficient mice would block tumor development. Inducible expression of ectopic NOTCH^{IC} before tumor appearance or in established tumors resulted in effective tumor suppression and delayed HNSCC development (Fig. 2F and fig. S9, C to F). Thus, loss of *Adam10* promotes HNSCC tumorigenesis by impairing NOTCH signaling.

Our second hit, *AJUBA*, is mutated in 7% of HNSCC, 18% of cutaneous SCC, and 2 to 7% of esophageal SCC cases, but not in other human cancers (*3*, *21*). AJUBA is a scaffold protein and has been shown to regulate the Hippo (*22*), Wnt (*23*), and Aurora-A signaling pathways (*24*) and cell adhesion molecules (*25*), but its role in SCC pathogenesis is unclear (*1*). *Ajuba*-null mice were embryonically lethal. However, we found that loss of one *Ajuba* allele triggered HNSCC development in *Pik3ca*^{H1047R}, *HRas*^{G12V}, or K14-*HPV16* mice but retained expression of the wild-type *Ajuba* allele (Fig. 3A and fig. S10, A to D). These data suggest that *Ajuba* also functions as a haploinsufficient tumor suppressor.

Transcriptional profiling and GSEA of keratinocytes isolated from *sgAjuba*;*Pik3ca*^{H1047R} mice 4 days after birth showed no difference in Hippo or Wnt pathway activation but significant down-regulation of "NOTCH signaling" (fig. S11, A to D). To corroborate this finding, we investigated the effect of *Ajuba* loss on NOTCH pathway activation in primary keratinocytes. We used either the canonical NOTCH ligand JAG1 or EDTA treatment, which forces dissociation of the heterodimeric NOTCH receptor, thereby allowing release of NOTCH^{IC}. Loss of Ajuba impaired ligand- and EDTA-induced up-regulation of NOTCH target genes such as *Hes1* and *Hey1* (Fig. 3B and fig. S11E). To test whether these changes are directly mediated by NOTCH^{IC}, we performed Rbpj CUT&RUN (cleavage under targets and release using nuclease) sequencing (*26*). Most Rbpj-binding sites in primary keratinocytes are in enhancers and promoters, where Rbpj is dynamically recruited upon NOTCH activation (fig. S12, A and B) (*27*). NOTCH pathway inhibition by pharmacological ADAM10 inhibition or loss of *Ajuba* markedly impaired Rbpj recruitment to and the expression of canonical NOTCH target genes such as *Hes1*, *Heyl*, and *Nrarp* and other genes such as *Id1*, *Egfr*, *Tgfβ1*, *Krt10*, and *JunB* (fig. S12, A to D, and table S4). Ten of the 15 primary screening hits are themselves NOTCH target genes, including *Ajuba*, *Adam10*, *Ripk4*, *Notch1/3*, *Jag1/2*, and *Irf6*, indicating the existence of a feed-forward loop to enhance NOTCH signaling and thus tumor suppression (figs. S12C and S17 and table S3).

To investigate Ajuba's function, we identified vicinal proteins by proximity-dependent biotinylation coupled to mass spectrometry (BioID) in human 293 cells. AJUBA BioID enriched for Hippo components (LATS1/2, AMOT, and PTPN14) and several new proximity interactors including NOTCH1, NOTCH2, and

