Does Earth hold a wealth of clean hydrogen? p. 630

Tsetse flies produce volatile pheromones pp. 638 & 660

Mechanistic insight in immunotherapy wins Michelson & *Science* Prize p. 654

Science

$15
17 FEBRUARY 2023
science.org

AAAS

SIGHT UNSEEN

Hiding pigmented eyes from predators

pp. 643 & 695

EDITORIAL

Blind spots in biodefense

In October, the Biden administration released its National Biodefense Strategy (NBS-22), the first update since the COVID-19 pandemic began. Although the document notes that one of the lessons of the pandemic is that threats originating anywhere are threats everywhere, it frames threats as largely external to the United States. NBS-22 focuses primarily on bioterrorism and laboratory accidents, neglecting threats posed by routine practices of animal use and production inside the United States. NBS-22 references zoonotic disease but assures readers that no new legal authorities or institutional innovations are needed. Although the US is not alone in failing to confront these risks, its failure to comprehensively address them echoes across the globe.

More zoonotic diseases originated in the United States than in any other country during the second half of the 20th century. In 2022, the US processed more than 10 billion livestock, the largest number ever recorded and an increase of 204 million over 2021. Risks occur across the supply chain, from facilities where animals are born to homes where they are consumed. The ongoing H5N1 avian influenza outbreak has left 58 million animals dead in backyard chicken coops and industrial farms. It has infected animals in one of the dozens of live poultry markets in New York City (elsewhere called "wet markets"). Of the many agencies that govern food animal production, the US Department of Agriculture is the most important, but even it has no authority to regulate on-farm animal production.

Since 2011, the US has recorded more swine-origin influenza infections than any other country. Most occurred at state and county fairs, where an estimated 18% of swine have tested positive. These fairs attract 150 million visitors each year. In 2012, H3N2v influenza jumped from pigs to humans at livestock exhibitions and infected 306 people across 10 states, with suspected human-to-human transmission. Still, animal fairs remain largely unregulated and exempt from federal oversight.

Zoonotic risks also arise from interactions with free and captive wildlife. Each year, the US consumes an estimated 1 billion pounds of "game" (elsewhere called "bushmeat"). Yet, most hunter-harvested meat is not inspected, and no sanitary measures are required. Avian influenza has spread from wild birds to hunters and also appeared in captive game farms, where 40 million birds are raised annually. Three million mink live on US fur farms in long rows of wire cages where their waste falls onto the floor or onto other animals below. Though mink are slaughtered on-site, no federal laws govern them, and few states even require licenses. In some states, regulators were unaware that fur farms existed within their borders until animals in them began contracting COVID-19. In Michigan, mink generated a new strain of the virus, transmitting it to workers; in Utah, health department officials were denied access to an infected farm, unable to carry out containment efforts or even testing.

> "The US is still very far from... recognizing its responsibility in generating these global risks."

The US is the largest importer of wildlife in the world. More than 200 million live wild animals enter the US each year, most undergoing no health and safety checks when they arrive. Live animal imports are governed by a range of agencies, many with ill-defined jurisdictions. The US Fish and Wildlife Service, which oversees most of these imports, has stated that it does not have independent authority to detain shipments of sick animals. Mpox arrived in the US in 2003 in one of these shipments, destined for the pet trade.

These examples illustrate a regulatory system in urgent need of reconstruction. What is needed is not simply for agencies to do their jobs better or to paper over the gaps, but a fundamental restructuring of the way that human–animal interfaces are governed. A One Health approach, which NBS-22 claims as its guiding principle, would take the health of other living things not merely as the occasional means or obstacles to human health, but as continuous with it. The first step in implementing such an approach would be to create a high-level process for integrating the broken mosaic of multiple agencies, with their unclear and sometimes competing mandates, into an effective, comprehensive regime. The US is still very far from taking such decisive action, or even recognizing its responsibility in generating these global risks.

–Ann Linder and Dale Jamieson

Ann Linder
is at the Brooks McCormick Jr. Animal Law and Policy Program, Harvard Law School, Cambridge, MA, USA. alinder@law.harvard.edu

Dale Jamieson
is at the Center for Environmental and Animal Protection, Department of Environmental Studies, New York University, New York, NY, USA. dale.jamieson@nyu.edu

10.1126/science.adg9237

NEWS

IN BRIEF

Edited by
Jeffrey Brainard

ASTRONOMY

Chile radio telescope array gains keener vision

ALMA, one of the world's biggest radio telescope arrays, is getting hardware and software upgrades to allow it to collect much more data and produce sharper images. Announced last week, the latest modernizations to ALMA—officially the Atacama Large Millimeter/submillimeter Array, which sits high in the Chilean Andes—will cost $37 million and take 6 years to complete. Workers will upgrade the data transmission from the dishes to a central processor. They will also replace the heart of that processor, known as the correlator, a supercomputer that combines the input from individual dishes into composite images. The new correlator's speed will effectively create 1000 more hours of observing time every year. And in an additional, previously funded upgrade project now underway, ALMA is getting new detectors sensitive to radio waves from 1.1 millimeters to 1.4 millimeters wavelength on each of its 66 dishes, allowing it to gain a clearer view of objects from the Solar System and the far reaches of the universe.

Japan's science council under fire

POLICY | Science groups in Japan are lining up against a potential government move to remake the Science Council of Japan (SCJ), the country's science academy, arguing the changes would weaken the council's independence. More than five dozen academic societies, as well as SCJ itself, have raised concerns that the government will soon ask the Diet to approve legislation giving government officials greater control over who serves on SCJ's governing board, selections traditionally made by council members. In 2020, however, then–Prime Minister Yoshihide Suga sparked controversy by blocking the appointment of six of 105 nominees, and late last year the current government suggested that a vaguely defined outside panel help select nominees. SCJ has asked the government to reconsider that idea, saying it would put the council "under political and administrative control or pressure." Observers say the government has objected to several of SCJ's reports, including one urging universities to be cautious in accepting funding for research on dual-use, civilian-military technology.

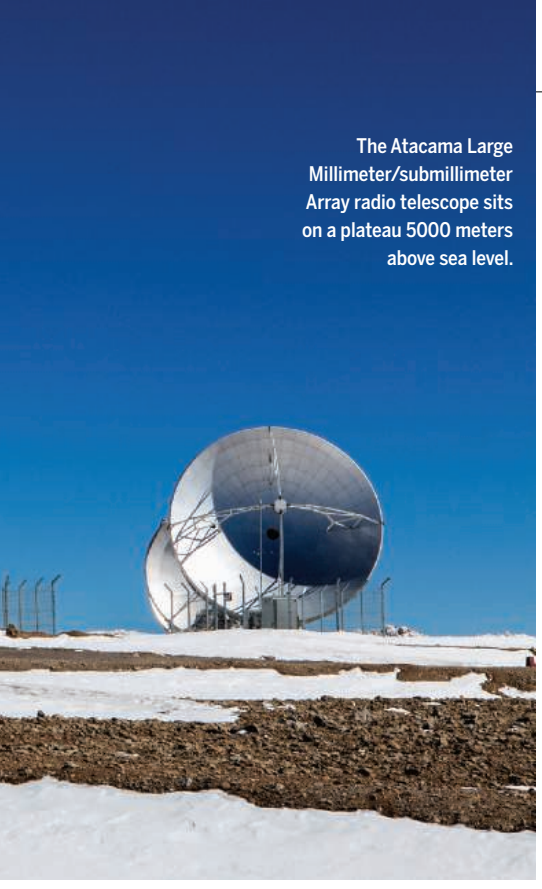
Protecting LGBTQ+ researchers

WORKPLACE | Research mentors and institutions should take precautions to safeguard LGBTQ+ ecologists from discrimination, sexual harassment, and violence when they do fieldwork in isolated areas, a working group says in a recent paper. The analysis, published last month in the *Journal of Applied Ecology*, describes LGBTQ+ researchers' heightened risks and offers recommendations to minimize them. One is to develop field safety plans that, for example, outline inclusive housing and bathroom arrangements. Another is providing scientists with radios to help them stay in contact with one another in case a dangerous situation arises. In addition, research leaders should be aware whether researchers on a fieldwork team are out of the closet so that information can be protected as needed, especially in places where disclosure may be risky. A *Science* interview with the study's authors is available at https://scim.ag/LGBTQfieldwork.

Darwin's letters go online

HISTORY OF SCIENCE | For 50 years, the Darwin Correspondence Project gathered, curated, and digitized more than 15,000 letters between the father of evolutionary biology and about 2000 correspondents. The project officially ended last month,

PHOTO: RONALD PATRICK/GETTY IMAGES



The Atacama Large Millimeter/submillimeter Array radio telescope sits on a plateau 5000 meters above sea level.

PHOTO: SHILONG YANG

and last week the 30th and final volume of *The Correspondence of Charles Darwin* rolled off the presses. The fruits of the project's labor will live online, freely accessible through the Cambridge Digital Library. In addition to providing a peek into Darwin's personal life, the letters illuminate the vast network of people who shaped, expanded, and critiqued the 19th century zoologist's ideas. Gathering all the letters from dozens of repositories—and deciphering their handwriting—was a huge task, says James Secord, a science historian emeritus at the University of Cambridge who directed the effort since 2006. He says the mixture of serious scientific discussions and personal details, some quirky, made him and his team "feel like you've been living with someone most of your working life." An exhibit showcasing the project begins on 4 May at the New York Public Library.

U.S. fraud unit gets new chief

LEADERSHIP | The U.S. Office of Research Integrity (ORI), which handles scientific fraud involving federally funded biomedical research, will soon have its first permanent leader in nearly 2 years. Sheila Garrity, an attorney with master's degrees in public health and business, will take the helm in late March. Garrity has 18 years of experience leading research integrity efforts at Johns Hopkins University and George Washington University. ORI oversees investigations of fabrication, falsification, and plagiarism and recommends sanctions, such as temporary funding bans for offenders. The office has often gone long stretches without a permanent director; its last one, Elisabeth Handley, served less than 2 years before assuming another federal position in June 2021. ORI recently began to review the Department of Health and Human Services's 18-year-old scientific misconduct regulations and ORI's review process, which some observers think is slow and inefficient.

Academies seek antiracist action

DIVERSITY | To advance diversity and inclusion within the scientific community, leaders must dismantle the power structures that lead to racial inequities, a U.S. national science academies panel said this week. Its report points to "institutional cultures that, intentionally or otherwise, create exclusionary and discriminatory environments" and issues 12 recommendations. One urges decision-makers to look for patterns of bias in data on graduate admissions, hiring, promotion, and awards. Other recommendations include conducting regular culture audits and drawing lessons from minority-serving institutions about "providing intentional and culturally responsive ... support." The report, *Advancing Antiracism, Diversity, Equity, and Inclusion in STEMM Organizations*, was issued by the National Academies of Sciences, Engineering, and Medicine.

Centipede uses heat to 'see'

PHYSIOLOGY | Chinese red-headed centipedes, which are sightless, use an unusual thermal sensor to detect sunlight, a study has found. The adaptation allows the venomous arthropod, *Scolopendra subspinipes mutilans,* to shelter in dark spots beneath forest leaves, lurking for prey and avoiding predators. The species is among the first arthropods identified as having this ability, although researchers have separately discovered it in some sightless species of fish and shrimp. A research team confirmed the trait in the centipedes, which lack light-sensing proteins, by covering their antennae in foil. This reduced their ability to move from bright to dark experimental chambers. The researchers also found the thermal receptor protein BRTNaC1 in their antennae, which the team concluded helps the animal detect the warmth of sunlight, they report this week in the *Proceedings of the National Academy of Sciences.*

The sensitive antennae of the Chinese red-headed centipede contain a heat sensor that reacts to light.

THEY SAID IT

"Look, there's America."

Chinese aeronautical scientist Wu Zhe, in a state-media video about a high-altitude airship he said his team launched in 2019. In the video, described in *The New York Times*, he points to a red line on a computer screen tracing the airship's path near the U.S. southern border. Wu has helped lead China's efforts to create balloons for civilian and military monitoring. The 2019 flight was not widely disclosed until the United States shot down a Chinese balloon off the South Carolina coast on 4 February.

Journal rejects retraction

MISCONDUCT | The *Proceedings of the Royal Society B: Biological Sciences* will not retract a 2016 paper on anemone fish behavior even though a lengthy university investigation found its data were fabricated. The paper, authored by marine ecologists Danielle Dixson of the University of Delaware (UD) and Anna Scott of Southern Cross University in Australia, is one of three studies that UD last year asked journal editors to retract after an independent investigation. *Science* retracted a paper in August 2022 as a result. But *Proceedings B* said in a 1 February editor's note that its own investigation did not turn up enough evidence of fraud, in part because a correction by the authors solved a major discrepancy in the study's timeline. Fish physiologist Timothy Clark of Deakin University, one of the whistleblowers in the case, calls the journal's decision "infuriating." Dixson and Scott did not respond to requests for comment. UD's website still lists Dixson as a faculty member; a UD spokesperson did not answer *Science's* questions about the case.

SOCIAL SCIENCES

Twitter's threat to curtail free data access angers scientists

Social media giant wants to monetize its global data set

By **Kai Kupferschmidt**

When Twitter announced on 2 February that the social media platform would end free access to its application programming interface (API) 7 days later, a clock began ticking for Jean-Philippe Cointet. Like other researchers interested in topics such as political polarization or how misinformation spreads, the social scientist at Sciences Po in Paris uses the API to freely gather data on the hundreds of millions of tweets sent daily. If Twitter tries to charge a significant price for such information, some of his projects will not be possible. "When we got the news, we started to rush on a few data collection projects," he says.

Twitter has since extended its deadline twice and had not revealed its new policy as *Science* went to press, but Cointet and other researchers remain anxious. "A lot of people are worried about what this means for their research on political campaigning, disinformation and other topics," says Michael Zimmer, a social media researcher from Marquette University. Meanwhile hundreds of scientists and organizations have signed an open letter calling on Twitter to "ensure that APIs for studying public content on the platform remain easily accessible for journalists, academics, and civil society."

Twitter currently gives researchers much more access than other online platforms such as Facebook or YouTube, notes Casey Fiesler, an information researcher at the University of Colorado, Boulder. It has become the equivalent of a model organism in the world of those who study social media. "Twitter has been a hugely important source of data for researchers on basically any topic you can think of—from public health to political science to digital humanities to misinformation," Fiesler says.

Researchers came to rely too heavily on the platform, according to Deen Freelon, a political communications researcher at the University of North Carolina, Chapel Hill. "I think this [API policy change] is really going to force people to look at other platforms that have been understudied," he says, pointing to Reddit as an example. "Reddit has more users than Twitter, the data are actually equally easy to access if not easier, but people haven't been paying as much attention to it."

Many researchers don't believe Twitter's new owner, billionaire Elon Musk, is targeting them. "Based on conversations I've had with a number of folks who have access to information inside the company, I think that Musk wasn't considering academics whatsoever," says Rebekah Tromble, a computational social scientist at George Washington University. Companies and other nonacademic accounts use data they can freely access through Twitter's API for their own products. Musk is seeking new revenue sources for the company and likely wants to make money off those users. "It's an ill-considered decision to try to monetize the free API," Tromble says.

Musk is mercurial, having already floated several new Twitter policies and then backtracked or revised them after outcries. Initially, the company simply announced it would no longer support free data access. "A paid basic tier will be available instead," it said. In its first deadline extension tweet, Twitter then said it would offer a "low level of API usage" for $100 a month, but the details and what kind of access researchers would have remain unclear.

Stephan Lewandowsky, a psychologist at Bristol University who mines Twitter data to study what type of misinformation people share, says a lot will depend on how much money Twitter asks for. "If researchers can buy access for a flat fee, e.g. $100/month, then many (but not all!) researchers may be able to absorb that," Lewandowsky says. "If it's based on volume [of data mined] and if it ends up being loads more than [$100], then it will disable some important research because many people couldn't afford charges of multiple thousands of dollars."

By charging for its data, Twitter will advantage established researchers and

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

wealthy institutions, Freelon says. "Who's going to lose out are going to be grad students, people without institutional affiliations, people whose institutions are lower wealth." Even if researchers ultimately retain access to Twitter's data, the uncertainty caused by the API announcement highlights larger issues, they say. Because social media influences society as a whole, social media companies should not be able to control research on the impact of their products, argues Philipp Lorenz-Spreen of the Max Planck Institute for Human Development, who studies social media networks. "We do not depend on the oil industry to be able to measure CO_2, but we are dependent on Facebook to measure polarization on Facebook," he says. "That is a bad situation."

In Europe, the Digital Services Act, whose rules will apply from early 2024 on, seeks to address this issue. One provision allows national authorities to compel access to social media companies' data for researchers studying "systemic risks." But there is still a lot of uncertainty about how this will be handled, how onerous applications for such research will be, and how fast companies will have to respond to the mandate.

The irony is that as Twitter curbs data access, other companies are actually expanding theirs, says Tromble, partly spurred by EU regulations. "YouTube has expanded its API, TikTok is developing new APIs," she says. "Twitter was the industry leader and now they are actually falling behind."

For now, scientists such as Cointet are rushing to finish projects as they await details from Twitter. In an ongoing project, he and colleagues aim to understand how political debates are structured on Twitter, in part by analyzing how likely people following one politician are to also follow others.

In France, they found a surprise. Instead of a simple left-right spectrum, the pattern that most accurately predicted whom people followed was an axis from "global to local," Cointet says: "Basically, people who are against Europe versus for Europe, people who are denouncing elites versus people who are not, people who distrust institutions versus people who don't." The Sciences Po team has been striving to reproduce this work for 10 other countries but may not be able to gather all the data before their access is cut off, Cointet says. "It's too bad we need to do that in such a rush."

In the end, the debate highlights once more how much control Musk can exert over public goods, Lewandowsky says. "At an abstract overarching level, this reveals the fundamental problem of whether democratic societies can allow eccentric billionaires to control public spaces." ■

ENERGY

Laser fusion success sparks hope of new route to fusion power

Startups lay plans for power plants that would trigger tiny, rapid-fire fusion blasts

By **Daniel Clery**

Last year, when the National Ignition Facility (NIF) fired its 192 laser beams at a gold cylinder enclosing a tiny sphere of hydrogen isotopes, it did more than spark a historic fusion reaction. The shot—the first to produce more energy than the lasers delivered—also triggered a burst of optimism among some fusion scientists that the same general approach could one day lead to a commercial power plant.

If so, the shape of fusion power could be very different from the giant magnetic furnaces known as tokamaks that are the focus of most current hopes, including the international megaproject called ITER. Fusion power based on the NIF approach, known as inertial confinement fusion (ICF), would instead resemble the rapid-fire explosions of an internal combustion engine. The technical challenges remain huge. NIF's shot was a one-off, but an ICF-based power plant may need to fire up to 10 laser shots per second, consuming almost 1 million targets a day, and produce far more energy per shot than NIF did.

> **"The technology is all available, it just has to be integrated."**
> **Ed Moses**, Longview Fusion

The optimists at a handful of ICF startups are undaunted. Ed Moses, a former NIF director, has founded a company, Longview Fusion, that hopes to start building a test plant in 5 years with twice as many laser beams as NIF. It would shoot 10 targets per second—sheathed in lead instead of gold to lower costs—into a reaction chamber and blast them to produce a string of quick-fire fusion explosions. "It's not hard," Moses says. "The technology is all available, it just has to be integrated."

The companies pursuing ICF are small, with typically a few tens of millions of dollars each in backing. But they hope NIF's breakthrough will bring more investors to their doors. "There's general excitement from everyone you talk to, it's just fantastic," says Debra Callahan, who worked at NIF for 20 years before joining the startup Focused Energy, based in the United States and Germany.

Others are far more cautious. "We don't know how to build a power plant," says Tammy Ma, who heads the inertial fusion energy effort at Lawrence Livermore National Laboratory, the home of NIF. Late last month, Livermore published a report outlining a long program of research that the Department of Energy (DOE) should do to develop power plants based on ICF.

Although the December 2022 shot at NIF produced a record-breaking 3.15 megajoules of energy from a 2.05 megajoule laser pulse, a gain of about 1.5, generating that pulse consumed hundreds of megajoules of electricity, and NIF can only do one shot per day (*Science*, 16 December 2022, p. 1154). Most fusion experts still give better odds to tokamaks, doughnut-shaped devices that use powerful magnetic fields to trap an ionized gas of the hydrogen isotopes deuterium and tritium and heat it to 100 million degrees Celsius so the nuclei crash together with enough kinetic energy to fuse. But one big advantage ICF facilities have over tokamaks is that their different elements—lasers, reaction chambers, targets—can be developed and tested separately before they need to be combined. Building a tokamak is all or nothing and extremely difficult to change once completed.

For DOE's researchers, more ICF studies with NIF and Omega, another laser fusion facility at the University of Rochester, are first on the agenda. "We need to understand the fundamental physics," Ma says, and, most important, increase gain. Meanwhile, DOE scientists will assess which lasers or alternatives, such as ion beams, are worth pursuing and hone target designs as well.

Ultimately, Ma says, a new facility will be needed—perhaps lower power than NIF but with a higher repetition rate—so the field can "learn faster." NIF's lasers rely on neodymium-doped phosphate glass, which is pumped up with energy by slow and energy-hungry xenon flash lamps. Since NIF was designed in the 1990s, other kinds of lasers have been developed, such as ones using krypton-fluoride or argon-fluoride

wealthy institutions, Freelon says. "Who's going to lose out are going to be grad students, people without institutional affiliations, people whose institutions are lower wealth." Even if researchers ultimately retain access to Twitter's data, the uncertainty caused by the API announcement highlights larger issues, they say. Because social media influences society as a whole, social media companies should not be able to control research on the impact of their products, argues Philipp Lorenz-Spreen of the Max Planck Institute for Human Development, who studies social media networks. "We do not depend on the oil industry to be able to measure CO_2, but we are dependent on Facebook to measure polarization on Facebook," he says. "That is a bad situation."

In Europe, the Digital Services Act, whose rules will apply from early 2024 on, seeks to address this issue. One provision allows national authorities to compel access to social media companies' data for researchers studying "systemic risks." But there is still a lot of uncertainty about how this will be handled, how onerous applications for such research will be, and how fast companies will have to respond to the mandate.

The irony is that as Twitter curbs data access, other companies are actually expanding theirs, says Tromble, partly spurred by EU regulations. "YouTube has expanded its API, TikTok is developing new APIs," she says. "Twitter was the industry leader and now they are actually falling behind."

For now, scientists such as Cointet are rushing to finish projects as they await details from Twitter. In an ongoing project, he and colleagues aim to understand how political debates are structured on Twitter, in part by analyzing how likely people following one politician are to also follow others.

In France, they found a surprise. Instead of a simple left-right spectrum, the pattern that most accurately predicted whom people followed was an axis from "global to local," Cointet says: "Basically, people who are against Europe versus for Europe, people who are denouncing elites versus people who are not, people who distrust institutions versus people who don't." The Sciences Po team has been striving to reproduce this work for 10 other countries but may not be able to gather all the data before their access is cut off, Cointet says. "It's too bad we need to do that in such a rush."

In the end, the debate highlights once more how much control Musk can exert over public goods, Lewandowsky says. "At an abstract overarching level, this reveals the fundamental problem of whether democratic societies can allow eccentric billionaires to control public spaces." ■

ENERGY

Laser fusion success sparks hope of new route to fusion power

Startups lay plans for power plants that would trigger tiny, rapid-fire fusion blasts

By **Daniel Clery**

Last year, when the National Ignition Facility (NIF) fired its 192 laser beams at a gold cylinder enclosing a tiny sphere of hydrogen isotopes, it did more than spark a historic fusion reaction. The shot—the first to produce more energy than the lasers delivered—also triggered a burst of optimism among some fusion scientists that the same general approach could one day lead to a commercial power plant.

If so, the shape of fusion power could be very different from the giant magnetic furnaces known as tokamaks that are the focus of most current hopes, including the international megaproject called ITER. Fusion power based on the NIF approach, known as inertial confinement fusion (ICF), would instead resemble the rapid-fire explosions of an internal combustion engine. The technical challenges remain huge. NIF's shot was a one-off, but an ICF-based power plant may need to fire up to 10 laser shots per second, consuming almost 1 million targets a day, and produce far more energy per shot than NIF did.

> **"The technology is all available, it just has to be integrated."**
> **Ed Moses**, Longview Fusion

The optimists at a handful of ICF startups are undaunted. Ed Moses, a former NIF director, has founded a company, Longview Fusion, that hopes to start building a test plant in 5 years with twice as many laser beams as NIF. It would shoot 10 targets per second—sheathed in lead instead of gold to lower costs—into a reaction chamber and blast them to produce a string of quick-fire fusion explosions. "It's not hard," Moses says. "The technology is all available, it just has to be integrated."

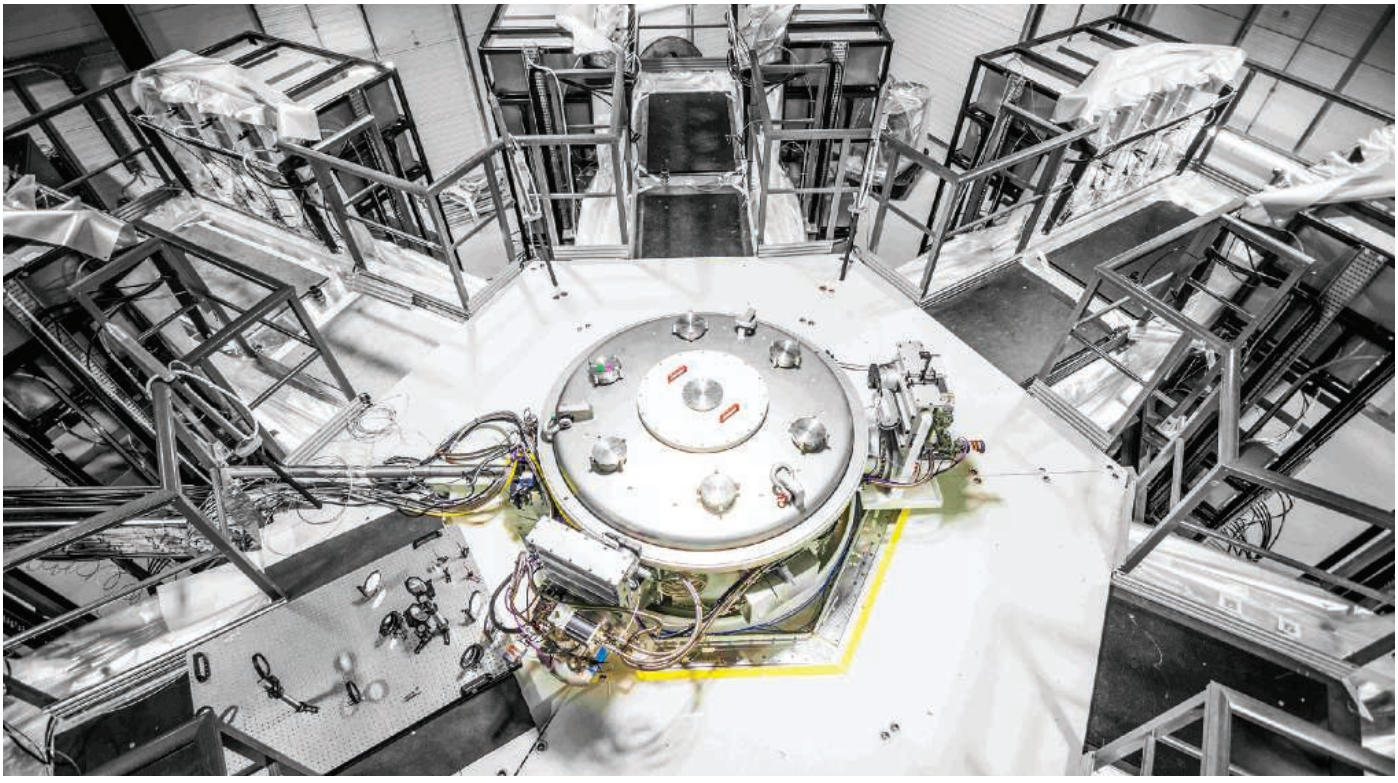
The companies pursuing ICF are small, with typically a few tens of millions of dollars each in backing. But they hope NIF's breakthrough will bring more investors to their doors. "There's general excitement from everyone you talk to, it's just fantastic," says Debra Callahan, who worked at NIF for 20 years before joining the startup Focused Energy, based in the United States and Germany.

Others are far more cautious. "We don't know how to build a power plant," says Tammy Ma, who heads the inertial fusion energy effort at Lawrence Livermore National Laboratory, the home of NIF. Late last month, Livermore published a report outlining a long program of research that the Department of Energy (DOE) should do to develop power plants based on ICF.

Although the December 2022 shot at NIF produced a record-breaking 3.15 megajoules of energy from a 2.05 megajoule laser pulse, a gain of about 1.5, generating that pulse consumed hundreds of megajoules of electricity, and NIF can only do one shot per day (*Science*, 16 December 2022, p. 1154). Most fusion experts still give better odds to tokamaks, doughnut-shaped devices that use powerful magnetic fields to trap an ionized gas of the hydrogen isotopes deuterium and tritium and heat it to 100 million degrees Celsius so the nuclei crash together with enough kinetic energy to fuse. But one big advantage ICF facilities have over tokamaks is that their different elements—lasers, reaction chambers, targets—can be developed and tested separately before they need to be combined. Building a tokamak is all or nothing and extremely difficult to change once completed.

For DOE's researchers, more ICF studies with NIF and Omega, another laser fusion facility at the University of Rochester, are first on the agenda. "We need to understand the fundamental physics," Ma says, and, most important, increase gain. Meanwhile, DOE scientists will assess which lasers or alternatives, such as ion beams, are worth pursuing and hone target designs as well.

Ultimately, Ma says, a new facility will be needed—perhaps lower power than NIF but with a higher repetition rate—so the field can "learn faster." NIF's lasers rely on neodymium-doped phosphate glass, which is pumped up with energy by slow and energy-hungry xenon flash lamps. Since NIF was designed in the 1990s, other kinds of lasers have been developed, such as ones using krypton-fluoride or argon-fluoride



First Light Energy's reaction chamber is surrounded by power storage banks ready to supply a 14-million-amp current to hurl a projectile toward a fuel target.

gases, that are more efficient and can fire more frequently, but none are yet ready for prime time.

Scientists will also have to figure out the best design for the fuel targets. Livermore has always favored enclosing the peppercorn-size hydrogen-filled capsules inside metal cylinders, which convert laser light into x-rays that then implode the fuel. In contrast, the Rochester laser lab has pioneered "direct drive": firing the laser straight onto the capsule. This approach is more efficient, but it requires more perfect laser beams to get the fuel capsules to implode symmetrically.

Longview is playing it safe by sticking close to NIF's indirect-drive approach—"the only proven fusion process on Earth," Moses says. Like NIF, the company's scheme would use glass lasers, but a more efficient design that can fire at a higher rate. The system will be "hyper-modular," Moses says, with each laser beam generated by a 10-meter-long box that can be replaced for upgrading or repair. With twice as many beams as NIF, its laser pulse would have 50% more energy, allowing for higher gain.

Focused Energy is pursuing another variant of ICF, known as fast ignition, which uses two separate lasers to perform the functions of NIF's single laser pulse: compressing the fuel and igniting the fusion burn. NIF's fuel pellet must implode symmetrically to create a central spot hot enough to spontaneously ignite. Removing the need for a hot spot means the imploding drive can be gentler and requires "less stringent symmetry," says Focused Energy's Pravesh Patel, another NIF veteran. Fast ignition instead uses a second laser to kindle the burn once the fuel reaches maximum density. The second laser pulse, blasting in from one side, will hit a curved metal foil, generating a beam of protons that provides the spark.

Patel says the team wants to build a demonstrator in 8 to 9 years with a 100-beam compression laser producing 500 kilojoule pulses, one-quarter the energy of NIF's. For the short ignition pulses, they plan to use lasers developed for Europe's Extreme Light Infrastructure, a research facility. Each produces pulses of 1.5 kilojoules so Focused Energy will need 100 of them, combining their output into a single beam to set the fuel alight.

First Light Fusion is taking an even more radical approach: dispensing with lasers entirely and relying on a high-speed projectile to implode the fuel. Founder Nick Hawker was inspired by his Ph.D. on the physics of the pistol shrimp, which snaps its oversized claw to emit shock waves capable of incapacitating fish bigger than itself. The company, based in the United Kingdom near Oxford, is firing projectiles from an electromagnetic gun at up to 20 kilometers per second to smash a small metal cube with a fuel capsule embedded inside. The cube has a complex internal structure of different metals that speeds and channels the resulting shock wave, wrapping it around the central capsule to compress the fuel.

"It's a geometry optimization problem," Hawker says, noting the targets don't require any unusual or expensive materials. With modeling and experiment, the team is working to get the best spherical compression from its unidirectional driver. The company announced last month it will start building a new machine next year. The goal is to demonstrate energy gain by hammering targets with projectiles traveling at 60 kilometers per second.

Steven Cowley, director of Princeton Plasma Physics Laboratory, DOE's premier fusion lab, calls NIF's record-breaking shot a "spectacular result," but is skeptical that ICF can quickly be reinvented as an energy source. To produce true excess energy, NIF would need to achieve a gain of about 100, which would produce an "unbelievably destructive bang." An ICF power facility would have to withstand that, 10 times a second, and clear away the debris between each fusion explosion. "There's an awful lot of work to be done," Cowley says. ■

PHOTO: FIRST LIGHT FUSION LTD.

PUBLIC HEALTH

Japan moves to bolster its vaccine R&D sector

Major funding effort aims to address weaknesses exposed by COVID-19 pandemic

By **Dennis Normile**

Three years after the COVID-19 pandemic began, many countries—including the United States, Germany, India, Cuba—have deployed their own vaccines. But Japan's are still not ready, a lag that is ringing alarms about the feeble state of the country's vaccine research and development capabilities. Now Japan is ramping up a 1.1 trillion yen ($8.5 billion) initiative to enable it to react faster to the next crisis, by having the capability to develop a vaccine for a new virus in 100 days.

That "very ambitious" push "is definitely a welcome development," says Diane Griffin, a virologist at Johns Hopkins Bloomberg School of Public Health. The funding, first outlined by the government in 2021, is now starting to flow to research, clinical trials, and an expansion of industry's capacity to manufacture vaccines, especially for neglected diseases.

But some say the initiative doesn't address deeper problems, including a dearth of stable research jobs in all areas but particularly in infectious diseases and vaccines, which has deterred young scientists from entering key research fields. And it's not clear whether the government is committed to sustaining funding after the first tranche expires in March 2027.

The decline of Japan's vaccine sector was decades in the making. "For the last 15 or 20 years, funding for infectious disease research has been getting lower and lower," says Yoshihiro Kawaoka, a virologist at the University of Tokyo and University of Wisconsin, Madison. Before the pandemic, the amount Japan's government spent in the field amounted to less than 2% of that spent by the United States, and also trailed the United Kingdom, Germany, and China, according to a 2021 report by Deloitte Tohmatsu Consulting. The country ended up without a critical mass of infectious disease experts, and too few young scientists in the field, Kawaoka says.

Vaccine research suffered. In the late 2010s, Ken Ishii, a vaccinologist at Japan's National Institute of Biomedical Innovation in Osaka, adopted the then-emerging messenger RNA (mRNA) technology to create a vaccine for Middle East respiratory syndrome (MERS), the deadly, camel-borne coronavirus disease. But the government and industry declined to fund human safety trials. "So the project was frozen," says Ishii, now at the University of Tokyo. The decision may have made commercial sense given that MERS affected few people, Ishii says. But it also meant no one in Japan was working on the mRNA vaccine technology that proved crucial in fighting SARS-CoV-2.

A pachinko parlor patron in Japan receives the Moderna coronavirus vaccine in 2021. The country could soon approve homegrown COVID-19 vaccines.

PHOTO: CARL COURT/GETTY IMAGES

The COVID-19 pandemic only made the weaknesses more obvious. Just two scientists in Japan, for example, appeared on a 2021 list of the 300 most cited authors of COVID-19 research. By contrast, Italy and Hong Kong, with much smaller research establishments, had 18 and 14 authors, respectively, on that tally, which appeared in a study of COVID-19 papers by a team led by Stanford University statistician John Ioannidis that was published in *Royal Society Open Science*.

The pandemic also highlighted weaknesses in Japan's vaccine industry. Early on, amid fears that the United States and other nations would hoard vaccines for their own populations, Japanese officials worried "that we couldn't save people we are responsible for," says Michinari Hamaguchi, director general of the new Strategic Center of Biomedical Advanced Vaccine Research and Development for Preparedness and Response (SCARDA). Japan ultimately managed to buy the vaccines it needed. But only now are two Japanese companies seeking approval to sell the country's first homegrown COVID-19 vaccines. Hamaguchi says the scare prompted officials to push for the "reconstitution" of domestic vaccine capabilities.

After the government approved the initiative in June 2021, agencies began rolling out detailed spending plans. SCARDA was set up in March 2022 and funds started flowing to researchers last fall. The agency, which is still reviewing grant applications, will provide $1.1 billion to investigators for work on vaccines for coronaviruses, influenza, Zika, dengue, Nipah, and smallpox, among others. "It is the first time Japan has awarded grants to cover development of vaccines" for neglected diseases, says Chieko Kai, a University of Tokyo virologist who recently won a 2-year, $15 million SCARDA grant to develop a Nipah vaccine.

An additional $400 million is going to new vaccine R&D centers at several universities. The flagship center, the University of Tokyo Pandemic Preparedness, Infection, and Advanced Research Center, aims to be "competitive at the top level of infectious disease research," says Kawaoka, who heads the effort. Four centers at other universities will work on a range of projects, including coronavirus vaccines that can be sprayed into the throat or nostrils—an approach many researchers believe could provide better protection than today's shots.

Other agencies will provide $2.7 billion to vaccine-related startup firms, $1.7 billion for vaccine manufacturing, and $2 billion to support large clinical trials and purchase COVID-19 vaccines.

Long-term funding for the initiative is uncertain, however, which could make it hard for labs to hire and retain talent. "It will take time and money to rebuild research teams, and they have to be maintained," says Hiroaki Mitsuya, director general of the National Center for Global Health and Medicine Research Institute in Tokyo. Mitsuya worries that in 5 years the government could have different priorities. Restoring Japan's vaccine infrastructure, Mitsuya says, must start "with more funding and positions for entry-level researchers." ■

Instruments aboard a modified bomber will count and analyze natural particles at altitudes of 20 kilometers.

ATMOSPHERIC SCIENCE

United States tiptoes into solar geoengineering research

Campaign seeks to understand reflective particles in the stratosphere, which planet-cooling schemes would enhance

By **Paul Voosen**

Any work on solar geoengineering—the notion of artificially making the atmosphere more reflective to cool an overheated planet—is fraught with controversy. Last year, for example, a tech entrepreneur claimed he launched two weather balloons from Baja California into the stratosphere, where they may have released a puff of sulfur dioxide that gave rise to a small patch of reflective sulfate particles. The stunt drew widespread condemnation. But for researchers, it prompted a question: If a rogue actor had conducted a larger release, would they be able to detect it—or know with any certainty what it would do?

The U.S. National Oceanic and Atmospheric Administration (NOAA) is venturing into the fray to answer these questions, with a bid to understand the types, amounts, and behavior of particles naturally present in the stratosphere. Unlike the Mexico caper, the balloons and high-altitude aircraft in the program aren't releasing any particles or gases. But the large-scale field campaign is the first the U.S. government has ever conducted related to solar geoengineering. It's very basic research, says Karen Rosenlof, an atmospheric scientist at NOAA's Chemical Sciences Laboratory. "You have to know what's there first before you can start messing with that."

Research on solar geoengineering—also called solar radiation management—has long been anathema to some climate scientists and activists. They fear it could distract from emissions cuts, could have unforeseen risks, and would not address some impacts of rising carbon dioxide, including ocean acidification. Federal agencies have largely steered clear of the work, even after a report from the National Academies of Sciences, Engineering, and Medicine (NASEM) in 2021 recommended a $200 million research program (*Science*, 2 April 2021, p. 19).

But in an unusual move, Congress directly ordered NOAA to develop a program in a 2020 spending bill. Support for the program, innocuously dubbed Earth's Radiation Budget, has grown to nearly $10 million annually. SilverLining, an organization supportive of solar geoengineering research, lobbied for it, and gained the support of key legislators, says Kelly Wanser, SilverLining's executive director, who came to climate advocacy from a career in technology. She says she told the lawmakers that emissions cuts weren't happening fast enough, and that solar geoengineering might—or might not—be needed to reduce impacts in the next 40 years. "What do we need to know to evaluate those things in a very rigorous and honest-with-ourselves way?" Wanser asks. "I want to know."

One focus of the NOAA program is sulfates, the reflective particles that would be lofted into the stratosphere in many solar geoengineering proposals. Volcanoes and industrial emissions are the main natural sources of sulfur dioxide, one precursor of sulfates; carbonyl sulfide, a gas emitted by microbes in the oceans, is another. The updrafts in severe storms in the tropics are thought to pump the gases into the stratosphere, but massive forest fires are emerging as another important mechanism for lifting them (*Science*, 19 November 2021, p. 921).

Because of a temperature inversion at the stratosphere's base, sulfate particles linger there for years, which encourages geoengineering advocates. But researchers need a clearer picture of natural stratospheric particles before they can contemplate supplementing them artificially, says Gregory Frost, the NOAA atmospheric chemist who oversees the program.

In 2020, NOAA began regularly launching sensor-equipped weather balloons to gather baseline data on the size and concentration of the particles. The record should allow NOAA to detect shifts in particle distributions, which could indicate a new source, such as an eruption—or a clandestine geoengineering intervention, Frost says.

The measurements could also help climate modelers represent stratospheric particles more realistically. "The only way to know if our models are doing the right thing is to have data to evaluate them," says Simone Tilmes, a modeler at the National Center for Atmospheric Research (NCAR). At the moment, most models handle particles in a coarse way, categorizing them into a few arbitrary size ranges. With NOAA funding, Tilmes's team has now upgraded NCAR's flagship climate model with 40 possible bins for sulfate size, and so far it seems to more reliably capture events like the sulfur-rich eruption of Mount Pinatubo in 1991.

Now, the project is launching its next phase, the largest stratospheric aircraft campaign in the past 2 decades, "if not ever," Frost says. One of NASA's WB-57 research jets—a heavily modified bomber from the 1960s—has been outfitted with 17 instruments, many of which have never flown to the stratosphere before. One can measure sulfur dioxide levels down

PHOTO: ROBERT MARKOWITZ/NASA-JSC

Corrected 16 February 2023. See full text.

to 2 parts per trillion, and another can distinguish different particles by their chemical makeup.

Next week, flights will begin out of Houston, Texas, to altitudes of up to 20 kilometers, and later this month, the team plans to move to Fairbanks, Alaska, where they will fly until late March. At that time of the year, air that entered the stratosphere a half-decade earlier in the tropics descends in the Arctic. "We're going to get into really old stratospheric air," says Troy Thornberry, the NOAA atmospheric chemist who leads the flights. By studying aging sulfate particles, the team hopes to witness the chemical reactions that break them apart and release sulfur at the end of their lifetimes. They want to study how such sulfur interacts with organic particles such as soot and the dust of meteorites. Rosenlof says they will also study how soot absorbs the Sun's heat, causing air parcels to rise and prolonging particle lifetimes in the stratosphere.

Additional flights are planned for Costa Rica in 2024 and the Southern Hemisphere in 2025. But in the meantime, the team is open to missions of opportunity. If a massive volcanic eruption occurs, they've assembled the perfect payload to explore its impact, Frost says. "That would be an event we'd want to study if at all possible." The same goes for any wildfires similar in scale to those in Australia in 2020.

Some researchers want the U.S. government to put together a bigger solar geoengineering research program. Although the NOAA program is well designed, it's a small piece of the puzzle, says Chris Field, a climate scientist at Stanford University and chair of the NASEM report. "It's really important to find a way to ask bold questions about where something unexpected could go wrong." In March 2022, Congress ordered the White House's Office of Science and Technology Policy to develop such a plan, but it still has not been released. "Federally funded research is a very good idea—and the only way this work ideally should be funded," says Sikina Jinnah, who studies geoengineering governance at the University of California, Santa Cruz.