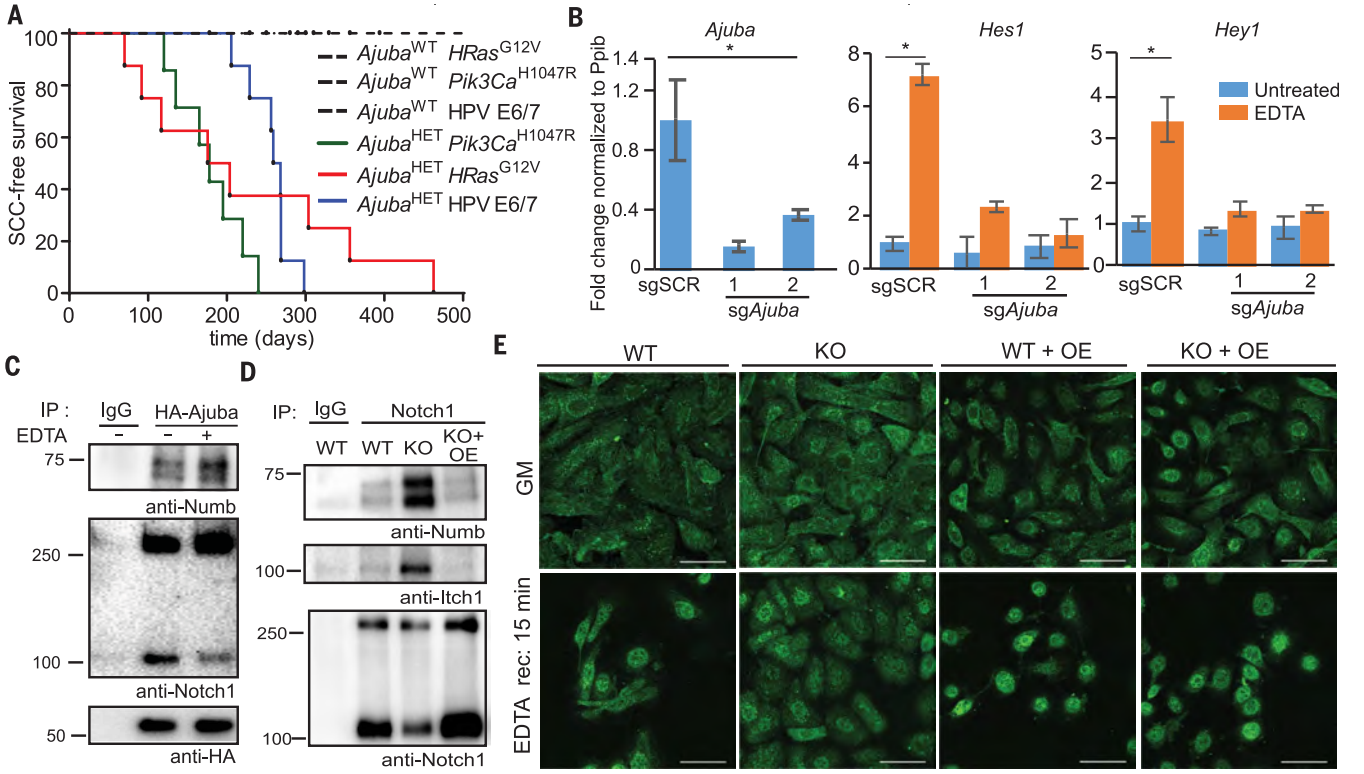


Fig. 3. Ajuba is a new NOTCH regulator. (**A**) Tumor-free survival of Ajuba$^{+/-}$ and Ajuba$^{+/+}$ mice in *Pik3ca*^{H1047R}, *HRas*^{G12V}, or K14-HPV16 mouse backgrounds. (**B**) RT-PCR results showing effects of CRISPR/Cas9-mediated ablation of *Ajuba* on EDTA activation of the canonical NOTCH targets *Hes1* and *Hey1* in primary mouse keratinocytes. Data are shown as means ± SEM (n = 3). *P < 0.05. (**C**) Co-IP of HA-Ajuba with endogenous Numb, full-length Notch1 (300 kDa), and NOTCH1^{IC} (120 kDa) in primary keratinocytes, which changes upon NOTCH pathway activation (30 min EDTA; 15 min recovery). (**D**) Co-IP of endogenous Notch1 with Numb and Itch1 in wild-type (WT), Ajuba-knockout (KO), and Ajuba-overexpressing (KO+OE) primary keratinocytes. (**E**) Genetic ablation of Ajuba impairs nuclear accumulation of NOTCH^{IC}. Immunofluorescence of Ajuba WT and KO as well as Ajuba-overexpressing (OE) WT and KO primary keratinocytes is shown. Cells were treated with EDTA (30 min) to stimulate NOTCH receptor activation, allowed to recover for 15 min (EDTA rec), and stained for NOTCH1. Scale bar, 50 μm.

NUMB (fig. S13A). NUMB recruits the E3 ubiquitin ligase ITCH to NOTCH, thereby facilitating NOTCH ubiquitination and degradation (*28*). In accordance with the BioID results, FLAG-tagged Ajuba coimmunoprecipitated (co-IP) endogenous Lats1, Notch1, Notch2, and Numb from primary mouse keratinocytes (fig. S13B and Fig. 3C). The reciprocal co-IP also showed that endogenous Notch1 and Notch2 interacted with endogenous Ajuba (fig. S13C).

Next, we explored how Ajuba regulates NOTCH signaling. NOTCH activation increased the interaction between Ajuba and Numb but reduced binding between AJUBA and cleaved NOTCH1^{IC} (Fig. 3C), implicating a dynamically regulated process. To test functional relevance, we genetically ablated Ajuba in primary keratinocytes and analyzed Notch-Numb-Itch complex formation. Loss of Ajuba resulted in (i) increased association between Numb and Notch1, (ii) aberrant recruitment of Itch to Notch1, and (iii) concomitant increased Notch1 ubiquitination, all of which was reversed by reexpressing Ajuba (Fig. 3D and fig. S13, D and E). Consistent with these findings, nuclear accumulation of NOTCH1^{IC} and NOTCH2^{IC} after receptor activation was markedly reduced in *Ajuba*-knockout cells, which was reversed upon reexpressing Ajuba (Fig. 3E and fig. S14, A to D). Last, we tested whether Numb, which is amplified in ~37% of HNSCC patients, could promote HNSCC development. Indeed, overexpression of Numb in *Pik3ca*^{H1047R} mice triggered rapid HNSCC and cSCC development (fig. S14E). These data suggest that Ajuba binds and sequesters Numb, thereby promoting NOTCH signaling and HNSCC tumor suppression.

Our third hit, *Ripk4* (Receptor Interacting Protein Kinase 4), was reported to be a NOTCH target gene (*29*), further implicating NOTCH signaling in HNSCC. Indeed, NOTCH pathway activation led to increased Rbpj CUT&RUN footprints in promoter and enhancer regions of *Ripk4* and transcriptional up-regulation of *Ripk4* (figs. S11E and S12, B and C). Conditional *Ripk4*$^{fl/fl}$;*Pik3ca*^{H1047R} but not heterozygous *Ripk4*$^{fl/+}$;*Pik3ca*^{H1047R} compound mutant mice formed tumors upon lentiviral Cre transduction, corroborating our CRISPR results (fig. S15A). Ripk4 promotes the differentiation of oral and epidermal keratinocytes by phosphorylating and activating the Irf6 (Interferon Regulatory Factor 6) transcription factor (*30*). Irf6 also scored in our screen and constitutes another direct NOTCH target gene (table S3 and fig. S12C). These data indicate that the Ripk4–Irf6 axis is an important downstream effector of NOTCH.

To identify additional downstream NOTCH targets that suppress HNSCC, we generated a sgRNA library targeting 80 genes downregulated upon *Adam10* ablation (fig. S15, B and C, and table S6). Transduced *Pik3ca*^{H1047R};*Cas9* mice developed tumors enriched for sgRNAs targeting the integrin subunit *Itgb5*, the Hippo component Angiomotin-Like 2 (*Amotl2*), Endonuclease Domain Containing 1 (*Endod1*), and Sushi Domain Containing 2 (*Susd2*) (Fig. 4, A and B, and fig. S15D). Independent sgRNAs targeting *Itgb5* or *Amotl2* confirmed our findings and induced highly aggressive SCC in the *Pik3ca*^{H1047R} mice (Fig. 4C and figs. S5E and S15E). CUT&RUN sequencing and real-time polymerase chain reaction (RT-PCR) confirmed that *Itgb5* or *Amotl2* are direct NOTCH target genes (Fig. 2D and fig. S12C) (*27*).