Although some climate scientists disagree that such research should go forward or have even called for a ban on it, shutting it down would be a mistake, says James Hurrell, a climate scientist at Colorado State University who previously led NCAR. "If we could help avoid the worst impacts [of climate change] and buy more time for the world to reduce concentrations, don't we want to know that?" And if it's a bad idea, he adds, "the best way to prove that is through research." ■

INFRASTRUCTURE

U.S. measurement institute's labs said to need major upgrade

National academies report finds that disrepair at NIST facilities threatens U.S. economic growth

By **Robert F. Service**

The U.S. government needs to spend $6.6 billion over the next 12 years to repair and upgrade research facilities at the National Institute of Standards and Technology (NIST), according to a blue-ribbon panel.

Leaks, floods, power outages, poor humidity control, and other problems are hobbling the productivity of NIST's laboratories in Gaithersburg, Maryland, and Boulder, Colorado, according to a report released last week by a committee assembled by the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM). "NIST facilities are not world class and therefore a growing impediment to attracting and retaining staff," says Ross Corotis, an emeritus engineering professor at the University of Colorado, Boulder, who chaired the committee. "We saw some labs that were nearly uninhabitable," notes committee member Kent Rochford, executive director of SPIE, the international society for optics and photonics, and former head of NIST's intramural program.

> **"We saw some labs that were nearly uninhabitable."**
>
> **Kent Rochford**, SPIE, the international society for optics and photonics

NIST, part of the Department of Commerce, is charged with promoting U.S. innovation and industrial competitiveness through research on measurement standards. Such standards are essential for everything from computer chip manufacturing and internet communication to electricity distribution and ensuring the reliability of COVID-19 tests.

The new report cites several examples of the toll that deteriorating infrastructure has taken. A lack of humidity and dust controls in one lab led to repeated delays in providing radiation sensors to two national laboratories for identifying nuclear materials. A power interruption at a Boulder lab destroyed a $6 million electron microscope used for research on 3D printing. And recent flooding in a basement lab in Gaithersburg permanently damaged a Kibble balance used to help redefine the standard kilogram in 2019. The committee found 63% of the institute's research facilities fail to meet established standards for acceptable building conditions, and that the shortcomings reduce researchers' productivity by an estimated 10% to 40%.

The NASEM panel notes that over the past 20 years, annual reports from outside experts have "consistently and unequivocally reported that the root cause of NIST's progressive facilities decline is grossly inadequate funding." The reports have repeatedly called on Commerce officials to budget for upgrades and Congress to pay for them.

The report endorsed an agency blueprint to spend roughly $550 million annually for the next 12 years to modernize a dozen existing facilities. "We'd love to do it more quickly, but this is the time frame we feel NIST can get it done," Corotis says.

In the past, each dollar invested in NIST facilities has returned $9 to the U.S. economy, according to independent studies. That ratio makes the proposed spending "one of the best investments the government can make in the nation's research," says committee member Eric Dillinger, a vice president of strategic consulting at Woolpert, an architecture and engineering firm.

Dillinger says the panel has scheduled briefings with White House science and budget officials working on the president's fiscal year 2024 budget request, which will be presented to Congress next month. But the new Republican majority in the U.S. House of Representatives has already demanded steep cuts in federal spending.

Prospects may be brighter over the long term, says science policy analyst Mitch Ambrose of the American Institute of Physics. "Congress at times has [supported] long-term, bipartisan campaigns to upgrade agencies' physical infrastructure," says Ambrose. He cites current efforts to modernize facilities at the National Institutes of Health and the nation's nuclear weapons facilities. ■

Corrected 16 February 2023. See full text.

sulfur dioxide levels down to 2 parts per trillion, and another can distinguish different particles by their chemical makeup.

Next week, flights will begin out of Houston, Texas, to altitudes of up to 20 kilometers, and later this month, the team plans to move to Fairbanks, Alaska, where they will fly until late March. At that time of the year, air that entered the stratosphere a half-decade earlier in the tropics descends in the Arctic. "We're going to get into really old stratospheric air," says Troy Thornberry, the NOAA atmospheric chemist who leads the flights. By studying aging sulfate particles, the team hopes to witness the chemical reactions that break them apart and release sulfur at the end of their lifetimes. They want to study how such sulfur interacts with organic particles such as soot and the dust of meteorites. Rosenlof says they will also study how soot absorbs the Sun's heat, causing air parcels to rise and prolonging particle lifetimes in the stratosphere.

Additional flights are planned for Costa Rica in 2024 and the Southern Hemisphere in 2025. But in the meantime, the team is open to missions of opportunity. If a massive volcanic eruption occurs, they've assembled the perfect payload to explore its impact, Frost says. "That would be an event we'd want to study if at all possible." The same goes for any wildfires similar in scale to those in Australia in 2020.

Some researchers want the U.S. government to put together a bigger solar geoengineering research program. Although the NOAA program is well designed, it's a small piece of the puzzle, says Chris Field, a climate scientist at Stanford University and chair of the NASEM report. "It's really important to find a way to ask bold questions about where something unexpected could go wrong." In March 2022, Congress ordered the White House's Office of Science and Technology Policy to develop such a plan, but it still has not been released. "Federally funded research is a very good idea—and the only way this work ideally should be funded," says Sikina Jinnah, who studies geoengineering governance at the University of California, Santa Cruz.

Although some climate scientists disagree that such research should go forward or have even called for a ban on it, shutting it down would be a mistake, says James Hurrell, a climate scientist at Colorado State University who previously led NCAR. "If we could help avoid the worst impacts [of climate change] and buy more time for the world to reduce concentrations, don't we want to know that?" And if it's a bad idea, he adds, "the best way to prove that is through research." ■

INFRASTRUCTURE

U.S. measurement institute's labs said to need major upgrade

National academies report finds that disrepair at NIST facilities threatens U.S. economic growth

By **Robert F. Service**

The U.S. government needs to spend $6.6 billion over the next 12 years to repair and upgrade research facilities at the National Institute of Standards and Technology (NIST), according to a blue-ribbon panel.

Leaks, floods, power outages, poor humidity control, and other problems are hobbling the productivity of NIST's laboratories in Gaithersburg, Maryland, and Boulder, Colorado, according to a report released last week by a committee assembled by the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM). "NIST facilities are not world class and therefore a growing impediment to attracting and retaining staff," says Ross Corotis, an emeritus engineering professor at the University of Colorado, Boulder, who chaired the committee. "We saw some labs that were nearly uninhabitable," notes committee member Kent Rochford, executive director of SPIE, the international society for optics and photonics, and former head of NIST's intramural program.

> **"We saw some labs that were nearly uninhabitable."**
>
> **Kent Rochford**, SPIE, the international society for optics and photonics

NIST, part of the Department of Commerce, is charged with promoting U.S. innovation and industrial competitiveness through research on measurement standards. Such standards are essential for everything from computer chip manufacturing and internet communication to electricity distribution and ensuring the reliability of COVID-19 tests.

The new report cites several examples of the toll that deteriorating infrastructure has taken. A lack of humidity and dust controls in one lab led to repeated delays in providing radiation sensors to two national laboratories for identifying nuclear materials. A power interruption at a Boulder lab destroyed a $6 million electron microscope used for research on 3D printing. And recent flooding in a basement lab in Gaithersburg permanently damaged a Kibble balance used to help redefine the standard kilogram in 2019. The committee found 63% of the institute's research facilities fail to meet established standards for acceptable building conditions, and that the shortcomings reduce researchers' productivity by an estimated 10% to 40%.

The NASEM panel notes that over the past 20 years, annual reports from outside experts have "consistently and unequivocally reported that the root cause of NIST's progressive facilities decline is grossly inadequate funding." The reports have repeatedly called on Commerce officials to budget for upgrades and Congress to pay for them.

The report endorsed an agency blueprint to spend roughly $550 million annually for the next 12 years to modernize a dozen existing facilities. "We'd love to do it more quickly, but this is the time frame we feel NIST can get it done," Corotis says.

In the past, each dollar invested in NIST facilities has returned $9 to the U.S. economy, according to independent studies. That ratio makes the proposed spending "one of the best investments the government can make in the nation's research," says committee member Eric Dillinger, a vice president of strategic consulting at Woolpert, an architecture and engineering firm.

Dillinger says the panel has scheduled briefings with White House science and budget officials working on the president's fiscal year 2024 budget request, which will be presented to Congress next month. But the new Republican majority in the U.S. House of Representatives has already demanded steep cuts in federal spending.

Prospects may be brighter over the long term, says science policy analyst Mitch Ambrose of the American Institute of Physics. "Congress at times has [supported] long-term, bipartisan campaigns to upgrade agencies' physical infrastructure," says Ambrose. He cites current efforts to modernize facilities at the National Institutes of Health and the nation's nuclear weapons facilities. ■

HIDDEN HYDROGEN

Does Earth hold vast stores of a renewable, carbon-free fuel?

By **Eric Hand**

In the shade of a mango tree, Mamadou Ngulo Konaré recounted the legendary event of his childhood. In 1987, well diggers had come to his village of Bourakébougou, Mali, to drill for water, but had given up on one dry borehole at a depth of 108 meters. "Meanwhile, wind was coming out of the hole," Konaré told Denis Brière, a petrophysicist and vice president at Chapman Petroleum Engineering, in 2012. When one driller peered into the hole while smoking a cigarette, the wind exploded in his face.

"He didn't die, but he was burned," Konaré continued. "And now we had a huge fire. The color of the fire in daytime was like blue sparkling water and did not have black smoke pollution. The color of the fire at night was like shining gold, and all over the fields we could see each other in the light. … We were very afraid that our village would be destroyed."

It took the crew weeks to snuff out the fire and cap the well. And there it sat, shunned by the villagers, until 2007. That was when Aliou Diallo, a wealthy Malian businessman, politician, and chair of Petroma, an oil and gas company, acquired the rights to prospect in the region surrounding Bourakébougou. "We have a saying that human beings are made of dirt, but the devil is made of fire," Diallo says. "It was a cursed place. I said, 'Well, cursed places, I like to turn them into places of blessing.'"

In 2012, he recruited Chapman Petroleum to determine what was coming out of the borehole. Sheltered from the 50°C heat in a mobile lab, Brière and his technicians discovered that the gas was 98% hydrogen. That was extraordinary: Hydrogen almost never turns up in oil operations, and it wasn't thought to exist within the Earth much at all. "We had celebrations with large mangos that day," Brière says.

Within a few months, Brière's team had installed a Ford engine tuned to burn hydrogen. Its exhaust was water. The engine was hooked up to a 300-kilowatt generator that gave Bourakébougou its first electrical benefits: freezers to make ice, lights for evening prayers at the mosque, and a flat-screen TV so the village chief could watch soccer games. Children's test scores also improved. "They had the lighting to learn their lessons before going to class in the morning," Diallo says. He soon gave up on oil, changed the name of his company to Hydroma, and began drilling new wells to ascertain the size of the underground supply.

Malian businessman and former presidential candidate Aliou Diallo casts his vote in 2018 (top). Now, his company is preparing to extract hydrogen from underground deposits near the village of Bourakébougou (bottom).

The Malian discovery was vivid evidence for what a small group of scientists, studying hints from seeps, mines, and abandoned wells, had been saying for years: Contrary to conventional wisdom, large stores of natural hydrogen may exist all over the world, like oil and gas—but not in the same places. These researchers say water-rock reactions deep within the Earth continuously generate hydrogen, which percolates up through the crust and sometimes accumulates in underground traps. There might be enough natural hydrogen to meet burgeoning global demand for thousands of years, according to a U.S. Geological Survey (USGS) model that was presented in October 2022 at a meeting of the Geological Society of America.

"When I first heard about it, I thought it was crazy," says Emily Yedinak, a materials scientist who devoted a fellowship at the Advanced Research Projects Agency-Energy (ARPA-E) to drumming up interest in natural hydrogen. "The more that I read, the more I started to realize, wow, the science behind how hydrogen is produced is sound … I was kind of like, 'Why is no one talking about this?'"

Since 2018, however, when Diallo and his colleagues described the Malian field in the *International Journal of Hydrogen Energy*, the number of papers on natural hydrogen has exploded. "It's absolutely incredible and really exponential," says geologist Alain Prinzhofer, lead author on the Mali paper and scientific director of GEO4U, a Brazil-based oil and gas services company that is doing more and more hydrogen work. Dozens of startups, many in Australia, are snatching up the rights to explore for hydrogen. Last year, the American Association of Petroleum Geologists formed its first natural hydrogen committee, and USGS began its first effort to identify promising hydrogen production zones in the United States. "We're in the very beginning, but it will go fast," says Viacheslav Zgonnik, CEO of Natural Hydrogen Energy. In 2019, the startup completed the first hydrogen borehole in the United States, in Nebraska.

The enthusiasm for natural hydrogen comes as interest in hydrogen as a clean, carbon-free fuel is surging. Governments are pushing it as a way to fight global warming, efforts that were galvanized when Russia invaded Ukraine last year and triggered a hasty search, especially in Europe, for alternatives to Russian natural gas. At the moment, all commercial hydrogen has to be manufactured, either in a polluting way, by using fossil fuels, or in an expensive way, by using renewable electricity. Natural hydrogen, if it forms sizable reserves, might be there for the taking, giving the experienced drillers in the oil and gas industry a new, environ-

A methane and hydrogen seep on Mount Chimaera in Turkey has burned for centuries.

PHOTOS: (TOP TO BOTTOM) MICHELE CATTANI/AFP VIA GETTY IMAGES; PAUL CHOUINARD/VERSATILE ENERGY SERVICES; (OPPOSITE PAGE) SVITLANA BELINSKA/ALAMY STOCK PHOTO

mentally friendly mission. "I believe that it has the potential to replace all fossil fuels," Zgonnik says. "That's a very large statement, I know."

Critically, natural hydrogen may be not only clean, but also renewable. It takes millions of years for buried and compressed organic deposits to turn into oil and gas. By contrast, natural hydrogen is always being made afresh, when underground water reacts with iron minerals at elevated temperatures and pressures. In the decade since boreholes began to tap hydrogen in Mali, flows have not diminished, says Prinzhofer, who has consulted on the project. "Hydrogen appears, almost everywhere, as a renewable source of energy, not a fossil one," he says.

It is still early days for natural hydrogen. Scientists don't completely understand how it forms and migrates and—most important—whether it accumulates in a commercially exploitable way. "Interest is growing fast, but the scientific facts are still lacking," says Frédéric-Victor Donzé, a geophysicist at Grenoble Alpes University. Big Oil is hanging back, watching while wildcatters take on the risky exploratory work. Commercialization of the Mali field has run into snags, and elsewhere only a few exploratory wells have been drilled. Donzé, who has sworn off accepting industry money, worries about hype.

Yet some scientists have become true believers. Eric Gaucher, a geochemist at the University of Bern, left a career at French oil giant Total because it wasn't moving fast enough on hydrogen. He believes the Mali discovery might end up in the history books alongside one that happened 163 years ago in Titusville, Pennsylvania. At the time, the world knew about seeps of oil in places such as Iraq and California but was blind to the vast deposits that lay underground. Then on 27 August 1859, a nearly bankrupt prospector named Edwin Drake, working in Titusville with a steam engine and cast-iron drill pipes, struck black gold at a depth of 21 meters, and began collecting it in a bathtub. Before long, U.S. companies were harvesting millions of bathtubs of oil every day.

"I am thinking we are not very far from that with hydrogen," Gaucher says. "We have the concept, we have the tools, the geology. … We only need people able to invest."

EVEN THOUGH IT'S CARBON-FREE, hydrogen has its faults as an energy source. One kilogram of hydrogen holds as much energy as a gallon of gasoline (just under 4 liters). But at ambient pressures, that same kilogram of hydrogen occupies more space than the drum of a typical concrete mixing truck. Pressurized tanks can hold more but add weight and costs to vehicles. Liquefying hydrogen requires chilling it to −253°C—usually a disqualifying expense.

These storage issues—along with a lack of pipelines and distribution systems—are the main reasons why, in the race to electrify cars, batteries have won out over fuel cells, which convert hydrogen to electricity. Similarly, for domestic heating, most experts believe electric heat pumps make more sense than hydrogen furnaces.

Yet as much as half of the world's projected energy demand will remain hard to decarbonize by switching to electricity, says Dharik Mallapragada, an energy systems researcher at the Massachusetts Institute of Technology: "That's where hydrogen comes in." He sees room for hydrogen to replace hydrocarbons in heavy-duty vehicles that are ill-suited to batteries: trucks, ships, and perhaps even planes, all of which can handle larger tanks and fewer fueling stations. Industries such as steel that require high-temperature combustion are another likely market. And today's primary markets for hydrogen—it is needed to make ammonia fertilizers, for example—will continue to grow from the current 90 million tons a year.

The hydrogen rainbow

Researchers use colors to distinguish between different kinds of hydrogen.

- **Gray hydrogen** Made from fossil fuels, which release carbon dioxide and add to global warming.
- **Blue hydrogen** Same as gray hydrogen, but the carbon is captured and sequestered.
- **Green hydrogen** Made without carbon emissions by using renewable electricity to split water.
- **Gold hydrogen** Tapped from natural subsurface accumulations.
- **Orange hydrogen** Stimulated by pumping water into deep source rocks.

But to be climate-friendly, hydrogen needs to be produced cleanly. Today's hydrogen is "gray," made by reacting methane with steam at high pressures or using fossil fuels in other ways. Those processes emit some 900 million tons of carbon dioxide every year, almost as much as global aviation. In principle, that carbon could be captured and sequestered underground, yielding "blue" hydrogen. But most hopes rest on "green" hydrogen—using renewable solar or wind power to split water molecules into oxygen and hydrogen with electrolyzers.

Governments have embraced the concept. In September 2022, the U.S. Department of Energy (DOE) said it would spend $7 billion on at least half a dozen hydrogen "hubs": production sites for green or blue hydrogen. And in May 2022, the European Union called for 20 million tons a year of new green hydrogen—half imported, half domestic—by 2030.

But green hydrogen costs about $5 per kilogram, more than twice as much as gray hydrogen, which tends to track the price of natural gas. Cheaper electrolyzers will help—DOE is sponsoring a "moonshot" to reach $1 per kilogram within a decade. But green hydrogen would also require a huge scale-up of renewable electricity. Meeting the EU target, for instance, would require about 1000 terawatt-hours of new solar and wind installations, nearly double what Europe has now, Mallapragada says.

Pumping hydrogen out of the ground should be much cheaper, which is why proponents sometimes call the natural stuff "gold." Brière says extraction at the Mali site, which benefits from shallow wells and nearly pure hydrogen, could be as cheap as 50 cents per kilogram. Ian Munro, CEO of Helios Aragon, a startup pursuing hydrogen in the foothills of the Spanish Pyrenees, says his break-even costs might end up between 50 and 70 cents. "If it does work, it could revolutionize energy production," he says. "There's a big 'if' there. But you're not going to get that with green hydrogen, right? To me, that's a bottomless pit."

THE OIL AND GAS industry has punctured Earth with millions of wells. How could it have overlooked hydrogen for so long? One reason is that hydrogen is scarce in the sedimentary rocks that yield oil and gas, such as organic-rich shales or mudstones. When compacted and heated, the carbon molecules in those rocks consume any available hydrogen and form longer chain hydrocarbons. Any hydrogen the oil encounters as it migrates to a porous "reservoir" rock such as a sandstone tends to react to form more hydrocarbons. Hydrogen can also react with oxygen in rocks to form water or combine with carbon dioxide to form "abiotic" methane. Microbes gobble it up to make yet more methane.

Even if the hydrogen survives, geologists thought, it should not accumulate. Hydrogen is the smallest molecule of all: It can leak through minerals and even metals. If Earth were producing hydrogen, it seemed unlikely to hang around.

And so, historically, when well loggers cataloged their borehole emanations, they rarely bothered to measure for hydrogen. "The bottom line—they weren't really looking

for hydrogen," says Geoffrey Ellis, an organic geochemist at USGS. "We weren't looking in the right places with the right tools."

Yet the hints were there for those who did look. According to Zgonnik, a geochemist who recently published a review of natural hydrogen, the first scientific discussion of it dates to 1888, when Dmitri Mendeleev, the father of the periodic table, reported hydrogen seeping from cracks in a coal mine in Ukraine. Zgonnik, who was born and raised in Ukraine, says reports of hydrogen are relatively common throughout the former Soviet Union—because Soviet researchers were looking for it. They held to a now discredited theory that would have required significant amounts of natural hydrogen to produce oil from nonliving processes rather than from ancient life.

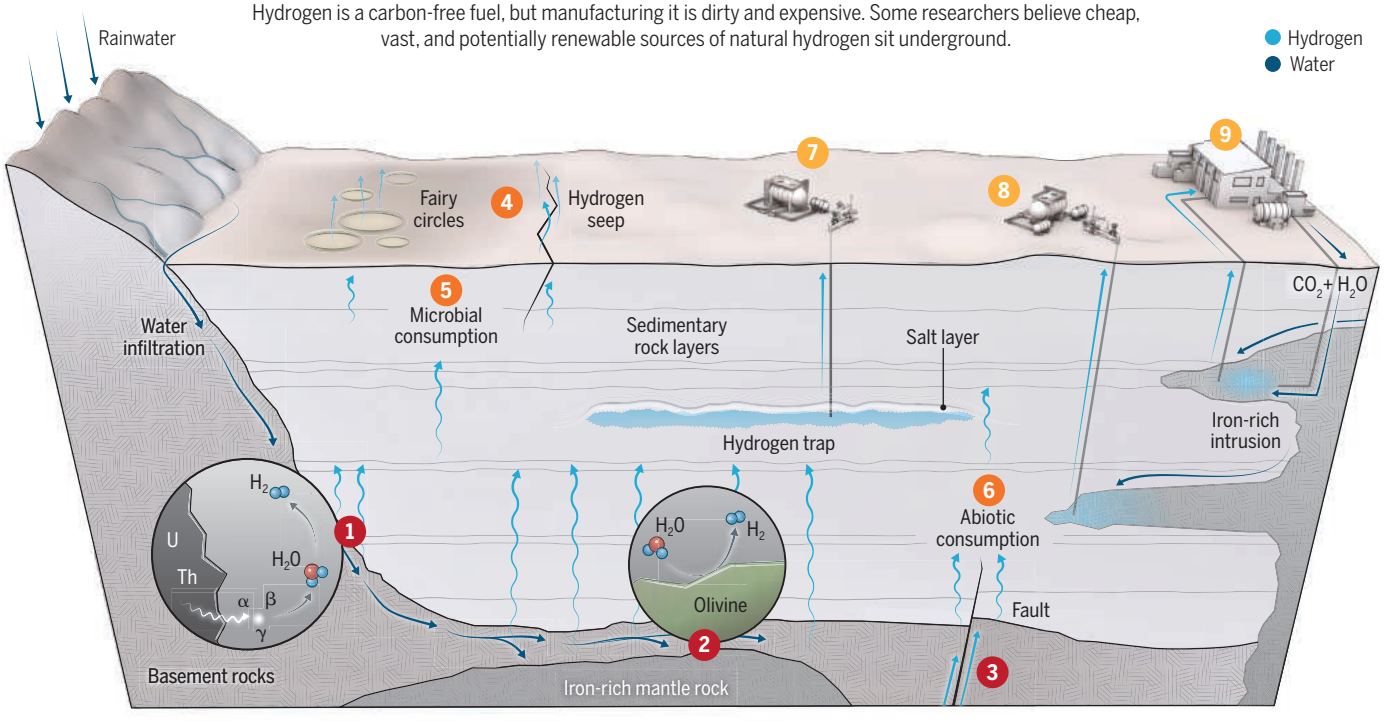
For Barbara Sherwood Lollar, a University of Toronto geochemist, the hydrogen revelation came in the 1980s, when she was a graduate student. She was taking data in mines in Canada and Finland, following up on evidence they sometimes contained flammable gases. She measured the expected hydrocarbons—some methane, some ethane—but they didn't add up to the total mass in her samples. Finally, she realized that some contained as much as 30% hydrogen. "We didn't even measure for it, because nobody expected hydrogen in the system," she says.

Hundreds of hydrogen seeps have now been documented around the world. Zgonnik has come to believe hydrogen can explain even more common features: hundreds of thousands of shallow, circular depressions in the land, tens or hundreds of meters across, that go by names like fairy circles, witch rings, or water basins. "They're widespread and very mysterious," he says. Surveying several of these features along the U.S. East Coast, where they are called Carolina bays, Zgonnik and his colleagues found they were leaking hydrogen, and that its concentration grew with depth. Zgonnik thinks that hydrogen dissolves away minerals in underlying rocks, leading to slumping at the surface.

Vegetation within the circles can sometimes be suppressed, notes Isabelle Moretti, a geologist at the University of Pau and the Adour Region who has documented hydrogen seeping from fairy circles in Brazil, Namibia, and Australia. She speculates it might have something to do

Earth's hydrogen factories

Hydrogen is a carbon-free fuel, but manufacturing it is dirty and expensive. Some researchers believe cheap, vast, and potentially renewable sources of natural hydrogen sit underground.



Generation

1 Radiolysis
Trace radioactive elements in rocks emit radiation that can split water. The process is slow, so ancient rocks are most likely to generate hydrogen.

2 Serpentinization
At high temperatures, water reacts with iron-rich rocks to make hydrogen. The fast and renewable reactions, called serpentinization, may drive most production.

3 Deep-seated
Streams of hydrogen from Earth's core or mantle may rise along tectonic plate boundaries and faults. But the theory of these vast, deep stores is controversial.

Loss mechanisms

4 Seeps
Hydrogen travels quickly through faults and fractures. It can also diffuse through rocks. Weak seeps might explain shallow depressions sometimes called fairy circles.

5 Microbes
In shallower layers of soil and rock, microbes consume hydrogen for energy, often producing methane.

6 Abiotic reactions
At deeper levels, hydrogen reacts with rocks and gases to form water, methane, and mineral compounds.

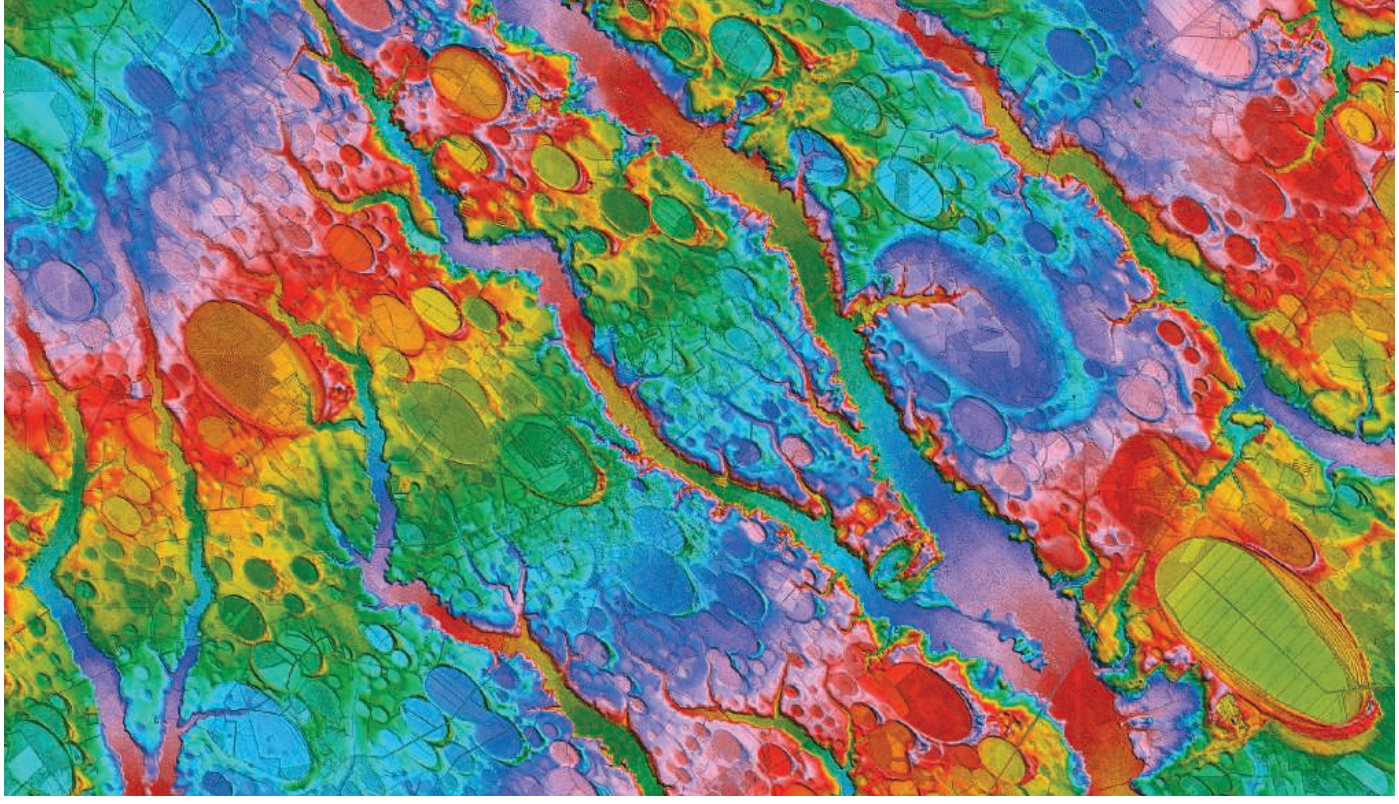
Extraction

7 Traps
Hydrogen might be tapped like oil and gas—by drilling into reservoirs trapped in porous rocks below salt deposits or other impermeable rock layers.

8 Direct
It might also be possible to tap the iron-rich source rocks directly, if they're shallow and fractured enough to allow hydrogen to be collected.

9 Enhanced
Hydrogen production might be stimulated by pumping water into iron-rich rocks. Adding carbon dioxide would sequester it from the atmosphere, slowing climate change.

CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) GEOFFREY ELLIS/USGS



Hydrogen seepages might explain mysterious depressions often called fairy circles. Some are more than 1 kilometer wide in this lidar image of coastal North Carolina.

with hydrogen-loving microbes consuming other nutrients. "Maybe there is nothing left," she says.

Most hydrogen seeps are too feeble to be commercially exploitable. But their very existence is promising, and "really a miracle," Gaucher says. "You have oxidant in the soil, oxygen in the atmosphere, and plenty of microbes that love to eat this hydrogen," he says. "That it exists at all, there must be more."

"There must be a deeper, bigger source."

THE MAIN ENGINE of natural hydrogen production is now thought to be a set of high-temperature reactions between water and iron-rich minerals such as olivine, which dominate Earth's mantle. One common reaction is called serpentinization, because it converts olivine into another kind of mineral called serpentinite. In the process the iron oxidizes, grabbing oxygen atoms from water molecules and releasing hydrogen.

Scientists diving in submersibles have seen this process up close at the volcanic Mid-Atlantic Ridge, where tectonic plates are tugged apart and mantle rocks rise up to create fresh slabs of ocean crust. At a site known as Lost City (for the towering "white smoker" chimneys gushing mineral-rich hot water), researchers measured high amounts of hydrogen spewing from the sea floor. And on Iceland, which straddles the Mid-Atlantic Ridge, Moretti and her colleagues have recorded comparable hydrogen flows at some of the hot springs and geothermal wells that dot the country.

But the hope for commercial natural hydrogen lies closer to potential customers on the continents. Prospectors are looking in cratons, the ancient cores of continents, says Owain Jackson, exploration director at H2Au, a U.K.-based hydrogen company. Trapped within them are bands of iron-rich rock, called greenstone belts, which are the remnants of ocean crust that got squeezed between the cratons in ancient continental collisions. Where olivine and other minerals are buried deep enough to be hotter than 200°C, and yet still exposed to water percolating from the surface, they can produce hydrogen. Jackson, who once helped evaluate lease blocks in a region of Mali several hundred kilometers away from Bourakébougou, believes greenstone belts deep in the West African craton are driving the hydrogen production there. "We're just a bit annoyed we gave the blocks back," he says.

> **"We weren't looking in the right places with the right tools."**
> **Geoffrey Ellis**, U.S. Geological Survey

Cratons hold a second major source of iron with hydrogen-producing potential, Prinzhofer says—one that dates to an evolutionary turning point about 2.4 billion years ago. At the time, the oceans were anoxic and saturated in dissolved iron. But then, in a revolution known as the Great Oxidation Event, ocean-living microbes evolved the ability to photosynthesize. The oxygen they emitted caused the iron to fall as rust to the ocean floor, where it eventually turned to stone. Like greenstone belts, some of these deposits wound up surviving in the cratons and are known today as banded iron formations. They're thought to hold some 60% of the world's iron reserves.

In a 2014 paper, Sherwood Lollar and colleagues considered the makeup of Earth's cratons and found that serpentinization should produce as much as 80% of Earth's hydrogen. A second mechanism, radiolysis, may generate the rest. As radioactive elements in the crust such as uranium and thorium decay, they emit beta particles, aka helium nuclei, along with other radiation that can split water molecules underground and generate an extra trickle of hydrogen.

Zgonnik favors a third and more deep-seated source: He thinks primordial hydrogen, trapped soon after the planet's birth in its iron core, is seeping to the surface through thousands of kilometers of rock. The evidence is spotty and Zgonnik acknowledges the theory is controversial. "It goes against many paradigms," he says.

For Prinzhofer, the question of where natural hydrogen comes from is academic. "Maybe we are all completely wrong," he says. "It doesn't matter for the industry." The oil industry sprang up long before it understood oil's origins, he says. Similarly, what matters for the natural hydrogen in-

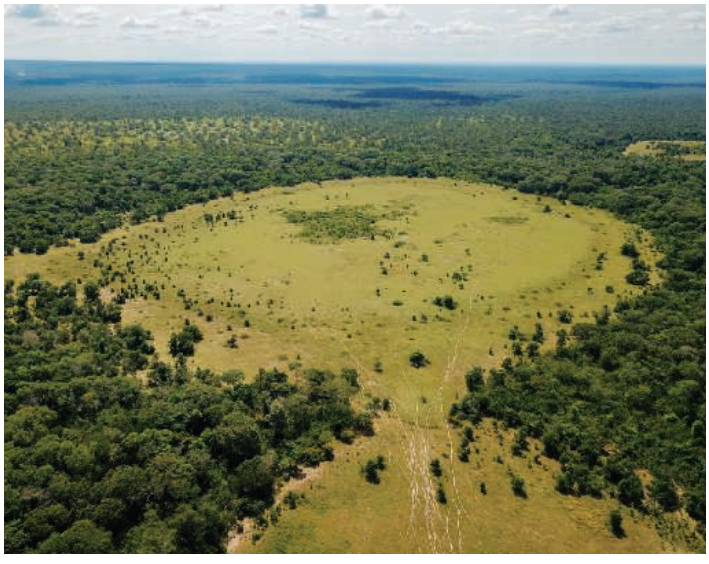
IMAGE: VIACHESLAV ZGONNIK

dustry is simply whether there is enough of the stuff to go after.

At USGS, Ellis is working on answering that question. He thinks Earth produces orders of magnitude more hydrogen each year than the 90 million tons that humans manufacture. But it's not only that flow that matters—it's the size of the underground stock. "How much can be trapped in the subsurface that we can actually go after?" Ellis asks. "That's a much more difficult question to answer."

He and his USGS colleague Sarah Gelman gave it a try using a simple "box" model borrowed from the oil industry. The model accounted for impermeable rock traps of different kinds, the destructive effect of microbes, and the assumption—based on oil industry experience—that only 10% of hydrogen accumulations might ever be tapped economically. Ellis says the model comes up with a range of numbers centered around a trillion tons of hydrogen. That would satisfy world demand for thousands of years even if the green-energy transition triggers a surge in hydrogen use.

Ellis acknowledges that much of this global resource could end up being too scattered to be captured economically, like the millions of tons of gold that are dissolved in the oceans at parts per trillion levels. But that worry hasn't stopped the hydrogen hunters.



A large fairy circle in Brazil that leaks hydrogen is curiously devoid of vegetation.

PHOTO: ALAIN PRINZHOFER

WHILE CONFINED by one of Australia's COVID-19 lockdowns in November 2020, Luke Titus found himself reading an obscure 1944 report: *Bulletin Number 22* from the Department of Mines of the Geological Survey of South Australia. It contained an analysis of data from farmers who had banded together to search for oil, using divining rods and other questionable techniques. "There's even reports of them dipping their hands in kerosene," Titus says. "It's all rather amusing." But Titus, co-founder of a company called Gold Hydrogen, wasn't laughing when he saw the data from one borehole, drilled in 1921 on Kangaroo Island. It had produced as much as 80% hydrogen. Another well, on the nearby Yorke Peninsula, was close to 70%.

In February 2021, when South Australia expanded its oil regulations to allow drilling for hydrogen, Titus pounced. That same month, he submitted an application to explore nearly 8000 square kilometers on the Yorke Peninsula and Kangaroo Island. He created two other paper companies to lodge applications on thousands more square kilometers. Within weeks, he had competitors. "A bunch of other businesses got wind of it," he says.

Now, South Australia is in the middle of a hydrogen boom, at least on paper. The state is blessed with favorable geology. It's covered by the ancient Gawler Craton, and its iron and uranium mines point to the source rocks needed for both serpentinization and radiolysis. With the ocean so close, Titus says, the rocks are sure to be water-saturated. This year he plans to conduct an airborne geophysical survey to delineate what he believes is the source rock on the Yorke Peninsula, just 1.8 kilometers down. In January, in an initial public offering on the Australian Stock Exchange, the company raised $20 million, enough to drill an exploratory well. "We're working at the bleeding edge," he says.

In Spain, Munro is waiting for regulations to catch up. Like Gold Hydrogen, his company Helios Aragon was founded on old but promising data: a "show" of 25% hydrogen in the Monzon-1 well, drilled in 1963 to a depth of 3.7 kilometers by the National Petroleum Company of Aragon. And like Titus, Munro believes he's got an ideal site for hydrogen. In the core of the Pyrenees are iron-rich marine rocks, squeezed and lifted up when the Iberian Plate closed an ocean and rammed into France some 65 million years ago. Munro says deep faults channel hydrogen produced in those rocks up into a porous sandstone layer, which is capped by a tight shale.

Munro plans to drill an exploratory well in late 2024. "We believe we'll be Europe's first natural hydrogen well," he says. But because his lease was awarded under Spain's oil laws, and a 2021 climate law has since put a moratorium on new operations, he won't be able to produce commercially until Spain carves out an exemption for hydrogen.

In the United States, the birthplace of fracking and the shale gas boom, the regulatory environment is looser. Yet, Ellis says, "For unknown reasons to me really, it hasn't taken off in North America yet." At USGS, he and a couple other employees are the only staff focusing on natural hydrogen. At ARPA-E, it was just Yedinak and one other person—until she left the agency a few months ago to join a clean energy startup.

Ellis is now using geophysical data to assess promising U.S. terrain for hydrogen generation. He says the United States likely sits on two rich veins. One is about 10 to 20 kilometers off the Eastern Seaboard, where iron-rich mantle rocks lie about 10 kilometers beneath the seabed. He believes hydrogen created in those rocks may be migrating up and toward shore through porous sediments—perhaps explaining why Carolina bays are found all along the East Coast. Another potential hot spot is in the Midwest, where a volcanic rift failed to split North America a billion years ago. It brought iron-rich mantle rocks close to the surface in a band from Minnesota to Kansas.

That's the target for Zgonnik. In 2019, Natural Hydrogen Energy completed its 3.4-kilometer-deep well in the middle of a "water basin"—the local term for a fairy circle—and surrounded by corn and soybean fields. The well, near Geneva, Nebraska, sits close to deep faults that might connect it to the rocks of the failed rift zone. Zgonnik declined to say how much hydrogen the well produces, but in April 2022, the company HyTerra bought a stake in the operation. A HyTerra presentation says gas from the well "burned with a clear flame"—a sign that hydrogen is predominant.

Gaucher believes the first target for natural hydrogen explorers should be shallow accumulations that sit under impermeable caps within a kilometer or two of the surface. But if the source rocks themselves are

In 2019, the startup Natural Hydrogen Energy drilled the first U.S. hydrogen well amid corn and soybean fields in Nebraska.

within reach, he says, hydrogen could be collected from them directly, like oil from fracked shale; water could even be injected into the iron-rich rock to stimulate production. While collecting hydrogen, the well could also tap the geothermal energy in the heated water that returns to the surface. Best of all, if carbon dioxide were dissolved in the injected water, it could react with magnesium and calcium in the iron-containing rocks and be locked up permanently as limestone. "You'd be sequestering carbon dioxide and producing hydrogen at the same time," Yedinak says.

The prospects are exciting. But the enthusiasm is all hypothetical at the moment. No one anywhere in the world will produce hydrogen commercially anytime soon—except, perhaps, in Mali.

ON A NOVEMBER EVENING, Diallo has arrived on a late flight into Dakar, Senegal, where he maintains a home. He lounges in a spartan room in a white tunic, taking swigs from a water bottle as he tells his story on a video call. He is one of Mali's wealthiest citizens, having built his fortune from a gold mine and trading risky African government debt. He is also a former, and perhaps future, presidential candidate in a country now ruled by a military junta and roiled by years of struggle with Islamist terrorists.

But hydrogen is what he wants to talk about. It has been a passion project since he acquired the rights to the Bourakébougou field more than 15 years ago. "He's a gambler," Prinzhofer says. "He always told me, 'I like to make incredible bets, and this is so exciting.'"

The intervening years have been difficult, the path to commercialization slowed by the pandemic and the military coups. International sanctions meant to put pressure on the junta make it hard for Hydroma to import equipment. And few Western drilling companies are willing to work in Mali, given the constant security concerns.

Nevertheless, Hydroma is close to pumping commercial hydrogen, Diallo claims, though he won't say how close. "Equipment is being put together now as we speak," Brière says. Diallo says the priority is to use Bourakébougou as a filling station for fuel cells that could help electrify Mali, a country where half the population still lacks access to power.

"We'll be taking care of our generation and our children's children's generation."

Denis Brière,
Chapman Petroleum Engineering

But why stop there? Diallo wants to expand into hydrogen buses, trucks, and even trains. After that might come a fertilizer factory. Hydroma has created subsidiaries in Senegal, Mauritania, Niger, and Guinea-Bissau, and Diallo sees hydrogen driving prosperity across a region that holds 400 million people. He calls it the West African Big Green Deal. "We have proved that, beyond geological curiosity, [hydrogen] is a real natural resource that must be relied on in the future," he says.

With 30 wells drilled across the Bourakébougou field, Brière says he can now formally assess "the prize"—oil industry jargon for the recoverable quantity in a reservoir. The field is large, he says: It contains some 60 billion cubic meters of hydrogen, or about 5 million tons, trapped under expansive horizontal sills of ancient volcanic rock.

But the size of the prize may understate the promise. Because Earth makes hydrogen so much faster than oil, the volume of a reservoir is less meaningful, Brière says. "We don't see that it's a confined volume, we see that it's always being filled and flowing and continuous." It might be possible to tap the Bourakébougou field and others like it for many decades without depleting them. "We'll be taking care of our generation and our children's children's generation," Brière says.

The people of Bourakébougou certainly hope so. The Ford engine ran until its spark plugs gave out a few years ago, and a newly installed fuel cell—quieter and more efficient—has not yet been hooked into the village grid. Bourakébougou is dark, waiting for a hydrogen future to arrive. ■

With reporting by Tania Rabesandratana.

PHOTO: VIACHESLAV ZGONNIK

NEUROSCIENCE

Chemical notes of tsetse fly mating

Volatile pheromones offer a means to control flies that spread disease

By **Zainulabeuddin Syed**

As described nearly half a century ago, the tsetse fly lies at the center of a vast and complex biological system extending across Africa, "destroying man and his animals and stifling the economy" [(*1*), p.5]. These words still resonate today in sub-Saharan Africa, where tsetse flies continue to threaten millions of people and livestock by transmitting life-threatening diseases and destroying the economy (*2*, *3*). Indeed, a goal for global health outlined by the World Health Organization includes the sustainable elimination of tsetse-borne diseases by 2030 and emphasizes controlling flies as a primary strategy. Tsetse flies, like many other vectors of disease, display robust olfactory-driven behaviors to navigate toward a host for feeding and to a suitable mate (*4*). On page 660 of this issue, Ebrahim *et al.* (*5*) report the identification of volatile pheromones that control mating behavior in tsetse flies. This could bolster the arsenal of tools to manage the spread of disease by these insects.