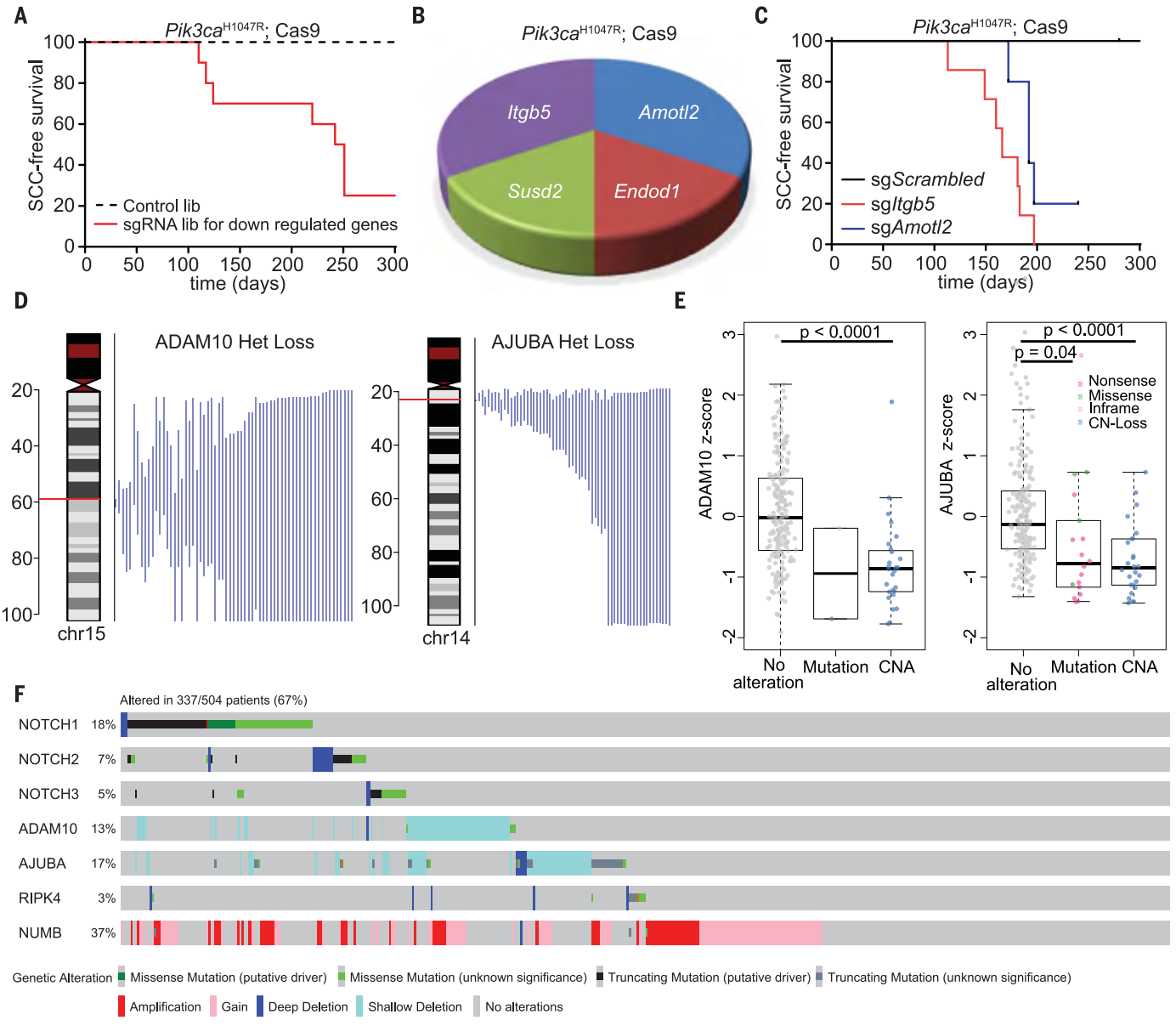


Fig. 4. NOTCH target genes and NOTCH pathway mutations in human tumors. (**A**) Tumor-free survival for *Pik3ca*^{H1047R} mice transduced with a lentiviral Cre-sgRNA library targeting genes down-regulated in *Adam10*-deficient tumors ($P < 0.0001$, log-rank test). (**B**) Pie chart showing top-scoring sgRNAs enriched in tumor DNA from (A). (**C**) Tumor-free survival for *Pik3ca*^{H1047R};Cas9 mice transduced with sgRNAs targeting *Itgb5* and *Amotl2* or scrambled control sgRNAs ($P < 0.0001$, log-rank test). (**D**) Patterns of monoallelic minimal deleted regions (MDRs) in 504 HNSCC TCGA samples. (**E**) Expression of *ADAM10* or *AJUBA* in tumors carrying *ADAM10* or *AJUBA* mutations or copy number alterations, respectively. (**F**) Genetic alterations of *NOTCH* receptors 1 to 3, *ADAM10*, *AJUBA*, *RIPK4*, and *NUMB* in human HNSCC samples ($n = 504$) from TCGA presented as a cBioPortal OncoPrint displaying a trend toward mutual exclusivity among mutations.

To extend our findings from mouse to human cancers, we analyzed 504 HNSCC cases from The Cancer Genome Atlas (*2*). The *ADAM10*, *AJUBA*, and *RIPK4* genes were mutated in 0.8%, 7.5%, and 3.0% of tumors, respectively, a frequency expected for long tail mutations. However, an additional 12.7% and 11.1% of human HNSCCs showed heterozygous *ADAM10* and *AJUBA* loss, respectively (fig. S16A). Analysis of overlapping allelic loss segments showed that chromosome 14 exhibited focal, monoallelic losses encompassing *AJUBA*, whereas *ADAM10* exhibited preferential monoallelic loss of broad chromosome segments (Fig. 4D). We also found allelic imbalance of heterozygous single nucleotide polymorphisms in the chromosomal region of *AJUBA*, further confirming heterozygous loss of *AJUBA* (fig. S16, B and C). Mutations and allelic copy number loss coincided with reduced expression of *ADAM10* and *AJUBA* (Fig. 4E). In addition, mutations and allelic loss of *ADAM10*, *AJUBA*, or *RIPK4* showed a trend toward mutual exclusivity with mutations in *NOTCH1/2/3* receptors (Fig. 4F, fig. S16D, and table S7). Altogether, ~27% of HNSCC patients carry inactivating *NOTCH1/2/3* receptor mutations, and an additional ~40% of HNSCC samples show inactivating alterations of *ADAM10* or *AJUBA* or amplification of NUMB (Fig. 4F). Although mutual exclusivity is neither a necessary nor a sufficient condition for genes involved in the same pathway, our functional

studies suggest that these alterations converge on inactivating NOTCH signaling in NOTCH receptor wild-type HNSCC patients (fig. S17). Thus, NOTCH is one of the most commonly dysregulated pathways in HNSCC.

As in the case of HNSCC, the identification of long tail genes that drive tumorigenesis in other cancer types may show convergence on specific pathways and reveal new regulators of those pathways and could inform on cancer biology and tumor evolution. This information may even provide translational opportunities and increase the number of patients who benefit from a pathway-specific treatment. In addition, our work indicates that many long tail genes function in a haploinsufficient manner. Because most algorithms used to delineate driver from passenger mutations are based on statistical enrichment of somatic point mutations, amino acid conservation, or homozygous deletions and amplification and never take heterozygous deletions into account, many haploinsufficient tumor suppressors might have been overlooked. This could be of special interest in the postgenomic era, in which most, if not all, mutations and copy number alterations have been identified, but their functional annotation lags far behind. In summary, this study shows the power of integrating cancer genomics with mouse modeling using in vivo CRISPR screens to uncover tumor-suppressive pathways in the long tail of cancer-associated mutations.

REFERENCES AND NOTES

1. C. R. Leemans, P. J. F. Snijders, R. H. Brakenhoff, *Nat. Rev. Cancer* **18**, 269–282 (2018).
2. Cancer Genome Atlas Network, *Nature* **517**, 576–582 (2015).
3. G. P. Dotto, A. K. Rustgi, *Cancer Cell* **29**, 622–637 (2016).
4. L. A. Garraway, E. S. Lander, *Cell* **153**, 17–37 (2013).
5. I. Martincorena *et al*., *Cell* **171**, 1029–1041.e21 (2017).
6. F. Castro-Giner, P. Ratcliffe, I. Tomlinson, *Nat. Rev. Cancer* **15**, 680–685 (2015).
7. I. Bozic *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18545–18550 (2010).
8. F. Sanchez-Vega *et al*., *Cell* **173**, 321–337.e10 (2018).
9. B. Vogelstein *et al*., *Science* **339**, 1546–1558 (2013).
10. H. Al-Ahmadie *et al*., *Cancer Discov.* **4**, 1014–1021 (2014).
11. G. Iyer *et al*., *Science* **338**, 221 (2012).
12. S. Beronja, G. Livshits, S. Williams, E. Fuchs, *Nat. Med.* **16**, 821–827 (2010).
13. J. Ablain, E. M. Durand, S. Yang, Y. Zhou, L. I. Zon, *Dev. Cell* **32**, 756–764 (2015).
14. Y. Bian *et al*., *Oncogene* **31**, 3322–3332 (2012).
15. M. Mullooly, P. M. McGowan, J. Crown, M. J. Duffy, *Cancer Biol. Ther.* **17**, 870–880 (2016).
16. U. Sahin *et al*., *J. Cell Biol.* **164**, 769–779 (2004).
17. S. Weber *et al*., *Development* **138**, 495–505 (2011).
18. C. S. Nowell, F. Radtke, *Nat. Rev. Cancer* **17**, 145–159 (2017).
19. A. Murthy *et al*., *Immunity* **36**, 105–119 (2012).
20. J. Guinea-Viniegra *et al*., *J. Clin. Invest.* **122**, 2898–2910 (2012).
21. J. D. Campbell *et al*., *Cell Rep.* **23**, 194–212.e6 (2018).
22. M. Das Thakur *et al*., *Curr. Biol.* **20**, 657–662 (2010).
23. K. Haraguchi *et al*., *Oncogene* **27**, 274–284 (2008).
24. T. Hirota *et al*., *Cell* **114**, 585–598 (2003).
25. S. Nola *et al*., *J. Cell Biol.* **195**, 855–871 (2011).
26. P. J. Skene, S. Henikoff, *eLife* **6**, e21856 (2017).
27. D. Castel *et al*., *Genes Dev.* **27**, 1059–1071 (2013).
28. M. A. McGill, C. J. McGlade, *J. Biol. Chem.* **278**, 23196–23203 (2003).
29. F. Meier-Stiegen *et al*., *PLOS ONE* **5**, e11481 (2010).
30. N. Oberbeck *et al*., *Nature* **574**, 249–253 (2019).

ACKNOWLEDGMENTS

We thank all members of our laboratories for helpful comments; Y. Q. Lu, D. Dervovic, and G. Mbamalu for assistance; Z. Y. Lin for mass spectrometry assistance; C. Go and J. D. R. Knight for access to the cell-map resource; The Centre for Phenogenomics and Network Biology Collaborative Centre at LTRI; and D. Durocher, J. Wrana, L. Pelletier, R. Bremner, J. McGlade, and J. Woodgett for critically reading the manuscript. **Funding:** This work was supported by a project grant to D.S. from the Canadian Institute of Health Research (CIHR 365252) and the Krembil Foundation. D.S. is the recipient of a career development award from HFSP (CDA00080/2015). S.K.L. is the recipient of a Canadian Cancer Society fellowship (BC-F-16#31919). A.C.G. was supported by a Terry Fox Research Institute program grant. **Author contributions:** S.K.L. performed all experiments. K.S. performed Ajuba mouse experiments, immunohistochemistry, and, together with K.T., helped in immunofluorescence experiments. E.L. helped to prepare the viral library. R.T. performed quantitative RT-PCR and CUT&RUN experiments. R.H.O. helped with the random genes library. A.M. performed all bioinformatic analysis. B.R. and A.-C.G. performed and analyzed the mass spectrometry experiments. P.S.-T. helped with the design of the mass spectrometry experiments. R.Q. and T.J.P. performed bioinformatics analysis on human TCGA data. D.S. coordinated the project and, together with S.K.L., designed the experiments and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All mass spectrometry data have been deposited in the MassIVE repository (MSV000083405 and PXD012600). Data for the 192 different BioID baits targeting the major subcellular compartments of a human cell used to control for AJUBA interactome specificity are available at https://humancellmap.org. All code and the manifests used to analyze the human HNSCC TCGA data are available under the R package SchramekLOH v1.0.0 (https://github.com/pughlab/SchramekLOH). All RNA-sequencing and CUT&RUN data are available at the Gene Expression Omnibus (GEO) (GSE140495, GSE140496, and GSE140497). All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6483/1264/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S7
References (*31*–*51*)