Tsetse flies transmit trypanosomes, the protozoan parasites that cause sleeping sickness in humans [human African trypanosomiasis (HAT)] and in animal livestock [African animal trypanosomiasis (AAT) in domestic animals]. There are recent indications of HAT's decline as the result of effective disease surveillance and patient treatment. However, AAT continues to represent the greatest animal health constraint to livestock production in sub-Saharan Africa, killing ~3 million cattle each year and inflicting a direct loss of up to US $1.2 billion annually. Disease transmission occurs when these obligate blood feeders find a host and transfer the parasites while feeding. Thus, a single fly bite can be lethal. A review of centuries-old disease control efforts—especially those directed at exploiting tsetse fly behavior, such as destruction of host animals, clearance of habitats for these insects (e.g., bushes and hedges), and application of insecticides—showed mixed results (*6*). However, insect traps (including targets that attract the flies but are treated with insecticides) have been shown to be quite effective. Because some of the most successful vector-borne disease management strategies have targeted the vector, the discovery of volatile sex attractants in the tsetse fly expands on the possibilities for tsetse management.

Department of Entomology, University of Kentucky, Lexington, KY 40546, USA.
Email: zainulabeuddin.syed@uky.edu

PHOTO: GEOFFREY ATTARDO/UNIVERSITY OF CALIFORNIA, DAVIS

Volatile pheromones have been identified in the tsetse fly *Glossina morsitans*, a vector for disease in humans and animals.

Pheromones are chemicals that mediate attraction between conspecifics (members of the same species) and belong to a broader category of semiochemicals, which modify behavior. The ability to sense chemicals (chemosensation) and respond to them is pronounced in the insect world. For example, a female moth can attract males from hundreds of meters away when it is ready to mate.

Although many insect pheromones have been identified and their mechanisms of action continue to be elucidated, chemical communication in tsetse flies has not been well understood. About 25 years ago, three compounds were isolated from the body washings of female tsetse flies and described as short-range, nonvolatile pheromones that elicit male mating behavior (*7*). Of the three chemicals, 15,19,23-trimethylheptatriacontane (called morsilure) was the most potent. Ebrahim *et al.* have now identified volatile constituents from the body washings of female *Glossina morsitans*, a species of tsetse fly that is a major vector of AAT. Using gas chromatography–mass spectrometry, they identified methyl palmitoleate (MPO), methyl oleate (MO), and methyl palmitate (MP) as constituents in the body washes that elicit a strong mating response. The authors used either artificial tsetse fly–sized decoys perfumed with each of the candidate compounds or virgin tsetse flies doused with each compound to assess mating responses in males (the latency for initiating mating in the presence of conspecific mates was 0.15 ± 0.02 min, which was used as a baseline).

Ebrahim *et al.* further systematically challenged a large variety of olfactory sensory units (sensilla) in the tsetse fly antennae with each volatile compound. Different types of sensilla on insect antennae house olfactory receptor neurons that detect different odor stimuli. Olfactory sensilla of the trichoid category normally respond to pheromones in insects. By recording the electrical activity of the olfactory receptor neurons from individual sensilla, the authors identified those that detect the major volatile constituent of the body washes, including MPO. Intriguingly, the olfactory receptor neurons that responded to MPO also strongly responded to 1-octen-3-ol and 4-methylphenol, the major chemostimuli identified decades ago from cow odors. The chemosensation of structurally diverse molecules by a single olfactory receptor neuron further supports the notion of robust olfactory coding principles among insects (*8*). It will be interesting to see whether the MPO-detecting neurons in tsetse flies express the 9-exon subfamily of odorant receptors, which are the proposed detectors of hydrocarbons present in the cuticle (the extracellular layer that covers the external body surface of insects). Ebrahim *et al.* identified MPO only when flies were soaked in solvent for 24 hours. The typical cuticular extraction regime of soaking for no more than 10 min did not yield MPO. Therefore, where MPO is produced or stored in the flies is yet to be determined.

Ebrahim *et al.* also report that infection of female tsetse flies with trypanosome parasites altered their cuticular chemical profile such that infected flies yielded richer profiles—quantitatively and qualitatively—and were less attractive to male conspecifics. It should be noted that major constituents in the altered chemical profiles were low–molecular weight terpenes with shorter elution times (between 3 and 7 min) compared with those of MPO, MO, and MP. These terpenes can act as strong chemostimuli in tsetse flies (*9*, *10*). It is conceivable, as the authors point out, that infections are metabolically expensive and, as such, flies fighting infection would invest less physiological resources in reproduction. This is then conveyed to males by changes in volatile chemical profiles.

Understanding the basis of chemical communication in tsetse fly mating behavior may guide fly-control strategies.

It has been nearly 65 years since the first pheromone was identified—bombykol, in the silk moth (*11*). The classical method used then—isolating the compound from glands of hundreds of thousands of moths and assessing their effect on male mating responses—is generally still followed today, but it is now combined with modern tools that allow the molecular, genetic, and cellular elements of insect olfaction to be studied. Once this foundational knowledge was established in tsetse fly chemosensation (*12*), attention turned to understanding the signals, or chemical stimuli.

Leveraging chemical attractants to manage blood-feeding, disease-transmitting insects originally emerged from tsetse fly research. This seminal work was built upon developments in three areas of research. The study of cattle—the animal most affected by these insects—has been key in identifying semiochemicals from cattle odors. As a result, a suite of chemicals, such as 1-ocetn-3-ol, as well as a range of methyl, ethyl, and propylphenols (*9*) that have since been shown to influence the behavior of every known blood-feeding arthropod, are constituents of all current field-deployable baits. Other key research emerged from the serendipitous discovery that plant odors and their constituent chemicals can be formulated in appropriate blends that attract tsetse flies (*10*, *13*). The invasive plant *Lantana camara* was long suspected to have enabled the expansion of tsetse flies in Africa. Recently, this plant was implicated as a refuge for mosquitoes, potentially contributing to the year-round maintenance of otherwise seasonal protozoan malaria and other arboviral diseases (*14*). A major constituent identified from *L. camara* odors is 1-octen-3-ol (*10*). Olfactory neurons in tsetse flies and mosquitoes detect this compound with extreme sensitivity. The third category of essential research has been on sex pheromones in tsetse flies (*7*). This now includes the study of Ebrahim *et al.*

Insects are among the most abundant and diverse organisms on Earth and display robust olfactory behaviors. Therefore, it is not surprising that a better understanding of their sensory biology is widening possibilities to exploit their behavior against them to control disease spread. ■

PHOTO: GEOFFREY ATTARDO/UNIVERSITY OF CALIFORNIA, DAVIS

REFERENCES AND NOTES

1. T. A. M. Nash, *Africa's Bane: The Tsetse Fly* (Collins, 1969).
2. World Health Organization (WHO), "Trypanosomiasis, human African (sleeping sickness)" (WHO, Fact Sheets, 2022).
3. M. Alsan, *Am. Econ. Rev.* **105**, 382 (2015).
4. Z. Syed, *Curr. Opin. Insect Sci.* **10**, 83 (2015).
5. S. A. M. Ebrahim *et al.*, *Science* **379**, eade1877 (2023).
6. S. J. Torr, G. A. Vale, *Trends Parasitol.* **31**, 95 (2015).
7. D. A. Carlson *et al.*, *Science* **201**, 750 (1978).
8. M. Herre *et al.*, *Cell* **185**, 3104 (2022).
9. D. R. Hall *et al.*, *Identification of Host Odour Attractants for Tsetse Flies: Final Report 1986-1990* (Natural Resources Institute, 1990).
10. Z. Syed, thesis, Université de Neuchâtel (2002).
11. P. Karlson, A. Butenandt, *Annu. Rev. Entomol.* **4**, 39 (1959).
12. J. S. Chahda *et al.*, *PLOS Genet.* **15**, e1008005 (2019).
13. Z. Syed, P. M. Guerin, *J. Insect Physiol.* **50**, 43 (2004).
14. S. B. Agha *et al.*, *Viruses* **13**, 32 (2020).

ACKNOWLEDGMENTS

Z.S. is supported by the National Institutes of Health (R21AI163886) and the National Institute of Food and Agriculture, US Department of Agriculture (under HATCH project 2353077000).

10.1126/science.adg2817

PLANETARY SCIENCE

The first dedicated ice giants mission

The mysteries of the Uranus system can be unlocked through interdisciplinary exploration

By **Kathleen E. Mandt**

Uranus and Neptune are commonly called ice giants because they have a greater percentage of heavy elements compared to the gas giants, Jupiter and Saturn. The only spacecraft to visit Uranus and Neptune was Voyager 2 with brief flybys in 1986 and 1989—more than 30 years ago. These encounters presented a captivating view of two exotic planetary systems, leaving more questions than previously imagined. A recent survey conducted by the National Academies of Sciences, Engineering, and Medicine identified a Uranus Orbiter and Probe (UOP) as the planetary science community's top priority for the next NASA large-scale mission. It will explore how Uranus formed; how much it migrated after formation; the planet's interior structure, atmosphere, magnetosphere, and ring system; and whether any moons have or once had subsurface liquid water oceans. This mission will serve to inspire and educate multiple generations about solar system history and the mysteries at its farthest reaches.

Each decade, NASA tasks the National Academies to conduct a "decadal survey" of the planetary science community to define priorities for the next 10 years. NASA and Congress view these surveys as the US space science community's formal statement of priority. They cost millions of dollars to execute and include independently evaluated mission concept studies. This large investment is critical to ensure that the survey recommends a technologically feasible program with realistic cost assumptions. It is a consensus report representing the voice of the community who conduct the survey and provide input through white papers. Without this process, the community would have limited influence with NASA in communicating top science priorities.

The planetary science decadal survey released in 2022, *Origins, Worlds, and Life* (OWL) (*1*), was one of the most unusual conducted to date because it took place during the COVID-19 pandemic. Although conditions created by the pandemic made the effort an extraordinary sacrifice for the many volunteers working on the survey, participants were determined to provide a true consensus. Over a period of 1 year, panelists held 176 meetings, carefully reviewed over 500 white papers, and sought further input through more than 300 presentations by external speakers given in open session. The survey was built on a framework of 12 priority science questions each outlined in individual chapters. Through these chapters, the "dearth of knowledge on the ice giants" was identified as a problem of highest priority for resolving in the coming decade.

To begin addressing this issue, a UOP was recommended as the next planetary flagship, or large-scale, mission because of the importance of full system science. This recommendation is the culmination of over 20 years of mission concept studies. UOP was first identified as a high priority in the 2003–2013 decadal survey *New Frontiers in the Solar System* (*2*). It was ranked as the third-highest priority flagship in the 2013–2023 decadal survey, *Visions and Voyages* (V&V) (*3*), following Mars Sample Return and the Europa Clipper, two missions now well advanced in their development. OWL reiterated the community consensus from V&V that UOP should be the next step. UOP will address science goals ranging from the origin and evolution of the Solar System to understanding processes only existing in this mysterious planetary system, and will provide results to help answer fundamental questions about one of the most common types of known exoplanets (planets that exist outside the Solar System).

Understanding how the giant planets formed and then migrated has broad implications for explaining the distribution of small bodies in our Solar System; the delivery through these small bodies of water and life-supporting elements to the inner Solar System; and, extending beyond the Solar System, the formation of exoplanets and their system architectures (*1*). Solar System formation and evolution are evaluated by connecting current measurements to conditions in the protosolar nebula (PSN)—the disk of gas, dust, and ice in which Solar System objects formed (*4*). Studying all four giant planets (Uranus, Neptune, Jupiter, and Saturn) is vital for reconstructing Solar System history (*5, 6*).

The giant planets formed by collecting gas and solids from the PSN, migrating to their current locations, and then collecting more solid materials through postformation impacts by icy and rocky planetesimals. Models predict different scenarios for formation and migration (*7*), but require measurements of isotope ratios and noble gas abundances in each planet's atmosphere to verify which modeled scenario is most accurate (*1, 4–6*). An atmospheric probe is required to measure them.

The nitrogen isotope ratio, $^{14}N/^{15}N$, is one of the more important isotopic measurements because it reveals the planet's source of nitrogen. Jupiter's $^{14}N/^{15}N$, the only giant planet measurement currently available, is similar to the Sun's. This suggests that Jupiter obtained its nitrogen either from PSN gas that was enriched in N_2 through heating of PSN ices, or from building blocks formed in extremely cold conditions (*6*). A lower limit for Saturn's tropospheric $^{14}N/^{15}N$ suggests a similar origin (*8*), but an atmospheric probe is needed to confirm a similarly high ratio in the well-mixed atmosphere. No measurement of $^{14}N/^{15}N$ is currently available for the ice giants.

The heavy noble gases provide another constraint for formation and evolution of the giant planets. Comparing noble gas abundances in giant planet atmospheres with abundances in analogs for solid materials in the PSN, comets and chondrites, can establish the relative contributions of rocky and icy material to the planet's formation and reveal how these solid materials were distributed within the PSN. Noble gas abundances are only available for Jupiter, where they also imply either enriched PSN gas or solid building blocks formed in very cold conditions (*5*). Similar measurements are needed for Saturn and the ice giants. The UOP mission concept design includes a probe that will measure the noble gas abundances and $^{14}N/^{15}N$ in the atmosphere (see the figure).

UOP also includes an orbiter that is designed for full system science. The Uranus system is especially interesting because it could bear testimony to a massive collision that tilted Uranus on its side (*9*), resulting in a system with rings and moons at an obliquity of 98°. This orientation causes extreme atmospheric seasonal variation over its 84–Earth-year orbit, but observations of haze and clouds from Earth cannot provide enough information to explain atmospheric

Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA. Email: kathleen.mandt@jhuapl.edu

circulation and wind patterns. The UOP probe will measure wind and temperature at one location, and the orbiter will make global observations that will determine how this distinct atmosphere works.

The system's extreme obliquity also limits visibility of the moons to one hemisphere during southern and northern summers. Voyager 2 could only image the moons' southern hemispheres, but what was seen was unexpected. The five largest moons, predicted to be cold dead worlds, all showed evidence of recent resurfacing (*10*), suggesting that geologic activity might be ongoing. One or more of these moons could have potentially habitable liquid water oceans under an ice shell, making them "ocean worlds." Ariel, the most extensively resurfaced moon, is a strong ocean worlds candidate along with the two largest moons, Titania and Oberon (*10*) (see the figure). UOP will image and measure the composition of the full surfaces of the moons to search for ongoing geologic activity, and measure whether magnetic fields vary in their interiors owing to the presence of liquid water.

Uranus has nine very dense narrow rings that would require resonances with a fleet of shepherding satellites to prevent them from rapidly spreading out and losing their sharp edges (*11*). Although Voyager 2 found resonances with small moons Cordelia, Ophelia, and Cressida, they are not sufficient to explain all of the ring features. Either a large number of moons too small for Voyager 2 detection are present, or resonances are created by the internal structure of Uranus (*12*). Furthermore, the composition of the surprisingly dark ring particles is not known and is clearly different from the surface composition of the moons (*11*). UOP will search for shepherding moons, monitor the motions of ring particles, and measure their composition.

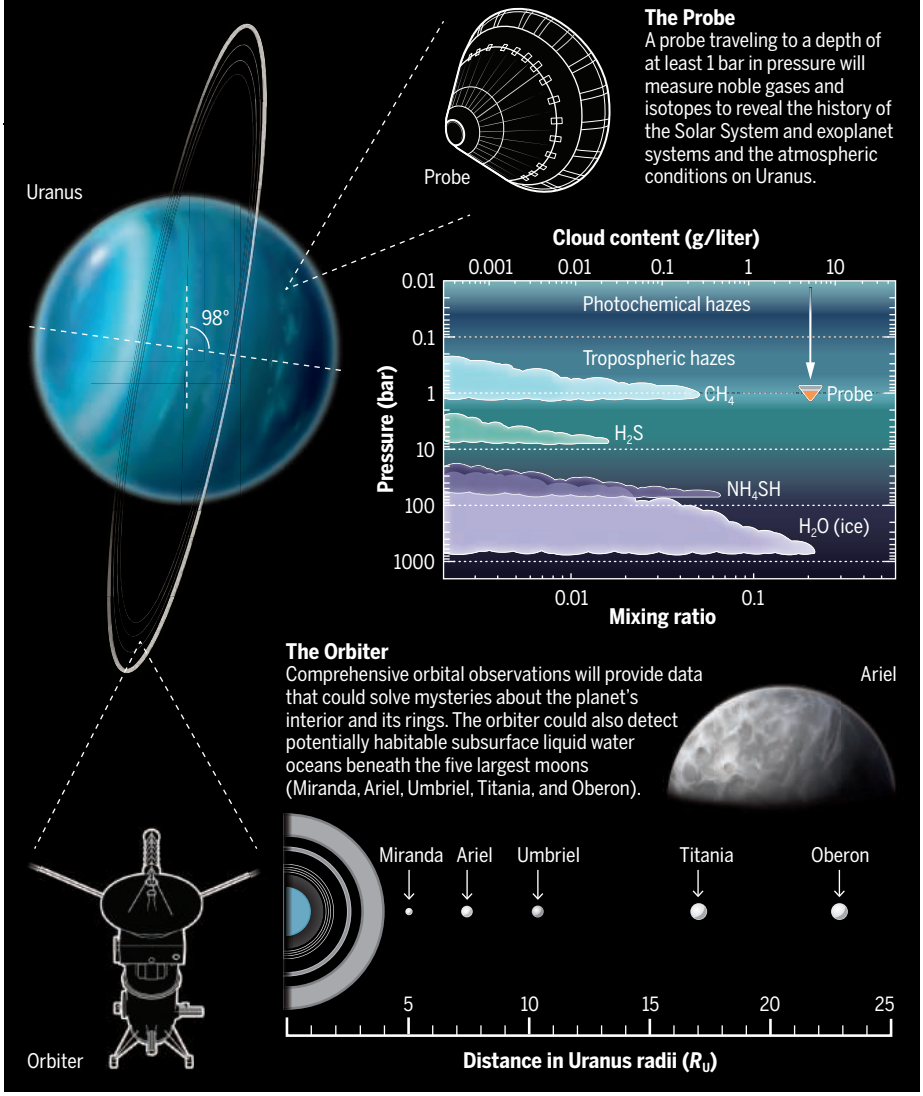
The magnetosphere of Uranus is also very unusual. The magnetic field is not only tilted 60° from the rotation axis, it is also offset from the center of the planet (*13*). It is not clear how such a field can be produced, and this unusual orientation, combined with the extreme obliquity of the planet, causes seasonal and diurnal variations that might alter the planet's atmosphere and moons. UOP will monitor the magnetosphere over the duration of the mission, recommended by OWL to be at least 5 years.

The interior structure of Uranus is a great mystery made more compelling by the discovery that the cores of Jupiter and Saturn are not solid with a clear boundary, but are instead diffuse and exchange material with the atmosphere. This could be the result either of planet-formation processes in which the core reaches a critical mass, vaporizes, and produces composition gradients; or from core erosion, in which heavy elements dissolve and erode into the atmosphere after formation (*14*). Determining whether the core of Uranus is solid or diffuse, and whether it is dominated by rock or ice, is important not only for understanding the ice giants, but also has implications for the origin and internal structure of similarly sized exoplanets. UOP will determine the interior structure of Uranus with atmospheric probe and orbiter gravity measurements.

The ideal next steps will build on past studies (*1–3, 15*), particularly the OWL study (*1*). OWL recommended launching by 2032 to allow the spacecraft to use Jupiter's massive gravity to gain speed out to Uranus. This also guarantees arrival well before northern autumn equinox in 2050, ensuring full visibility of the moons. As an added benefit, the recommended trajectory slingshots the spacecraft over the ecliptic, providing a polar view of the Sun—a perspective of interest to the heliophysics community for making measurements to better understand solar wind differences between the poles and the equator. Trajectories after 2032 that do not use Jupiter's gravity but still arrive before equinox are possible. However, they either deliver a smaller spacecraft into orbit carrying fewer instruments, or take longer to arrive. OWL had limited time to evaluate the millions of possible trajectories, so it will be necessary to use NASA-provided constraints on possible launch years to re-

Investigating the Uranus system

The proposed Uranus Orbiter and Probe (UOP) will help to determine how Uranus formed and its migration since it formed—the planet's interior structure, atmosphere, magnetosphere, and ring system—and whether any moons have habitable liquid water oceans.



GRAPHIC: A. FISHER/*SCIENCE*

fine trajectory options and determine specific mass constraints. This allows development of a spacecraft design optimized for the mass constraints and selection of instruments. Evaluating other contributions is also of great importance. The European Space Agency (ESA) has done several studies of ice giants missions and represents a community prepared to make essential scientific and technical contributions.

As preparation begins for UOP, it is important to recognize that although Neptune is also an ice giant, its system is very different from that of Uranus and presents questions that cannot be answered by UOP (*15*). Neptune's moon system is dominated by Triton, a Kuiper Belt object that reconfigured the entire system when it was captured. Additionally, the heat balance of Neptune and Uranus differs by an order of magnitude, potentially indicating substantial differences in internal structure (*14*). Because of its greater distance from the Sun, Neptune is more challenging to reach. Technology development and focused mission studies should begin in this decade to prepare to explore this equally interesting planetary system. The space science community has waited more than 30 years to explore the ice giants, and missions to them will benefit many generations to come. ■

REFERENCES AND NOTES

1. National Academies of Sciences, Engineering, and Medicine, Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023-2032 (National Academies Press, 2022).
2. National Research Council, New Frontiers in the Solar System: An Integrated Exploration Strategy (National Academies Press, 2003).
3. National Research Council, Vision and Voyages for Planetary Science in the Decade 2013-2022 (National Academies Press, 2011).
4. K. E. Mandt *et al.*, *Space Sci. Rev.* **197**, 297 (2015).
5. K. Mandt *et al.*, *Space Sci. Rev.* **216**, 1 (2022).
6. O. Mousis *et al.*, *Exp. Astron.* 10.1007/s10686-021-09775-z (2021).
7. D. Nesvorný, *Annu. Rev. Astron. Astrophys.* **56**, 137 (2018).
8. L. Fletcher *et al.*, *Icarus* **238**, 170 (2014).
9. R. Rufu, R. Canup, *Astrophys. J.* **928**, 123 (2022).
10. P. Schenk, J. Moore, *Philos. Trans. R. Soc. A* **378**, 20200102 (2020).
11. P. Nicholson *et al.*, in *Planetary Ring Systems: Properties, Structure, and Evolution*, M. S. Tiscareno, C. D. Murray, Eds. (Cambridge Planetary Science, 2018), pp. 93–111.
12. J. A'Hearn, M. M. Hedman, C. R. Mankovich, H. Aramona, M. S. Marley, *Planet. Sci. J.* **3**, 194 (2022).
13. C. Arridge, C. Paty, in *Magnetospheres in the Solar System*, R. Maggiolo *et al.*, Eds. (American Geophysical Union, 2021), pp. 515–534.
14. R. Helled, J. Fortney, *Philos. Trans. R. Soc. A* **378**, 20190474 (2020).
15. M. Hofstadter *et al.*, *Planet. Space Sci.* **177**, 104680 (2019).

ACKNOWLEDGMENTS

K.E.M. served on the Pre-decadal Ice Giants Science Definition Team and the OWL Giant Planet Systems panel. Thanks to members of the science community, especially M. Hofstadter and A. Simon for insightful discussions.

10.1126/science.ade8446

NEUROSCIENCE

Possible psychedelic therapeutic mechanism

Psychedelics act on intracellular serotonin receptors that are not accessible by serotonin alone

By **Evan M. Hess**[1] *and* **Todd D. Gould**[1,2,3,4]

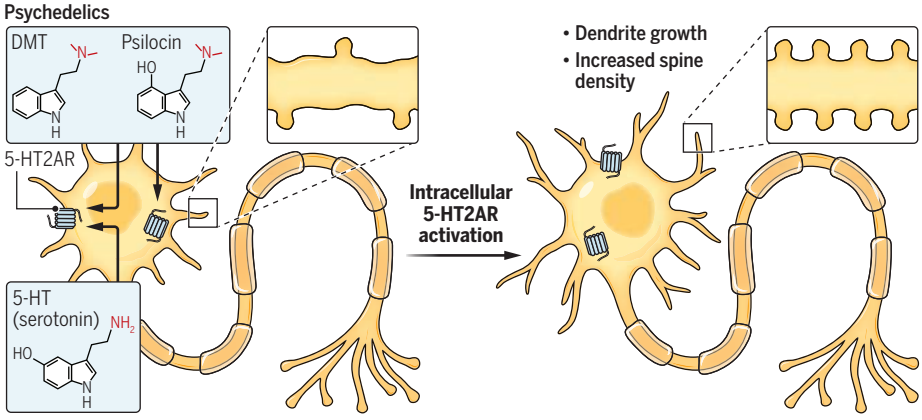
Classical psychedelic compounds, which include lysergic acid diethylamide (LSD), mescaline, psilocybin, and *N*,*N*-dimethyltryptamine (DMT), are broadly defined by their ability to produce altered states of consciousness and mood. They activate the 5-hydroxytryptamine (serotonin) 2A receptor (5-HT2AR), which can positively or negatively affect how an organism engages with its environment (*1*). Medically supervised psychedelic-assisted therapy may be efficacious for posttraumatic stress disorder, substance use disorders, and treatment-resistant depression (*2*–*5*). However, there are gaps in our understanding of the neurobiological mechanisms that are engaged by psychedelic compounds, their therapeutic potential at scale, and whether their positive outcomes can be separated from undesirable effects (*6*, *7*). On page 700 of this issue, Vargas *et al.* (*8*) reveal a potentially therapeutically relevant mechanism of action for DMT and psilocin (the active form of psilocybin) that involves 5-HT2AR, which, when activated intracellularly, promotes dendritic growth and increases the spine density of rat cortical neurons.

DMT is endogenously synthesized in mammals from the essential amino acid L-tryptophan and has been measured in the rat cortex at levels similar to those of other monoamine neurotransmitters such as serotonin (*9*). DMT is the main psychoactive component of ayahuasca, a potent hallucinogenic plant-based brew that has been consumed for thousands of years by indigenous people of South America for spiritual and healing ceremonies (*10*). Despite its long history of consumption and presence in the mammalian brain, the endogenous function of DMT is unknown at present. A key focus of Vargas *et al.* is the importance of how the chemical structures of DMT and the structurally similar psilocin permit them to easily pass through the cell membrane, unlike serotonin.

5-HT2AR is expressed intracellularly on multiple organelles, including endosomes and the Golgi apparatus. Vargas *et al.* found that the chemical properties of serotonin and DMT produce altered signaling according to the cellular localization of 5-HT2ARs in cortical neurons to produce distinct effects. Using cultured cells, they determined

Psychedelics activate intracellular signaling

Serotonin readily activates 5-hydroxytryptamine 2A receptors (5-HT2ARs) on the cell membrane, but its unmethylated amino group (red) restricts its membrane permeability. This functional group is methylated (red) in classical psychedelics such as *N*,*N*-dimethyltryptamine (DMT) and psilocin (produced from psilocybin), which increases membrane permeability. Activation of intracellular 5-HT2ARs by these psychedelics leads to sustained increases in dendrite growth and spine density.



GRAPHIC: A. MASTIN/*SCIENCE*

fine trajectory options and determine specific mass constraints. This allows development of a spacecraft design optimized for the mass constraints and selection of instruments. Evaluating other contributions is also of great importance. The European Space Agency (ESA) has done several studies of ice giants missions and represents a community prepared to make essential scientific and technical contributions.

As preparation begins for UOP, it is important to recognize that although Neptune is also an ice giant, its system is very different from that of Uranus and presents questions that cannot be answered by UOP (*15*). Neptune's moon system is dominated by Triton, a Kuiper Belt object that reconfigured the entire system when it was captured. Additionally, the heat balance of Neptune and Uranus differs by an order of magnitude, potentially indicating substantial differences in internal structure (*14*). Because of its greater distance from the Sun, Neptune is more challenging to reach. Technology development and focused mission studies should begin in this decade to prepare to explore this equally interesting planetary system. The space science community has waited more than 30 years to explore the ice giants, and missions to them will benefit many generations to come. ■

REFERENCES AND NOTES

1. National Academies of Sciences, Engineering, and Medicine, Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023-2032 (National Academies Press, 2022).
2. National Research Council, New Frontiers in the Solar System: An Integrated Exploration Strategy (National Academies Press, 2003).
3. National Research Council, Vision and Voyages for Planetary Science in the Decade 2013-2022 (National Academies Press, 2011).
4. K. E. Mandt *et al.*, *Space Sci. Rev.* **197**, 297 (2015).
5. K. Mandt *et al.*, *Space Sci. Rev.* **216**, 1 (2022).
6. O. Mousis *et al.*, *Exp. Astron.* 10.1007/s10686-021-09775-z (2021).
7. D. Nesvorný, *Annu. Rev. Astron. Astrophys.* **56**, 137 (2018).
8. L. Fletcher *et al.*, *Icarus* **238**, 170 (2014).
9. R. Rufu, R. Canup, *Astrophys. J.* **928**, 123 (2022).
10. P. Schenk, J. Moore, *Philos. Trans. R. Soc. A* **378**, 20200102 (2020).
11. P. Nicholson *et al.*, in *Planetary Ring Systems: Properties, Structure, and Evolution*, M. S. Tiscareno, C. D. Murray, Eds. (Cambridge Planetary Science, 2018), pp. 93–111.
12. J. A'Hearn, M. M. Hedman, C. R. Mankovich, H. Aramona, M. S. Marley, *Planet. Sci. J.* **3**, 194 (2022).
13. C. Arridge, C. Paty, in *Magnetospheres in the Solar System*, R. Maggiolo *et al.*, Eds. (American Geophysical Union, 2021), pp. 515–534.
14. R. Helled, J. Fortney, *Philos. Trans. R. Soc. A* **378**, 20190474 (2020).
15. M. Hofstadter *et al.*, *Planet. Space Sci.* **177**, 104680 (2019).

ACKNOWLEDGMENTS

K.E.M. served on the Pre-decadal Ice Giants Science Definition Team and the OWL Giant Planet Systems panel. Thanks to members of the science community, especially M. Hofstadter and A. Simon for insightful discussions.

10.1126/science.ade8446

NEUROSCIENCE

Possible psychedelic therapeutic mechanism

Psychedelics act on intracellular serotonin receptors that are not accessible by serotonin alone

By **Evan M. Hess**[1] *and* **Todd D. Gould**[1,2,3,4]

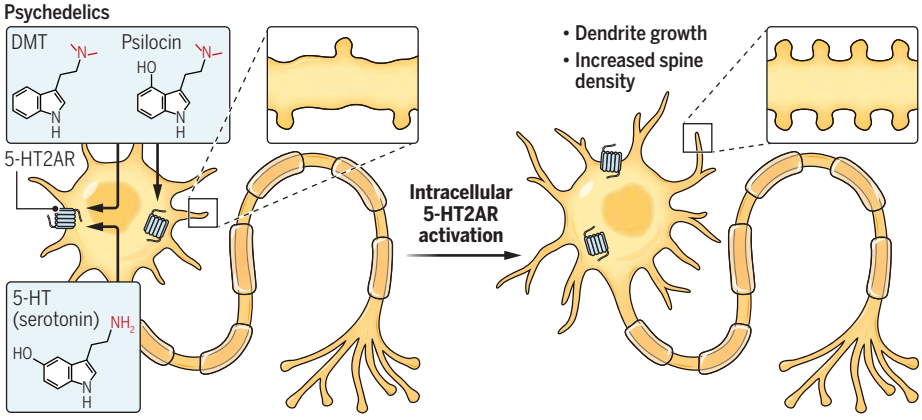
Classical psychedelic compounds, which include lysergic acid diethylamide (LSD), mescaline, psilocybin, and *N*,*N*-dimethyltryptamine (DMT), are broadly defined by their ability to produce altered states of consciousness and mood. They activate the 5-hydroxytryptamine (serotonin) 2A receptor (5-HT2AR), which can positively or negatively affect how an organism engages with its environment (*1*). Medically supervised psychedelic-assisted therapy may be efficacious for posttraumatic stress disorder, substance use disorders, and treatment-resistant depression (*2*–*5*). However, there are gaps in our understanding of the neurobiological mechanisms that are engaged by psychedelic compounds, their therapeutic potential at scale, and whether their positive outcomes can be separated from undesirable effects (*6*, *7*). On page 700 of this issue, Vargas *et al.* (*8*) reveal a potentially therapeutically relevant mechanism of action for DMT and psilocin (the active form of psilocybin) that involves 5-HT2AR, which, when activated intracellularly, promotes dendritic growth and increases the spine density of rat cortical neurons.

DMT is endogenously synthesized in mammals from the essential amino acid L-tryptophan and has been measured in the rat cortex at levels similar to those of other monoamine neurotransmitters such as serotonin (*9*). DMT is the main psychoactive component of ayahuasca, a potent hallucinogenic plant-based brew that has been consumed for thousands of years by indigenous people of South America for spiritual and healing ceremonies (*10*). Despite its long history of consumption and presence in the mammalian brain, the endogenous function of DMT is unknown at present. A key focus of Vargas *et al.* is the importance of how the chemical structures of DMT and the structurally similar psilocin permit them to easily pass through the cell membrane, unlike serotonin.

5-HT2AR is expressed intracellularly on multiple organelles, including endosomes and the Golgi apparatus. Vargas *et al.* found that the chemical properties of serotonin and DMT produce altered signaling according to the cellular localization of 5-HT2ARs in cortical neurons to produce distinct effects. Using cultured cells, they determined

Psychedelics activate intracellular signaling

Serotonin readily activates 5-hydroxytryptamine 2A receptors (5-HT2ARs) on the cell membrane, but its unmethylated amino group (red) restricts its membrane permeability. This functional group is methylated (red) in classical psychedelics such as *N*,*N*-dimethyltryptamine (DMT) and psilocin (produced from psilocybin), which increases membrane permeability. Activation of intracellular 5-HT2ARs by these psychedelics leads to sustained increases in dendrite growth and spine density.



GRAPHIC: A. MASTIN/*SCIENCE*

that intracellular receptors engaged by DMT are involved in rapidly enhancing structural neuroplasticity, specifically increased dendritic growth and spine density, relative to membrane-localized receptors (see the figure). Bestowing cortical neurons in vivo with serotonin transporters allowed serotonin to access the intracellular receptors, mimicking the changes in structural neuroplasticity that are observed when cells were treated with DMT. The relevance of this phenomenon to changes in mouse behavior was assessed using a forced swim test (FST), a widely used approach that is thought to have predictive value of antidepressant effects in humans, although the actual clinical predictive validity is questionable (*11*). Providing serotonin access to intracellular 5-HT2ARs reduced immobility in the FST, which is considered an antidepressant-like effect. Given that serotonin cannot normally access intracellular 5-HT2ARs in cortical excitatory neurons owing to a lack of serotonin transporters, a role of endogenous DMT may be to engage the intracellular receptor pool and facilitate neuroplastic changes. Further assessment of the role of this mechanism in mediating circuit activity and resulting behaviors relevant to depression is necessary.

The lifetime prevalence of depression is ~20%, and it is a leading cause of disability worldwide. First-line treatments such as selective serotonin reuptake inhibitors are taken daily for weeks before therapeutic effects are potentially observed and lead to sustained remission in only ~30% of cases; they also have side effects that affect patient adherence (*12*). This has led to a shift in the search for depression therapeutics in favor of rapid-acting compounds such as ketamine, which can result in sustained therapeutic effects for days or longer after a single dose (*13*). Similarly, recent clinical trials have suggested that DMT and psilocybin are rapidly efficacious in treating major depression, with psilocybin appearing to have sustained therapeutic effects for weeks after treatment sessions that consist of a combination of psilocybin with supportive therapy (*2*, *3*).

Unfortunately, clinical trials with psychedelics are commonly statistically underpowered and mostly do not represent common patient populations. Moreover, it is impossible to adequately include a placebo control given the profound effects that these compounds have on perception, which can easily unblind both study participants and experimenters (*6*, *7*). As a result, it is difficult to predict how the existing clinical findings will be extrapolated to a wider population. Thus, psychedelic-assisted therapy should be embraced with cautious optimism; although the thousands of years of human psychedelic use and the positive outcomes of modern, though early stage, clinical trials point toward therapeutic potential, scaling psychedelic use to potentially millions of people with depression, for example, requires a careful and rigorous scientific approach.

The study conducted by Vargas *et al.* is a key achievement in the understanding of the mechanism of action of psychedelics. However, much needs to be done to affirm the sufficiency of signaling through intracellular 5-HT2ARs as mediators of the properties in humans, either hallucinogenic or medical, and to determine whether positive therapeutic outcomes can be separated from the classical psychedelic properties of such drugs. This will require interrogation of the specific signaling cascades engaged by intracellular 5-HT2ARs that lead to increased dendritic growth and spine density and defining the mechanistic relationship of such changes to cognitive processes. Moreover, understanding whether the findings for DMT are fully shared with all psychedelics (such as the structurally dissimilar mescaline), which was only partially explored by Vargas *et al.*, will be needed before the proposed mechanism can be described as a common psychedelic mechanism. Nevertheless, these findings are an important step forward for a rapidly expanding and much-needed field of study. ■

REFERENCES AND NOTES

1. D. E. Nichols, *Pharmacol. Rev.* **68**, 264 (2016).
2. A. K. Davis *et al.*, *JAMA Psychiatry* **78**, 481 (2021).
3. D. C. D'Souza *et al.*, *Neuropsychopharmacology* **47**, 1854 (2022).
4. A. K. Davis, L. A. Averill, N. D. Sepeda, J. P. Barsuglia, T. Amoroso, *Chronic Stress* **4**, 2470547020939564 (2020).
5. M. P. Bogenschutz *et al.*, *JAMA Psychiatry* **79**, 953 (2022).
6. J. Phelps, R. N. Shah, J. A. Lieberman, *JAMA Psychiatry* **79**, 189 (2022).
7. D. B. Yaden, J. B. Potash, R. R. Griffiths, *JAMA Psychiatry* **79**, 943 (2022).
8. M. V. Vargas *et al.*, *Science* **379**, 700 (2023).
9. J. G. Dean *et al.*, *Sci. Rep.* **9**, 9333 (2019).
10. E. Frecska, P. Bokor, M. Winkelman, *Front. Pharmacol.* **7**, 35 (2016).
11. D. A. Slattery, J. F. Cryan, *Nat. Protoc.* **7**, 1009 (2012).
12. A. J. Rush *et al.*, *Am. J. Psychiatry* **163**, 1905 (2006).
13. E. M. Hess, L. M. Riggs, M. Michaelides, T. D. Gould, *Biochem. Pharmacol.* **197**, 114892 (2022).

ACKNOWLEDGMENTS

T.D.G. received research funding from Allergan Pharmaceuticals over the past 3 years. He is listed as co-author on patents and patent applications related to the pharmacology and use of ketamine metabolites in the treatment of depression, anxiety, anhedonia, suicidal ideation, and posttraumatic stress disorders.

[1]Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA. [2]Department of Pharmacology, University of Maryland School of Medicine, Baltimore, MD, USA. [3]Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD, USA. [4]Baltimore Veterans Affairs Medical Center, Veterans Affairs Maryland Health Care System, Baltimore, MD, USA. Email: tgould@som.umaryland.edu

10.1126/science.adg2989

PHYSIOLOGY

Photonic tinkering in the open ocean

Light-manipulating materials are discovered in the eyeglitter of pelagic crustaceans

By **Kate Feller** *and* **Megan Porter**

Billions of animals living in open water, or pelagic habitats, can disappear into their surroundings using a variety of light-manipulating camouflage solutions. These include transparent, antireflection, and glittery reflective structures. Although such photonic camouflage allows these animals to vanish into their surroundings, they still need to eat (and avoid being eaten), which requires the ability to detect their invisible neighbors. Thus, an arms race exists between predators and prey for the ability to see and yet not be seen (*1*). Evolutionary tinkering across the diversity of pelagic animals has produced multiple solutions for controlling the transmission, reflection, and detection of light. On page 695 of this issue, Shavit *et al.* (*2*) report the discovery of photonic glass materials that form the basis of sparkly "eyeglitter" in the larvae of pelagic crustaceans and allows for both reflective camouflage and vision. These findings present a mechanism for producing salient, tunable coloration and light manipulation in space-limited tissues.

When peering into a bucket of seawater scooped from the ocean, the tiny animals and larvae that live in open water, or zooplankton, often can only be spotted by their pigmented eyes. These black dots travel side by side through the water, giving no indication that they are part of a much larger creature. This is because the bodies of many crustacean zooplankton are almost as transparent as glass. This is achieved both through lack of pigmentation in most body tissues and by photonically minimizing the amount of light scattered or reflected from internal and external structures. How internal structures such as muscles, organs, and blood become transparent is still an open investigation, al-

Department of Biological Sciences, Union College, Schenectady, NY, USA. Email: fellerk@union.edu

that intracellular receptors engaged by DMT are involved in rapidly enhancing structural neuroplasticity, specifically increased dendritic growth and spine density, relative to membrane-localized receptors (see the figure). Bestowing cortical neurons in vivo with serotonin transporters allowed serotonin to access the intracellular receptors, mimicking the changes in structural neuroplasticity that are observed when cells were treated with DMT. The relevance of this phenomenon to changes in mouse behavior was assessed using a forced swim test (FST), a widely used approach that is thought to have predictive value of antidepressant effects in humans, although the actual clinical predictive validity is questionable (*11*). Providing serotonin access to intracellular 5-HT2ARs reduced immobility in the FST, which is considered an antidepressant-like effect. Given that serotonin cannot normally access intracellular 5-HT2ARs in cortical excitatory neurons owing to a lack of serotonin transporters, a role of endogenous DMT may be to engage the intracellular receptor pool and facilitate neuroplastic changes. Further assessment of the role of this mechanism in mediating circuit activity and resulting behaviors relevant to depression is necessary.

The lifetime prevalence of depression is ~20%, and it is a leading cause of disability worldwide. First-line treatments such as selective serotonin reuptake inhibitors are taken daily for weeks before therapeutic effects are potentially observed and lead to sustained remission in only ~30% of cases; they also have side effects that affect patient adherence (*12*). This has led to a shift in the search for depression therapeutics in favor of rapid-acting compounds such as ketamine, which can result in sustained therapeutic effects for days or longer after a single dose (*13*). Similarly, recent clinical trials have suggested that DMT and psilocybin are rapidly efficacious in treating major depression, with psilocybin appearing to have sustained therapeutic effects for weeks after treatment sessions that consist of a combination of psilocybin with supportive therapy (*2*, *3*).

Unfortunately, clinical trials with psychedelics are commonly statistically underpowered and mostly do not represent common patient populations. Moreover, it is impossible to adequately include a placebo control given the profound effects that these compounds have on perception, which can easily unblind both study participants and experimenters (*6*, *7*). As a result, it is difficult to predict how the existing clinical findings will be extrapolated to a wider population. Thus, psychedelic-assisted therapy should be embraced with cautious optimism; although the thousands of years of human psychedelic use and the positive outcomes of modern, though early stage, clinical trials point toward therapeutic potential, scaling psychedelic use to potentially millions of people with depression, for example, requires a careful and rigorous scientific approach.

The study conducted by Vargas *et al.* is a key achievement in the understanding of the mechanism of action of psychedelics. However, much needs to be done to affirm the sufficiency of signaling through intracellular 5-HT2ARs as mediators of the properties in humans, either hallucinogenic or medical, and to determine whether positive therapeutic outcomes can be separated from the classical psychedelic properties of such drugs. This will require interrogation of the specific signaling cascades engaged by intracellular 5-HT2ARs that lead to increased dendritic growth and spine density and defining the mechanistic relationship of such changes to cognitive processes. Moreover, understanding whether the findings for DMT are fully shared with all psychedelics (such as the structurally dissimilar mescaline), which was only partially explored by Vargas *et al.*, will be needed before the proposed mechanism can be described as a common psychedelic mechanism. Nevertheless, these findings are an important step forward for a rapidly expanding and much-needed field of study. ■

REFERENCES AND NOTES

1. D. E. Nichols, *Pharmacol. Rev.* **68**, 264 (2016).
2. A. K. Davis *et al.*, *JAMA Psychiatry* **78**, 481 (2021).
3. D. C. D'Souza *et al.*, *Neuropsychopharmacology* **47**, 1854 (2022).
4. A. K. Davis, L. A. Averill, N. D. Sepeda, J. P. Barsuglia, T. Amoroso, *Chronic Stress* **4**, 2470547020939564 (2020).
5. M. P. Bogenschutz *et al.*, *JAMA Psychiatry* **79**, 953 (2022).
6. J. Phelps, R. N. Shah, J. A. Lieberman, *JAMA Psychiatry* **79**, 189 (2022).
7. D. B. Yaden, J. B. Potash, R. R. Griffiths, *JAMA Psychiatry* **79**, 943 (2022).
8. M. V. Vargas *et al.*, *Science* **379**, 700 (2023).
9. J. G. Dean *et al.*, *Sci. Rep.* **9**, 9333 (2019).
10. E. Frecska, P. Bokor, M. Winkelman, *Front. Pharmacol.* **7**, 35 (2016).
11. D. A. Slattery, J. F. Cryan, *Nat. Protoc.* **7**, 1009 (2012).
12. A. J. Rush *et al.*, *Am. J. Psychiatry* **163**, 1905 (2006).
13. E. M. Hess, L. M. Riggs, M. Michaelides, T. D. Gould, *Biochem. Pharmacol.* **197**, 114892 (2022).