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

20 February 2019; accepted 14 February 2020
10.1126/science.aax0902

DukeNUS Medical School

SingHealth DukeNUS ACADEMIC MEDICAL CENTRE

Director
Programme in Emerging Infectious Diseases

Duke-NUS Medical School (Duke-NUS), Singapore's only graduate-entry medical school was established as a landmark collaboration between two world-ranking institutions of higher education – Duke University and National University of Singapore. Duke-NUS provides innovative education and impactful research to enhance the practice of medicine in Singapore and beyond. Through its strategic partnership with Singapore Health Services (SingHealth) which includes Singapore General Hospital, the School is able to leverage its joint capabilities and infrastructure to develop outstanding clinical education programmes and cutting-edge research collaborations that translate fundamental science into better health. Duke-NUS and its partners have created an academically-based Programme in Emerging Infectious Diseases (EID) designed to serve as a national and international resource of excellence in emerging infectious diseases. The mission of the Programme faculty is to conduct high-level basic and applied research, and to train graduate students, postdoctoral fellows and clinician-scientists in the disciplines relevant to emerging infectious diseases. Duke-NUS and its partners provide state-of-the-art research facilities, including a standalone ABSL3 unit.

We are seeking an individual with exceptional scientific credentials and leadership skills to head the EID Programme. The Programme Director will provide leadership, including engagement with the broader biomedical community in Singapore and with Duke University; strategic hiring and programme development; medical school and graduate education; faculty mentoring; budgetary and space planning. The School will provide the new Programme Director with the resources to support the highest level of research.

Interested candidates should send a CV and the names of three references to: **Chair, Search Committee on Emerging Infectious Diseases, Duke-NUS Medical School, Singapore by email to: hr@duke-nus.edu.sg**

Applications will be accepted until the position is filled. We anticipate to begin interviewing candidates in early April 2020.

More information on the Programme can be found at www.duke-nus.edu.sg.

NIH National Institute of Neurological Disorders and Stroke

Postdoctoral research position at National Institute of Neurological Disorders and Stroke, Bethesda, MD and surrounding area
Molecular Neuroscience

The lab of Dr. Katherine Roche at the National Institutes of Health (NIH) is seeking a motivated postdoctoral researcher with an interest and relevant experience in molecular mechanisms of synaptic transmission and neurodevelopmental disorders to join their group. The fellow will have the opportunity to drive meaningful projects in a collaborative and supportive environment that values high intellectual curiosity and creativity. The fellow will be supported by a generous funding package and benefits consistent with guidelines for NIH intramural research training awards.

Qualifications: Doctorate Degree

To apply the application should include: (1) Cover letter with a brief description of candidate's interest in the position, research experience and long-term career goals; (2) Curriculum Vitae; (3) Representative publications; (4) Contact information of three references.

Please email all materials as one PDF file to Dr. Roche (**rochek@ninds.nih.gov**), Investigator, National Institute of Neurological Disorders and Stroke, Receptor Biology Section, Building 35, Room 2C903, 35 Convent Drive, Bethesda, MD 20892. Review of applications will begin immediately and will continue until the position is filled.

The NIH is dedicated to building a diverse community in its training and employment programs.

THE OHIO STATE UNIVERSITY
COLLEGE OF MEDICINE

Faculty Positions in the Department of Neuroscience
College of Medicine at The Ohio State University (OSU), Wexner Medical Center

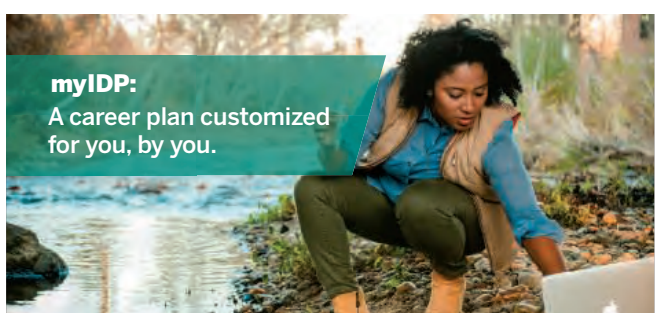
The College of Medicine, Department of Neuroscience is looking to hire two tenure-track faculty with research expertise focused in **traumatic central nervous system injury (brain or spinal cord)**. These positions may be at the rank of Assistant, Associate or Full Professor.

As part of an ongoing commitment to place OSU at the forefront of patient care, research and innovation, The College of Medicine is committing significant resources to build cutting-edge and internationally recognized basic and translational neuroscience programs. These two positions are partially funded by the Chronic Brain Injury program https://discovery.osu.edu/focus-areas/brain-injury/), part of Ohio State's Discovery Themes Initiative, a significant faculty hiring investment in key thematic areas in which the university can build on its culture of academic collaboration to make a global impact. Through this program, The Ohio State University's Colleges of Medicine, Engineering, Arts & Sciences and Nursing have already recruited 22 tenure-track Assistant Professors and tenured Associate/Full Professors who study various aspects of traumatic CNS injury, aging or neurodegenerative diseases.

Candidates must hold a Ph.D. or equivalent and at least 3 years postdoctoral experience with a strong record of scholarly activity. Candidates for appointment as Associate or Full Professors should also be nationally or internationally recognized scholars with consistent externally funded research programs.

Prospective candidates should send a statement of research interests, vita and a list of three references (PDF format) to **recruitosuneurosci@osumc**.edu. Applications are being accepted now and formal review will begin April 1, 2020. The search will remain open until both positions are filled.

The Ohio State University is an Equal Opportunity, Affirmative Action Employer and as such, women and minorities are encouraged to apply. Unless confidentiality is requested in writing, information regarding the applicants must be released upon request.



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.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:

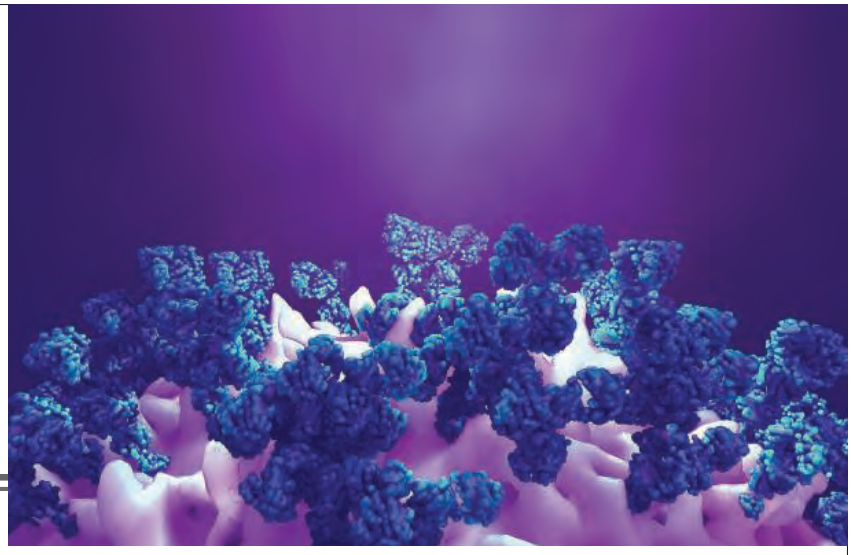
FASEB · University of Massachusetts Medical School · UCSF University of California San Francisco · UIC · Burroughs Wellcome Fund

SPECIAL JOB FOCUS:

Cancer Research

Issue date: April 10
Book ad by March 26

Ads accepted until April 3 if space allows



Deliver your message to a global audience of targeted, qualified scientists.

129,566
subscribers in print every week

26,776
yearly unique active job seekers searching for cancer research jobs

25,690
yearly applications submitted for cancer research positions

What makes *Science* the best choice for recruiting?

- Read and respected by 400,000 readers around the globe
- Your ad dollars support AAAS and its programs, which strengthens the global scientific community.

Why choose this job focus for your advertisement?

- Relevant ads lead off the career section with a special cancer research banner
- Bonus distributions:
  American Association for Cancer Research, April 24-29, San Diego, CA

  AACR Career Fair, April 25, San Diego, CA.

Expand your exposure by posting your print ad online:

- Additional marketing driving relevant job seekers to the job board.

Produced by the *Science*/AAAS Custom Publishing Office.

To book your ad, contact: advertise@sciencecareers.org

The Americas
202 326 6577
Europe
+44 (0) 1223 326527
Japan
+81 3 6459 4174
China/Korea/Singapore/ Taiwan
+86 131 4114 0012

Science Careers
AAAS
SCIENCECAREERS.ORG

FOR RECRUITMENT IN SCIENCE, THERE'S ONLY ONE *SCIENCE*.

The **Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC)** is an internationally renowned biomedical research center. Researchers at the MDC investigate the molecular mechanisms of health and disease, and translate those findings into medical application. The MDC is a research institution funded by the German Federal Government and the State of Berlin, and is a member of the Helmholtz Association of German Research Centers.

MDC researchers collaborate closely with colleagues at Berlin universities, Charité – Universitätsmedizin Berlin, the Berlin Institute of Health and many other institutions. The MDC is a member institution of the German Center for Cardiovascular Research (DZHK) and the National Cohort. Additionally, the MDC works with biotech companies and the pharmaceutical industry in developing applications from its research.

The MDC has a current annual budget of approximately €180 million. Around 1,250 employees and 500 guests currently work at the Center's sites in Berlin-Buch and Berlin-Mitte. Employee diversity is one of its strengths, enhancing the creativity and innovative spirit at the MDC.

The **MDC** invites applications for the position of

Chair of the Board and Scientific Director (m/f/x)

to be filled as soon as possible.

The Chair of the Board simultaneously serves as the Scientific Director and officially represents the Center and its research, especially at the intersection of science and politics. The candidate is expected to advance the development of the MDC within the Helmholtz Association and the collaboration with strategic partners, such as Charité – Universitätsmedizin, the Berlin Institute of Health and universities in Berlin. The candidate shares responsibility with the Administrative Director for the operative management and sustainable success of the Center.

Due to the responsibilities and size of the Center as well as the close collaboration with partners in universities, clinics, corporations and other national and international research centers, the candidate will need excellent leadership and communication skills, as well as experience in the scientific management of an international scientific institution or in similar areas. Additional prerequisites are an ability to develop strategic plans, to present and implement these persuasively and assertively, and to integrate staff members effectively. The candidate is also expected to have experience in international cooperation and in the transfer of knowledge from research and science into society.