ACKNOWLEDGMENTS

T.D.G. received research funding from Allergan Pharmaceuticals over the past 3 years. He is listed as co-author on patents and patent applications related to the pharmacology and use of ketamine metabolites in the treatment of depression, anxiety, anhedonia, suicidal ideation, and posttraumatic stress disorders.

[1]Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA. [2]Department of Pharmacology, University of Maryland School of Medicine, Baltimore, MD, USA. [3]Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD, USA. [4]Baltimore Veterans Affairs Medical Center, Veterans Affairs Maryland Health Care System, Baltimore, MD, USA. Email: tgould@som.umaryland.edu

10.1126/science.adg2989

PHYSIOLOGY

Photonic tinkering in the open ocean

Light-manipulating materials are discovered in the eyeglitter of pelagic crustaceans

By **Kate Feller** *and* **Megan Porter**

Billions of animals living in open water, or pelagic habitats, can disappear into their surroundings using a variety of light-manipulating camouflage solutions. These include transparent, antireflection, and glittery reflective structures. Although such photonic camouflage allows these animals to vanish into their surroundings, they still need to eat (and avoid being eaten), which requires the ability to detect their invisible neighbors. Thus, an arms race exists between predators and prey for the ability to see and yet not be seen (*1*). Evolutionary tinkering across the diversity of pelagic animals has produced multiple solutions for controlling the transmission, reflection, and detection of light. On page 695 of this issue, Shavit *et al.* (*2*) report the discovery of photonic glass materials that form the basis of sparkly "eyeglitter" in the larvae of pelagic crustaceans and allows for both reflective camouflage and vision. These findings present a mechanism for producing salient, tunable coloration and light manipulation in space-limited tissues.

When peering into a bucket of seawater scooped from the ocean, the tiny animals and larvae that live in open water, or zooplankton, often can only be spotted by their pigmented eyes. These black dots travel side by side through the water, giving no indication that they are part of a much larger creature. This is because the bodies of many crustacean zooplankton are almost as transparent as glass. This is achieved both through lack of pigmentation in most body tissues and by photonically minimizing the amount of light scattered or reflected from internal and external structures. How internal structures such as muscles, organs, and blood become transparent is still an open investigation, al-

Department of Biological Sciences, Union College, Schenectady, NY, USA. Email: fellerk@union.edu

though internal transparency is sometimes an actively maintained, physiological process. For example, a loss of transparency can occur after repeated activity or stress and be regained after a brief rest (*3*).

In contrast to internal transparency, external transparency of pelagic crustacean cuticles (exoskeleton surface) is achieved with static physical structures. Tiny bumps, or nanoprotuberances, covering the animal's surfaces serve to minimize reflections or glare that would make them more noticeable. In just one group of pelagic arthropods, the hyperiid amphipods, up to four variations of these antireflecting structures have been identified (*3*). Because transparent camouflage is under strong selection, convergent evolution has likely devised additional photonic solutions that are yet to be described.

Although it is an effective camouflage, total body transparency is not possible in animals that require vision. For eyes to properly detect a visual scene, the light-sensing cells must be isolated from one another by dark pigments. This explains why the eyes of crustacean zooplankton are so easy to spot when viewed in a bucket of seawater. Larvae of several groups of crustaceans have independently evolved a solution to this dilemma: reflective camouflage. The dark bits of the eyes are enveloped with a sparkly eyeglitter—photonic structures that reflect light matched to the color of the surrounding water (*4*).

Shavit *et al.* delve into the chemical and structural basis of eyeglitter camouflage in larvae of the freshwater shrimp *Machrobrachium rosenbergi*. They found a previously unknown type of compact, photonic glass formed by nanospheres in the eyeglitter structure that are smaller than the wavelengths of visible light (400 to 750 nm). This newly discovered material is tunable and can reflect colors ranging from deep blue to yellow, depending on the camouflage needs of the animal at different depths and times of day. Even more intriguing, they found that the glittery eyeshine of two additional groups of crustacean larvae (decapod crabs and stomatopods, or mantis shrimp) appear to arise from similar nanosphere assemblies, although through the use of a different crystalline material that remains to be identified. This suggests that different lineages have evolved convergent mechanisms to cloak the dark retina from view. These discoveries highlight the vast potential of crustacean zooplankton for discovery of photonic materials that either reflect or transmit light in ways that can render objects invisible.

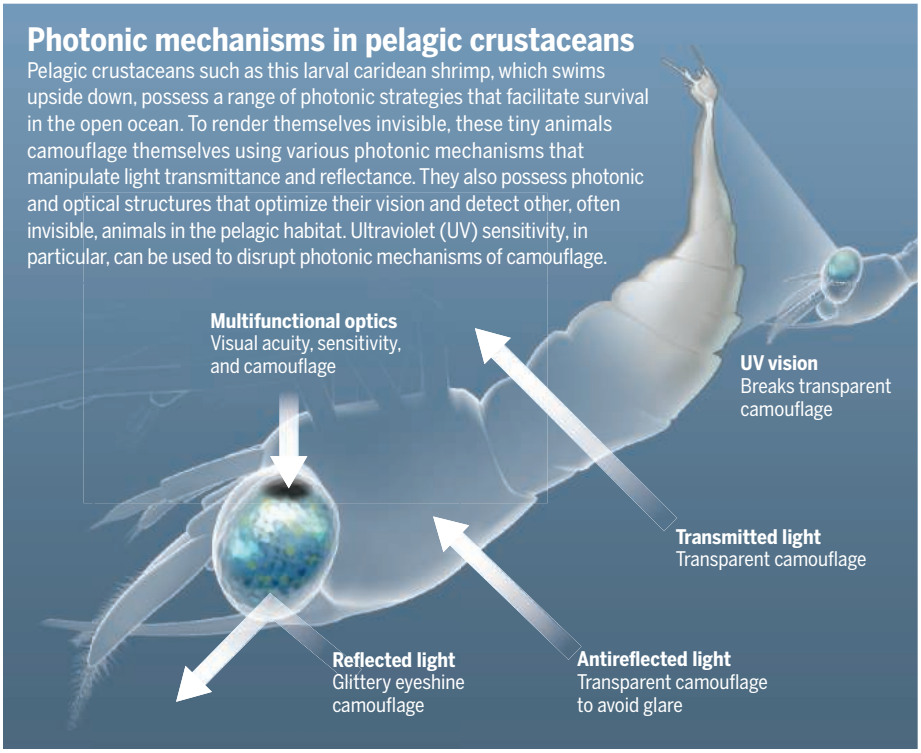
In the game of open-ocean hide and seek, animals also possess eyes that are specialized to meet their visual needs without compromising invisibility. In the study of Shavit *et al.*, the same reflective structures in *M. rosenbergi* used for eyeshine may play an additional role in enhancing visual sensitivity in the dark or isolating photoreceptors for improved resolution. Given their presence in other larvae, multifunctional reflective structures seem to be a common theme within such tiny eyes. For example, reflecting structures inside mantis shrimp larval photoreceptors filter light arriving at some photoreceptors, while reflecting light to enhance the sensitivity of other cells (*5*).

Another emerging theme in the visual ecology of small, pelagic crustaceans is the prevalence of ultraviolet (UV) sensitivity. UV light scatters more than visible wavelengths, even in the most transparent tissues. Thus, this scattering can make an animal appear as a shadow against the background of open water when viewed in the UV range, effectively breaking transparent or reflective camouflage (*1*). Although molecular, physiological, and anatomical studies of pelagic crustacean vision are still in their infancy, larvae of some mantis shrimp larvae (*Neogonodactylus oerstedii*) may take this invisibility-detecting strategy to the extreme because they have photoreceptors with the capability of detecting up to three distinct wavelengths of UV light (*6*). Further examination of open water crustacean visual systems may reveal innovative multifunctional optical solutions for both sensing and manipulating light, particularly UV wavelengths (see the figure).

Why are such exciting photonic discoveries only being made now? The vastness of the pelagic realm, and the relatively sparse distribution of animals within it, makes this habitat difficult to access and sample effectively (*1*). Larval crustaceans are numerous and easier to collect than larger pelagic species, yet they remain understudied owing to an historic lack of descriptions that link larval stages to the adult form for identification. The development of DNA barcoding to molecularly identify crustacean larvae has allowed species level comparisons like never before, especially for the study of biological photonics (*4*–*6*). The diversity of optical solutions discovered in just a few recently investigated pelagic crustaceans spotlights the untapped potential for photonic innovation yet to be found in animals living in the open ocean. By mimicking nature's solutions, humans can optimize and develop better photonic materials for solar energy, communications, remote sensing, and other light-dependent technologies. Given the diverse applications of photonics, studies of how these minute creatures both hide and seek in the open ocean is a burgeoning resource of bioinspiration for human detection, control, and manipulation of light. ■

Photonic mechanisms in pelagic crustaceans

Pelagic crustaceans such as this larval caridean shrimp, which swims upside down, possess a range of photonic strategies that facilitate survival in the open ocean. To render themselves invisible, these tiny animals camouflage themselves using various photonic mechanisms that manipulate light transmittance and reflectance. They also possess photonic and optical structures that optimize their vision and detect other, often invisible, animals in the pelagic habitat. Ultraviolet (UV) sensitivity, in particular, can be used to disrupt photonic mechanisms of camouflage.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. S. Johnsen, *Annu. Rev. Mar. Sci.* **6**, 369 (2014).
2. K. Shavit *et al.*, *Science* **379**, 695 (2023).
3. L. E. Bagge, *Integr. Comp. Biol.* **59**, 1653 (2019).
4. K. D. Feller, T. W. Cronin, *J. Exp. Biol.* **217**, 3263 (2014).
5. K. D. Feller *et al.*, *Curr. Biol.* **29**, 3101 (2019).
6. M. S. McDonald, S. Palecanda, J. H. Cohen, M. L. Porter, *J. Exp. Biol.* **225**, jeb243256 (2022).

10.1126/science.adf2062

RETROSPECTIVE

C. David Allis (1951–2023)

Champion of modern chromatin biology

By **Shelley L. Berger**

Molecular biologist C. David Allis, whose discoveries ushered in a new era of epigenetics research, died on 8 January. He was 71. Allis identified enzymes that chemically modify histones, transforming our understanding of the chromatin fiber. Composed of DNA and histone proteins, chromatin was previously thought to be an inert structure, but Allis showed that it dynamically regulates DNA processes. He led this expanding field of research with electrifying discoveries, extensive collaborations, inspiring mentorship, and a boundless passion for unearthing the biological impacts of histone biochemistry.

Born in Cincinnati, Ohio, Allis graduated with a bachelor's degree in biology in 1973 from the University of Cincinnati. He initially planned to go to medical school but became enthralled with research during his senior year in Michael Bharier's lab. Studying with Anthony Mahowald at Indiana University Bloomington, Allis attained his PhD in biology in 1978. During postdoctoral studies with Martin Gorovsky, Allis was exposed to chromatin biology. He took academic positions at Baylor College of Medicine in Houston, Syracuse University, the University of Rochester, and the University of Virginia as the impact of his work on chromatin grew. Moving institutions suited Allis's collaborative nature and his desire to seek out outstanding colleagues and trainees. In 2003, he was recruited to the Rockefeller University, where he remained for two decades.

Allis focused on connecting chromatin biochemistry with cellular function. His early chromatin research centered on the rarely studied organism *Tetrahymena*, a protozoan with two separate nuclei. Focused on the acetylation modification of the amino acid lysine, Allis showed that acetylation had distinct patterns on lysines of the histones isolated from the two nuclei, thereby linking histone modifications to divergent functional roles.

However, identifying modifying enzymes was difficult. To overcome this obstacle, Allis's lab invented a biochemical method to track such enzymes. In 1996, he purified the first nuclear histone-modifying enzyme, an acetylating enzyme. Serendipitously, this enzyme was a known gene regulatory factor recruited to genes to turn on expression—a model my lab had proposed, hence, my initial connection with Allis.

Meanwhile, another lab, led by Stuart Schreiber, discovered the first enzyme that erases acetylation, and it was a known gene repressor. These revelations deeply altered our understanding of DNA regulation in the nucleus. A model emerged: Gene activation (turning genes on) involves the association of a histone acetylation enzyme, and gene repression (turning genes off) involves histone deacetylation. Thus, the modern era of chromatin biology was born, with Allis at the vanguard, fully appreciating the implications.

PHOTO: ZACH VEILLEUX/THE ROCKEFELLER UNIVERSITY

Allis next turned his attention to histone methylation. In collaborative studies with Thomas Jenuwein, Allis uncovered the first methylation pathway, portended by findings in *Drosophila*. Certain mutants exhibit generation-to-generation eye color switching—a classic epigenetic-based toggle—and the new biochemical findings identified the methylation enzyme and the specific histone modification. Parallel to the earlier seminal findings of acetylation, this extension of histone modification to methylation spurred universal interest in histone biochemistry and chromatin function.

During the 2000s, as his lab and others were uncovering new modifications and their cognate enzymes, Allis would flash his crooked grin as he illustrated the field's progression with side-by-side pictures of an empty beach on the left and the same beach packed with midday crowds on the right. He thoroughly enjoyed the popularity and energy of this vibrant field that was once static, like the early models of passive chromatin. In a culmination of the bountiful histone modifications, Allis and various collaborators proposed a combinatorial histone "code," or language. He hypothesized that specific constellations of histone modifications signal different processes on the DNA—a controversial but galvanizing concept that triggered a new wave of experiments.

Throughout the 2010s, Allis continued with a steady stream of discoveries linking chromatin biology with disease, principally cancer, because many of the enzymes mutate in tumors. Proving the importance of single histone modifications, others discovered mutations in histones at some of the critical lysine amino acids in certain brain cancers. Allis immediately turned his attention to these so-called oncohistone mutations, and his lab generated provocative evidence that oncohistones were altering the activity of their cognate enzymes, leading to broad disruption of cell growth in cancer.

Allis was devoted to research and mentorship. He admired his own mentors' ability to fashion a stimulating and fun environment, and those early experiences informed his mentoring style. He famously had a whiteboard in his office that was filled with colorful writing documenting the intricacies of the histone tails and their enzyme "coders."

My lab group collaborated extensively with Allis, and we also worked in parallel tracks in those profoundly exciting early days. A disarming enthusiast, Allis proselytized the amazing properties of histones and deeply enjoyed the series of discoveries that followed. Phone calls with Allis to discuss collaborations were replete with his illuminating and passionate ideas. At conferences, when he described newly discovered histone enzymes and emerging paradigms, he would joyously exclaim, "Wow! I never expected that!" His talks made audience members wish they worked in his lab.

Allis was elected to the National Academy of Sciences in 2005 and was the recipient of the Albert Lasker Basic Medical Research Award in 2018, among numerous other accolades. The world has lost a visionary scientist, a caring mentor, and a tremendous spokesman for chromatin biology. ■

Departments of Cell and Developmental Biology, Biology, and Genetics; Epigenetics Institute; and Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: bergers@pennmedicine.upenn.edu

10.1126/science.adg7738

POLICY FORUM

ARTIFICIAL INTELLIGENCE

Leveraging IP for AI governance

Copyleft AI with Trusted Enforcement (CAITE) can support an adaptable soft law approach for ethics in AI

By **C. D. Schmit[1], M. J. Doerr[2], J. K. Wagner[3,4,5,6,7,8]**

The rapidly evolving and expanding use of artificial intelligence (AI) in all aspects of daily life is outpacing regulatory and policy efforts to guide its ethical use (*1*). Governmental inaction can be explained in part by the challenges that AI poses to traditional regulatory approaches (*1*). We propose the adaptation of existing legal frameworks and mechanisms to create a new and nuanced system of enforcement of ethics in AI models and training datasets. Our model leverages two radically different approaches to manage intellectual property (IP) rights. The first is copyleft licensing, which is traditionally used to enable widespread sharing of created content, including open-source software. The second is the "patent troll" model, which is often derided for suppressing technological development. Although diametric in isolation, these combined models enable the creation of a "troll for good" capable of enforcing the ethical use of AI training datasets and models.

The growing regulatory deficiency begets legitimate concerns that institutions are increasingly guilty of "blind reliance on AI," a tool capable of becoming "a high-tech pathway to discrimination" (*2*). Concerning uses of AI are many, including its commercial and Big Tech use; its use in employment, policing, and medicine; and its bias and discrimination, negatively affecting communities worldwide (*2*, *3*). Problems such as opacity, inaccuracy, inappropriate application, and lack of community engagement or meaningful consent are hallmarks of AI's current application. Some in industry have gone as far as to implore for effective AI regulation to curb the "existential threat" posed by unregulated AI (*4*).

Although governments across the globe are wrestling with these issues, AI policy efforts have been slow to materialize. At least 60 countries and territories have proposed at least 700 AI-focused policy initiatives (*5*). In the United States, 17 states introduced general AI bills or resolutions in 2021 (*6*), and at least 164 bills mentioning AI have been introduced in the 117th Congress alone (for example, S.3035 Good AI Act of 2021, S.3175 Advancing American AI Innovation Act of 2021, and H.R. 3723 Consumer Safety Technology Act). Despite this flurry of legislative activity, many only convened task forces or commissions to study the use of AI and did not provide guidance for its use. The Blueprint for an AI Bill of Rights issued by the White House Office of Science and Technology Policy (OSTP) is a laudable development in the United States but ultimately a nonbinding white paper (*3*). And although the risk-based framework of the European Union's proposed AI Act could become a de facto global approach, it has not yet been enacted (*7*).

Traditional regulatory approaches face many challenges with AI: Technology has rapidly outpaced oversight, there is considerable unpredictability about future AI applications, social values are dynamic, and anticipating future risks is challenging. After surveying the available governance options, Johnson and Bowman conclude, "AI will require a flexible and responsive governance framework to maximize its capacity to respond to shifting risks and problems" [(*1*), p. 72]. They argue that soft law approaches—such as nonbinding guidelines and recommendations, supported by multistakeholder bodies and transnational standards—provide the most promising institutions and instruments for near-term global AI governance capable of addressing these challenges.

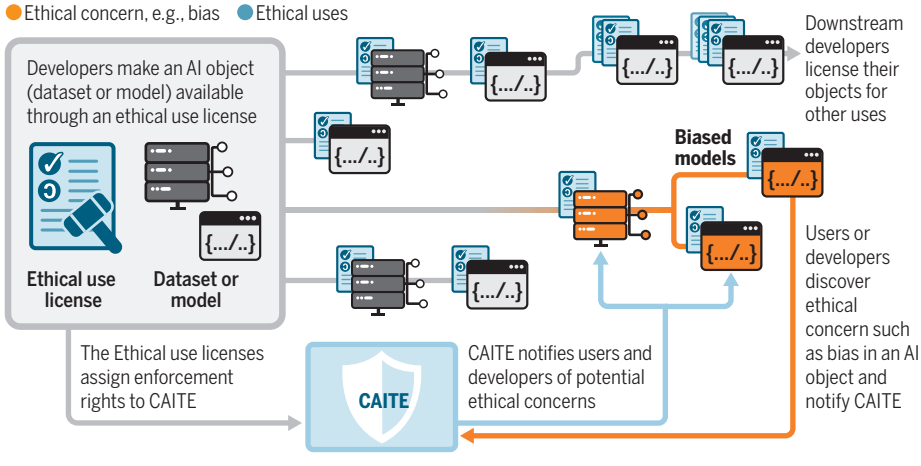
Ideally—and most daunting—AI governance should maximize technological benefits while minimizing risks. Stone *et al.* argue "[r]ather than 'more' or 'stricter' regulation, policies should be designed to encourage helpful innovation, generate and transfer expertise, and foster broad corporate and civic responsibility for addressing critical societal issues raised by these technologies" [(*8*), p. 42]. Traditional regulatory approaches (national laws or international treaties) are often too blunt, clumsy, and slow for this task. Increasingly, scholars have highlighted potential strengths of collaborative governance approaches (such as coregulation and responsive regulation) (*9*).

OBJECTIVE

The laws currently applicable to AI are insufficient to impel ethical conduct (*1*). Increasingly outdated laws—such as data privacy laws—favor the most well-resourced actors who can leverage technical and legal personnel to understand and comply with—

The CAITE governance model for ethical AI

Under a copyleft license, all derivative works of the original objects are bound by the terms and conditions of the ethical use license. By pooling enforcement authority in a single entity, Copyleft AI with Trusted Enforcement (CAITE) can facilitate a system of rapid notice of potentially harmful artificial intelligence (AI) applications and objects.



GRAPHIC: N. CARY/*SCIENCE*

[1]Department of Health Policy and Management, Texas A&M University, College Station, TX, USA. [2]Sage Bionetworks, Seattle, WA, USA. [3]School of Engineering Design and Innovation, Pennsylvania State University, University Park, PA, USA. [4]Institute for Computational and Data Sciences, Pennsylvania State University, University Park, PA, USA. [5]Department of Biomedical Engineering, Pennsylvania State University, University Park, PA, USA. [6]Rock Ethics Institute, Pennsylvania State University, University Park, PA, USA. [7]Penn State Law, University Park, PA, USA. [8]Huck Institutes for the Life Sciences, Pennsylvania State University, University Park, PA, USA. Email: schmit@tamu.edu

or circumvent—legal obligations. Moreover, legal obligations that fail to address AI risks could leave many well-meaning AI developers and users unaware of the potential harms to which their AI models cause or substantially contribute.

Our proposal aims to promote adherence at scale while providing flexibility as the ethical principles that guide appropriate AI develop and evolve (*10*). Similarly, our model incorporates a self-reporting mechanism combined with a sentinel alerting system to incentivize rapid reporting of unanticipated ethical issues, highlighting specific, contextual ethical issues to AI developers, users, and others. Such monitoring would instill confidence and trust in AI and clarify bases for good-faith AI use. Last, our proposal provides several mechanisms to allow an (eventually) self-sustaining system of enforcement and long-term governance. We call this model Copyleft AI with Trusted Enforcement (CAITE).

Copyleft licensing

Copyleft licensing is a way to facilitate sharing of copyrightable material under limited conditions. A creator who shares their work under a copyleft license allows anyone to use it to create derivative works as long as the new creations are subject to the same—or compatible—license as that of the original. For example, if the creator of X makes it available under a copyleft license, another person can use X to make X1, X2, and X3; however, X1, X2, and X3 must all be made available under the same license terms as the original work X. The capacity of copyleft licenses to permit derivative works to be licensed under updated licenses that are compatible with the original license and to phase out older license versions adds further flexibility to this system (*11*). Traditional copyleft approaches typically impose few use restrictions, which are usually limited to attribution or noncommerical use clauses. The new Responsible AI Licenses (RAIL) initiative, however, provides an example of efforts to use copyleft to ensure the ethical use of protected works (*12*). However, these traditional copyleft applications have limited enforcement potential because individual creators must enforce their license, making enforcement atomistic and vulnerable to David versus Goliath dynamics.

Nonpracticing entities and patent trolls

A nonpracticing entity is an organization that owns substantial IP rights but does not manufacture goods or otherwise commercialize their IP. Instead, a nonpracticing entity monetizes their IP rights through other activities, such as licensing. A "patent troll" is one type of nonpracticing entity that enforces their IP rights through private legal actions. Patent trolls are widely criticized because they typically do not manufacture or invent things; instead, patent trolls' aggressive legal maneuvers can discourage technological progress through legal threats. For example, P is a patent troll that purchases all the patents from a bankrupt technology company. P then sues another technology firm for allegedly using a technology covered by one of these patents without authorization. Nevertheless, patent trolls' ability to subsist on these enforcement actions makes them a promising vehicle for ensuring that AI is used ethically.

CAITE

Our approach to the ethical governance of AI uses a copyleft approach to create an enforcement "troll," leveraging an ethical use license defined by six core legal terms.

(i) Terms describing restrictions and conditions that would permit only ethical use of an AI model or training dataset. For example, these terms could require the promotion of equitable outcomes, informed consent fidelity, independent ethical review of AI applications, periodic algorithmic bias assessments, or public transparency.

(ii) Terms requiring that any derivative product of the AI training dataset or AI model be subject to the prior license's terms, including the ethical use limitations.

(iii) Terms assigning enforcement rights to a designated entity.

These terms accomplish two critical tasks. First, they create standards for ethical AI that perpetuate as licensed objects are reused. This first function is similar to other efforts to use licensing to promote responsible AI (such as the RAIL initiative) (*12*). Second, these terms pool enforcement rights in a single trusted entity, the CAITE host (see the figure). This second function creates a troll-like nonpracticing entity that exists to enforce the terms of the ethical use licenses. Collectively, however, ethical use licenses can create a quasi-governmental regulator capable of leveraging an aggregation of individual rights to further social objectives as a single and coherent regulator.

The ethical use license underlying the CAITE model provides a flexible vehicle to enhance AI governance with additional license terms.

(iv) Terms requiring adherence to a specific code of conduct.

(v) Terms requiring developers to register licensed AI objects with the CAITE host. Such registration would have the dual function of notifying the host of their new designated enforcement authority and enabling the creation of a library of AI objects available under ethical use licenses.

(vi) Terms imposing a duty to report ethical issues or violations of the code of conduct.

These additional terms allow the CAITE host to act as a quasi-government regulator, identifying and removing dangerous AI objects from use and encouraging improvements through transparent corrective action plans. The CAITE host could administer a system analogous to the Fast Track Product Recall Program of the US Consumer Product Safety Commission, which helps remove potentially hazardous products from the marketplace in a timely fashion (*13*). A similar recall mechanism for the CAITE host could be mutually beneficial for AI developers (by demonstrating bona fide corrective steps to avoid severe penalties) and regulatory agencies (by supplementing existing oversight capacity). In this way, the CAITE model could become a mechanism for ensuring meaningful AI governance through community partnership to promptly and effectively identify problems, notify stakeholders, and establish a pathway for corrective action.

THE CAITE HOST

The implementation of CAITE hinges on a trusted entity fulfilling administrative and enforcement responsibilities. We propose a community-designated, nongovernment host composed of delegated proxies of AI developers and users. The host could also include community members affected by AI applications to deepen and sharpen the oversight focus of the group and engender broader trust in CAITE as an oversight system. Notably, the CAITE host's authority and long-term relevance depends on the ongoing voluntary use of licenses that assign enforcement rights to the host. Thus, there are inherent incentives within the CAITE system to ensure that the host is fulfilling the community's needs for ethical oversight. Moreover, our model envisions the AI community as having a central role in working with the host to develop the code of conduct, revising ethical use licenses to meet evolving needs, and determining self-reporting norms upon discovery of any substantial flaw, defect, bias, or other attribute conferring risk of harm.

COMPLIANCE AND ENFORCEMENT TAILORED TO CONDUCT

The CAITE model's license approach enables substantial enforcement flexibility for a range of compliance issues and can accommodate different enforcement options. License terms can describe consequences for particular conduct or processes for conflict resolution (arbitration). Critically, license terms, enforced through the CAITE host, can also establish conditions for leniency, enable restorative justice remedies,

or define safe harbors from penalties. For example, the host could incentivize due diligence before release of AI objects and prompt self-reporting of issues by defining conditions for safe harbor from penalties. Similarly, the host can provide safe harbor to AI users and developers who act promptly in response to a notice of concern pertaining to one of their AI objects. The host could provide advisory opinions and guidance to communicate expectations and guide ethical conduct within the AI community. These incentives could improve quality, safety, and effectiveness of AI and promote confidence and trust in AI.

Conversely, if a user of an AI object subject to an ethical use license subsequently does not adhere to the reporting requirements or take reasonable, appropriate corrective actions once problems have been discovered, the host would have discretion to act on penalty clauses included within the contract to compel compliance (for example, through court orders for monetary penalties and injunctive relief or settlements) and to report wrongdoing to government authorities able to enforce requirements that exist in law to protect consumers. Whistleblower protections and incentives (such as Qui Tam provisions) could also be leveraged to encourage those with knowledge of wrongdoing to come forward. Through these various enforcement options, the ethical use licenses' assignment of enforcement rights enables the CAITE host to operate with case-by-case discretion as a traditional governmental enforcement agency might.

POTENTIAL BENEFITS AND RISKS

Benefits of the CAITE model are diverse and far reaching, especially in comparison with the current regulatory void. The CAITE model is agnostic; it can be designed to implement any ethical AI guidance or set of principles. Further, just as Creative Commons licenses periodically update to address new issues with new versions, so too could CAITE. Because this is a scalable, bottom-up governance solution, the CAITE model is feasible for broad implementation and customizable for narrower, more specific contexts with particular, nuanced expectations for ethical conduct. Another benefit is that the CAITE model could nurture trust without excessive regulatory burdens.

If implemented at scale, the CAITE model would have increased benefits. CAITE could reduce AI entry costs by making AI models and training datasets widely available—provided that users comply with the ethical use license—which addresses a key barrier posed by traditional copyright protections (*14*). Moreover, an ever-increasing body of ethical use license–bound AI objects could make it increasingly disadvantageous for entities to practically opt out of compliance because opting out would require recreating and re-engineering equivalent datasets and models. Additionally, CAITE could become self-sustaining through figurative and literal "buy in" from AI developers and users. For example, AI developers could, in exchange for a modest fee, market their objects with "certificates of conformity" or other symbolic indicators of their adherence or commitment to this governance framework. Users could gain confidence by relying on AI objects with such "certificates of conformity" for a similar modest fee. These fees—combined with reinvestment of settlement funds or penalties collected after "enforcement" actions—could offset administrative costs of oversight. Notably, a potential risk or area of uncertainty with such a system is that if it is not carefully designed, it could have a chilling effect on the AI development ecosystem (for example, fear of punishment, or fees as barriers). Initiating the CAITE model will require seed funding for piloting and iterative improvements, including to assess self-sustaining design elements.

Soft-governance approaches often supplement traditional government oversight. Frequently, government agencies use adherence to soft-governance rules when making enforcement discretion decisions, being more willing to "drop the hammer" on actors unwilling to conform with emergent ethical AI best practices and expectations for professional responsibilities. In this way, CAITE could enable collaborative governance approaches such as coregulation and responsive regulation (*9*). For example, governments could require or incentivize entities to opt in to the CAITE model through traditional regulatory approaches in a coregulatory approach.

Ultimately, the CAITE model requires sufficient participation from the AI community and industry to see real-world impact. We have outlined many incentives (or coregulatory mandates) that could be used to achieve sufficient participation. The consequences of inaction, however, underlie all this. Industry is not naïve to existential risks posed by unregulated AI (*2*–*4*). Voluntary adoption of a community-driven, soft-law governance approach—such as CAITE—is likely preferable to industry members than the likely inevitable alternative: a blunt, inflexible, and stringent traditional regulatory framework (*1*, *8*, *9*).

DEVELOPING THIS VISION

As others have highlighted for emerging biotechnologies (*15*), existing legal mechanisms provide a rich and diverse a la carte menu of possibilities for different self-regulatory carrots and sticks to encourage ethical adherence in AI, which otherwise exists in a regulatory lacuna. To realize CAITE's potential, additional foundational research, and funding support for such research, is needed to advance our understanding of ethical AI standards, restorative justice practices in ethical AI, and detection or prediction mechanisms not only for unethical AI use but also harm from AI use. Moreover, a pilot implementation of this model would help to determine the required resources, enforcement costs, fees, penalties, and incentives that would be necessary to meet sustainability goals. Such efforts should involve a diverse and interdisciplinary working group of subject matter experts, practitioners, and community members to address challenges to implementation; iteratively examine, modify, and select features that could be integrated into the CAITE model; and synthesize, frame, and pilot CAITE with ready partners. ■

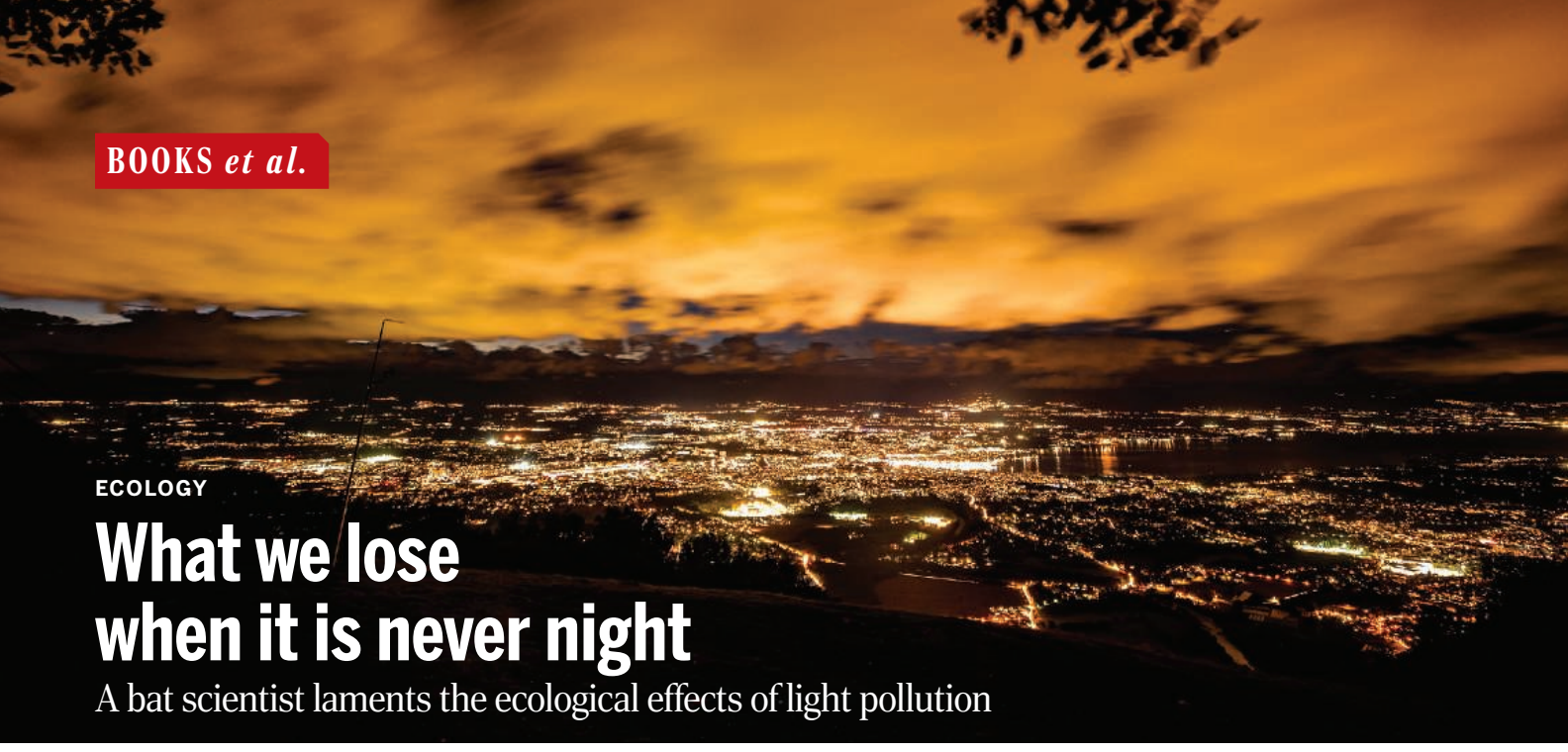
REFERENCES AND NOTES

1. W. G. Johnson, D. M. Bowman, *IEEE Technol. Soc. Mag.* **40**, 68 (2021).
2. K. Johnson, Feds warn employers against discriminatory algorithms, *Wired* 16 May 2022; https://www.wired.com/story/ai-hiring-bias-doj-eocc-guidance.
3. The White House, "Blueprint for an AI bill of rights: making automated systems work for the american people" (OSTP, 2022).
4. C. Domonoske, "Elon Musk warns governors: Artificial intelligence poses existential risk," National Public Radio, 17 July 2017; https://www.npr.org/sections/thetwo-way/2017/07/17/537686649/elon-musk-warns-governors-artificial-intelligence-poses-existential-risk.
5. OECD.AI, National AI policies and strategies (2021); https://oecd.ai/en/dashboards.
6. National Conference of State Legislatures, "Legislation related to artificial intelligence," 5 January 2022; https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx.
7. Artificial Intelligence Act, European Commission, 21 April 2021.
8. P. Stone *et al*., "One hundred year study on artificial intelligence: Report of the 2015–2016 study panel" (Stanford Univ., 2016); http://ai100.stanford.edu/2016-report.
9. M. E. Kaminsky, *S. Cal. L. Rev.* **92**, 1529 (2019).
10. A. Jobin, M. Ienca, E. Vayena, *Nat. Mach. Intell.* **1**, 389 (2019).
11. Creative Commons, Attribution-ShareAlike 4.0 International — CC BY-SA 4.0.
12. D. Contractor *et al*., 2022 ACM Conf. Fair. Account. Transp. 10.1145/3531146.3533143 (2022).
13. D. H. Carpenter, "The Consumer Product Safety Act: A Legal Analysis (R45174, Version 3), (Congressional Research Service, updated 24 April 2018; https://crsreports.congress.gov/product/pdf/R/R45174.
14. S. M. Fiil-Flynn *et al*., *Science* **378**, 951 (2022).
15. C. J. Guerrini, M. A. Curnutte, J. S. Sherkow, C. T. Scott, *Nat. Biotechnol.* **35**, 22 (2017).

ACKNOWLEDGMENTS

The authors thank the National Institutes of Health Office of Data Science Strategy for hosting the Innovation Lab, "A Data Ecosystems Approach to Ethical AI for Biomedical and Behavioral Research," on 14 to 18 March 2022, which stimulated early collaborative conversations. The authors thank participants at that Innovation Lab who provided constructive feedback and encouragement.

10.1126/science.add2202



ECOLOGY

What we lose when it is never night

A bat scientist laments the ecological effects of light pollution

By **Christopher Kemp**

One summer midnight several years ago, standing outside a wooden cabin in Michigan's Upper Peninsula, I looked up. It was a revelation. The sky was filled with thousands of stars. There were constellations I had never seen before, except in books. The spiral arm of our home galaxy, the Milky Way, unfurled across the sky. The sight of thousands of stars was almost enough to make me, a nonbeliever, offer a word of gratitude up into the star-filled sky. But to whom? Perhaps to Johan Eklöf, author of *The Darkness Manifesto*.

A bat scientist, Eklöf works in the night shadows in Bollebygd in western Sweden. His work on bats requires a specific kind of darkness—the absolute kind, unpolluted by light. But this category of darkness is threatened. In the 1980s, Eklöf tells readers, two-thirds of the churches in Sweden's southwest housed bat colonies. Not any longer. "Today, forty years later, research I've done with my colleagues shows that this number has been reduced by a third due to light pollution and other factors. Because the churches all glow like carnivals in the night," he writes. "We are blasting ourselves with light."

We have all noticed it when driving through a city—any city—at night. Empty places are floodlit. Sometimes, on a quiet night, you can even hear the light. How bright must light be to hear the processes that make it?

The reviewer is at the Department of Translational Neuroscience, Michigan State University, Grand Rapids, MI 49503, USA, and is the author of *Dark and Magical Places: The Neuroscience of Navigation* (Norton, 2022). Email: cjkemp@gmail.com

Inhabitants of Hong Kong live beneath a night sky that is 1200 times brighter than an unlit sky. Citizens of some large cities, writes Eklöf, have never allowed their eyes to adapt to true night vision (it takes half an hour of complete darkness for our eyes to make the shift). We are only now beginning to understand the effects of all this light.

A third of vertebrates and two-thirds of invertebrates are nocturnal. Excess light is incredibly destructive to the complex ecosystems these animals inhabit. It scares away the bats that Eklöf studies. It stuns light-sensitive moths, leaving them vulnerable to predation or flying endlessly into porch lights that will never return their love. Newly hatched turtles crawl away from the shoreline toward the lights of distant coastal cities, corals no longer mate on schedule, reef fish eggs go unhatched. Birds do not migrate on time—they forget to even sing. Modern advancements such as LED lights could significantly reduce some of the worst impacts, but they have not. At least, not yet.

Eklöf's excellent book covers a lot of ground. It shaped and sharpened my thinking on a subject I knew I was supposed to feel something about. In the week after I read it, I found myself saying, again and again, at the dinner table (six bulbs above it where one would suffice), "Yes, children, but did you know…?" I can think of no better praise than that.

It is worth mentioning that middle-aged white male writers like Eklöf can advocate for a darker world—for darkened campuses and unlit parking lots. And middle-aged white male book reviewers can agree with him. But darkness is not safe for everyone. We need to address the social issues that make lighted places so appealing in the first place.

The bottom line: We can change if we want to. Parts of Denmark have been designated "Dark Sky Communities." In 2019, France passed legislation limiting how much light can be emitted into the sky. In Vienna, Austria, the city's lights are turned off at 11 p.m. In contrast, China has announced plans to launch several moonlike satellites that will reflect the Sun's light after dark. These new moons will bathe cities in even more light.

Some of the solutions to light pollution—motion-detecting lights, shielded lights that do not reflect light upward, artificial light with wavelengths that mimic natural light—are already within our grasp, if we just reach for them. "We could just turn it all off, but I guess we don't want to," said Eklöf in a recent interview. "So it's vital we find a middle way."

Right now, it is hard to know what that middle way might look like. In 50 years, every city could be lit by an array of programmed and environmentally low-impact LED lights, or we might have completely forgotten what darkness is—the sky filled with little moons. ■

10.1126/science.adg2297

The Darkness Manifesto
Johan Eklöf, translated by Elizabeth DeNoma
Scribner, 2023. 272 pp.

Artificial light illuminates the skies above Geneva, Switzerland, on 24 September 2019.

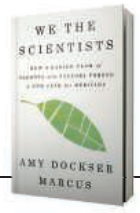
PHOTO: FABRICE COFFRINI/AFP VIA GETTY IMAGES

MEDICINE

Science and its stakeholders

Patients' families became research partners in a quest to understand a rare genetic disease

We the Scientists: How a Daring Team of Parents and Doctors Forged a New Path for Medicine
Amy Dockser Marcus
Riverhead Books, 2023. 256 pp.



By **Christi J. Guerrini**

For centuries, science was largely conducted by amateurs with an interest in a topic and ample time and means to study it. Beginning in the late 1800s, however, rigorous academic programs and professional standards and guidelines were established that expanded scientific knowledge production while limiting participation in it. The problem, writes Pulitzer Prize–winning journalist Amy Dockser Marcus in *We the Scientists*, is that these efforts "eventually went too far": Scientists became separated from people with personal connections to the problems under investigation, and these people's interests became secondary to other agendas pursued behind closed doors. The book describes what happens when nonscientists convince scientists to let them back in.

An absorbing narrative spanning 10 years, *We the Scientists* follows a group of US-based parents as they race to save the lives of their children diagnosed with Niemann-Pick disease type C (NPC). NPC is a rare genetic condition characterized by the inability to transport cholesterol inside cells, which causes neurological problems that are eventually fatal, usually between the ages of 10 and 25.

The gene responsible for most cases of NPC was identified in the 1990s. However, there is no cure or treatment approved for the disease. This is in part due to its rarity; the small market makes it difficult for companies to justify investing the substantial resources needed to develop, test, and obtain approval for novel interventions. In addition, the symptoms and progression of NPC can vary considerably from person to person, making it difficult to identify outcomes one might measure to demonstrate a treatment's efficacy.

Whereas the conventional approach to drug development is to sideline patients until the clinical trial stage, Dockser Marcus recounts how a handful of NPC families and researchers partnered at the beginning of the process to identify drug leads for further study and eventual development. Described as "a remarkable social experiment," the partnership was founded on the belief that the parents' involvement as equals would bring a needed sense of urgency and focus as well as a valuable and complementary expertise to the process.

The partnership launched in 2007 when more than a dozen parents and leading scientists on NPC convened at the National Institutes of Health (NIH) at the invitation of Chris Austin, an NIH scientist who had built a robotic system to screen pharmaceutical compounds for therapeutic potential. At the meeting, the group decided to use the system to test thousands of drugs on tissue samples from NPC patients with the goal of generating leads and testing the most promising ones in clinical trials within 5 years—a major compression of the usual drug development timeline.

The Hempels partnered with researchers to help their daughters, Addi and Cassi.

PHOTO: HUGH AND CHRIS HEMPEL

A conflict eventually arose over how to move forward with one promising drug, cyclodextrin. When the scientists were slow to test it in a clinical trial because of safety concerns, some families took matters into their own hands and obtained permission from the US Food and Drug Administration (FDA) to try the drug by way of compassionate use requests, a move that risked the success of future trials and eventual FDA approval. The families also confronted the FDA about the value of real-world observations and experiences in an effort to secure approval of a different drug. Their efforts raise fundamental questions about what counts as scientific evidence, how much risk is too much to bear and who gets to decide this, and how to navigate the life-and-death interests of current versus future patients.

Frustration is a major theme throughout the book. For both patients with NPC and those caring for them, managing symptoms can feel like a never-ending game of medical whack-a-mole. Navigating regulatory systems that are inherently slow and not designed for patient engagement is no less frustrating. The same, too, with reconciling the different primary commitments of stakeholders and ensuring that all data collected in the care of NPC patients are put to use.

Still, Dockser Marcus is careful not to cast any person, entity, or industry as the bad guy in this story. Acts of kindness are emphasized. There is the researcher's spouse who bakes chocolate chip cookies for a young patient every time he travels to the NIH for cyclodextrin transfusions. There is the pharmaceutical executive who, after learning of a family's blog post pleading for access to proprietary data to support a compassionate use request, wakes them with a phone call offering to help. (Access is provided within days.) There is the FDA official who takes seriously the mother determined to obtain an orphan drug designation for cyclodextrin—and does not blink when she submits the application in a pink binder, her sick daughters' favorite color. (The designation is approved 3 months later.)

Through details of perseverance, courage, and grace, frustration is ultimately eclipsed in *We the Scientists* by another theme. It is the name for the NIH drug discovery robots proposed by one of the parents: Hope. ■

The reviewer is at the Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA. Email: guerrini@bcm.edu

10.1126/science.adg5240

PHOTO: XIHUI WU

Agriculture in China's dry regions, such as semi-arid Gansu province, undermines water security.

Edited by **Jennifer Sills**

Farming in arid areas depletes China's water

In recent years, agriculture in China's arid areas has developed rapidly (*1*). The additional crops have improved China's food security (*2*), but they also require irrigation, increasing water consumption (*3*) and exacerbating water shortages. Therefore, China must urgently address water security in its arid regions.

From 2000 to 2021, agricultural water consumption accounted for more than 80% of total water consumption each year in China's arid areas, and agricultural water consumption increased by 1.47 billion m^3 during that time (*4*). Groundwater, the main water source for agricultural production in these regions (*5*), decreased by more than 11.08 billion m^3 (*4*), resulting in large-scale aquifer depletion (*5*). Irrigation water consumption per unit area of cropland in arid areas is about five times the national average, but the yield produced per unit of irrigation water use on cultivated land is only 15% of the national average (*6*), indicating that the benefits of farming do not outweigh the risks to water security in dry climates. Furthermore, global climate change and extreme drought events in arid areas intensify water shortages (*7*).

To alleviate water resource pressure, China should limit agricultural expansion and irrigation that depends on groundwater in arid areas (*8*). Water resources should be regulated through pricing—higher water prices will incentivize farmers to save water and optimize the planting area of different crops. In addition, China's arid regions should import agricultural products that require high water consumption from areas with abundant water resources to reduce the need for local planting (*9*). Given the geographic differences in China's arid areas, the government should guide farmers to choose crops based on regional heterogeneity and promote new varieties of drought-tolerant crops. Finally, China should establish sustainable agricultural irrigation water-saving systems (*10*) such as precision irrigation technology to improve irrigation water use efficiency.

Yongyong Song[1]*, Dongqian Xue[1], Beibei Ma[1], Siyou Xia[2,3], Hao Ye[1]
[1]School of Geography and Tourism, Shaanxi Normal University, Xi'an, China. [2]Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. [3]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China.
*Corresponding author.
Email: syy2016@snnu.edu.cn

REFERENCES AND NOTES

1. National Bureau of Statistics, "China statistical yearbook" (2022); https://navi.cnki.net/knavi/yearbooks/YINFN/detail?uniplatform=NZKPT&language=chs [in Chinese].
2. X. X. Q. Han *et al.*, *South-to-North Wat. Transfers Wat. Sci. Technol.* **18**, 82 (2020) [in Chinese].
3. X. X. Qi *et al.*, *One Earth* **5**, 1139 (2022).
4. Ministry of Water Resources of the People's Republic of China, "China water resources statistics bulletin 2000–2021" (2021) [in Chinese].
5. W. K. Wang *et al.*, *Hydrogeol. J.* **29**, 521 (2021).
6. W. Xiang *et al.*, *Environ. Impact. Assess.* **97**, 106904 (2022).
7. X. Lian *et al.*, *Nat. Rev. Earth Environ.* **2**, 232 (2021).
8. Y. Guang *et al.*, *Sci. Tot. Environ.* **691**, 506 (2019).
9. T. L. An *et al.*, *J. Clean Prod*, **288**, 125670 (2021).
10. J. P. Rafael *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2214291119 (2022).

10.1126/science.adg4780

A shift to English in Algerian education

In September, Algeria will change the language used in higher education from French to English (*1*). If successful, the new policy could facilitate access to English literature and promote international exchange and collaboration. However, Algeria's plan to implement the change immediately will undermine its benefits.

In Algeria, primary and secondary education are taught in Arabic with some linguistic courses in French and English (*2*). Historically and culturally, Algeria has been more closely tied to France than to any English-speaking country (*3*). As a result, French fluency is common (*4*), whereas English proficiency is limited. In 2022, Education First ranked Algeria 78 out of 111 countries in English proficiency (*5*).

Switching to English will likely decrease the quality of teaching, evaluation, and learning (*6*), reducing the efficiency of graduated students in the workforce. Students will experience more frustration and anxiety (*7*) and will require more time to finish writing assignments, including theses (*8*). Professors may find learning English challenging and look for other jobs as a result.

The change will also disrupt established programs. The demography of international students, who have typically come from French-speaking African countries, will change, affecting historical strategic partnerships. Updating libraries with English versions of the French and Arabic textbooks that have been accumulated over decades will require substantial financial, logistical, and staff resources, which could otherwise be directed toward research or equipment.

Intensive English training for professors (*9*) started in December 2022, but less than a year of language courses will likely be insufficient to ensure effective teaching when the policy takes effect. No training will be provided to students. As a result, private English training will likely increase, deepening the gap between students from low- and high-income families (*10*).

Before full implementation of the language policy change at the national level, pilot studies should be conducted. For example, select universities could implement a gradual shift to English coursework. Institutions could continue to teach undergraduate courses in French or Arabic, and intense mandatory training in English could be added to the requirements. After

at least 2 years of English study, students could switch to a partial or a full course load in English. Throughout the process, evaluations could determine the effects of the program on students, professors, and staff, as well as the resources needed to provide materials and support.

Shifting higher education to English could have benefits in the long term. However, the change should be informed by both pilot studies and the experiences of other countries. Rwanda made a similar language transition 15 years ago (*11*) and both teachers and students in Rwanda's universities continue to face challenges (*12*). Algeria should implement the language shift gradually to minimize disruption to its education system.

Rassim Khelifa
Department of Biology, Concordia University, Montreal, QC, Canada.
Email: rassim.khelifa@concordia.ca

REFERENCES AND NOTES

1. M. Zerrouky, "Remplacer le français par l'anglais à l'université? Polémique linguistique en Algérie" *Le Monde* (2019) [in French].
2. K. Taleb Ibrahimi, *Revue Int. Educ. Sèvres* **70**, 53 (2015) [in French].
3. M. Mohammed, M. Brahim, in *Towards an Arab Higher Education Space: International Challenges and Societal Responsibilities*, B. Lamine, Ed. (UNESCO Regional Bureau for Education in the Arab States, Beirut, 2010), pp. 267–280.
4. F. Aitsiselmi, D. Marley, in *Studies in French Applied Linguistics: Language Learning & Language Teaching*, D. Ayoun, Ed. (John Benjamins Pub Co., Philadelphia, PA, 2008), vol. 6, pp. 185–222.
5. EF English Proficiency Index, "The world's largest ranking of countries and regions by English skills" (2022); www.ef.com/epi.
6. O. Effiong, *TESOL J.* **7**, 132 (2016).
7. L. P. F. Ma, *High. Educ. Res. Dev.* **40**, 1176 (2021).
8. V. Ramírez-Castañeda, *PLOS One 15*, e0238372 (2020).
9. A. Boumaza, "Programme de formation en anglais pour les professeurs d'université," Algérie 360° (2022) [in French].
10. S. Mazzella, *La Mondialisation Étudiante: Le Maghreb entre Nord et Sud* (Karthala Editions, 2009) [in French].
11. P. Borg, *J. Gifu Keizai Univ.* **48**, 65 (2015).
12. T. Kral, *Glob. Soc. Educ.* 10.1080/14767724.2022.2033613 (2022).

10.1126/science.adg3508

Ethical and legal wastewater surveillance

As researchers on molecular methods of wastewater analysis (*1*, *2*), we agree with J. I. Levy *et al.* ("Wastewater surveillance for public health," Perspectives, 6 January, p. 26) that community-level monitoring can be an efficient way of detecting new outbreaks of disease and activating or prioritizing local public health actions. This approach to monitoring also serves to assess the effectiveness of mitigation and includes communities with minimal individual testing. However, Levy *et al.* say little about the ethical or legal considerations that must be considered when expanding wastewater surveillance to new targets or communities. Although wastewater monitoring itself is not new, recent pandemic-stimulated growth in monitoring infrastructure and personnel, technological innovations, and proposals for wider application have made the need for ethical review and oversight urgent. Given that interest in and applications of wastewater surveillance continue to grow, the scientific community and government officials have an obligation to use the technology ethically and legally, ensuring that personal data remains private and vulnerable groups are protected (*2*, *3*).

When deciding which pathogens (or other targets) and sewersheds to surveil, researchers and public health officials must minimize the possibility of stigmatizing communities identified as suffering from an outbreak. In addition, they must commit to making use of actionable information for public health purposes. Furthermore, policymakers must ensure that wastewater surveillance infrastructure and data are not used for other types of surveillance without appropriate deliberation and transparency. For example, guardrails should be put in place to prevent wastewater surveillance infrastructure and data from being shared with law enforcement (*3*).

Wastewater monitoring typically preserves individual anonymity, but reports of infections among groups has the potential to stigmatize entire neighborhoods or cultural groups (*4*, *5*). In one case, COVID-19 dormitory wastewater monitoring led to a testing requirement (*6*). Is detection in a neighborhood sufficient cause to test all individuals in the sewershed?

Such questions must be answered before wastewater surveillance expands further. As Levy *et al.* observe, wastewater can identify more than just infectious diseases. Indeed, government-sponsored wastewater surveillance projects are already underway or in the research stage to monitor wastewater for opioids and cannabinoids (*7*), stress hormones (a proxy measure of community mental health) (*8*), and asthma medications (*9*). A private company selling wastewater surveillance services has indicated that it has monitored wastewater for meat products at a school with a strict vegetarian policy (*10*).

Researchers, policymakers, and regulators must work together to develop and plan the wastewater monitoring campaigns of the future. Policies must include plans for the ethical application of the data, criteria for their continuation or termination, and restrictions on additional applications to which the established infrastructure and data can be applied. Regulations should also be put in place to control the use of wastewater surveillance in the private sector.

Natalie Ram[1], William Shuster[2], Lance Gable[3], Jeffrey L. Ram[4]*
[1]Carey School of Law, University of Maryland, Baltimore, MD, USA. [2]Civil and Environmental Engineering, Wayne State University, Detroit, MI, USA. [3]Wayne State University Law School, Detroit, MI, USA. [4]Department of Physiology, Wayne State University School of Medicine, Detroit, MI, USA.
*Corresponding author. Email: jeffram@wayne.edu

REFERENCES AND NOTES

1. N. W. West *et al.*, *Sci. Tot. Environ.* **847**, 1 (2022).
2. N. Ram, L. Gable, J. L. Ram, *Univ. Richmond Law Rev.* **56**, 911 (2022).
3. National Academies, "Community wastewater-based infectious disease surveillance" (2023).
4. C. Smith-Morris, *Soc. Sci. Med.* **179**, 106 (2017).
5. S. Bhatnagar *et al.*, *Indian J. Psychol. Med.* **43**, 428 (2021).
6. J. Kaiser, "Poop tests stop COVID-19 outbreak at University of Arizona," *Science* (2020).
7. N. Sims, B. Kasprzyk-Hordern, *Environ. Int.* **139**, e105689 (2020).
8. E. M. Driver, A. J. Gushgari, J. C. Steele, D. A. Bowes, R. U. Halden, *Sci. Tot. Environ.* **838**, 155961 (2022).
9. E. Fattore *et al.*, *Environ. Res.* **150**, 106 (2016).
10. M. Molteni, "Sewage sleuths helped an Arizona town beat back Covid-19. For wastewater epidemiology, that's just the start," *STAT News* (2021).

10.1126/science.adg7147

TECHNICAL COMMENT ABSTRACTS

Comment on "Sexual selection promotes giraffoid head-neck evolution and ecological adaptation"

Sukuan Hou, Qinqin Shi, Michael J. Benton, Nikos Solounias

Wang *et al*. (Research Articles, 3 June 2022, eabl8316) reported an early Miocene giraffoid that exhibited fierce head-butting behavior and concluded that sexual selection promoted head-neck evolution in giraffoids. However, we argue that this ruminant is not a giraffoid and thus that the hypothesis that sexual selection promoted giraffoid head-neck evolution is not sufficiently supported.

Full text: dx.doi.org/10.1126/science.add9559

Response to comment on "Sexual selection promotes giraffoid head-neck evolution and ecological adaptation"

Shi-Qi Wang, Jin Meng, Bastien Mennecart, Loïc Costeur, Jie Ye, Chunxiao Li, Chi Zhang, Ji Zhang, Manuela Aiglstorfer, Yang Wang, Yan Wu, Wen-Yu Wu, Tao Deng

Hou *et al*. challenged the giraffoid affinity of Discokeryx and its ecology and behavior. In our response we reiterate that Discokeryx is a giraffoid that, along with Giraffa, shows extreme evolution of head-neck morphologies that were presumably shaped by selective pressure from sexual competition and marginal environments.

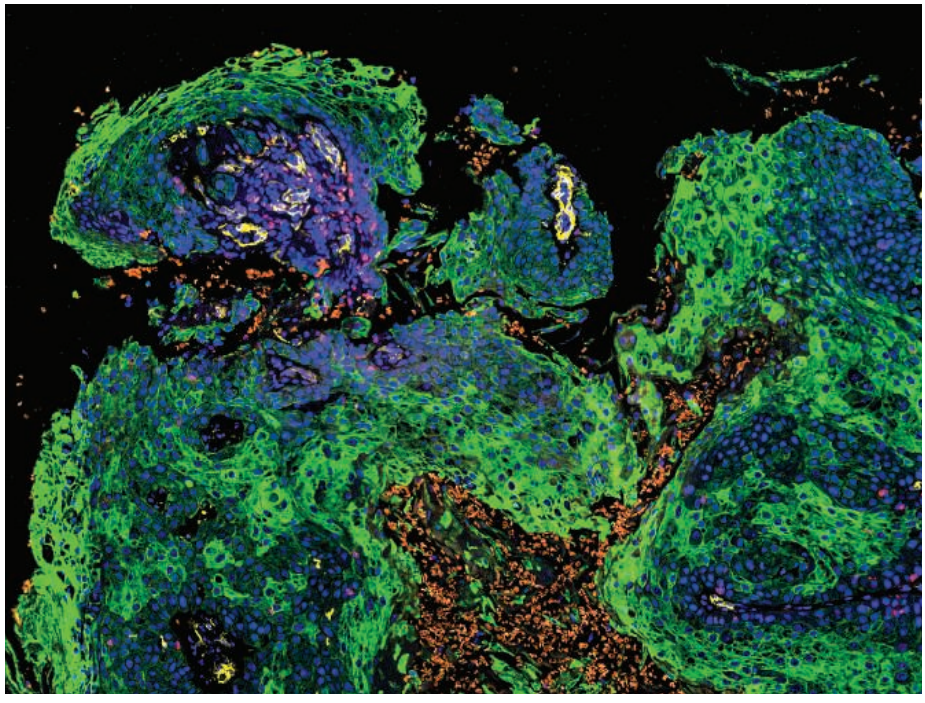
Full text: dx.doi.org/10.1126/science.ade3392

PRIZE ESSAY

GRAND PRIZE WINNER
Aleksandar Obradovic

Aleksandar Obradovic received undergraduate degrees in Biology and Computer Science from Columbia University, where he continued as an MD and PhD and received his PhD in the Department of Systems Biology in 2021. While working to complete his MD training, Aleksandar started his laboratory with an appointment as an associate research scientist in the Department of Medicine at Columbia in 2022. His research focuses on the development of innovative approaches for the analysis of single-cell transcriptional, T cell receptor sequencing, and spatial data, which are applied toward an improved understanding of immunological mechanisms of checkpoint inhibitor resistance and the discovery of synergistic combination therapies. **www.science.org/doi/10.1126/science.adg5585**

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY



IMMUNOLOGY

Precision immunotherapy

A mechanistic approach to overcoming treatment resistance reveals new targets

By **Aleksandar Obradovic**

In the past decade, cancer treatment has been revolutionized by the rise of checkpoint inhibitor immunotherapies. Traditional approaches to treating cancer have been focused on developing an increasingly sophisticated arsenal of drugs that are targeted against various mechanisms of tumor cell growth. With immunotherapy, the capacity of the human immune system to recognize self and nonself has been harnessed to activate the antitumor immune response in a way that can be broadly applied across many tumor types, regardless of their growth mechanisms (*1*). Nonetheless, many patients remain nonresponsive to immunotherapy and reliable predictors of therapy failure are lacking (*1, 2*), which have motivated a shotgun approach to improving outcomes. At present, hundreds of clinical trials attempt to combine immunotherapy with other treatments and boost its efficacy, with mixed results. In my research, I focus on a more systematic approach for both discovering the cell-cell interactions involved in immunotherapy resistance within different tumor types and identifying candidate therapies to overcome treatment resistance.

My approach hinges on the fact that all therapies are, to some extent, immune therapies that exert often incompletely understood or off-target effects on a patient's immune cells in addition to the intended effect on tumor cells. This realization has critical implications for the combination of existing drugs with checkpoint inhibitors, and an understanding of these effects may be harnessed to prioritize the testing of synergistic combinations. For example, my colleagues and I have shown that androgen-deprivation therapy for the treatment of prostate cancer induces a transient CD8 T cell infiltrate in what is otherwise thought to be an "immune-cold" tumor, which is rapidly counteracted by a corresponding regulatory T cell (T_{reg}) infiltrate (*3*). This finding has spurred neoadjuvant trials of combination androgen-deprivation plus immunotherapy, with encouraging initial results (*4*). A deeper understanding of the effects that both immunotherapy and conventional therapies exert on the immune microenvironment will enable more precise and rational combinations

Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA.
Email: azo2104@columbia.edu

PHOTOS: (LEFT TO RIGHT) COURTESY OF ALEKSANDAR OBRADOVIC; ALEKSANDAR OBRADOVIC AND YOUNG KIM

GeoMx Digital Spatial Profiler imaging of a human head and neck squamous cell carcinoma section from a patient treated with anti–PD-1 immunotherapy.

of complementary drugs in patients who are nonresponsive to current treatments.

The development of high-throughput single-cell RNA sequencing technology has opened an unprecedented window into the tumor microenvironment (*5*). Yet, a limited number of RNA molecules are available to capture from a single cell such that more than 90% of genes "drop out" in any given cell, which is particularly problematic for transcriptional regulators of cell state and signaling molecules that tend to have high biological relevance but low expression (*6*, *7*). To address this issue, my colleagues and I have developed a protein activity inference approach for identifying drivers of cell state in single-cell data upstream of noisy gene expression. This approach hinges on two algorithms: the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) and Virtual Inference of Protein activity by Enriched Regulon analysis (VIPER) (*5*, *8*, *9*). ARACNe uses mutual information between genes at the expression level to build and prune gene regulatory networks, such that each transcription factor and signaling protein has an inferred "regulon" of downstream targets (*9*). VIPER can then be used to infer protein activity as the normalized enrichment of each regulon in each single cell.

Because ARACNe infers many downstream targets for each protein, VIPER is less sensitive to gene dropout, is less affected by transcriptional noise, and correlates more robustly with measured protein abundance, comparable with and (for some proteins) better than flow cytometry (*6*, *7*). VIPER can be leveraged to rapidly match unfavorable cell states to candidate therapeutics by identifying the subset of active "druggable" proteins as well as cell-by-cell drug sensitivities that are inferred from a paired database of protein activity changes in response to an array of hundreds of cancer drugs across dozens of tumor cell lines (*10*). Using a broad systems biology approach that leverages protein activity inference, I have been able to discover that different immune and nonimmune cell clusters associate with outcome in different tumor types, such that protein activity markers of these populations can serve not only as biomarkers for treatment prioritization but also as targets in combination with immunotherapy to potentially sensitize nonresponders and improve outcomes in future clinical trials. This analysis pipeline represents a substantial improvement on previous approaches to the analysis of single-cell data and has yielded actionable therapy targets across a broad range of tumor and cell types (*4*, *6*, *10*, *11*).

By applying these tools systematically, I identified cell subpopulations that are associated with clinical outcomes in different tumor types: (i) a population of $C1q^{+}TREM2^{+}APOE^{+}$ macrophages that are associated with postsurgical tumor recurrence in clear cell renal carcinoma (*6*), which may risk-stratify patients for prioritization of more aggressive up-front treatment; and (ii) a subpopulation of fibroblasts that are associated with improved response to immunotherapy in head and neck squamous cell carcinoma, which are both prognostic and found to directly modulate T cell function ex vivo in a contact-dependent manner (*12*). Most notably, analyses have shown tumor-specific T_{regs} to be a mechanism of immunotherapy resistance that is shared across a broad range of tumor types (*6*, *11*), with T_{reg}-specific regulators of tumor-infiltration being identifiable by VIPER (*11*). A high-throughput drug screening approach has allowed rapid identification and repurposing of US Food and Drug Administration–approved drugs with beneficial off-target effects on the T_{reg} transcriptional profile, which has resulted in reduced tumor infiltration (*11*). Gemcitabine was identified by this approach as a drug that at very low doses differentially depletes tumor-infiltrating T_{regs} but not proinflammatory T cells, which improves response rates to anti–programmed cell death protein 1 (PD-1) checkpoint immunotherapy (*11*). As we have observed, different cancers have different populations of cells that drive therapy response and resistance, and this approach to drug prioritization may be readily extended in future work to target other immunosuppressive cell populations for clinical development and trials.

Taken together, the analytical and experimental tools that make up my research represent an opportunity to tailor future immunotherapies at the single-cell level. This can be extended even beyond cancer treatment to identify drugs with complementary immune effects across any number of medical contexts, for example, to suppress alloimmune organ transplant rejection, inhibit autoimmune disease, or prioritize adjuvants to improve the efficacy of vaccines. These benefits all result from leveraging protein activity inference and drug sensitivity prediction algorithms to deepen our understanding of treatment effects on recruitment and activation of critical immune cell phenotypes. By mechanistically selecting optimal combinations of drugs to overcome treatment resistance, one may begin to approach true precision immunotherapy: disease by disease, and patient by patient. ■

PHOTOS: (TOP TO BOTTOM) COURTESY OF ROSER VENTO-TORMO; COURESY OF JOSHUA TAN

FINALIST

Roser Vento-Tormo

Roser Vento-Tormo received a BS from the Polytechnic University of Valencia and a PhD from the University of Barcelona. After completing her postdoctoral fellowship at the Wellcome Sanger Institute, Vento-Tormo started her laboratory there in the Cellular Genetics program in 2019. Her research focuses on understanding how cell-cell communication and the tissue microenvironment regulate cell identity and function in the context of immunity and development. **www.science.org/doi/10.1126/science.adg6260**

FINALIST

Joshua Tan

Joshua Tan received his BS from Monash University and a PhD in Infection, Immunology, and Translational Medicine from the University of Oxford. After completing his postdoctoral fellowship—which was jointly performed at the Institute for Research in Biomedicine, Bellinzona; the University of Oxford; and the National Institutes of Health (NIH)—he started his laboratory in the National Institute of Allergy and Infectious Diseases at the NIH in 2020. His research focuses on understanding the human antibody response to infectious pathogens—including *Plasmodium falciparum*, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and *Mycobacterium tuberculosis*—through the study of monoclonal antibodies. **www.science.org/doi/10.1126/science.adg6261**

REFERENCES AND NOTES

1. G. T. Gibney, L. M. Weiner, M. B. Atkins, *Lancet Oncol.* **17**, e542 (2016).
2. G. Schepisi *et al.*, *J. Oncol.* **2019**, 7317964 (2019).
3. A. Z. Obradovic *et al.*, *Clin. Cancer Res.* **26**, 3182 (2020).
4. J. E. Hawley *et al.*, bioRxiv 10.1101/2022.05.06.490968 (2022).
5. G. X. Zheng *et al.*, *Nat. Commun.* **8**, 14049 (2017).
6. A. Obradovic *et al.*, *Cell* **184**, 2988 (2021).
7. L. Vlahos *et al.*, bioRxiv 10.1101/2021.05.20.445002 (2021).
8. M. J. Alvarez *et al.*, *Nat. Genet.* **48**, 838 (2016).
9. A. Califano, M. J. Alvarez, *Nat. Rev. Cancer* **17**, 116 (2017).
10. A. Obradovic *et al.*, bioRxiv 10.1101/2022.02.28.482410 (2022).
11. A. Obradovic *et al.*, bioRxiv 10.1101/2022.02.22.481404 (2022).
12. A. Obradovic *et al.*, *Clin. Cancer Res.* **28**, 2094 (2022).

10.1126/science.adg5585

PRIZE ESSAY

FINALIST

Roser Vento-Tormo

Roser Vento-Tormo received a BS from the Polytechnic University of Valencia and a PhD from the University of Barcelona. After completing her postdoctoral fellowship at the Wellcome Sanger Institute, Vento-Tormo started her laboratory there in the Cellular Genetics program in 2019. Her research focuses on understanding how cell-cell communication and the tissue microenvironment regulate cell identity and function in the context of immunity and development. **www.science.org/doi/10.1126/science.adg6260**

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY

IMMUNOLOGY

Decoding foreign antigen tolerance

Cell atlases of human tolerogenic milieus guide transformative immunotherapies

By **Roser Vento-Tormo**[1,2]

The ability to distinguish between self and nonself is a fundamental function of the immune system. Yet, in a few special situations, nonself antigens must be safely harbored in the body. This phenomenon of immunotolerance is crucial for the placenta, where the fetal genome is safe from attack by the maternal immune system. It is also relevant in the testis, where the synthesis of meiosis-associated neoantigens occurs during puberty after the establishment of self-tolerance.

Immunotolerance in the placenta and testis is achieved through a complex and still poorly understood regulation of immune cells and their signaling. Insights gained into this process are particularly crucial for immunotherapies, where the patient's immune cells are used to fight cancer cells. Immunotherapy is often complicated by tolerogenic mechanisms present in the tumor microenvironment, which lead to immune evasion or the failure of engineered cells to access the tumor tissue. By understanding the complex immunoregulatory processes that modulate immune cell activity within the tissue microenvironment, it is possible to guide the precision engineering of immune cells tailored to avoid tumor immunotolerance.

To study the process of immunoregulation, scientists have traditionally relied on model organisms, such as mice. However, various human-specific features as well as the complexity of the immunoregulatory microenvironments are not reflected in mouse models. In my laboratory, we take a conceptually distinct approach: We examine the formation of physiological immunoregulatory environments in human tissues to understand how tolerance is established in situ. We then use advanced in vitro models to mechanistically dissect the processes that we discover, with the goal of engineering cells that can overcome the immunoregulatory barriers of the tumor microenvironment.

The use of human tissues comes with its own set of challenges, not least the scarcity of the material. To make the most out of the precious material, we first use single-cell genomics to obtain a "parts list" of the full spectrum of immune cell types in a tissue through gene expression profiles (*1*). To understand how cell types fit within their tissue microenvironments and interact with other cells, we are aided by a more recent tool—spatial transcriptomics (*2*). This technology enables the assignment of spatial coordinates to cells in tissues, which is essential to reconstruct tissue architecture and define cellular neighborhoods.

This single-cell toolbox provides a systems-level view of the immune system; however, to move toward an integrated vision of the immunoregulatory environment, there is a need to better understand cell-cell communication. I have taken a computational approach and developed a tool called CellPhoneDB that allows inference of cell communication (*3*). CellPhoneDB uses single-cell data to measure the number of ligands and receptors expressed in spatially adjacent positions in the tissue and reads out signaling activity from transcriptomic signatures. This tool has been widely used by the immunology community to reconstruct the dynamic processes of the immune system from relatively static datasets.

I have applied these single-cell experimental and computational approaches to study the human placenta (*4*) and, more recently, the developing testis (*5*) and have provided insights into how their immunotolerant environments are established. In the developing testes, two specific populations of fetal testicular macrophages (*TREM2*$^+$ ftMs and *SIGLEC15*$^+$ ftMs) that are likely to play a key role in shaping the immunoregulatory landscape have been defined. *TREM2*$^+$ ftMs resemble microglia—specialized macrophages that were thought to be restricted to the central nervous system. Unexpectedly, we found *TREM2*$^+$ ftMs inside the testicular cords—closed structures where gametes develop, which were thought to not have immune cells. When interactions with the

[1]Wellcome Sanger Institute, Cambridge, UK. [2]Centre for Trophoblast Research, University of Cambridge, Cambridge, UK. Email: rv4@sanger.ac.uk

PHOTO: COURTESY OF ROSER VENTO-TORMO

surrounding microenvironment were quantified, we found that *TREM2*$^{+}$ ftMs are likely to oversee the removal of damaged cells and minimization of inflammation in the testis. *SIGLEC15*$^{+}$ ftMs closely resemble osteoclasts (macrophages found in the bone) and may aid in the endothelial remodeling characteristic of fetal testis development while dampening the associated inflammatory immune response triggered by this process. We are currently analyzing various cancer cell atlases to look for the presence of these cells and their transcriptomic programs in the context of cancer.

It is possible that the signatures of immunotolerance that have been defined in physiological contexts [e.g., the presence of known immunoregulatory molecules in ftMs (e.g., ENTPD1, LGALS9, TIM3, etc.)] will also explain cancer immunotolerance and thus help guide new immunotherapies. As the molecular mechanisms of tolerance in the human physiological context begin to be uncovered, there remains the question of how best to take advantage of this knowledge to drive future clinical advances. One of the key challenges in the field is the successful design of specific immune cell types in vitro. Such cells can be used either as research reagents and tools for modeling disease or be transplanted back into the patient to perform a particular function (e.g., killing tumor cells in immunotherapy).

The rich developmental maps that we and others have generated as part of the Human Developmental Cell Atlas (*6*) offer a transformative opportunity to inform protocols to differentiate human induced pluripotent stem cells (hiPSCs) into a variety of immune cells. These protocols allow us to generate high numbers of cells from a specific donor and to mechanistically dissect the process of differentiation with genetic engineering (e.g., CRISPR-Cas9 genome editing) or chemical factors. We have shown that myeloid cells produced from hiPSCs accurately recapitulate the early events of human embryonic hematopoiesis. We have also leveraged the information contained in these dense atlases to generate dendritic cells, which play a key role in mobilization and activation of the adaptive immune response (*7*). Our recent research therefore opens avenues to the development of myeloid-based therapies that can overcome the immunoregulatory mechanisms of the tumor microenvironment.

Our single-cell, computational, and in vitro approaches are aiding understanding of the mechanisms of immunoregulation and tolerance in the human body. This rich body of knowledge can next be used to guide engineering efforts and to design specific immune identities for modeling therapeutic interventions. Is it possible to engineer cells that penetrate the barrier around the tumor microenvironment? Ultimately, by elucidating the events that mediate immunotolerance, our integrated analysis of physiological tissues will be a key tool in this future endeavor. ■

REFERENCES AND NOTES

1. V. Svensson, R. Vento-Tormo, S. A. Teichmann, *Nat. Protoc.* **13**, 599 (2018).
2. A. Rao, D. Barkley, G. S. França, I. Yanai, *Nature* **596**, 211 (2021).
3. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, *Nat. Protoc.* **15**, 1484 (2020).
4. R. Vento-Tormo *et al.*, *Nature* **563**, 347 (2018).
5. L. Garcia-Alonso *et al.*, *Nature* **607**, 540 (2022).
6. M. Haniffa *et al.*, *Nature* **597**, 196 (2021).
7. C. Alsinet *et al.*, *Nat. Commun.* **13**, 2885 (2022).

10.1126/science.adg6260

PRIZE ESSAY

FINALIST

Joshua Tan

Joshua Tan received his BS from Monash University and a PhD in Infection, Immunology, and Translational Medicine from the University of Oxford. After completing his postdoctoral fellowship—which was jointly performed at the Institute for Research in Biomedicine, Bellinzona; the University of Oxford; and the National Institutes of Health (NIH)—he started his laboratory in the National Institute of Allergy and Infectious Diseases at the NIH in 2020. His research focuses on understanding the human antibody response to infectious pathogens—including *Plasmodium falciparum*, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and *Mycobacterium tuberculosis*—through the study of monoclonal antibodies. **www.science.org/doi/10.1126/science.adg6261**

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY

IMMUNOLOGY

Searching for common ground

The conserved coronavirus fusion peptide is a target of broadly neutralizing antibodies

By **Joshua Tan**

In the mid-1960s, while studying patients with respiratory symptoms, two pairs of scientists discovered a new type of virus unlike any they had seen before in humans (*1, 2*). These viruses were later given the name "coronaviruses" owing to spike-like projections on the surface that were reminiscent of the solar corona. In the following decades, other human-infecting coronaviruses were identified, along with a wide variety of coronaviruses that infect birds and mammals (*3*). During this time, coronaviruses were generally only associated with mild symptoms, especially in comparison to more lethal viruses such as HIV and influenza. All this changed with the emergence of three deadly coronaviruses in the 21st century: severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2. The spread of COVID-19, caused by SARS-CoV-2, has been the most severe pandemic in the past 100 years, resulting in more than 6.5 million deaths (*4*).

The emergence of SARS-CoV-2 triggered a flurry of research activity across the globe to develop vaccines and therapeutic antibodies that target the SARS-CoV-2 spike protein. My laboratory joined the effort to investigate this novel virus by leveraging our experience with the discovery of monoclonal antibodies (mAbs) against another potentially deadly pathogen, *Plasmodium falciparum*, which causes malaria (*5*–*7*). We identified neutralizing mAbs targeting the receptor-binding domain (RBD) and N-terminal domain of the spike protein and found that combining specific mAbs in a bispecific antibody format greatly improved potency and increased resistance to mutations found in SARS-CoV-2 variants of concern (*8*).

Shortly after we completed this study, the course of the COVID-19 pandemic shifted with the emergence of a new and distinct variant of concern. The Omicron BA.1 variant had a staggering number of mutations in the RBD, which resulted in reduced efficacy of most COVID-19 vaccines and therapeutic antibodies (*9, 10*). Subsequent Omicron subvariants have increased the diversity of mutations in this region (*11*), which further highlights the challenge of relying on RBD-specific tools to target all SARS-CoV-2 variants, including those that will emerge in the future. Therefore, we started a search for regions of the spike protein that are conserved among all SARS-CoV-2 variants of concern but are still targetable by neutralizing antibodies. We also asked a more ambitious question: Are there antibody-targetable regions that are common to all coronaviruses?

To answer this question, we screened plasma from a cohort of COVID-19 convalescent donors to identify those with broad reactivity to spike proteins from the seven known human coronaviruses (SARS-CoV-2, SARS-CoV-1, MERS-CoV, HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E) (*12*). We identified 19 donors of interest and performed an epitope-agnostic screen of B cells from these donors to isolate antibodies that bound to spike protein from different coronaviruses without limiting ourselves to any specific epitope. We anticipated that broadly reactive antibodies, if present, were likely to be uncommon, and we developed a two-step screening approach that combined flow cytometry and optofluidics to identify rare B cells of interest. From ~670,000 B cells, we isolated 60 immunoglobulin G mAbs that showed some degree of broad reactivity. However, only six were able to bind to spike protein from all seven coronaviruses (*12*). Peptide-mapping experiments revealed that all six mAbs targeted the same epitope—the fusion peptide—which is adjacent to the conserved S2′ cleavage site in the spike S2 subunit. During viral fusion, cleavage of the S2′ site by transmembrane protease serine 2 (TMPRSS2) or cathepsins unleashes the fusion peptide for insertion into the host membrane. Consistent with the broad reactivity of these mAbs, this site is highly conserved in all four coronavirus genera (Alpha, Beta, Gamma, and Delta). Moreover, the sequence of the fusion peptide is identical across all SARS-CoV-2 variants of concern that have been identified to date (*12*).

We next evaluated the function of the fusion peptide–specific mAbs. Two of these

Antibody Biology Unit, Laboratory of Immunogenetics, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD, USA. Email: tanj4@nih.gov

COURTESY OF JOSHUA TAN

mAbs, COV44-62 and COV44-79, broadly neutralized both alpha- and betacoronaviruses, including all SARS-CoV-2 variants of concern, albeit with lower potency than RBD-specific mAbs (*8, 12*). We also tested COV44-62 and COV44-79 for the ability to prevent SARS-CoV-2–mediated disease in a Syrian hamster model. We found that COV44-79 reduced weight loss and lung pathology in these hamsters relative to a control group. Crystal structures of the antibody-peptide complexes and alanine scan mutagenesis experiments revealed that COV44-62 and COV44-79 specifically targeted the arginine 815 (R815) residue—the exact location of the S2′ cleavage site. In line with this finding, these mAbs inhibited fusion between cells expressing the SARS-CoV-2 spike and those expressing the angiotensin-converting enzyme 2 (ACE2) receptor (*12*). Analysis of the crystal structures indicated that these mAbs are unable to bind to the fusion peptide in the prefusion conformation of the spike protein, which is the main form on intact virions (*13*). Therefore, a conformational shift of the spike protein, perhaps during the fusion process, is probably required to allow mAbs to bind.

These findings, in conjunction with work performed by other groups on mAbs against the fusion peptide (*14, 15*), suggest that this peptide should be considered as a target for therapeutic antibodies and vaccines that aim to confer broad protection against coronaviruses. Whether these tools are potent enough to prevent disease in humans will need to be evaluated, but therapeutics and vaccines targeting the fusion peptide are likely to retain their function against new SARS-CoV-2 variants. Furthermore, these tools could be used as rapidly deployable stopgaps if other concerning coronaviruses emerge while more pathogen-specific vaccines and therapeutics are developed. The epitope-agnostic approach that we optimized to identify new targets of functional antibodies is applicable to other pathogens, especially given recent improvements in technology that have allowed for the identification of B cells with rare specificities. Instead of focusing on immunodominant sites that are under strong immune pressure, it may be worth searching for conserved, functionally essential epitopes that evade most antibody responses but are still targetable by rare antibodies. ■

REFERENCES AND NOTES

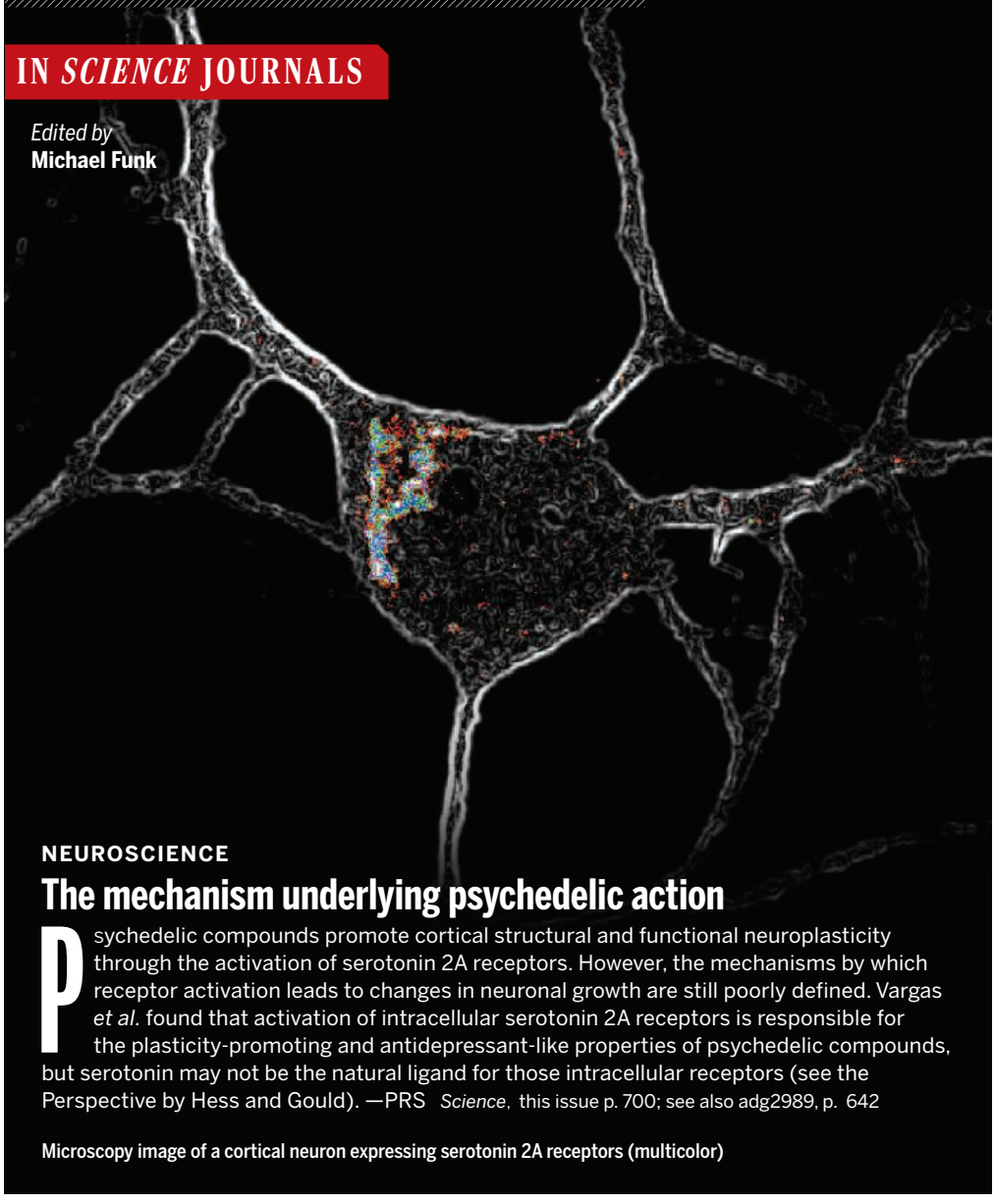
1. D. Hamre, J. J. Procknow, *Exp. Biol. Med.* **121**, 190 (1966).
2. D. A. Tyrrell, M. L. Bynoe, *BMJ* **1**, 1467 (1965).
3. P. V'Kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, *Nat. Rev. Microbiol.* **19**, 155 (2021).
4. E. Dong, H. Du, L. Gardner, *Lancet Infect. Dis.* **20**, 533 (2020).
5. J. Tan *et al.*, *Nature* **529**, 105 (2016).
6. K. Pieper *et al.*, *Nature* **548**, 597 (2017).
7. J. Tan *et al.*, *Nat. Med.* **24**, 401 (2018).
8. H. Cho *et al.*, *Sci. Transl. Med.* **13**, eabj5413 (2021).
9. L. Liu *et al.*, *Nature* **602**, 676 (2022).
10. N. Andrews *et al.*, *N. Engl. J. Med.* **386**, 1532 (2022).
11. H. Tegally *et al.*, *Nat. Med.* **28**, 1785 (2022).
12. C. Dacon *et al.*, *Science* **377**, 728 (2022).
13. Z. Ke *et al.*, *Nature* **588**, 498 (2020).
14. J. S. Low *et al.*, *Science* **377**, 735 (2022).
15. X. Sun *et al.*, *Nat. Microbiol.* **7**, 1063 (2022).

10.1126/science.adg6261

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk



NEUROSCIENCE

The mechanism underlying psychedelic action

Psychedelic compounds promote cortical structural and functional neuroplasticity through the activation of serotonin 2A receptors. However, the mechanisms by which receptor activation leads to changes in neuronal growth are still poorly defined. Vargas *et al.* found that activation of intracellular serotonin 2A receptors is responsible for the plasticity-promoting and antidepressant-like properties of psychedelic compounds, but serotonin may not be the natural ligand for those intracellular receptors (see the Perspective by Hess and Gould). —PRS *Science*, this issue p. 700; see also adg2989, p. 642

Microscopy image of a cortical neuron expressing serotonin 2A receptors (multicolor)

ELECTROCHEMISTRY

Protons from H_2 for ammonia synthesis

Electrochemical synthesis of ammonia from nitrogen (N_2) and hydrogen (H_2) could advantageously decentralize the current mass production of fertilizer. One promising method being explored involves lithium ion cathodic reduction in an organic solvent electrolyte, followed by reaction of the lithium with N_2. However, conventional H_2 oxidation catalysts for the complementary anodic process are unstable in these conditions. Fu *et al*. report that a gold–platinum alloy can robustly catalyze this oxidation and thus steadily produce the protons for ammonia under continuous flow conditions. —JSY

Science, adf4403, this issue p. 707

ORGANIC CHEMISTRY

Nickel's dual role in ketone transformation

Adding carbon nucleophiles to ketones is a common route to tertiary alcohols. When those ketones have a chiral center adjacent to the carbonyl, it is possible in principle to use a base to dynamically epimerize that center and thereby achieve enantioconvergent production of just one tertiary alcohol diastereomer. That approach has nonetheless proven challenging in practice. Ruan *et al*. report that a nickel catalyst can accomplish the epimerization through a distinct, base-free mechanism and then introduce an aryl or alkenyl nucleophile enantioselectively to deliver a broad range of chiral tertiary alcohols. —JSY

Science, ade0760, this issue p. 662

ELECTRON FLUIDS

Manipulating electron flow

Electron flow in conductors becomes viscous when electron-electron collisions dominate over collisions with defects and other sources of resistance. One way to study this viscosity is by physically fabricating narrow channels for electrons to flow through, but the rough edges of such channels make comparison with theory tricky. Krebs *et al*. instead used electrostatics to form tunable barriers to electron flow in graphene. Using scanning tunneling potentiometry, they observed the electron fluid flowing around the barriers in the viscous regime. —JS

Science, abm6073, this issue p. 671

SOLAR CELLS

Through thick but not thin

To maintain high charge carrier conductivity in perovskite solar cells, the passivating layer is usually very thin (~1 nanometer) to enable electron tunneling. However, this approach limits efficiency because it creates a trade-off between open-circuit voltage and fill factor and challenges in fabricating thin films from solution over large areas. Peng *et al*. grew a thick (~100 nanometer) dielectric mask formed by depositing alumina nanoplates and thus created random nanoscale openings for carrier transport. This layer reduced nonradiative recombination

IMAGE: DAVID OLSON/UC DAVIS

and boosted power conversion efficiencies from 23 to 25.5% compared with a conventional passivation layer. —PDS

Science, ade3126, this issue p. 682

GEOPHYSICS

Silently not able to heal

Faults on the Earth rupture over time but do not always produce earthquakes. Such aseismic, or slow slip, events are an important way to release stress. Shreedharan *et al*. determined the frictional healing properties of a slow slip portion of the Hikurangi fault in New Zealand. They found that the ability of the material to strengthen after failure was limited, unlike for earthquake-producing events. These observations could explain why these shallow, slow slip events happen frequently and at low stress. —BG

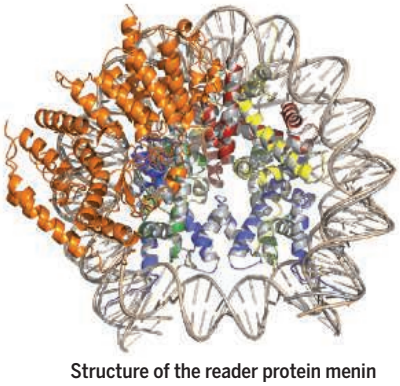
Science, adf4930, this issue p. 712

EPIGENETICS

A reader for H3K79me2

Decoding a histone modification requires identification of proteins that "read" the mark. Methylation of histone H3 lysine-79 (H3K79) plays key roles in gene regulation, but it remains a mystery how this histone mark is recognized and interpreted by specific readers. Lin *et al*. identified the protein menin as a bona fide reader of H3K79me2 using a multifunctional nucleosome-based photoaffinity probe. Biochemical and structural studies using cryo–electron microscopy revealed that menin engages with the H3K79me2 nucleosome through multivalent contacts. —DJ

Science, adc9318, this issue p. 717



CREDITS (LEFT TO RIGHT): LI ET AL.; BEN CURTIS/AP IMAGES

Structure of the reader protein menin (orange) bound to a nucleosome bearing the H3K79me2 modification

NEUROSCIENCE

Pathogenic palmitoylation in Parkinson's

The proteins synaptotagmin-11 (Syt11) and α-synuclein are associated with Parkinson's disease. Ho *et al*. uncovered a link between the two proteins that may contribute to the cellular pathology of the disease. Syt11 was palmitoylated, which enhanced its stability and membrane association and stimulated an increase in the aggregation-prone monomeric form of α-synuclein. These effects were seen in primary neurons from mice and in induced neurons derived from induced pluripotent stem cells from a patient with familial Parkinson's disease. —LKF

Sci. Signal. **16**, eadd7220 (2023).

HIV

Reverse engineering antibodies

Efforts to develop HIV-1 vaccines have included the identification of HIV-1 Env epitopes that can induce broadly neutralizing antibodies targeting the CD4-binding site. One such antibody, IOMA, is considered a good candidate for guiding the development of Env immunogens. Gristick *et al*. used a yeast display library screen and structure-based sequential immunization to evaluate Env immunogens in transgenic mice expressing germline-reverted IOMA. Antibody responses specific to the CD4-binding site and with heterologous neutralization capacity were induced by vaccination in both transgenic mice and in animals with polyclonal antibody repertoires, illustrating the potential of such Env immunogens as an HIV-1 vaccine strategy. —CNF

Sci. Immunol. **8**, eade6364 (2023).

IN OTHER JOURNALS

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

PROTEIN SYNTHESIS

Gregarious ribosomes in gregarious locusts

A solidary locust is like any other grasshopper: camouflaged to blend into the background and hopping about in search of food. However, when they are forced together by food shortage, the insects become colorful and attracted to each other. It is in this gregarious phase that locusts form swarms that can blacken the skies and obliterate crops. Many changes take place during this behavioral transition, and Li *et al.* show that protein synthesis is one thing that is altered. The authors found that in gregarious locusts, more ribosomes, which read the genetic code and assemble amino acids into protein, hop onto the mRNA than in solitary locusts. It appears that protein synthesis has to be ratcheted up in gregarious forms to meet the greater demand for proteins to sustain their highly migratory and aggressive lifestyle. —DJ

Proc. Natl. Acad. Sci. U.S.A. **120**, e2216851120 (2023).

Migratory locusts ramp up protein synthesis to enable transition into their colorful swarming form.

NEUROSCIENCE

Anxiety and irrelevant information

An increased sensitivity to threat-related information is advantageous in the context of immediate and actual threat avoidance. However, it is maladaptive if neutral cues in the environment acquire aversive associations based on irrelevant threat-related information and these associations subsequently guide behavior. Using behavioral modeling combined with brain scanning, Aberg *et al.* investigated the neurocomputational bases of learning relevant and irrelevant information. Anxious individuals avoided a neutral stimulus after it was paired with relevant neutral feedback and irrelevant fearful faces, although participants were explicitly instructed that the faces were

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PLANETARY SCIENCE

A new probe in the outer Solar System

The ice giants Uranus and Neptune have barely been explored. The only spacecraft to visit them was Voyager, which went on flybys in 1986 and 1989. As a result, the Uranus Orbiter and Probe (UOP) has been identified by the academic community as a priority for the next large-scale mission to be undertaken by NASA. In a Perspective, Mandt discusses the many unknowns about Uranus and what we could learn from UOP about how the planet was formed, its composition and structure, its atmosphere, and its ring and moon systems. Although Neptune is distinct from Uranus, this mission could also pave the way for future exploration. —GKA

Science, ade8446, this issue p. 640

CORONAVIRUS

CoronaVac and elderly immunity

The inactivated CoronaVac vaccine has been one of the most globally distributed vaccines during the COVID-19 pandemic. However, the persistence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants has resulted in the need for vaccine boosting. Filardi *et al*. characterized SARS-CoV-2–specific antibody responses in 293 nonhospitalized adults in Brazil and the Dominican Republic who had been initially vaccinated with two doses of CoronaVac, followed by a homologous CoronaVac booster or heterologous vaccination with the messenger RNA vaccine BNT162b2 or the adenovirus vector-based vaccine ChAdOx1. All vaccine types increased virus-specific immunoglobulin G (IgG) 28 days after boosting, but spike and receptor binding domain–specific IgG titers and neutralizing antibody concentrations were markedly lower against Omicron subvariants in individuals over 50 years old boosted with CoronaVac. These findings highlight the potential benefit of heterologous vaccine boosters for older individuals who were initially vaccinated with CoronaVac. —CNF

Sci. Transl. Med. **15**, eade6023 (2023).

IMMUNOLOGY

Immune cells in the extracellular matrix

The extracellular matrix is a dynamic tissue support network that is made up of components including fibrillar proteins, glycosaminoglycans, proteoglycans, and mucus. There is growing knowledge that this network has an intricate relationship with immune cells, which has important implications for our understanding of host defense, wound repair, fibrotic disease, and aging. Sutherland *et al.* reviewed the latest research on the interconnectedness of these systems with an emphasis on how this cross talk is an important element in the success (or failure) of immune cell–based therapies. —STS

Science, abp8964, this issue p. 659

NEUROSCIENCE

The scent of tsetse

Volatile sex attractants can be used in traps to monitor and control insect spread. Tsetse flies spread disease among humans and livestock in sub-Saharan Africa, and methods for controlling them are lacking. Ebrahim *et al*. identified fatty acid methyl esters as the volatile sex attractants in the tsetse fly *Glossina morsitans,* and showed that mating behavior was promoted for male files in response to these compounds (see the Perspective by Syed). Mating behavior in this species was also modulated by infection with parasites that cause African trypanosomiasis. These results suggest that sex attractants might be used in traps for controlling tsetse spread in sub-Saharan Africa. —MM

Science, ade1877, this issue p. 660; see also adg2817, p. 638

CANCER IMMUNOLOGY

Dendritic cells save the day

Dendritic cells are a key part of the immune system and are critical for antitumor immunity because they help to induce T cell responses against tumor antigens. CD5 is a glycoprotein expressed on some immune cells, including a subset of dendritic cells, but its function had not been well understood. He *et al*. demonstrated that CD5-expressing dendritic cells help to prime T cell responses against melanoma and are associated with survival in human patients and with immune rejection of tumors in mouse models. Moreover, these cells were needed for optimal results with immune checkpoint blockade, a type of immunotherapy that promotes reactivation of the body's own immune responses to treat cancer. —YN

Science, abg2752, this issue p. 661

SOLAR CELLS

Phosphorus stabilization of perovskites

Lewis base molecules that contain electron-donating atoms such as oxygen or sulfur can bind to undercoordinated lead atoms and passivate defects at interfaces and grain boundaries in perovskite films. Li *et al*. used density functional theory to screen potential Lewis bases and found that phosphorus-containing molecules showed the strongest binding to lead. A small amount of 1,3-bis(diphenylphosphino) propane stabilized inverted perovskite solar cells. The solar cells could maintain a power conversion efficiency of about 23% for more than 1500 hours under open-circuit conditions at 85°C. —PDS

Science, ade3970, this issue p. 683

BIOPHOTONICS

A photonic glass to camouflage crustaceans

Eye pigments are usually required for vision, but for creatures that want to stay hidden, they make it easier to be seen by others. This raises the question how marine invertebrates avoid being detected. Studying larval crustaceans, Shavit *et al*. found that many of these animals use an ultracompact reflector that overlies the opaque eye pigments and conceals them from view (see the Perspective by Feller and Porter). The reflection of light from the inner surface of the eye is produced by a photonic glass comprising crystalline isoxanthopterin nanospheres. The size and ordering of the nanospheres is varied to tune the reflectance from deep blue to yellow-green, enabling the organisms to blend in with different habitats. The use of a photonic glass may also enhance the vision of the eye. —MSL

Science, add4099, this issue p. 692; see also adf2062, p. 643

MOLECULAR BIOLOGY

Methylation suppression

Methylation of the N^6 position of adenine (m^6A) is a chemical tag for many messenger RNAs (mRNAs) added by m^6A "writer" proteins. Tagged mRNAs are marked for regulation through m^6A "reader" proteins, which bind methylated mRNAs and alter their expression. However, how the cell selects specific regions on mRNAs to be marked has remained unclear. He *et al*.

identified a new function for the exon junction complex as an m^{6}A "suppressor" that packages nearby mRNA and protects these regions from m^{6}A marking. Exon architecture controls mRNA accessibility to m^{6}A methylation, as well as potentially a broader set of regulatory complexes. —DJ

Science, abj9090, this issue p. 677

and boosted power conversion efficiencies from 23 to 25.5% compared with a conventional passivation layer. —PDS

Science, ade3126, this issue p. 682

GEOPHYSICS

Silently not able to heal

Faults on the Earth rupture over time but do not always produce earthquakes. Such aseismic, or slow slip, events are an important way to release stress. Shreedharan *et al*. determined the frictional healing properties of a slow slip portion of the Hikurangi fault in New Zealand. They found that the ability of the material to strengthen after failure was limited, unlike for earthquake-producing events. These observations could explain why these shallow, slow slip events happen frequently and at low stress. —BG

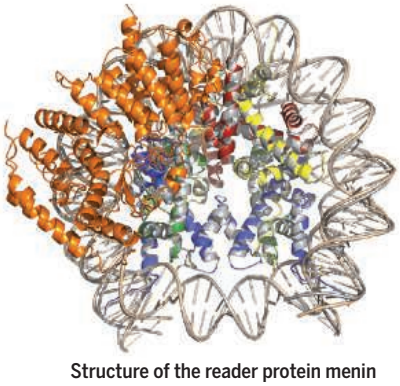
Science, adf4930, this issue p. 712

EPIGENETICS

A reader for H3K79me2

Decoding a histone modification requires identification of proteins that "read" the mark. Methylation of histone H3 lysine-79 (H3K79) plays key roles in gene regulation, but it remains a mystery how this histone mark is recognized and interpreted by specific readers. Lin *et al*. identified the protein menin as a bona fide reader of H3K79me2 using a multifunctional nucleosome-based photoaffinity probe. Biochemical and structural studies using cryo–electron microscopy revealed that menin engages with the H3K79me2 nucleosome through multivalent contacts. —DJ

Science, adc9318, this issue p. 717



CREDITS (LEFT TO RIGHT): LI ET AL.; BEN CURTIS/AP IMAGES

Structure of the reader protein menin (orange) bound to a nucleosome bearing the H3K79me2 modification

NEUROSCIENCE

Pathogenic palmitoylation in Parkinson's

The proteins synaptotagmin-11 (Syt11) and α-synuclein are associated with Parkinson's disease. Ho *et al*. uncovered a link between the two proteins that may contribute to the cellular pathology of the disease. Syt11 was palmitoylated, which enhanced its stability and membrane association and stimulated an increase in the aggregation-prone monomeric form of α-synuclein. These effects were seen in primary neurons from mice and in induced neurons derived from induced pluripotent stem cells from a patient with familial Parkinson's disease. —LKF

Sci. Signal. **16**, eadd7220 (2023).

HIV

Reverse engineering antibodies

Efforts to develop HIV-1 vaccines have included the identification of HIV-1 Env epitopes that can induce broadly neutralizing antibodies targeting the CD4-binding site. One such antibody, IOMA, is considered a good candidate for guiding the development of Env immunogens. Gristick *et al*. used a yeast display library screen and structure-based sequential immunization to evaluate Env immunogens in transgenic mice expressing germline-reverted IOMA. Antibody responses specific to the CD4-binding site and with heterologous neutralization capacity were induced by vaccination in both transgenic mice and in animals with polyclonal antibody repertoires, illustrating the potential of such Env immunogens as an HIV-1 vaccine strategy. —CNF

Sci. Immunol. **8**, eade6364 (2023).

IN OTHER JOURNALS

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

PROTEIN SYNTHESIS

Gregarious ribosomes in gregarious locusts

A solidary locust is like any other grasshopper: camouflaged to blend into the background and hopping about in search of food. However, when they are forced together by food shortage, the insects become colorful and attracted to each other. It is in this gregarious phase that locusts form swarms that can blacken the skies and obliterate crops. Many changes take place during this behavioral transition, and Li *et al.* show that protein synthesis is one thing that is altered. The authors found that in gregarious locusts, more ribosomes, which read the genetic code and assemble amino acids into protein, hop onto the mRNA than in solitary locusts. It appears that protein synthesis has to be ratcheted up in gregarious forms to meet the greater demand for proteins to sustain their highly migratory and aggressive lifestyle. —DJ

Proc. Natl. Acad. Sci. U.S.A. **120**, e2216851120 (2023).

Migratory locusts ramp up protein synthesis to enable transition into their colorful swarming form.

NEUROSCIENCE

Anxiety and irrelevant information

An increased sensitivity to threat-related information is advantageous in the context of immediate and actual threat avoidance. However, it is maladaptive if neutral cues in the environment acquire aversive associations based on irrelevant threat-related information and these associations subsequently guide behavior. Using behavioral modeling combined with brain scanning, Aberg *et al.* investigated the neurocomputational bases of learning relevant and irrelevant information. Anxious individuals avoided a neutral stimulus after it was paired with relevant neutral feedback and irrelevant fearful faces, although participants were explicitly instructed that the faces were

CHEMISTRY OF ART

Whence lead formate?

The goals of applying modern analytical techniques to centuries-old paintings are twofold: first, to understand how past artists derived their pigments and, second, to ascertain the best means of preserving the art for generations to come. Gonzalez *et al.* applied x-ray powder diffraction and infrared spectroscopy to Rembrandt's masterpiece *The Night Watch* and unexpectedly detected lead formate. Although the metal is a well-known paint component, this counterion is not. Experiments with linseed oil and lead oxide implicated a possible formation pathway, although revarnishing may also have played a role. —JSY *Angew. Chem. Int. Ed. Engl.* 10.1002/anie.202216478 (2023).

Paint from the masterpiece "The Night Watch" by the Dutch artist Rembrandt van Rijn has been found to contain lead formate, a compound never before seen in historical paintings.

unrelated to task performance. These results indicate that anxiety is associated with an increased integration of irrelevant threat-related information during feedback processing. —PRS

J. Neurosci. **43**, 656 (2023).

PSYCHOLOGY

Misinformed overconfidence

Why do some people hold strong opinions about well-evidenced science yet others remain neutral? A probability survey of more than 2000 UK adults addressed this question by focusing on attitudes toward specific genetic technologies such as genetically modified foods and the subjects' beliefs about the accuracy of their own understanding. Fonseca *et al.* found that an important factor was self-confidence: Neutral people were less confident in their own understanding of science, whereas people on the extremes (i.e., either those very supportive of science or very distrustful) were more confident that they accurately understood science, even when this was not warranted. Moreover, some misinformed yet confident people held very negative attitudes toward science. To address discrepancies between what people think they know versus what they actually know about science requires replacing inaccurate beliefs with an accurate understanding of science. —EEU

PLOS Biol. 10.1371/journal.pbio.3001915 (2023).

PSYCHIATRY

Tagging early trauma

Early life traumas result in increased activation of neurons in the paraventricular nucleus of the thalamus (PVT) in mice and increase the risk of developing mood and substance use disorders. To prevent these negative consequences, a clear understanding of the brain areas affected by early traumas is critical. Kooiker *et al.* used rodent models of early life adversity and genetic tagging to show that neurons expressing the receptor for the stress-related peptide corticotropin-releasing hormone in the PVT are overactivated in mice that experienced adversities compared with normally reared mice. Therefore, the PVT, which is known to be involved in reward circuits, is also critical for encoding adverse experiences occurring in the early postnatal period. —MM

Biol. Psychiatry Glob. Open Sci. 10.1016/j.bpsgos.2023.01.002 (2023).

EXTREMOPHILES

Yellowstone's primary producers

Geothermal springs present an ideal environment for autotrophic microbial communities because water rich in nutrients, minerals, and hydrogen sulfide meets oxidized surface fluids, providing a chemical potential that can be harvested by metabolism. Fernandes-Martins *et al*. studied microbial communities in three springs within Yellowstone National Park that are in close proximity but have differing hydrologic features. Microbial diversity, composition, and metabolic activity were strikingly different in these three sites, which the authors attribute to differences in the provision of reduced or oxidized sulfur compounds to the springs, which influences pH and the availability of substrates for metabolism. —MAF

Environ. Microbiol. 10.1111/1462-2920.16340 (2023).

GALAXIES

Spectroscopy of a galaxy at redshift 12

The expansion of the Universe causes light from distant objects to be redshifted to longer wavelengths, so measuring a redshift can determine the object's distance and age. Rough estimates of a galaxy's redshift can be made using imaging alone, but these are only candidate detections until they are confirmed by spectroscopy. Bakx *et al.* performed submillimeter spectroscopy of a candidate high-redshift galaxy that was previously identified by infrared imaging. They detected a spectral line and determined a spectroscopic redshift of 12.1, which is among the highest ever measured. The results place the galaxy only 367 million years after the Big Bang. —KTS

Mon. Not. R. Astron. Soc. 10.1093/mnras/stac3723 (2023).

IMAGE: REMBRANDT/WIKIMEDIA COMMONS

RESEARCH ARTICLE SUMMARY

CANCER IMMUNOLOGY

CD5 expression by dendritic cells directs T cell immunity and sustains immunotherapy responses

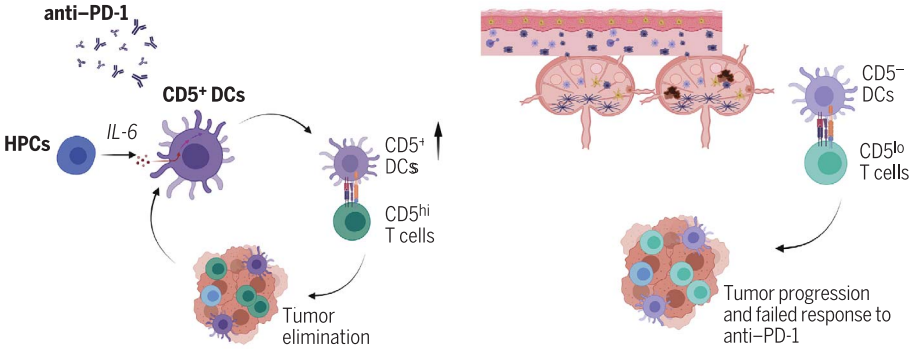
Mingyu He, Kate Roussak, Feiyang Ma, Nicholas Borcherding, Vince Garin, Mike White, Charles Schutt, Trine I. Jensen, Yun Zhao, Courtney A. Iberg, Kairav Shah, Himanshi Bhatia, Daniel Korenfeld, Sabrina Dinkel, Judah Gray, Alina Ulezko Antonova, Stephen Ferris, David Donermeyer, Cecilia Lindestam Arlehamn, Matthew M. Gubin, Jingqin Luo, Laurent Gorvel, Matteo Pellegrini, Alessandro Sette, Thomas Tung, Rasmus Bak, Robert L. Modlin, Ryan C. Fields, Robert D. Schreiber, Paul M. Allen, Eynav Klechevsky*

INTRODUCTION: Immune checkpoint blockade (ICB) therapy that blocks inhibitory T cell checkpoint molecules such as programmed cell death protein 1 (PD-1) and cytotoxic T lymphocyte–associated protein 4 (CTLA-4) has revolutionized cancer treatments. Despite its efficacy, half of the treated patients respond poorly and experience disease progression after therapy. To further improve immunotherapy outcome, a better understanding of the immune landscape and the tumor microenvironment (TME) is needed. In this context, dendritic cells (DCs), a functionally diverse system of antigen-presenting cells, play an instrumental role in eliciting antitumor responses during ICB therapy by inducing de novo $CD8^+$ cytotoxic and $CD4^+$ helper T cells against cancer-specific antigens. However, the specific DC features that are critical for driving effective T cell immunity and how subset composition in the tissue affects responsiveness to therapy remain unresolved.

RATIONALE: CD5, which is a transmembrane glycoprotein expressed on the surfaces of conventional T cells and some B cells, was recently recognized as a marker of a subset of DCs in both mice and humans. CD5 plays an important role in fine-tuning T cell receptor signaling during development and in their effector function in the periphery. However, the physiological roles of CD5 on DCs and the impact of such $CD5^+$ DCs on tumor immunity remain unclear. In our previous work, we showed that migratory $CD5^+$ DCs in the human skin prime helper T cells and multifunctional cytotoxic T cells more effectively than DCs that lack CD5. We hypothesized that the increased frequency of $CD5^+$ DCs in inflammatory settings correlates with their involvement in antitumor responses and responses to ICB therapy.

RESULTS: We investigated the mechanisms that underlie DC-mediated effector T cell priming, focusing on the role of CD5 expressed by these interacting cells. We analyzed the myeloid compartment of human skin draining lymph nodes and found that the frequency of the $CD5^+$ DC population within the DC2 compartment was reduced in human melanoma-affected lymph nodes compared with the same population in unaffected tissue. Accordingly, *CD5* expression as well as a $CD5^+$ DC gene signature correlated with greater survival and relapse-free survival in patients with a variety of cancers, including melanoma. We identified a critical immunostimulatory function for CD5 on DCs that potentiates the priming of tumor-reactive T cell activation, proliferation, effector function, and response to ICB therapy. The expression of CD5 on DCs correlated directly with the extent of effector helper $CD4^+$ and cytotoxic T cell priming in humans. Selective deletion of CD5 expression in DCs prevented an efficient response to ICB therapy in tumor-bearing mice, resulting in defective immune rejection of tumors. This defect correlated with the activation of $CD5^{lo}$ T cells, including $CD5^{lo}CD4^+$ T cells and neoantigen-specific $CD5^{lo}CD8^+$ T cells with poor effector function. In parallel, we examined biopsies of patients with melanoma and found that $CD5^{hi}$ T cell frequencies aligned with the $CD5^+$ DC density. Similarly, deletion of CD5 expression in T cells negatively affected $CD5^+$ DC-mediated T cell priming, antitumor immunity, and the response to ICB. Moreover, we found increased numbers of $CD5^+$ DCs after ICB in vivo, which corresponded to increased numbers of these cells in a patient with *PD-1* deficiency. Furthermore, interleukin-6 (IL-6), which was abundant in the cells of this patient, was also detected at higher levels within tumors after ICB therapy compared with baseline measurements. Correspondingly, we identified IL-6 as an important factor in $CD5^+$ DC differentiation and survival.

CONCLUSION: Our data provide insight into how immunotherapies like ICB work and identify CD5 on DCs as a new potential target for enhancing response rates among patients undergoing treatments. The role of DC CD5 in engaging effective antitumor immunity provides one explanation for the correlation of human DC2s, the only conventional DC subpopulation in humans that expresses CD5, with effector helper T cell density in tumors and with the patient responsiveness to ICB therapy. Overall, increased $CD5^+$ DCs in human lymphoid tissue aligns with markers of effector T cell quality, as well as with patient overall survival. Thus, harnessing CD5 on DCs may be a promising avenue for enhancing the efficacy of immunotherapy against multiple tumors. ■



DC CD5 directs the response to immunotherapy. $CD5^+$ DC frequency was reduced in tumor-affected lymph nodes, and their presence correlated with greater patient survival in multiple tumors. $CD5^+$ DC numbers increased during ICB therapy, and low IL-6 concentrations promoted their de novo differentiation. The requirement for CD5 on DCs was linked to the priming of effector $CD5^{hi}$ T cells and optimal ICB therapy. [Figure created with Biorender.com]

The list of author affiliations is available in the full article online.
***Corresponding author. Email: eklechevsky@wustl.edu**
Cite this article as M. He *et al.*, *Science* 379, eabg2752 (2023). DOI: 10.1126/science.abg2752

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg2752

RESEARCH ARTICLE

CANCER IMMUNOLOGY

CD5 expression by dendritic cells directs T cell immunity and sustains immunotherapy responses

Mingyu He[1], Kate Roussak[1], Feiyang Ma[2], Nicholas Borcherding[1], Vince Garin[1], Mike White[1], Charles Schutt[1], Trine I. Jensen[3], Yun Zhao[1], Courtney A. Iberg[1], Kairav Shah[1], Himanshi Bhatia[1], Daniel Korenfeld[1], Sabrina Dinkel[1], Judah Gray[1], Alina Ulezko Antonova[1], Stephen Ferris[1], David Donermeyer[1], Cecilia Lindestam Arlehamn[4], Matthew M. Gubin[1], Jingqin Luo[5], Laurent Gorvel[1], Matteo Pellegrini[2], Alessandro Sette[4,6], Thomas Tung[7], Rasmus Bak[3,8], Robert L. Modlin[9,10], Ryan C. Fields[4], Robert D. Schreiber[1], Paul M. Allen[1], Eynav Klechevsky[1]*

The induction of proinflammatory T cells by dendritic cell (DC) subtypes is critical for antitumor responses and effective immune checkpoint blockade (ICB) therapy. Here, we show that human CD1c$^+$CD5$^+$ DCs are reduced in melanoma-affected lymph nodes, with CD5 expression on DCs correlating with patient survival. Activating CD5 on DCs enhanced T cell priming and improved survival after ICB therapy. CD5$^+$ DC numbers increased during ICB therapy, and low interleukin-6 (IL-6) concentrations promoted their de novo differentiation. Mechanistically, CD5 expression by DCs was required to generate optimally protective CD5^{hi} T helper and CD8$^+$ T cells; further, deletion of CD5 from T cells dampened tumor elimination in response to ICB therapy in vivo. Thus, CD5$^+$ DCs are an essential component of optimal ICB therapy.

Immunotherapy has revolutionized cancer treatment. Melanoma is a prominent example of immune checkpoint blockade (ICB) success, with more than 50% of patients responding to dual anti–programmed cell death protein 1 (PD-1) and anti–cytotoxic T lymphocyte–associated protein 4 (CTLA-4) regimens (*1*, *2*). However, this leaves nearly 50% of patients as primary nonresponders, and most of these patients will experience disease progression after therapy. To improve the benefit of immunotherapy in patients with melanoma and to extend the success to other more immunotherapy-refractory cancers (for example, immunologically "cold" tumors such as pancreatic cancer), a better understanding of the immune landscape and tumor microenvironment (TME) is needed.

[1]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA. [2]Molecular Cell and Developmental Biology at University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA. [3]Department of Biomedicine, Aarhus University, 8000 Aarhus C, Denmark. [4]Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, La Jolla, CA 92037, USA. [5]Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA. [6]Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California San Diego (UCSD), La Jolla, CA 92037, USA. [7]Department of Surgery, Division of Plastic and Reconstructive Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA. [8]Aarhus Institute of Advanced Studies (AIAS), Aarhus University, 8000 Aarhus C, Denmark. [9]Division of Dermatology, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA. [10]Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA.
*Corresponding author. Email: eklechevsky@wustl.edu

Dendritic cells (DCs) play critical roles in antitumor immunity by inducing effector T cell responses against cancer-specific and cancer-associated antigens, many of which are relatively poorly immunogenic (*3*). Tumor cells, however, hijack the immune system, causing T cell exhaustion and DC dysfunction. Cancer-induced T cell exhaustion may be reversed through checkpoint blockade; however, this treatment fails to show clinical benefit in many patients, is unsuccessful for some cancer types, and can have severe side effects that limit its use.

Conventional DCs (cDCs) are broadly divided into cDC1 and cDC2 populations that arise through distinct pre-DC lineages (*4*). cDCs either reside in the lymph node (LN) or migrate in from peripheral tissues, such as the skin (*5*). Through their specific surface receptors and cytokines and through their differential use of antigen processing and presentation pathways, distinct DC subsets take specialized roles for activating different modes of immunity (*6*–*8*). For example, human skin harbors Langerhans cells in the epidermis and CD1a$^+$CD1c$^+$ or CD14-expressing DCs in the dermis. Whereas migratory Langerhans cells and dermal CD1a$^+$CD1c$^+$ DCs prime helper T cells (*6*, *9*) and cross-prime cytotoxic T lymphocytes (CTLs) (*6*, *7*, *10*), the dermal CD14$^+$ DCs promote humoral immunity (*6*, *11*) but inhibit CTLs (*6*, *7*) and prime regulatory T cells (*12*).

Both human and murine cDC1s comprise a relatively homogeneous population (*13*), whereas cDC2s are more heterogeneous. cDC2s are defined by the cell surface expression of CD11b, DCIR-2, and CD172a [signal regulatory protein alpha (SIRP-α)] (*14*, *15*) in mice and CD1c and Clec10A in humans (*13*), with differential surface expression of Esam, Mgl2 (CD301b, the mouse homolog to Clec10A) (*16*), or CLEC12A (*17*) in mice and CD5 in humans (*18*, *19*). Both cDC1s and cDC2s contribute to antitumor immunity by priming CTL responses and CD4$^+$ helper T (T_H) cells, respectively. Yet despite such understanding of the functions elicited by distinct DC subsets (*5*), our knowledge of the DC subset or subsets that promote the most potent antitumor responses in humans is less clear. We previously reported that migratory skin CD5$^+$ DCs, which we found on a subset of DC2s, prime inflammatory T_H cells and multifunctional CTLs more effectively than their CD5$^-$ DC2 counterparts (*18*, *20*). Their increased frequency in inflammatory settings led us to hypothesize that these cells are predominantly involved in the antitumor response.

CD5 is a type I transmembrane glycoprotein that is expressed on the surface of T cells and a subset of B cells. Additionally, it is now accepted as a marker of a subset of DC2s (CD1c$^+$ DCs) in humans (*18*, *19*, *21*). Although CD5 was initially detected on murine DCs in the early 1990s (*15*, *22*–*24*), its physiological role is not fully understood. Whereas some studies of murine cells indicated that CD5 functions as an inhibitory molecule (*25*–*29*), others suggested that when expressed on mature T cells, CD5 can promote T cell activation, interleukin-2 (IL-2) production (*30*–*33*), and pathogenic T_H17 cell differentiation (*25*, *34*, *35*) and can also confer a nuclear factor κB (NF-κB)–dependent survival advantage (*36*). Thus, the functional relevance of CD5 remains unresolved and may vary with cell type or microenvironment to promote a specific functional outcome (*37*).

Here, we identified a critical immunostimulatory function for CD5 on DCs that potentiates the priming of effector T cells. CD5 expression on DCs correlated with greater overall survival and relapse-free survival in patients with melanoma, lung squamous cell carcinoma, sarcoma, breast cancer, cervical squamous cell carcinoma, and endocervical adenocarcinoma, and the frequency of CD5$^+$ DCs was reduced 20-fold in human melanoma-affected LNs compared with unaffected tissue. Further, using conditional deletion of CD5 on DCs in mice, we found resistance to ICB therapy in sarcoma and colorectal cancer tumor models. The requirement for CD5 on DCs was linked to the activation of CD5^{hi}CD4$^+$ T cells and neoantigen-specific CD5^{hi}CD8$^+$ T cells. These data suggest that CD5 expression on cells of the myeloid compartment plays a key role in determining the outcomes of naturally occurring immune responses to cancer and that harnessing CD5 on DCs may be a promising avenue for enhancing the efficacy of ICB therapy against multiple tumors.

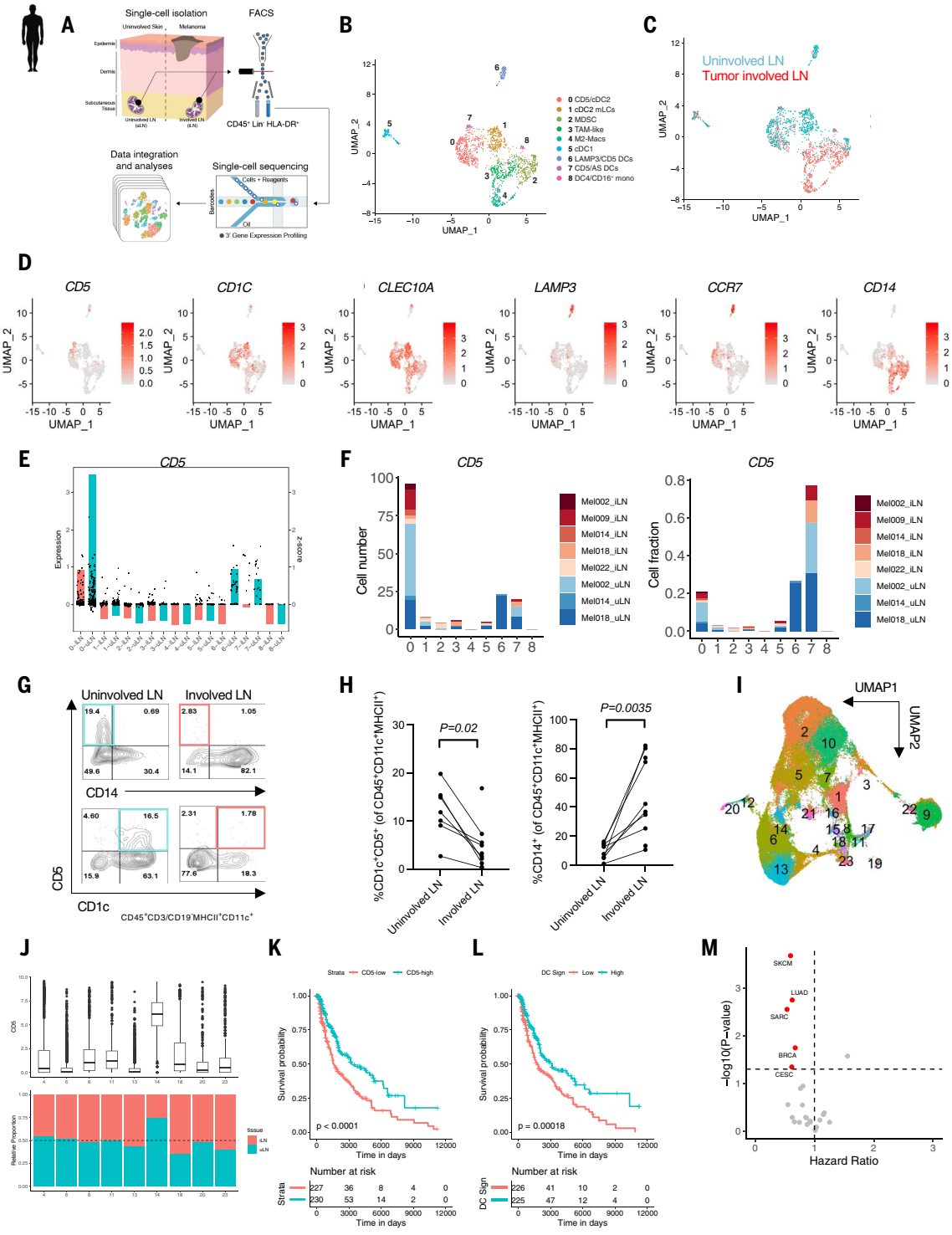


Fig. 1. Altered distribution of CD1c$^+$CD5$^+$ DCs in melanoma-affected human LNs and correlation with patient survival. (**A**) Workflow for single-cell analysis of myeloid cells from human melanoma that were isolated from five iLNs and three matched uLNs from patients with melanoma. (**B**) UMAP displaying nine clusters of myeloid CD45$^+$lineage$^-$HLA-DR$^+$ cells sorted from five iLNs and three matched uLNs from patients with melanoma and processed for scRNA-seq using 10× Genomics technology (see also fig. S1, A and B). MDSC, myeloid-derived suppressor cell; TAM, tumor-associated macrophage. (**C**) UMAP displaying CD45$^+$lineage$^-$ HLA-DR$^+$ cell distribution based on tissue type. (**D**) UMAP plots of six marker genes associated with various myeloid cell types in human tumor iLNs and uLNs as defined in Fig. 1, B and C. The color scale represents the normalized expression of the gene. (**E**) Scatter plot showing the *z*-score based on the total expression of CD5 for each cluster separated by iLN (red) and uLN (blue). The left *y* axis corresponds to the dots on the plot, which represent the expression of CD5 in individual cells. The right *y* axis corresponds to the bars. The height of each bar represents the *z*-score of the gene in the group of cells of each cluster. (**F**) Bar plots showing the number of cells expressing CD5 in each cluster (left) and the fraction of cells in each cluster and sample that expresses CD5 (right). Colors represent the samples from which the cells were derived. Shades of blue or red indicate uLNs or iLNs from different donors, respectively. (**G**) Expression of CD1c$^+$CD5$^+$ (bottom) and CD14$^-$CD5$^+$ (top) in iLNs and uLNs of representative patient Mel002. Cells were

gated on CD45$^+$CD3$^-$CD19$^-$MHCII$^+$CD11c$^+$. One representative plot of eight patients is shown. (**H**) CD1c$^+$CD5$^+$ (left) and CD14$^+$CD5$^-$ (right) in eight iLNs and matched uLNs, plus an additional two iLNs from patients with melanoma. Symbols indicate values from individual patients. (**I**) UMAP displaying CyTOF analysis of CD45$^+$ cells from four iLNs and matched uLNs of patients with melanoma. (**J**) Relative expression of CD5 in each myeloid cell–specific cluster (top) and distribution among the iLN and uLN cells (bottom). For the box-and-whisker plots, the bars are the default setting using interquartile ranges (IQR), where the lower whisker is quartile 1 minus 1.5 × IQR and the upper whisker is quartile 3 plus 1.5 × IQR. (**K**) Kaplan-Meier curve (top) showing the proportion of overall survival across the TCGA melanoma cohort by median *CD5* mRNA level, as generated by RNA-seq, comparing *CD5*-high (blue) versus *CD5*-low (red). Strata (bottom) refers to grouping, high versus low. (**L**) Kaplan-Meier curve (top) showing the proportion of overall survival across the TCGA melanoma cohort by median CD5$^+$ DC signature mRNA level comparing CD5$^+$DC-high (blue) versus CD5$^+$DC-low (red). DC signature (DC Sign) (bottom) was developed using the product of *CD5*, *CD1C*, *LAMP3*, and *CLEC10A* divided by the mRNA level of *CD3G*. (**M**) Volcano plot of the hazard ratios versus −log10(*P* value) across TCGA cohorts for the CD5$^+$ DC signature. Points highlighted in red are significantly ($P < 0.05$) associated with worse overall survival. SKCM, skin cutaneous melanoma; LUAD, lung adenocarcinoma; SARC, sarcoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma.

Results

CD5 expression correlates with improved disease prognosis in patients with melanoma

After identifying CD5$^+$ DCs in migratory skin DCs (*18*), we queried myeloid heterogeneity in the skin tumor-draining lymph nodes (TDLNs) of patients with metastatic melanoma. We sorted CD45$^+$lineage (CD3/CD56/CD19)$^-$major histocompatibility complex (MHC) II$^+$ antigen-presenting cells from five tumor-affected and three unaffected LNs that drain the skin of a melanoma patient (table S1) and performed single-cell RNA sequencing (scRNA-seq) using the 10× Genomics Chromium platform paired with deep sequencing. Uniform manifold approximation and projection (UMAP) analysis of 1596 TDLN myeloid cells from all the integrated samples (Fig. 1A and fig. S1A) yielded nine high-quality and distinct population clusters (designated clusters 0 to 8; Fig. 1B). Clusters 0, 1, 6, and 7 were expressed mainly in tumor-free tissue (Fig. 1, B and C, and fig. S1, A and B), whereas clusters 2, 3, 4, and 8 were principally present in tumor-affected tissue (Fig. 1, B and C, and fig. S1, A and B). We used gene overlays of individual canonical myeloid markers to further identify the cell clusters (Fig. 1D). Clusters 0, 1, and 6 contained CD1c- and CLEC10A-expressing DC populations, representing the cDC2 subset (*13*). Cluster 6 shared genes with lysosome-associated membrane glycoprotein 3 (LAMP3) DCs (*38*) as well as the recently described mature DCs enriched in immunoregulatory molecules (mreg) DCs, with the exception of *IL4R*, *STAT6*, and *AXL*, which were absent in LN cluster 6 DCs but present in mregDCs and which were required for their suppressive activity (*39*) (fig. S1, C and D). Cluster 7 contained progenitor pre-DC/AXL$^+$SIGLEC6$^+$ DCs, which express *AXL*, *SIGLEC6*, and *IL3RA*. We found that *CD5* transcript expression corresponded with cells in the tumor-uninvolved tissue (Fig. 1, E and F)—mainly the CD1c$^+$CLEC10A$^+$ cDC2s in clusters 0, 6, and 7—and was concordant with the expression of *LAMP3* and *CCR7* (Fig. 1D); these findings were in accordance with the known biology of migratory skin DCs (*40*). A subset of migratory Langerhans cells, identified by the expression of CD1a (*CD1A*), *CD1C*, and *CD207*, was present in cluster 1 (fig. S1E) and was also identified using cytometry by time of flight (CyTOF) (in myeloid cluster 16; fig. S1, F and G). XCR1$^+$CLEC9A$^+$ cells (cDC1s) occupied cluster 5 and were present in both tumor-involved and uninvolved tissues (Fig. 1, B and C, and fig. S1, E and H). Macrophages and monocyte-derived cells occupied clusters 2, 3, 4, and 8 and were enriched in melanoma-involved tissue, with cluster 2 cells expressing *THBS1* and the *S100A* family members *VCAN* and *FCN1*, which are characteristic of myeloid-derived suppressor cell–like cells. Clusters 3 and 4 contained triggering receptor expressed on myeloid cells 2 (*TREM2*)–expressing macrophages, tumor-associated macrophage–like and M2 macrophages (Fig. 1B and fig. S1, E and H), which also expressed *CD206/MRC1* (higher on cluster 3) and *C1QA*, *C1QB*, and *APOE* (higher on cluster 4) (fig. S1, E and H). Cluster 8 corresponded to nonclassical CD16$^+$ monocytes (DC4) that expressed CD16 (*FCGR3A*), *C5AR1*, *S100A4*, and *LILRB2* (fig. S1H).

We next assessed the presence of CD1c$^+$CD5$^+$ DCs in eight tumor-involved LNs (iLNs) and patient-matched uninvolved LNs (uLNs) by flow cytometry and CyTOF. We observed that these cells were restricted to a uLN and reduced in a metastatic tumor LN (mean percentage ± SEM; 13.5 ± 2.5% versus 4.0 ± 1.0%) [Fig. 1, G (bottom) and H (left)]. This contrasted with the CD14$^+$ cells, which preferentially occupied tumor-affected tissue and had no CD5 expression (9.2 ± 3.6% versus 55.2 ± 13.8%) [Fig. 1, G (top) and H (right)]. UMAP analysis showed that CD5-expressing DCs occupied a distinct myeloid cluster 14, which was enriched in uLNs (Fig. 1, I and J, and table S2).

We then examined the association of CD5 with disease prognosis based on The Cancer Genome Atlas (TCGA) melanoma cohort that is composed of 469 patients. *CD5* expression (Fig. 1K) and the CD5$^+$ DC signature (Fig. 1L) were associated with favorable prognosis for melanoma. There was no correlation between CD5 expression and patient age or disease staging, as measured by Clark level (fig. S2, A and B). Furthermore, the CD5$^+$ DC signature in tumors correlated with survival in an independent melanoma cohort (*41*) (fig. S2C). In addition to cutaneous melanoma, the CD5$^+$ DC signature was associated with favorable prognosis of lung squamous cell carcinoma, sarcoma, breast cancer, cervical squamous cell carcinoma, and endocervical adenocarcinoma (Fig. 1M). As expected, *CD5* correated with the expression of *CD1c* [coefficient of determination (R^2) = 0.53], *LAMP3* (R^2 = 0.51), *CLEC10A* (R^2 = 0.65), *CD40* (R^2 = 0.48), and *CLEC9A* (R^2 = 0.6) but was poorly correlated with *TREM2*, *THBD*, and *AXL* expression in the tumor (fig. S2D). Overall, CD5$^+$ DCs were enriched in the tumor-free LNs, and their presence was associated with better prognosis for patients with melanoma, lung squamous cell carcinoma, sarcoma, breast cancer, cervical squamous cell carcinoma, and endocervical adenocarcinoma.

CD5$^+$ DCs isolated from LNs and skin efficiently prime allogeneic naïve T cell responses

DCs are distinctive in their capacity to induce proliferation of allogeneic naïve CD4$^+$ T cells and CD8$^+$ T cells in a mixed lymphocyte reaction (*42*). Within this diverse cell system, we previously identified a subset of cells that express CD5 within the migratory dermal CD14$^-$CD141$^-$CD1c$^+$ DCs and epidermal Langerhans cells (*18*). The dermal CD5$^+$ DCs (fig. S3A) were powerful stimulators of naïve CD4$^+$ T and CD8$^+$ T cell proliferation (Fig. 2A). As few as 1000 of these cells were sufficient to induce nine and seven cell divisions within the CD4$^+$ T cell (Fig. 2B, left) and CD8$^+$ T cell (Fig. 2B, right) populations, respectively, whereas the CD5$^-$ dermal DCs induced only four and two divisions, respectively (Fig. 2B). Similarly, CD1c$^+$CD5$^+$ LN DCs (fig. S3B) were stronger stimulators of naïve CD4$^+$ and CD8$^+$ T cell proliferation than their CD1c$^+$CD5$^-$ counterparts (Fig. 2, A and B). The enhanced capacity of LN CD5$^+$ DCs to activate T cells was reflected in the amounts of cytokines that were produced in the primed T cell cultures (Fig. 2, C and D). Compared with LN CD5$^-$ DCs, CD5$^+$ DCs isolated from LNs induced more interferon-γ (IFN-γ) and tumor necrosis factor–α (TNF-α) production by naïve CD4$^+$ T cells (Fig. 2C, left) and CD8$^+$ T cells (Fig. 2C, right). Moreover, higher amounts of TNF-α were detected in cocultures of T cells activated by LN CD5$^+$ DCs (Fig. 2D, left) or dermal CD5$^+$ DCs (Fig. 2D, right) compared with those detected in cocultures of T cells and CD5$^-$ DCs or CD14$^+$ cells (Fig. 2D). CD5$^+$ DC–primed T cells also yielded higher amounts of IFN-γ than those primed by the other LN-specific

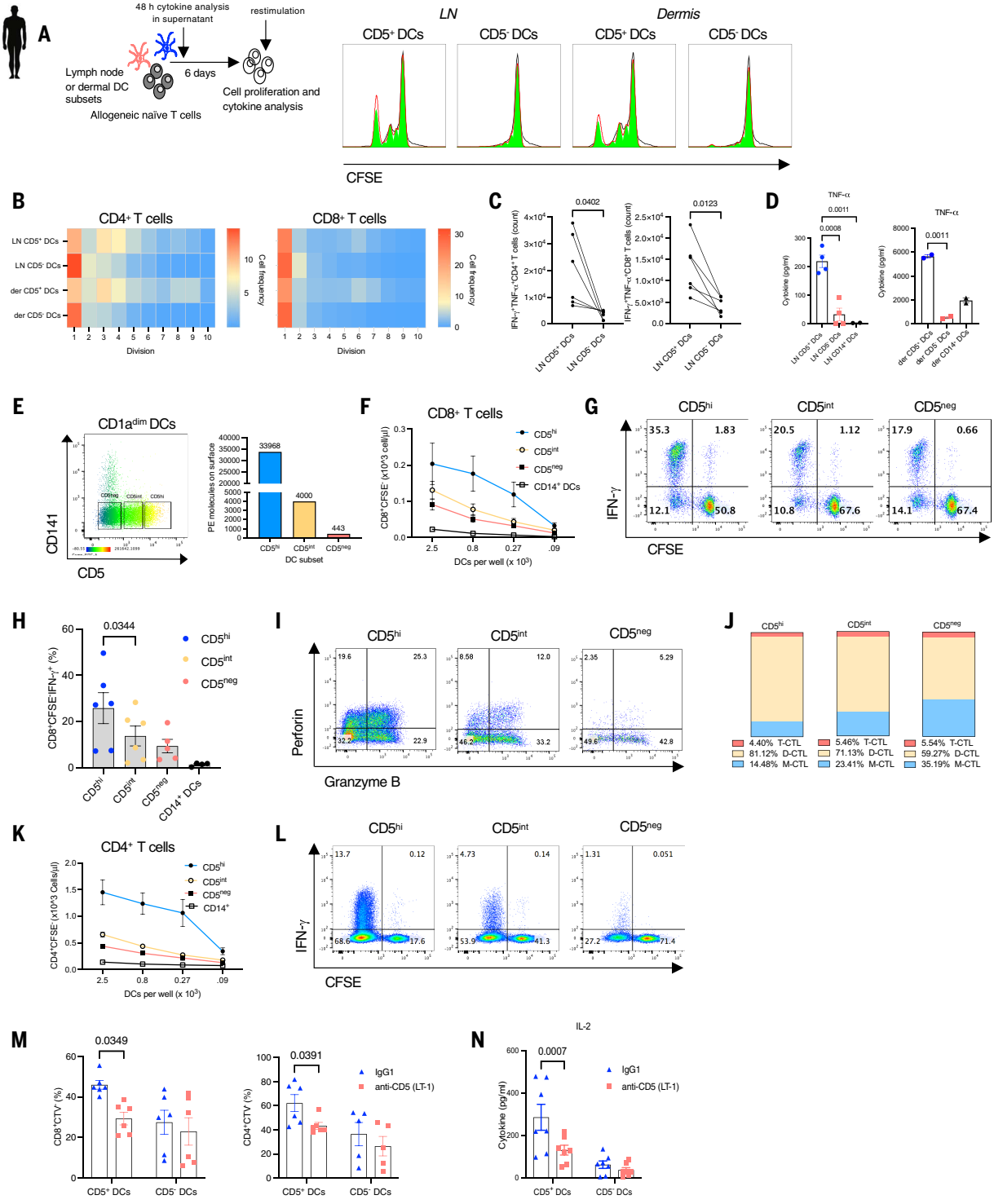


Fig. 2. CD5 expression on DCs correlates with CD4$^+$ and CD8$^+$ T cell activation. (**A**) CFSE dilution of allogeneic naïve T cells primed by either CD5$^+$ or CD5$^-$ DCs isolated from an uLN from a patient with melanoma or from healthy human dermis. Representative results of three melanoma LN donors and at least 10 dermal DC donors are shown. (**B**) Proliferation index of allogeneic naïve CD4$^+$ T cells (left) and CD8$^+$ T cells (right) primed by CD5$^+$ or CD5$^-$ LN DCs or CD5$^+$ and CD5$^-$ dermal (der) DCs. Representative results of three melanoma LN donors and at least 10 dermal DC donors are shown. The color scale represents the fraction of cells that have diluted CFSE in each of the indicated cell divisions. (**C**) Number of IFN-γ$^+$TNF-α$^+$ CD4$^+$ T cells (left) and CD8$^+$ T cells (right) primed by CD5$^+$ or CD5$^-$ LN DCs. Composite data of three experiments with three LN donors are shown. (**D**) TNF-α production measured in the cultures of T cells stimulated by CD5$^+$ or CD5$^-$ LN DCs (left) or CD5$^+$ or CD5$^-$ dermal DCs (right) or by control CD14$^+$ DCs for 48 hours. Composite data of two experiments with two LN donors and two different dermal donors are shown. Data represent the means ± SEM. (**E**) Flow cytometric gating scheme showing CD5^{hi}, CD5^{int}, and CD5^{neg} DCs (live, lineage$^-$HLA-DR$^+$CD11c$^+$CD14$^-$CD1a^{dim}) (left). See also fig. S3A. Quantification of surface CD5 expression (right) on CD5^{hi}, CD5^{int}, and CD5^{neg} dermal DCs using BD QuantiBrite beads. One representative quantification of three performed with DCs from three different donors is shown. PE, phycoerythrin. (**F**) Sorted

dermal CD5^{hi} (blue), CD5^{int} (yellow), and CD5^{neg} (red) DCs or control CD14$^+$ DCs (black) were cocultured with naïve CD8$^+$ T cells at different DC:T cell ratios for 6 days before CD8$^+$ T cell numbers were determined using BD TruCOUNT beads. Composite data of three experiments performed with three different donors are shown; data represent means ± SEM. (**G**) IFN-γ expression by the proliferating CD8$^+$ T cells primed by dermal CD1a$^{(dim)}$ CD5^{hi}, CD5^{int}, or CD5^{neg} DCs. One experiment of six performed with six different donors is shown. (**H**) Composite data of six experiments performed with six different donors are shown; data represent means ± SEM. (**I**) Granzyme B and perforin expression by proliferating (CFSE^{lo}) CD8$^+$ T cells primed by CD5^{hi}, CD5^{int}, or CD5^{neg} DCs. One experiment of three performed with three different donors is shown. (**J**) Proportions of T-CTLs, D-CTLs, and M-CTLs induced by dermal CD1a$^{(dim)}$ CD5^{hi}, CD5^{int}, or CD5^{neg} DCs. Composite data of three experiments with three different donors are shown. (**K**) Activated CD4$^+$ T cell proliferation in response to different numbers of dermal CD1a$^{(dim)}$ CD5^{hi} (blue), CD5^{int} (yellow), and CD5^{neg} (red) DCs and CD14$^+$ (black) DCs. CD4$^+$ T cells were determined on day 6 using BD Trucount beads. Composite data of four experiments with four different donors are shown; data represent means ± SEM. (**L**) Expression of IFN-γ by proliferating CD4$^+$ T cells primed by dermal CD1a$^{(dim)}$ CD5^{hi}, CD5^{int}, or CD5^{neg} DCs. One experiment of three performed with three different donors is shown. (**M**) Proliferation of CD8$^+$ (left) and CD4$^+$ T cells (right) after stimulation with CD5$^+$ and CD5$^-$ DCs and anti-CD5 blocking mAb. Composite data of two experiments performed with two different donors in triplicates are shown. CTV, CellTrace Violet proliferation dye. (**N**) IL-2 production measured in the culture supernatant of T cells after stimulation with CD5$^+$ or CD5$^-$ DCs in the presence or absence of anti-CD5 blocking mAb (LT-1). Composite data of three experiments performed with three different donors in duplicates or triplicates is shown; data represent means ± SEM. In (C), (D), (H), (M), and (N), the numbers over the brackets are *P* values.

DC subsets [CD5$^-$CD206$^-$, CD206$^+$CD5$^-$ DCs, or XCR1$^+$ DCs] (fig. S3, C and D). Other lymphoid organs tested, including spleen and tonsils, showed only scarce numbers of CD5$^+$ DCs compared with LNs (fig. S3, E and F). Thus, LN CD1c$^+$CD5$^+$ DCs may populate migrating skin DCs and share a functional similarity with the dermal CD1c$^+$CD5$^+$ DCs in potentiating CD4$^+$ and CD8$^+$ T cell priming.

CD5$^+$ DCs efficiently reactivate immunity

We next assessed the role of CD1c$^+$CD5$^+$ DCs in reactivating influenza matrix protein M1 (Flu-M1)–specific CD8$^+$ T cells. CD1c$^+$CD5$^+$ DCs or CD1c$^+$CD5$^-$ DCs, as well as CD14$^+$ DCs, were isolated from the dermis of human leukocyte antigen (HLA)–A2$^+$ donors and loaded with an 18–amino acid peptide that contains an HLA-A2–restricted epitope (amino acids 58 to 66) derived from Flu-M1. The DCs were cultured with autologous CD8$^+$ T cells isolated from blood, and the expansion of Flu-M1–specific CD8$^+$ T cells was analyzed after 8 days by using a specific MHC tetramer (fig. S4A). CD5$^+$ DCs were notably more efficient at inducing greater numbers of antigen-specific CD8$^+$ T cells than CD5$^-$ DCs or CD14$^+$ DCs (12.6 ± 6.57% versus 2.85 ± 1.45% versus 0.65 ± 0.3%) (fig. S4B). The enhanced capacity to reactivate CD8$^+$ T cells was not restricted to HLA-A2. CD5$^+$ DCs purified from the dermis and cultured with autologous CD8$^+$ T cells in the presence of MHC I–restricted peptides from Epstein-Barr irus and cytomegalovirus induced more effector (granzyme B–producing) CD8$^+$ T cells than did CD5$^-$ and CD14$^+$ DCs (fig. S4C, left). Measurement of IFN-γ production showed that dermal CD1a$^{(dim)}$ CD5$^+$ DCs activated twofold-stronger antigen-specific responses than their dermal CD1a$^{(dim)}$ CD5$^-$ counterparts (fig. S4C, right). Similar results were obtained upon measurement of IFN-γ release by CD8$^+$ T cells in response to an MHC I–restricted peptide pool [fig. S4, D (left) and E]. A related approach revealed the greater efficiency of CD5$^+$ DCs in activating MHC II–restricted viral-specific CD4$^+$ T cell responses [fig. S4, D (right) and E]. Collectively, CD5$^+$ DCs were potent inducers of antigen-specific memory CD8$^+$ and CD4$^+$ T cells against tumor-associated viral antigens.

CD5 expression on DCs correlates with T cell priming

To determine the impact of CD5 expression by DCs on the priming of effector T cells, we sorted the dermal CD1c$^+$ DCs based on CD5 expression (CD5^{hi}, CD5^{int}, and CD5^{neg} DCs) (Fig. 2E, left). Using a directly conjugated primary monoclonal antibody (mAb) and Quantibrite quantification beads, we estimated that CD5^{hi} DCs (blue) expressed ~10 times more CD5 molecules than the CD5^{int} DCs and ~100 times more than the CD5^{neg} DCs (Fig. 2E, right). After a 6-day coculture, we found that allogeneic naïve CD8$^+$ T cell expansion (percentages and numbers) was more effectively primed by CD5^{hi} DCs than the expansion achieved by CD5^{int} and CD5^{neg} DCs (Fig. 2, F and G). Further, the frequencies of CD8$^+$ T cells that produced IFN-γ (Fig. 2, G and H), as well as the frequencies of the effector molecules granzyme B and perforin (Fig. 2I), were higher after priming by CD5^{hi} DCs than the frequencies detected after priming by CD5^{int} and CD5^{neg} DCs (35.3 versus 20.5 versus 17.9%, respectively, for IFN-γ; 25.3 versus 12 versus 5.29%, respectively, for granzyme B and perforin).

We have previously defined three major subpopulations (monocytotoxic, dicytotoxic, and tricytotoxic) of CTLs (*10*); therefore, we also monitored the frequency of multifunctional CTLs primed by CD5^{hi}, CD5^{int}, and CD5^{neg} DCs. Monocytotoxic T lymphocytes (M-CTLs) express only a single effector molecule (granzyme B, perforin, or granulysin), whereas dicytotoxic T lymphocytes (D-CTLs) express such effector molecules in pairs and tricytotoxic T lymphocytes (T-CTLs) express all three molecules. CD5^{hi} DCs primed a higher frequency of T-CTLs and D-CTLs than CD5^{int} and CD5^{neg} DCs (Fig. 2J; 85.5 versus 76.6 and 64.8%, respectively). Similar to the effect on CD8$^+$ T cells, CD5^{hi} DCs promoted substantially greater CD4$^+$ T cell proliferation (Fig. 2K) and IFN-γ production (Fig. 2L) than CD5^{int} and CD5^{neg} DCs.

Next, we used an anti-CD5 mAb reported to block the immunogenicity of CD5 in a soluble state (*30*) to assess the importance of CD5 in inducing T cell activation. Indeed, the presence of this anti-CD5 mAb reduced the proliferation of both CD8$^+$ (Fig. 2M, left) and CD4$^+$ (Fig. 2M, right) T cells, as well as cytokine secretion in the culture supernatant (Fig. 2N). This effect was substantial in CD5$^+$ DC:T cell cocultures, which suggests a key role for CD5 on DCs in their interaction with T cells.

CD5 expression on human DCs directly enhances T cell priming

To assess the contribution of CD5 expressed on DCs to T cell priming, we used a CRISPR gene activation approach (CRISPRa) to up-regulate CD5 expression on CD5$^-$ DCs (*43*). CD5$^-$ DCs were nucleofected with single guide RNAs (sgRNAs) that target the CD5 promoter and dCas9-VP64-p65-Rta (VPR) mRNA or dCas9-VPR mRNA alone. We obtained a marked up-regulation of CD5 expression (CD5_CRISPRa DCs; 77%) on DCs compared with control cells as early as 24 hours after nucleofection (Fig. 3, A and B). This expression was sustained on a subset of cells for 2 days and then declined (Fig. 3B). CD5 up-regulation was not associated with up-regulation of costimulatory molecules CD72, CD40, CD70, CD80, or CD155, as detected by RNA-seq (Fig. 3C) and flow cytometry (Fig. 3D).

Next, we assessed the impact of CD5_CRISPRa on allogeneic naïve T cell priming (Fig. 3E). Carboxyfluorescein succinimidyl ester (CFSE) dilution analysis showed that CD5_CRISPRa DCs induced an increase in T cell proliferation (Fig. 3F) and primed higher numbers of T cells that produce IFN-γ and TNF-α than control DCs, as determined by intracellular staining (Fig. 3G). CD5 expression on DCs also led to increases in IFN-γ (Fig. 3H) and the T_H2 cell cytokine IL-13 in the culture supernatants (Fig. 3I).

To determine the impact of CD5 on T cell activation by other types of DCs, we induced monocytes to differentiate into DCs (moDCs) by incubation with granulocyte-macrophage colony stimulating factor (GM-CSF) and IL-4 for 5 days. Similar to dermal CD14$^+$ DCs, which

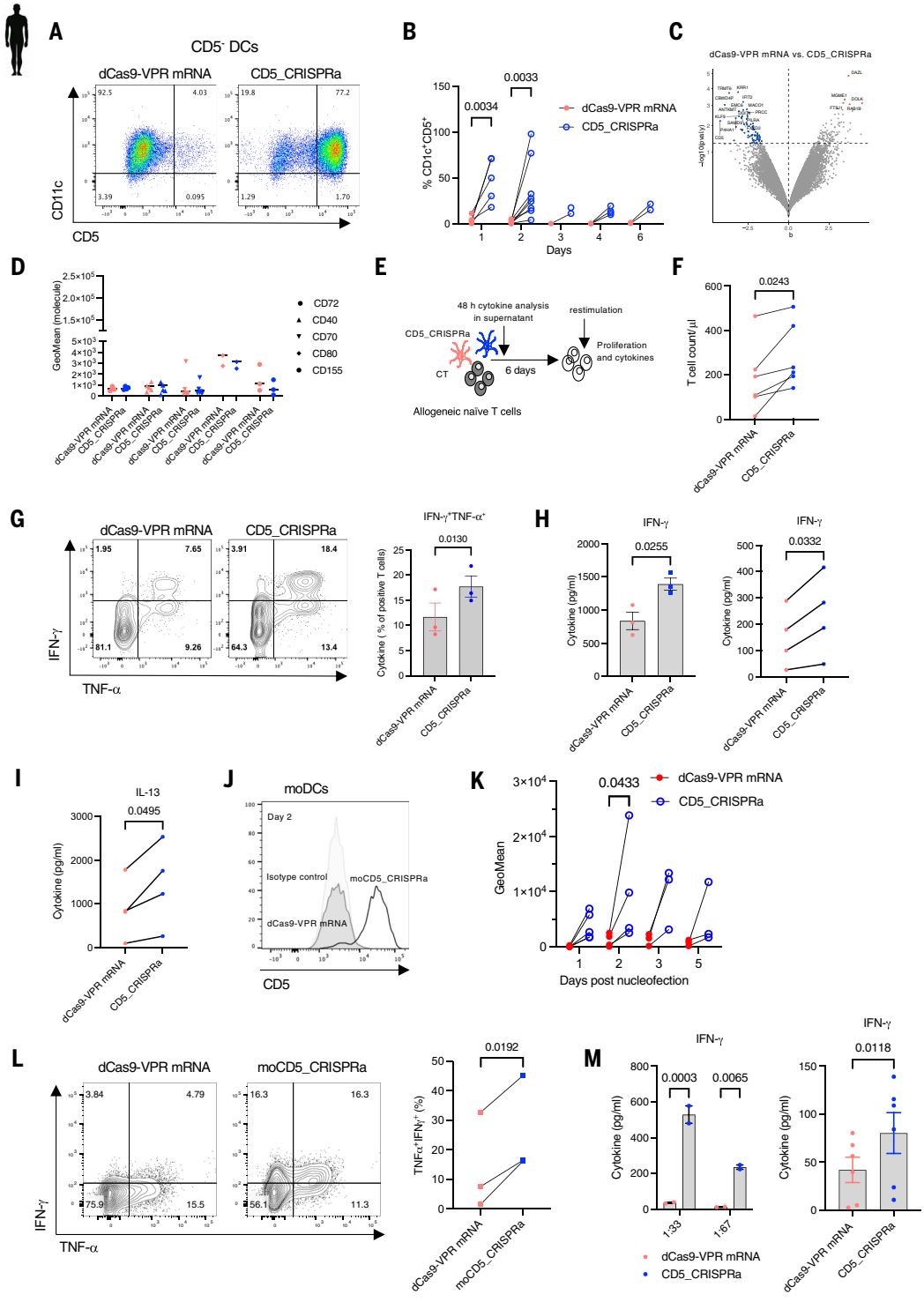


Fig. 3. Up-regulation of CD5 on human DCs promotes effector T cell activation. (**A**) Flow cytometric analysis of CD5 expression in CD5⁻ dermal DCs 48 hours after nucleoporation with or without dCas9-VPR mRNA and sgRNAs. One representative plot of eight experiments is shown. (**B**) Flow cytometric analysis of CD5 expression on dermal CD5⁻ DCs 1 to 6 days after nucleoporation with dCas9-VPR mRNA and CD5a sgRNAs or control dCas9-VPR. Symbols indicate values from individual donors. Composite data of CD5 expression analyzed 1 day (n = 7 donors), 2 days (n = 8 donors), 3 days (n = 2 donors), 4 days (n = 4 donors), and 6 days (n = 2 donors) after nucleofection are shown. (**C**) Volcano plot showing differentially expressed genes between control dCas9-VPR mRNA–electroporated CD5⁻ DCs and CD5_CRISPRa DCs. Genes up-regulated and down-regulated in CD5_CRISPRa DCs are shown in blue and red, respectively. (**D**) Flow cytometric analysis of the expression of CD72, CD40, CD70, CD80, and CD155 by control dCas9-VPR mRNA–electroporated CD5⁻ DCs and CD5_CRISPRa DCs 24 hours after nucleofection. Composite data of at least four donors analyzed for each marker are shown. (**E**) Workflow used in (F) to (I). CT, control. (**F**) Numbers of T cells with diluted CFSE (one division or more) as analyzed on day 6 of the coculture. Composite data of four performed experiments with five donors are shown. (**G**) CD5_CRISPRa or dCas9-VPR mRNA CT dermal DCs were cultured for 6 days with allogeneic naïve CD4⁺ and

$CD8^+$ T cells. Percentages of IFN-γ– and TNF-α–expressing T cells were determined by flow cytometry 6 hours after activation. One representative experiment out of three performed is shown (left). The graph shows composite data of three independent experiments; data represent means ± SEM (right). (**H**) IFN-γ production by T cells after activation for 48 hours with CD5_CRISPRa Langerhans cells or dCas9-VPR mRNA. One representative experiment of five performed with five different donors is shown. Data represent means ± SEM (n = 3) (left). Composite data of four additional experiments performed in the same way using CD5_CRISPRa dermal DCs or dCas9-VPR mRNA CT DCs (right). (**I**) T helper cytokine (IL-13) production after activation of T cells for 48 hours with CD5_CRISPRa or dCas9-VPR mRNA. The graph shows composite data of four independent experiments. (**J**) CD5 expression in moDCs 2 days after nucleoporation with or without dCas9-VPR mRNA and sgRNAs. One representative experiment of five performed with five donors is shown. (**K**) Composite data of CD5 expression analyzed 1 day (n = 5 donors), 2 days (n = 4 donors), 3 days (n = 3 donors), and 5 days (n = 3 donors) after nucleofection are shown. (**L**) Frequency of IFN-γ– and TNF-α–expressing T cells after 6 days of activation with CD5_CRISPRa or dCas9-VPR mRNA moDCs. One representative experiment of three performed is shown. Data represent means ± SEM (left). Composite data of three experiments performed with three different donors are shown (right). (**M**) IFN-γ production that was measured in the culture supernatant of naïve T cells after 2 days of activation with CD5_CRISPRa or dCas9-VPR mRNA moDCs at a T:DC ratio of 33:1 or 67:1. One representative experiment is shown. Data represent means ± SEM (left). Composite data of three additional experiments performed at a T:DC ratio of 40:1 in duplicates or triplicates with three different donors are shown (right). In (B), (F) to (I), and (K) to (M), the numbers over the brackets are P values.

are of monocytic origin, these cells do not express CD5 (*18*). We then induced CD5 expression by moDCs using CRISPRa and evaluated the stability of its expression on the cell surface by flow cytometry. The highest CD5 expression was detected on day 2 (Fig. 3, J and K). CD5 was reduced on day 3 and was further reduced by day 5 (Fig. 3K). Similar to the results obtained with skin DCs, CD5_CRISPRa moDCs induced higher naïve allogeneic T cell activation than nonactivated moDCs, which was indicated by the production of IFN-γ and TNF-α, as detected by intracellular staining (Fig. 3L) and in the culture supernatant (Fig. 3M). Moreover, CD5_CRISPRa moDCs were ~1.6 times more efficient than control moDCs at prompting reactivation of autologous effector Flu-M1–specific $CD8^+$ T cells, as measured by the binding of a specific MHC tetramer (fig. S5, A and B). Overall, our data indicate that CD5 on DCs functions to activate effector T cell responses.

CD5 is functional on mouse DCs

We verified earlier studies in mice (*15*, *22*, *24*) showing that CD5 is expressed on splenic and LN DCs (fig. S6A) and further identified them particularly on resident LN DCs (fig. S6, B and C). In contrast to human DCs, CD5 was not restricted to a certain DC lineage in mice but was instead detected on the two conventional DC subsets: cDC1 ($CD24^+XCR1^+$) and cDC2 (shown as CD172a/SIRP-α^+CD24^-) (fig. S6, D to F). To gain a better understanding of the intrinsic roles of CD5 on DCs in pathological conditions in vivo, we generated conditional CD5 knockout (KO) mice using CRISPR-Cas9 technology to delete CD5 specifically in DCs [CD5ΔDC; fig. S6, A, E, and F, and (*44*)]. Compared with DCs, T cells expressed higher levels of CD5, and the T cells' CD5 expression remained high in the CD5ΔDC mice, in which *Cd5* is deleted under the control of the *Zbtb46* promotor (fig. S6, A, G, and H). This was in contrast with the *Cd11c* (*Itgax*)–Cre line, which was not specific for DCs and led to CD5 deletion in 50% of T cells (fig. S6H). Quantification of the number of CD5 molecules on the surface of the DCs revealed higher expression of CD5 on cDC2 than cDC1 (fig. S6I). Deletion of CD5 on DCs in CD5ΔDC mice was not complete relative to mice with global deletions in CD5 (CD5KO mice) (fig. S6, J and K) but was greater than in mice missing one allele of CD5 (heterogeneous) (fig. S6L). We did not detect differences in gene expression in cDC1 or cDC2 isolated from CD5ΔDC mice versus control mice (fig. S6M), particularly related to immune checkpoint proteins such as CD86, CD40, and programmed cell death ligand 1 (PD-L1) (fig. S6N). The numbers of DCs (including cDC1, cDC2, and plasmacytoid DCs) and T cells ($CD8^+$ or $CD4^+$ T cells) in CD5ΔDC mice were also comparable to those in control mice in both spleens and LNs (fig. S6O).

To evaluate the potential requirement for expression of CD5 by DCs to mount immune responses to tumors in vivo, we adopted a mouse model of tumor generation using a modified 3-methylcholanthrene (MCA)–induced syngeneic sarcoma cell line (MCA1956) that expresses membrane-associated ovalbumin (OVA) (MCA1956-mOVA) (*45*) (Fig. 4A). In this model, tumor elimination is dependent on antigen presentation by DCs through the expression of mOVA that converts MCA1956 tumors into regressor tumors. Tumor elimination was observed in 14 out of 15 wild-type (WT) (control) mice (Fig. 4, B and C). By contrast, CD5ΔDC mice (9 out of 11) failed to reject mOVA tumors, thus confirming a requirement for CD5 expression by DCs for tumor rejection (Fig. 4, B and C). Of note, a single copy of CD5 was sufficient to induce tumor rejection (Fig. 4C). Furthermore, SIINFEKL-H-$2K^b$ tetramer$^+$ $CD8^+$ T cells expanded in control mice in response to MCA1956-mOVA but were reduced in CD5ΔDC mice (Fig. 4D). Overall, our data indicate that CD5 on DCs functions to activate antitumor effector $CD8^+$ T cell responses.

CD5 expression on DCs is required for de novo priming of $CD5^{hi}CD4^+$ conventional T cells by DCs

To determine whether CD5 on DCs also regulates $CD4^+$ T cell responses, we performed antigen-specific T cell activation assays by culturing CD5-deficient DCs (from CD5ΔDC mice or CD5KO mice) or control cDC2s with purified T cells from OT-II mice (Fig. 4E) in the presence of a long peptide containing the MHC II–restricted epitope. We determined the level of T cell proliferation and activation based on CFSE dilution and CD44 expression. After loading with antigenic peptide, CD5-deficient cDC2s from CD5ΔDC mice were less efficient at activating $CD4^+$ OT-II T cells than controls (Fig. 4E). In addition, consistent with the greater deletion of CD5 in CD5KO DCs, the proliferation and activation of OT-II T cells was lower than that of the CD5ΔDC2s for all tested peptide concentrations (Fig. 4E). Moreover, the amount of CD5 on the proliferating T cells correlated with the expression on the DCs (Fig. 4F); that is, control DC-activated T cells expressed higher amounts of CD5 compared with those activated by CD5-deficient DCs (CD5ΔDC or CD5KO).

To confirm these findings in vivo, we investigated the potential requirement for CD5 on DCs for the priming of $CD5^{hi}CD4^+$ conventional T cells. We examined early $CD4^+$ T cell proliferation in response to OVA synthetic long peptide (SLP) vaccine by transferring OVA-specific OT-II transgenic CD4 T cells into control or CD5ΔDC mice (Fig. 4G). OT-II cells underwent cell division in the spleen of control mice, consistent with recognition of $OVA_{323–339}$ peptide–MHC II complexes on DCs, and the proliferating cells expressed high levels of CD5 [Fig. 4, H and I (blue)]. By contrast, CD5 expression on proliferating OT-II cells was markedly reduced in CD5ΔDC mice wherein the CD5 expression on DCs is targeted for deletion and was absent on OT-II cells activated in CD5KO mice, which are characterized by complete deletion of CD5 [Fig. 4I (red and black)]. Overall, these data show that CD5 levels on DCs have a direct impact on CD5 expression on primed T cells.

Optimal benefit of ICB therapy requires $CD5^+$ DCs

We next evaluated the potential requirement for CD5 expression by DCs to mount immune responses to tumors in vivo using mouse tumor models that are dependent on DCs for tumor elimination by ICB therapy (Fig. 4J). We first analyzed responses to MCA1956 tumors

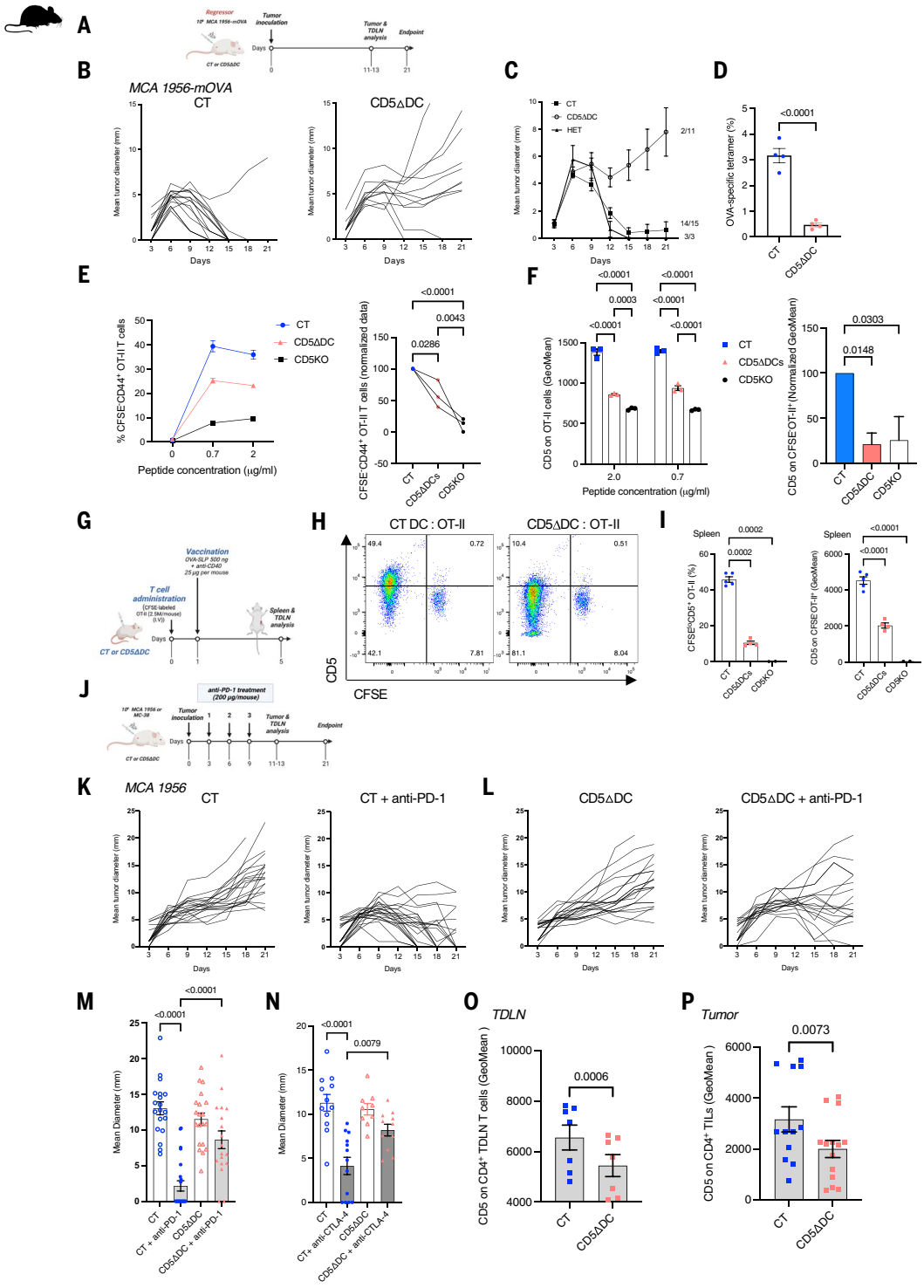


Fig. 4. Deletion of CD5 on DCs compromises the response to ICB therapy and modulates T cell immunity in vivo. (**A**) Workflow of MCA1956-mOVA tumor growth in WT control (CT) or CD5ΔDC mice. (**B**) MCA1956-mOVA tumor growth in CT or CD5ΔDC mice. Results depict tumor growth curves of individual mice from three pooled experiments: CT (WT) (n =15) and CD5ΔDC (n = 11). (**C**) MCA1956-mOVA tumor growth in CT, CD5ΔDC, or $CD5^{HET}$ heterozygous (HET) mice. Combined tumor growth curves for CT (n = 15), CD5ΔDC (n = 11), and HET (n = 3) mice are shown. Overall group difference average as measured across time for CD5ΔDC versus CT (P = 0.0055) and CD5ΔDC versus HET (P = 0.0087). There was no difference between HET and CT. Data represent means ± SEM. (**D**) Binding of OVA:I-A^d tetramer to TDLN cells isolated from MCA1956-mOVA–bearing CT and CD5ΔDC mice. Data represent means ± SEM (n = 4). (**E**) In vitro OT-II cell proliferation in response to cDC2s isolated from CT, CD5ΔDC, and CD5KO mice and cultured with SLP from the OVA protein (OVA-SLP) that contains the MHC II–restricted peptide. One representative experiment of three is shown. The plot shows means ± SEM of three experimental replicates (left). Proliferated OT-II cells in response to cDC2s loaded with 0.7 μg/ml SLP. The mean of three pooled experiments is displayed with normalization to the control (right). (**F**) CD5 mean expression on OT-II cells proliferating in response to cDC2s isolated from CT, CD5ΔDC, or CD5KO DCs (left). Data represent means ± SEM

of three pooled experiments displayed with normalization to the control (right). (**G**) Workflow of OVA-SLP vaccination and analysis of the immune response. (**H**) Representative plot showing CD5 expression on proliferating OT-II T cells in CT or CD5ΔDC mice in vivo. (**I**) Frequency of proliferating $CD5^+OT\text{-}II^+$ cells isolated from the spleen of CT (n = 5), CD5ΔDC (n = 4), or CD5KO (n = 2) mice (left). Mean CD5 expression on proliferating OT-II T cells in CT (n = 5), CD5ΔDC (n = 4), or CD5KO (n = 2) mice in response to SLP vaccine in vivo (right). (**J**) Workflow of the analysis of immune cell infiltration of MCA1956 tumors in mice treated with anti–PD-1 or IgG2a isotype control. Part of the illustration was created with Biorender.com. (**K**) Tumor growth from CT mice injected with immunoglobulin G2a (IgG2a) isotype control or anti–PD-1 antibodies (18 to 23 mice per group). Data represent the results of four pooled experiments. (**L**) Tumor growth from CD5ΔDC mice injected with IgG2a isotype control or anti–PD-1 antibodies (18 to 23 mice per group). Data represent the results of four pooled experiments. (**M**) Mean tumor diameter in CT or CD5ΔDC mice on day 21 after implantation and treatment with IgG2a isotype control or anti–PD-1 antibodies. Data represent the results of four pooled experiments. (**N**) Mean tumor diameter in CD5ΔDC and CT mice on day 24 after implantation and treatment with isotype control or anti–CTLA-4 antibodies. (**O**) CD5 expression by $CD4^+$ T cells in the lymphoid immune compartment in the TDLNs of CD5ΔDC or CT mice treated with isotype control or anti–PD-1 antibodies. (**P**) CD5 expression by $CD4^+$ T cells in the lymphoid immune compartment in the tumor of CD5ΔDC or CT mice treated with isotype control or anti–PD-1 antibodies. In (D) to (F), (I), and (M) to (P), the numbers over the brackets are P values. [Illustrations in (A), (G), and (J) created with Biorender.com]

in WT control and CD5ΔDC mice. MCA1956 tumors grew progressively in both control and CD5ΔDC mice [Fig. 4, K (left) and L (left)]. However, whereas control mice responded effectively to anti–PD-1 ICB therapy ($P > 0.0001$) (Fig. 4K, right), CD5ΔDC mice responded less efficiently and the tumor was not rejected (Fig. 4L, right). Whereas 18 out of 23 tumor-bearing control mice responded to anti–PD-1 treatment and 15 out of 23 completely rejected the tumor, only 2 out of 18 CD5ΔDC mice rejected tumors, with an average tumor size at day 21 after anti–PD-1 treatment of 2.1 ± 0.7 mm in control mice compared with 8.6 ± 1.2 mm in CD5ΔDC mice (Fig. 4M). Similar results were obtained using anti–CTLA-4 therapy (Fig. 4N). Compared with control mice, CD5ΔDC responded less efficiently to the treatment. The average tumor size at day 18 after anti–CTLA-4 treatment was 4.1 ± 0.9 mm in control mice compared with 8.2 ± 0.6 mm in CD5ΔDC mice. At this early time point, 5 out of 13 control mice had already rejected the tumors, whereas none of the CD5ΔDC mice showed complete rejection of the tumors (Fig. 4N).

The $CD5^{lo}$ T cell population is expanded in the TME of CD5ΔDC mice

To understand the impact of the absence of CD5 expression on DCs in shaping the T cell milieu in the TME, we examined the immune infiltrates in the tumor and TDLNs of MCA1956-bearing control or CD5ΔDC mice by flow cytometry (Fig. 4, O and P). We found that $CD8^+$ and $CD4^+$ T cells were equally represented in tumor infiltrates from control and CD5ΔDC mice after ICB therapy (fig. S7A), and no difference in regulatory T cell infiltrates was detected (fig. S7B). Although no difference in CD5 expression was observed on $CD4^+$ T cell subsets under steady-state conditions (fig. S7C), CD5 expression on the primed $CD4^+$ T cells in the TDLNs of CD5ΔDC mice was lower than that in the control mice after ICB therapy (5443 ± 436 versus 6555 ± 494). This further emphasizes that CD5 on DCs modulates CD5 expression by $CD4^+$ T cells, which are critical for tumor rejection, during an immune response at the site of activation (Fig. 4O) (*46*).

Reduced CD5 expression on $CD4^+$ T cells was also detected in the tumor infiltrates of CD5ΔDC mice relative to CT mice (2002 ± 337.3 versus 3161 ± 495) (Fig. 4P). CD5 expression on the $CD4^+$ T cells in the tumor was lower than that detected in the TDLNs (3161 ± 495 versus 6555 ± 494) [Fig. 4, O and P (blue)]. For control tumor-infiltrating lymphocytes, CD5 expression on the $CD8^+$ T cells was lower than that on the $CD4^+$ T cells (1851 ± 302 versus 3161 ± 495), with no difference in CD5 expression on $CD8^+$ T cells in control mice compared with that of the corresponding population in CD5ΔDC mice (fig. S7, D and E). Anti-CD3 induced CD5 up-regulation in T cells that were isolated from CD5ΔDC mice, which confirmed that the priming of $CD5^{lo}$ T cells was directly affected by CD5-deficient DCs (fig. S7F). Anti–PD-1 or anti–CTLA-4 alone did not modulate CD5 expression on T cells (fig. S7F), supporting the requirement for T cell receptor (TCR) stimulation by DCs. Additionally, the frequency of inducible T cell costimulator $(ICOS)^+CD4^+$ T cells was reduced in tumor infiltrates of CD5ΔDC mice compared with those of control mice (fig. S7G, left). By contrast, a higher frequency of T cell immunoreceptor with immunoglobulin and immunoreceptor tyrosine-based inhibitory motif domains $(TIGIT)^+CD4$ T cells was found in the tumors of CD5ΔDC mice after anti–PD-1 treatment (fig. S7G, right). The frequency of activated (fig. S7H, left) and granzyme B–expressing $CD8^+$ T cells was also reduced in tumor infiltrates of CD5ΔDC mice (fig. S7G, right).

We next used the MC-38 colorectal cancer model to determine the effect of the deletion of CD5 expression by DCs in the priming of tumor-specific $CD8^+$ T cells. Neoantigen-specific $CD8^+$ T cells that are reactive against the mutated epitopes from the adenosine diphosphate (ADP)–specific glucokinase (ADPGK) proteins were detected in control tumors. However, the frequency of these antigen-specific $CD8^+$ T cells was reduced in tumors from CD5ΔDC mice (fig. S8, A and B). Moreover, CD5 expression on the antigen-specific T cells was lower after priming by CD5ΔDCs relative to that of cells primed by control DCs (fig. S8C). Overall, these data show that deletion of CD5 expression on DCs favors the expansion of $CD5^{lo}$ T cells and the reduction of tumor-specific $CD8^+$ T cells.

CD5 deletion on T cells dampens tumor elimination in response to ICB therapy

To address the impact of low CD5 expression by T cells in restraining tumor rejection in CD5ΔDC mice, we developed mice with CD5-deficient T cells by crossing $Cd5^{flox/flox}$ ($Cd5^{f/f}$) mice with $Cd4^{cre+}$ mice (CD5ΔT) (fig. S9, A and B). In these animals, CD5 expression is maintained on DCs but is deleted in both $CD4^+$ and $CD8^+$ T cells (fig. S9, C and D). There was no significant difference in the frequency of T cells (fig. S9E) or DC frequency (fig. S9F) in CD5ΔT mice compared with control mice or in the expression of CD3, activating (ICOS), or inhibitory (PD-1) molecules (fig. S9G). MCA1956 tumor growth was comparable in control and CD5ΔT mice (Fig. 5, A and B). However, whereas the majority of the control mice (13 out of 16) responded to anti-PD1 treatment, with eight completely rejecting the tumor, none of the CD5ΔT mice (0 out of 14) responded to the treatment (Fig. 5, C and D). Furthermore, CD5ΔT mice (25 out of 27) failed to reject MCA1956-mOVA cells, which were rejected spontaneously in control mice (Fig. 5E; $P < 0.0001$). This was also reflected by the reduced frequencies of OVA-positive $CD8^+$ T cells in the tumor (Fig. 5F, left) and spleen (Fig. 5F, right) of CD5ΔT mice. This result was not restricted to the MCA1956 tumor because the failed tumor rejection by CD5ΔT mice in response to anti–PD-1 treatment was also observed in the murine MC-38, adenocarcinoma of the colon, tumor model (fig. S10). Control mice efficiently rejected the tumor in response to anti–PD-1 treatment (P = 0.0007); however, CD5ΔT mice failed to respond (fig. S10A), which correlated with reduced frequencies of tetramer-positive $CD8^+$ T cells specific for the neoantigens ADPGK and RALBP1 associated Eps domain containing 1 (REPS1) in the tumor (fig. S10, B and C). Collectively, these data suggest that high CD5 expression on T cells facilitates the control of tumor growth in response to ICB therapy.

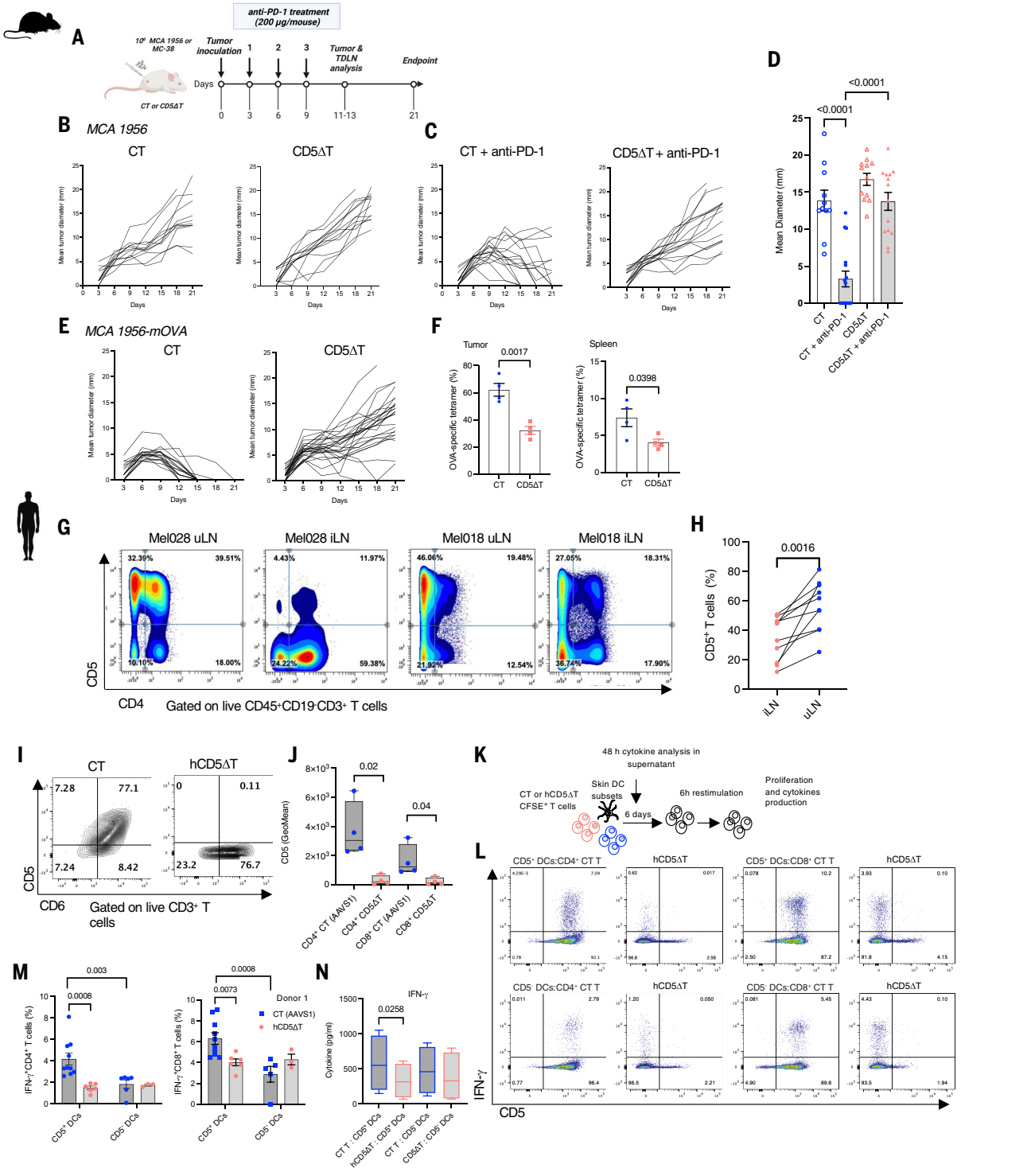


Fig. 5. Deletion of CD5 on T cells compromises effector T cell priming and response to ICB therapy. (**A**) Workflow of tumor growth in control (CT) or CD5ΔT mice. MCA1956 tumor growth in CT (WT) or CD5ΔT mice injected with IgG2a isotype control or anti–PD-1 antibodies. [Illustration created with Biorender.com] (**B**) Results depict tumor growth curves of individual mice from three pooled experiments for CT (n = 12) and CD5ΔT (n = 13). (**C**) Results depict tumor growth curves of individual mice from three pooled experiments for CT with anti–PD-1 (n = 13) and CD5ΔT with anti–PD-1 (n = 16). Overall group difference as measured across time for CD5ΔT versus CT (P < 0.0001). (**D**) Average tumor growth on day 21 in CT or CD5ΔT mice treated with anti–PD-1 or isotype control IgG2a. Data represent means ± SEM. (**E**) MCA1956-mOVA tumor growth in CT or CD5ΔT mice. Results depict tumor growth curves of individual mice from three pooled experiments for CT (n = 15) and CD5ΔT (n = 8). Overall group difference as measured across time for CD5ΔT versus CT (P < 0.0001). (**F**) Frequency of OVA-specific $CD8^+$ T cells that were detected in the tumor (left) and spleen (right) of CT and CD5ΔT mice 10 days after tumor cell inoculation. Data represent means ± SEM. (**G**) Expression of CD5 on T cells (gated on live, $CD45^+CD19^-CD3^+$; plots show CD4 expression versus CD5 expression) in iLNs and uLNs of two representative patients, Mel028 and Mel018. (**H**) Composite data showing the frequency of $CD5^+$ T cells in iLNs and uLNs of 10 patients with melanoma. Data represent means ± SEM. (**I**) CD5 and CD6

expression on CT or hCD5ΔT cells after coculture with DCs. (**J**) CD5 geometric mean expression on CT or hCD5ΔT CD4$^+$ or CD8$^+$ T cells before coculture with DCs. Composite data of four donors are shown. Data represent means ± SEM. (**K**) Experimental scheme for (L) to (N). (**L**) Frequency of IFN-γ expression by CD4$^+$ and CD8$^+$ T cells in CT or hCD5ΔT cultures with CD5$^+$ or CD5$^-$ dermal DCs. Data represent one experiment of four performed. (**M**) Flow cytometric frequency analysis of IFN-γ expressed by CT or hCD5ΔT CD4$^+$ (left) or CD8$^+$ (right) T cells cocultured with either CD5$^+$ or CD5$^-$ DCs. Data represent one experiment of four performed with four different donors (see also fig. S11E). Each dot indicates a technical replicate. Data represent means ± SEM. (**N**) IFN-γ production measured in the culture supernatant of either CT T cells or hCD5ΔT cells cocultured with either CD5$^+$ or CD5$^-$ DCs. Composite data of four experiments performed with four donors are shown. Data represent means ± SEM. In (D), (F), (H), (J), (M), and (N), the numbers over the brackets are *P* values.

CD5 deletion on human T cells mirrors CD5 deletion on DCs

To determine the correlation between CD5$^+$ DC density and CD5 levels on T cells in human tumors, we assessed the expression of CD5 on T cells in iLNs and patient-matched uLNs by CyTOF. We found that a larger fraction of T cells (both CD4$^+$ and CD8$^+$ T cells) in the uLNs expressed CD5 and that this proportion was reduced in the corresponding metastatic tumor LNs (Fig. 5, G and H). To determine the functional requirement for CD5 expression on human T cells, we designed an sgRNA for *CD5* disruption in primary human T cells using a nonviral CRISPR-Cas9–based protocol (*47*). We confirmed the gene deletion efficiency by insertion-deletion (indel) analysis (fig. S11A) and flow cytometry (Fig. 5I and fig. S11B). CD5 disruption was effective in both CD4$^+$ and CD8$^+$ T cells (Fig. 5J and fig. S11B). Similar to mice, the expression of CD5 was higher on CD4$^+$ T cells than on CD8$^+$ T cells (Fig. 5J and figs. S9A and S11B). CD6, the expression of which parallels CD5 expression in T cells (*48*, *49*), was not affected in the human CD5-deleted T cells (hCD5ΔT) (Fig. 5I and fig. S11C). In addition, there was no substantial difference in the expression of activating and inhibitory receptors, including CD69, PD-1, CCR7, and CD25 (fig. S11, C and D, and table S3). We then stimulated unedited control T cells or hCD5ΔT cells with CD5$^-$ or CD5$^+$ DCs and measured the production of effector cytokines (Fig. 5, K to N, and fig. S11, E and F). Compared with the control cells, hCD5ΔT cells (both CD4$^+$ and CD8$^+$) produced lower amounts of IFN-γ. Indeed, CD5 expression on T cells was important for T cell activation but was inhibited when CD5 was also expressed by the primed DCs, indicating a critical role for CD5 on DCs in the interaction of CD5 on T cells (Fig. 5, L to N, and fig. S11, E and F). This was determined by the amounts of cytokines detected by intracellular staining of IFN-γ in CD4$^+$ (Fig. 5M, left) and CD8$^+$ T cells (Fig. 5M, right, and fig. S11E) and in the culture supernatant amounts of IFN-γ and IL-5 (Fig. 5N and fig. S11F). Although gene expression analysis of CD5 ligation on T cells indicated no changes in expression (table S4), T cell CD5 potentiated the production of TCR-induced T cell cytokines IFN-γ and TNF-α when cross-linked by two clones of anti-CD5 mAb [#OKT1 (fig. S12A) and Novus #T1 (fig. S12B)] that function as agonists. Overall, these data demonstrate that engagement of CD5 on T cells in concert with TCR stimulation potentiates inflammatory T cell responses, which are a critical component of the response to ICB therapy in patients.

CD5$^+$ DCs are increased in response to ICB therapy

To better understand the mechanism by which CD5$^+$ DCs promoted a favorable response to ICB therapy in control mice, we examined the frequency of CD5$^+$ DCs in the tumors and TDLNs of mice 11 to 13 days after MCA1956 cell inoculation (Fig. 6A). Although the DC (fig. S13A) and macrophage (fig. S13B) frequencies were not altered in response to ICB therapy, such treatment up-regulated the CD5$^+$ DC frequency in both the tumor (Fig. 6B) and the TDLN (Fig. 6C) DC1 and DC2 compartments. This effect of anti–PD-1 was also seen with the MC-38 tumors (fig. S13C). Mechanistically, this effect was likely not due to enhanced migration. CD5$^+$ DCs were mainly found within resident LN DCs (identified by IE/IA^{lo}CD11c^{hi}; fig. S13D, left) and showed lower levels of CCR7 than the CD5$^-$ DCs (fig. S13D, right), albeit its expression was slightly increased after anti–PD-1 treatment, particularly on CD5$^+$ DC1s (fig. S13E). In addition, there was no impact on the CD5$^+$ DC proliferative state after anti–PD-1 treatment, although CD5$^+$ DCs expressed higher levels of K_i-67 than the CD5$^-$ DCs (fig. S13F). Importantly, however, CD5$^+$ DCs were more resistant than CD5$^-$ DCs to apoptosis in vitro (Fig. 6D), and CD5$^+$ DCs, including CD5$^+$ DC1s and CD5$^+$ DC2s but not CD5$^-$ DCs, were ~50% more resistant to apoptosis in the presence of anti–PD-1 in vivo (Fig. 6E and fig. S13G). Overall, these data indicate that CD5$^+$ DCs are targets of immune suppression in the TME, thus implicating the priming of new T cells by CD5$^+$ DCs as a critical step for the efficiency of ICB therapy.

Low levels of IL-6 promote the development of CD5$^+$ DCs

To determine factors that might promote CD5$^+$ DC development and persistence in response to anti–PD-1, we studied the properties of DCs from a patient with inherited PD-1 deficiency. This revealed that a higher proportion of DCs in this patient expressed CD5 compared with control subjects (Fig. 6F) (*50*). We thus hypothesized that this increased frequency of CD5$^+$ DCs might be caused by the excessive IL-6 production in this patient. Moreover, we detected an increase in IL-6 production by immune cells in the TME of mice treated with anti–PD-1 (Fig. 6G). Indeed, the presence of IL-6 during DC differentiation from bone marrow (BM) progenitors favored the development of CD5$^+$ DCs (Fig. 6, H and I). High concentrations of IL-6 (250 ng/ml) skewed the development from DC1 toward DC2 [SIRP-α$^+$; fig. S13H (dark blue)] (*51*, *52*); however, low IL-6 concentrations (below 20 ng/ml) increased the expression of CD5 on both of these DC subsets [Fig. 6, H and I, and S13H (light blue)]. The same pattern was observed in human DCs, with the presence of IL-6 found to promote the differentiation of CD5$^+$ DCs from cord blood CD34$^+$ hematopoietic progenitor cells (HPCs) [Fig. 6, J and K, and fig. S13, I, J (dark blue), and K]. To determine whether IL-6 up-regulates CD5 expression rather than increasing the differentiation of CD5-expressing murine DCs, purified splenic DCs were exposed to IL-6 for 5 days (fig. S13L). Similarly, sorted human dermal CD5$^+$ and CD5$^-$ DCs were exposed to IL-6 (fig. S13M). Under these conditions, CD5 remained on the surface of the positive cells, and its expression did not change on terminally differentiated mouse DCs (fig. S13L) or human DCs [fig. S13M, left (blue)]. Moreover, CD5 expression was not detected on IL-6–stimulated CD5$^-$ DCs [fig. S13M, left (red) and right]. IL-23, which was also abundant in the PD-1–deficient patient, did not affect CD5 expression on CD5$^-$ DCs (fig. S13N) nor did it have any impact on CD5$^+$ DC differentiation from mouse BM progenitors (fig. S13O) or human CD34$^+$ HPCs [fig. S13, J (orange) and P]. Moreover, ligation of CD5 on human DCs with an agonistic mAb (fig. S14, A and B) promoted the production of cytokines and up-regulation of gene pathways related to IL-6 and TNF-α signaling (fig. S14C), as well as other genes associated with DC differentiation [colony-stimulating factor 2 (*CSF2*), *GMCSF*], survival (*BCL2*), and tumor immune cell infiltrate (*SERPINE1*) (*53*, *54*) (fig. S14, D and E, and tables S5 and S6). Thus, low levels of IL-6 promote the differentiation of CD5$^+$ DCs and may further explain the reduced apoptosis and increased frequencies of CD5$^+$ DCs after PD-1 inhibition.

Discussion

In this study, we investigated the mechanisms that underlie DC-mediated effector T cell priming, focusing on the role of CD5 expressed by these interacting cell types. Our results

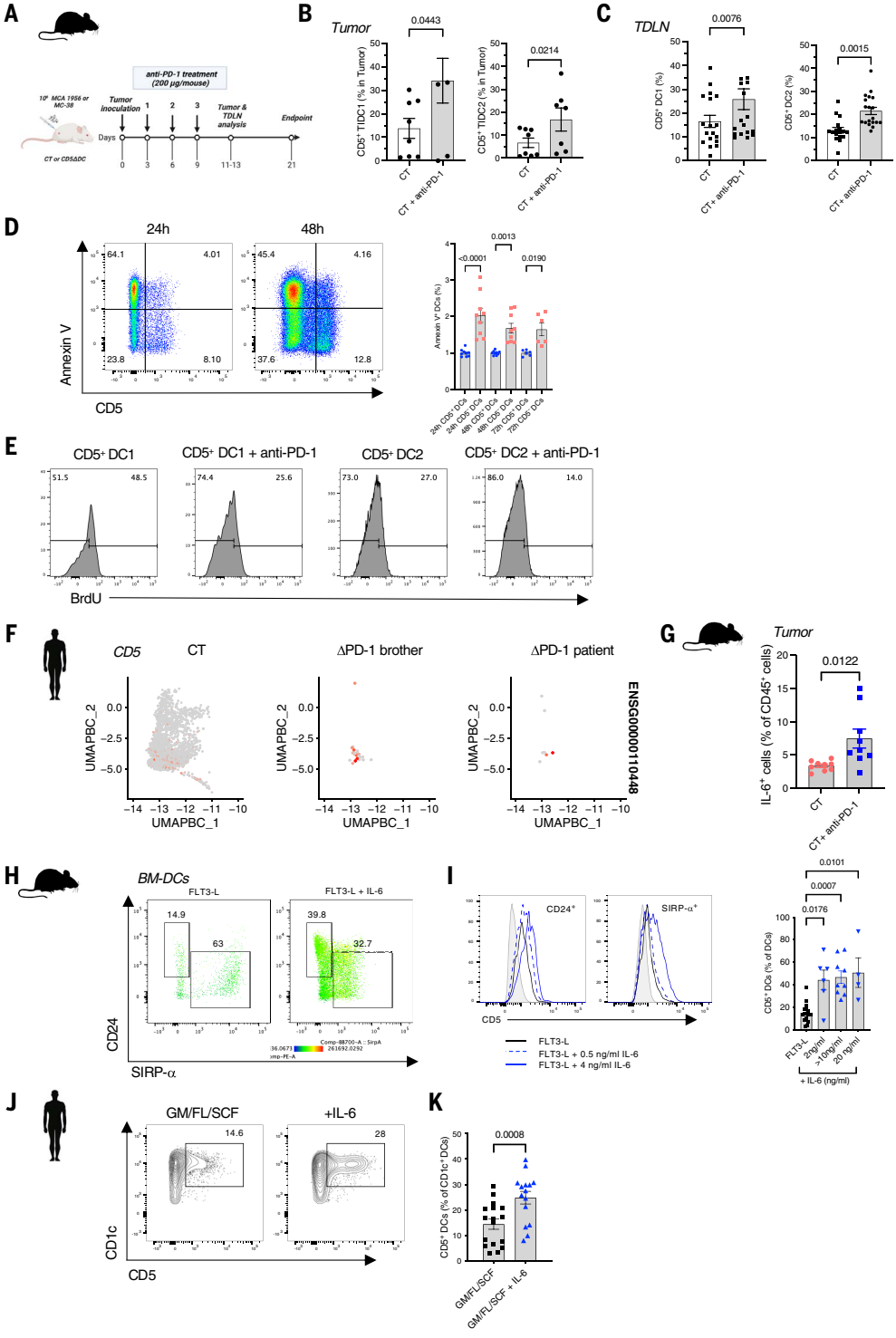


Fig. 6. CD5$^+$ DCs are modulated in the TME. (**A**) Workflow of tumor growth and immune cell infiltrate analysis in control (CT) or CD5ΔDC mice. [Illustration created with Biorender.com] (**B**) CD5 expression of tumor-infiltrating cDC1 (TIDC1) (left) or cDC2 (TIDC2) (right) from CT mice treated with anti–PD-1 (n = 8; three data points are outside the axis limits for TIDC1) or IgG2a isotype control (n = 7). Data represent means ± SEM. (**C**) Frequency of CD5$^+$ cDC1 (left) or CD5$^+$ cDC2 (right) in TDLNs of CT mice treated with IgG2a isotype control (n = 18) or anti–PD-1 antibodies (n = 19). Data represent means ± SEM. (**D**) Plots show the expression of Annexin V by CD5$^+$ and CD5$^-$ splenic DCs after 24 and 48 hours of incubation. One representative experiment is shown. Cells are gated on live, lineage$^-$F4/80$^-$CD64$^-$IE/IA$^+$CD11c$^+$ cells (left). Composite data of three experiments are shown. Each dot represents a different mouse. Annexin V expression was normalized to CD5$^+$ DCs for each indicated time point (right). Data represent means ± SEM. (**E**) CD5$^+$ cDC1s and CD5$^+$ cDC2s isolated from TDLNs of MCA1956 tumor–bearing mice treated with IgG2a isotype control or anti–PD-1 antibodies and stained with an apoptosis detection antibody (APO-BrdU TUNEL assay). Composite data of three mice per group from one of two independent experiments are shown. (**F**) UMAP shows *CD5* gene expression by the myeloid DC clusters isolated from scRNA-seq of the peripheral blood mononuclear cells from a patient with inherited *PD-1* deficiency and the patient's brother, as well as control subjects. The cells were analyzed per the authors'

Mendeley data report [(*50*); see https://data.mendeley.com/datasets/nb26v3mx3x/2]. (**G**) Frequency of IL-6–producing immune cells within tumor-infiltrating immune cells of mice bearing MCA1956 tumors and treated with IgG2a isotype control or anti–PD-1. Data represent means ± SEM (n = 9 mice per group). (**H**) DC1 ($CD24^+$) and DC2 ($Sirp\text{-}\alpha^+$) output from BM cells that were cultured with either Fms-like tyrosine kinase receptor 3 ligand (FLT3-L) or FLT3-L and 4 ng/ml IL-6 for 8 days. CD5 median expression is shown in heatmap statistic plots. (**I**) BM cells were cultured for 8 days with FLT3-L (black line), FLT3-L with 0.5 ng/ml IL-6 (dashed blue line), or FLT3-L with 4 ng/ml IL-6 (solid blue line). CD5 expression on output BM DC1s or DC2s (left) is shown. Composite data showing $CD5^+$ DC output from mouse BM DCs that were cultured with FLT3-L and the indicated amounts (0 to 20 ng/ml) of IL-6; each dot represents an independent experiment (right). Data represent means ± SEM. (**J**) CD5 expression on $CD1c^+$ DCs derived from human $CD34^+$ HPCs that were differentiated with GM-CSF, FLT3-L, and SCF (GM, FL, and SCF) in the presence or absence of IL-6 (100 ng/ml) for 7 days. One experiment is shown. (**K**) $CD5^+$ DC output within the CD34-derived $CD1c^+$ compartments. Cells are gated on live, $lineage^-HLA\text{-}DR^+CD11c^+CD1c^+$. The graph shows composite data of 10 independent experiments. In (B) to (D), (G), (I), and (K), the numbers over the brackets are P values.

demonstrate that CD5 on DCs serves as an immunostimulatory receptor that is important for de novo priming of antitumor T cells and the response to ICB therapy. The amount of CD5 expressed on DCs correlated directly with the extent of effector T cell priming in humans. Selective deletion of CD5 expression by DCs prevented an efficient response to ICB therapy in tumor-bearing mice, which resulted in defective immune rejection of tumors. The cause of this defect was likely that the activation of $CD5^{lo}$ T cells with poor effector function was favored. Similarly, deletion of CD5 expression by T cells negatively affected antitumor immunity and the response to ICB. Importantly, we found increased numbers of $CD5^+$ DCs after ICB in vivo and identified IL-6 as an important factor for $CD5^+$ DC differentiation and survival. Thus, based on our model, we predict that a high density of $CD5^+$ DC2 cells is likely to align with an increased frequency of tumor-specific T cells in patients as well as improvements in $CD5^{hi}$ T cell activity.

We previously identified CD5 on a subset of human $CD1c^+$ DCs (DC2) in migratory skin DCs, peripheral blood, and cord blood. These cells showed a superior capacity to prime T cell responses relative to their $CD5^-$ DC2 counterparts and their numbers were also elevated in inflamed skin (*18*). We hypothesized that in cancer, CD5 expression on DCs has a role in determining the fate of tumor-specific T cells, which renders lymphocytes more capable of efficiently recognizing and eliminating malignant cells. In this study, we indeed found that these cells are present in tumor-free LNs and are reduced in melanoma-affected LNs. Moreover, we showed that CD5 expression, and, specifically, a $CD5^+$ DC signature, are associated with a better prognosis for patients with melanoma, lung squamous cell carcinoma, sarcoma, breast cancer, cervical squamous cell carcinoma, and endocervical adenocarcinoma. CD5 is also an independent prognostic biomarker for overall survival in the early stages of non–small cell lung cancer (*55*). Analysis of TCGA datasets for the statistical significance of T cell–specific *CD5* expression on the overall survival of melanoma patients was not conclusive. This could be due to issues with the analysis or could indicate that the differences in CD5 protein expression do not correspond with the mRNA levels in T cells. Alternatively, the genetic variability of *CD5* might provide an explanation. Single-nucleotide variations s2241002 (C/T; Pro224Leu) and rs2229177 (C/T; Ala471Val) have been associated with severe forms of systemic lupus erythematosus, rheumatoid arthritis, Crohn's disease (*56*), and certain types of cancer (*57*). Furthermore, an inherited functional variant of CD5 influenced melanoma survival (*58*). Thus, different variants might affect $CD5^+$ DC function, interaction with T cells through CD5, signaling, T cell persistence, and, ultimately, patient survival and response to ICB therapy.

The absence of $CD5^+$ DCs in iLNs of patients with melanoma suggests that the TME negatively regulates the differentiation of these DCs from progenitors or, alternatively, reduces their survival. Indeed, we found that CD5 signaling (in humans and mice) enhanced DC survival. Additionally, the treatment of mice with anti–PD-1 increased the frequency of $CD5^+$ DCs in the tumor and lymphoid tissue. Given the apparent irreversibility of certain forms of exhaustion (*59*), it is possible that the efficacy of T cell–targeted immunotherapies is linked to ongoing de novo LN priming by these $CD5^+$ DCs rather than merely a blockade of checkpoint ligands in the tumor. Harnessing the CD5 costimulatory pathway in DCs during priming may increase the efficacy of immunotherapies in conventional T cells. Practically, we envision using these cells for immunotherapy applications. Although we could not identify $CD5^+$ DCs in tonsils, and though they were scarce in human spleen, $CD5^+$ DCs can be efficiently isolated from peripheral blood mononuclear cells and cord blood and can also be differentiated from HPCs (*18*). Moreover, antibodies against CD5 might be used therapeutically to inhibit T cell activation in autoimmunity by blocking their interaction with CD5 on DCs. Indeed, although they may engage CD5 on T cells, we found that antibodies against CD5 could only inhibit T cell activation when CD5 was also expressed by the primed DCs, which suggests that the interaction between CD5 on DCs and T cells is critical for T cell activation.

We help explain the mechanism through which effective ICB therapy alters these signals to favor $CD5^+$ DC differentiation and promote tumor rejection. This mechanism was revealed by investigating the DCs of a patient with inherited PD-1 deficiency (*50*). Such patients, who are susceptible to tuberculosis and autoimmunity, have an immunological profile that mirrors that of patients with signal transducer and activator of transcription 3 (STAT3) gain-of-function disease (*51*) and which is characterized by the production of a large amount of IL-6 by their T cells. In addition, we found that a large proportion of the DCs from one of these patients expressed CD5. Thus, we hypothesized that IL-6 plays a role in $CD5^+$ DC differentiation or survival. Indeed, we identified IL-6 as a factor that drives the differentiation of $CD5^+$ DCs and that could explain the reduced apoptosis of these cells after ICB. It appears that CD5 is induced early on during the differentiation of $CD1c^+$ DCs from $CD34^+$ HPCs (via the $CD14^+CD1c^+$ DC transitioning stage). Moreover, ligating CD5 with an agonistic mAb promoted the production of IL-6 from DCs, which can maintain high expression of CD5 through an autocrine mechanism, thereby inducing de novo differentiation of $CD5^+$ DCs or promoting their survival. CD5 expression by DCs is sensitive to low levels of IL-6, which is expressed in immune cells after ICB therapy. Indeed, we found that low levels of IL-6 were sufficient to promote CD5 expression on DCs without skewing their differentiation toward DC2 or macrophage lineages. Whether other factors could also regulate or promote $CD5^+$ DC differentiation from hematopoietic progenitors or monocytes, as well as the immune cell–specific source of IL-6 during ICB in vivo, remains to be established.

Based on studies performed primarily with murine T cells, CD5 was posited as a negative regulator of T cell function (*27*, *28*, *60*). However, emerging studies indicate its ability to induce T cell activation and pathogenic T_H17 cell differentiation (*34*, *35*). Thus, the functional relevance of CD5 expression on cells remains unresolved and may vary with cell type or microenvironment (*36*, *37*). CD5 expression by DCs was recognized in humans (*18*, *19*) only recently, although it was previously reported in mice (*15*, *22*–*24*). A study by Li *et al.* proposed that CD5 expression by DCs may be correlated with decreased effector T cell

activation and proinflammatory cytokine production (29) and a reduced antitumor response. Multiple factors may explain this discordance with our findings. The study by Li *et al.* mainly used transferred BM-derived DCs (BM-DCs) from CD5KO and control mice to drive their conclusion. The transferred cell subset composition may not be similar between CD5KO and control BM-DCs. In addition, the authors used a Toll-like receptor 4 (TLR4) agonist to trigger cell cytokine production; however, in such CD5-independent DC activation, the expression of this TLR4 on CD5KO and CT DCs may differ. As our data indicate, $CD5^-$ DCs are more sensitive to apoptosis than their $CD5^+$ DC counterparts; therefore, increased cell death of CD5KO DCs may lead to increased inflammation, antigen uptake, and activation of resident WT DCs (61). Further, an enhanced transfer of peptide or peptide-MHC complex (62, 63) from migratory DCs to LN resident DCs in the recipient mice might occur in these types of experiments and may explain the observed increase in T cell proliferation.

Here, we show that CD5 deletion on DCs leads to the priming and activation of $CD5^{lo}$ $CD4^+$ T cells in vivo. CD5 expression on the T cells did not correlate directly with proliferation, as indicated by the dilution of CFSE dye, but was rather affected by the CD5 signal provided by the DCs. In addition to studies in which CD5 was identified as a marker of T cell activation (27, 64), studies using polyclonal and TCR transgenic T cells showed that mature T cell effector functions and memory responses correlate with CD5 expression (65–68). Indeed, T cells that emerge from the thymus with a higher affinity for self-antigens (expressing high levels of CD5; $CD5^{hi}$) also have an increased affinity for a foreign antigen, with a distinct advantage in becoming engaged in both homeostatic and antigen-mediated responses compared with their $CD5^{lo}$ counterparts (66, 69). Moreover, the gene expression profiles of $CD5^{hi}$ and $CD5^{lo}$ naïve T cells suggest that $CD5^{hi}$ cells are transcriptionally poised to engage in both proliferative and effector functions far more rapidly than $CD5^{lo}$ cells of the same specificity (67, 70). CD5 expression is reduced on tolerized cells relative to effector T cells (71), and T cells isolated from the infiltrate of human lung carcinoma patient tumors have lower CD5 expression than T cells with the same antigen reactivity that are isolated from the peripheral blood of the same subject. Our analysis of T cells isolated from melanoma iLNs and uLNs reveals that a higher fraction of CD5-expressing T cells is confined to the uninvolved tissue. Based on our results, it can be speculated that the response to immunotherapy might be less efficient in these patients compared with those exhibiting mixed CD5 expression levels on their tumor-infiltrating lymphocytes (72).

To understand the functional impact of CD5 on T cell responses in vivo, we developed a mouse model in which CD5 expression is controlled in a cell type–specific manner. To delete CD5 in DCs, we used the *Zbtb46*-Cre line because of its high specificity for DCs and DC progenitors (73). By contrast, the *Cd11c (itgax)*–Cre line, which is not specific for DCs, resulted in CD5 deletion in 50% of T cells. The caveat of the *Zbtb46*-Cre line, however, is that one *Cd5* allele is deleted in all progeny. Nevertheless, in spite of the deleted copy in $Zbtb46^{cre+}Cd5^{f/-}$ (CD5ΔDC) mice, the CD5 expression on the T cells was comparable with that of control mice ($Cd5^{f/f}$ or WT), and CD5 expression on DCs was reduced as expected. The expression of *Zbtb46* in DC progenitors might explain the transient *Cre* expression in the germline.

The inhibition of tumor immunity was more pronounced upon deletion of CD5 when using a *Cd4*-Cre line compared with the *Zbtb46*-Cre line. This effect might be due to the complete deletion of CD5 obtained using this line as opposed to the partial deletion seen with the *Zbtb46*-Cre line. Alternatively, there might be an additive effect due to the reduction in CD5 on some DCs in addition to the deletion in T cells in the *Cd4*-Cre line. Indeed, we noted that DCs were affected in this model (because of their expression of *Cd4*) and saw some reduction in CD5 expression on the DCs, particularly in LN DCs.

The extracellular domain of CD5 consists of three subdomains that belong to the scavenger receptor cysteine-rich superfamily. The bona fide ligand or ligands for CD5 remain uncertain but may include the C-type lectin CD72 (74), an inducible receptor on T cells (64), or CD5 itself through homophilic interactions between cells (75). CD5 may also bind and respond to IL-6 on B cells (76), as well as to fungal cell wall components (77). In the CD5ΔDC mice, the T cells that were induced in response to tumor challenge expressed lower levels of CD5 than WT DCs. This suggests that high expression of CD5 on T cells might be maintained through a homotypic interaction with CD5 on DCs (75). Alternatively, CD5 on DCs might interact with an unknown ligand on T cells that indirectly regulates CD5 expression on T cells. CD5 expression on the T cells in our tumor-bearing CD5ΔDC mice was reduced on the antigen-specific $CD8^+$ T cells, as well as on $CD4^+$ T cells in both LNs and tumors. The impact of this low CD5 expression might translate directly into a reduction in the helper or effector functions that are required for successful tumor rejection in response to ICB therapy (46) or into poor licensing of cDC1s (45), which ultimately has a negative impact on CTL-mediated antitumor immunity.

Taken together, our findings highlight the requirement of the $CD5^+$ DC population for directing antitumor $CD4^+$ and $CD8^+$ T cell immunity. Furthermore, $CD1c^+CD5^+$ DC abundance in human lymphoid tissue might serve as a biomarker not only of T cell effector quality but also of responsiveness to ICB therapy. Classifying the TME based on the immune infiltrate has predictive power (78). Thus, in addition to recent efforts to identify distinctive patient TMEs (38, 79, 80), our unbiased, high-dimensional analytic approach has identified CD5 expression on DCs as a valuable component that could facilitate the design of treatment protocols and may help identify patients that are poised for a therapeutic response.

Materials and methods summary

Tissue specimens

Human tissues (including skin, blood, and peripheral lymphoid organs) were obtained from healthy volunteers and patients with melanoma who underwent cosmetic and therapeutic surgical procedures at Washington University in St. Louis and Barnes-Jewish Hospital in accordance with the guidelines of the institutional review board (IRB). Samples were processed immediately, and cell subsets were isolated. A written informed consent was obtained in accordance with the guidelines of the IRB from patients with melanoma as well as healthy donors who donated skin and whole blood (listed in table S1).

Human DC subsets

Human tissue DCs were isolated from tissues using mechanical and enzymatic methods followed by fluorescence-activated cell sorting (FACS) as needed. Human DCs were also generated in vitro from monocytes or from cord blood–derived $CD34^+$ HPCs with GM-CSF, Fms-like tyrosine kinase receptor 3 ligand (FLT3-L), and stem cell factor (SCF). Where noted, IL-6 and IL-23 were added at the indicated concentrations, as described previously (18) and in (44).

Correlating CD5 expression with disease prognosis

TCGA level 3 normalized RNA-seq–based expression for 26 cancer types was assessed using the University of California, Santa Cruz (UCSC) Xena Browser. Survival analysis was performed using mRNA data to obtain high and low binary variables for *CD5* and $CD5^+$ DC signature expression. Hazard ratios were calculated using the Cox proportional hazard model.

scRNA-seq of live $CD45^+HLA\text{-}DR^+$ cells, sorted from iLNs and uLNs by using the Chromium Single Cell 3′ Reagent Kit v3, was used to identify different myeloid cell types and specific gene signatures for each cluster. For such identification, the R package Seurat (version 3.1.2) was used to analyze the digital expression matrix generated from raw reads aligned to the human genome (hg38). Principal components analysis (PCA), t-distributed stochastic

neighbor embedding (t-SNE), and UMAP were used to reduce the dimensions of the data, and the first two dimensions were used in plots. The "FindClusters" function was later used to cluster the cells. To measure the relative abundance of *CD5* expression, the total expression was calculated for each myeloid cluster separated by tumor-involved and tumor-uninolved samples.

Human DC and T cell cocultures

To elucidate the effects of CD5-expression on DCs or T cells to the magnitude of the T cell response, we sorted DCs based on their CD5 expression; alternatively, CD5 expression was genetically altered using CRISPR and CRISPR-Cas9 technology. $CD5^+$ and $CD5^-$ DCs were then cocultured with autologous or allogeneic $CD4^+$ and $CD8^+$ T cells. Alternatively, we used hCD5ΔT cells or control T cells and cocultured with $CD5^+$ and $CD5^-$ DCs. Multiparameter flow cytometry was used for immunophenotyping and T cell proliferation assessment, and RNA-seq was used for gene expression evaluation for both DCs and T cells.

Murine CD5-deficient models

Using CRISPR-Cas9 technology, we generated mice with loxP insertions that flanked axons 2 and 9 of the *Cd5* locus. We then bred these $Cd5^{f/f}$ mice with mice that expressed Cre recombinase under the *Zbtb46*, *Itgax*, and *Cd4* promoters to delete CD5 in cDCs [CD5ΔDC (fig. S6)] and T cells [CD5ΔT (fig. S9)].

Tumor models and ICB therapy

MCA1956 (*81*) and MC-38 (*82*) lines were injected subcutaneously into the anatomical right flanks of mice (1×10^6 cells per mouse). Mice were treated intraperitoneally with anti–PD-1 or anti–CTLA-4 antibody (200 μg per mouse) every 3 days, starting at day 3 after tumor inoculation. The OVA-expressing MCA1956-mOVA line is spontaneously rejected by WT mice (*45*). Tumor growth was quantified by Vernier caliper measurements and expressed as the average of two perpendicular diameters. For tumor models, animals were injected at 8 to 10 weeks of age. All studies performed on mice were done in accordance with the Institutional Animal Care and Use Committee (IACUC) at Washington University in St. Louis. In accordance with our IACUC-approved protocol, maximal tumor diameter was 20 mm in one direction and in no experiments was this limit exceeded.

Full materials and methods are accessible in the supplementary materials (*44*).

REFERENCES AND NOTES

1. J. Larkin *et al*., Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. *N. Engl. J. Med.* **381**, 1535–1546 (2019). doi: 10.1056/NEJMoa1910836; pmid: 31562797
2. I. Mellman, G. Coukos, G. Dranoff, Cancer immunotherapy comes of age. *Nature* **480**, 480–489 (2011). doi: 10.1038/nature10673; pmid: 22193102
3. R. M. Steinman, J. Banchereau, Taking dendritic cells into medicine. *Nature* **449**, 419–426 (2007). doi: 10.1038/nature06175; pmid: 17898760
4. J. Lee *et al*., Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. *J. Exp. Med.* **212**, 385–399 (2015). doi: 10.1084/jem.20141442; pmid: 25687283
5. E. Klechevsky, Functional diversity of human dendritic cells. *Adv. Exp. Med. Biol.* **850**, 43–54 (2015). doi: 10.1007/978-3-319-15774-0_4; pmid: 26324345
6. E. Klechevsky *et al*., Functional specializations of human epidermal Langerhans cells and $CD14^+$ dermal dendritic cells. *Immunity* **29**, 497–510 (2008). doi: 10.1016/j.immuni.2008.07.013; pmid: 18789730
7. J. Banchereau *et al*., The differential production of cytokines by human Langerhans cells and dermal $CD14^+$ DCs controls CTL priming. *Blood* **119**, 5742–5749 (2012). doi: 10.1182/blood-2011-08-371245; pmid: 22535664
8. J. Banchereau *et al*., Immunoglobulin-like transcript receptors on human dermal $CD14^+$ dendritic cells act as a CD8-antagonist to control cytotoxic T cell priming. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18885–18890 (2012). doi: 10.1073/pnas.1205785109; pmid: 23112154
9. H. Fujita *et al*., Human Langerhans cells induce distinct IL-22-producing $CD4^+$ T cells lacking IL-17 production. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21795–21800 (2009). doi: 10.1073/pnas.0911472106; pmid: 19996179
10. S. J. Balin *et al*., Human antimicrobial cytotoxic T lymphocytes, defined by NK receptors and antimicrobial proteins, kill intracellular bacteria. *Sci. Immunol.* **3**, eaat7668 (2018). doi: 10.1126/sciimmunol.aat7668; pmid: 30171080
11. C. Caux *et al*., $CD34^+$