A strong track record in basic biomedical research and/or translational medicine and/or pharmaceutical research and development is also required.

Experience is expected in strategic corporate management as well as in interactions with governance bodies such as supervisory boards and international scientific advisory boards. Excellent English and German language skills are required, and knowledge of the German research system would be an advantage.

The position will be filled for five years, with the possibility of extension for additional terms. The salary will be based on the remunerative framework for prominent university professors, with the possibility of performance-related remuneration.

The members of the Helmholtz Association have set the goal of attracting more women into leadership and decision-making positions in science and research. Therefore, we especially encourage qualified female candidates to apply. Severely disabled applicants will be given preferential treatment in the case of equal qualification.

Contact: Specific questions regarding this position should be directed to:
Prof. Dr. Veronika von Messling, e-mail: veronika.vonMessling@bmbf.bund.de

Applications should be sent both electronically (by e-mail) and through the post. They must be received by **30.04.2020** and should be addressed to:

Frau MinDir'in Prof. Dr. Veronika von Messling
Vorsitzende des Aufsichtsrates des MDC und Vorsitzende der Findungskommission
– persönlich –
Bundesministerium für Bildung und Forschung
Kapelle Ufer 1
10117 Berlin
veronika.vonMessling@bmbf.bund.de

We would like to point out that your documents will be made accessible to the selection committee and to an external recruitment agency. The selection committee is composed of representatives from the Supervisory Board and the Scientific Advisory Board of the MDC, the Federal Ministry of Education and Research, as well as external researchers and guests. Please note the MDC's data privacy policies: https://www.mdc-berlin.de/data-privacy and https://www.mdc-berlin.de/content/data-protection-declaration-applicants (Data protection declaration for applicants)

By **Paul Bierman**

Teamwork, the Cuban way

I settled into my seat on a plane bound for Cuba feeling frustrated. When I planned the trip, I had assumed that my Cuban collaborators and I would hit the ground running, heading out into the field straight away to collect water and sediment samples from rivers. That's how I'd done fieldwork in Namibia, Bolivia, and Greenland. But not in Cuba, it seemed. Five days earlier, a Cuban scientist emailed to inform me that we'd only be meeting to talk about our planned project. Sampling would happen during a later trip, she wrote. That left me feeling impatient and unhappy. Why did I need to travel there to have a meeting? But I had something to learn in Cuba.

At the airport, one of my collaborators greeted me with a broad smile. "Welcome to Cuba!" he exclaimed in perfect English, giving me a strong handshake and a hug. The next day, we drove to the research center where he worked. As scorpions scurried across the floor of the conference room, each of us gave a presentation about our science and what we hoped to learn from the study of Cuban rivers.

Then, we toured every lab in the building. I met scientists, technicians, secretaries, students, and the cook. Some spoke English; others communicated to me in Spanish while my collaborator translated. I was impressed that I was introduced to each person in their center. The lack of hierarchy—the team atmosphere—was unlike anything I'd experienced before in academia.

"The lack of hierarchy ... was unlike anything I'd experienced before in academia."

The next day, we met again to brainstorm. Together, we pored over maps to plan how we were going to collect samples. Had it not been for the Cubans, I would have been unaware that the maps I had were outdated and wrong. They left out reservoirs, which was a problem because had we sampled downstream of them, our results would have been biased. Local involvement and knowledge were key—making me wonder what I'd missed working without such a team in Africa, South America, and the Arctic.

Six months later, I flew back to Cuba and—this time—we headed into the field. I was impressed, yet again, by the lengths to which my collaborators went to ensure that all team members were treated equally. We drove around Cuba in bright yellow vans, and we made sure that each van had a mix of Cubans and Americans at all seniority levels. In the field, students, faculty, and technicians all sweated together.

On the last night of the trip, we searched for a restaurant that could seat all 14 of us at one table—because that's what teams do, they sit together. When a restaurant couldn't accommodate the team without splitting us up, my collaborators insisted that we move on and find a place with a large enough table.

In 26 years as a professor, I've always tried my best to treat my students as valued collaborators. I have never been a fan of academia's hierarchy. I want everyone working with me to feel as though they are part of a team. But my Cuban collaborators take teamwork to another level entirely. They make it clear—through actions both big and small—that all team members are valued, that everyone is equal, and that true teamwork makes for better science.

I returned to the United States a changed scientist. Now, I spend more time listening and making sure that everyone's voice is heard. Four months ago, I took the Cuban approach to heart when I led a workshop for scientists from five countries. We met to discuss how we were going to analyze a few precious grams of rock collected from beneath the Greenland Ice Sheet. I made sure that every scientist had a voice in the discussions and that all 35 of us ate dinners together. The approach worked: We began as individuals, but after the workshop, we were a team.

Many people outside Cuba focus on its communist system or the bad blood between our two countries. In Cuba, my collaborators taught me about coming together. I learned that the best teams recognize that individual members bring different perspectives to the table. All voices have merit, and each and every person deserves respect. I hope this essay inspires others to recognize the power of real teamwork—even during routine moments such as dinner. ■

Paul Bierman is a Gund Fellow and professor at the University of Vermont in Burlington. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER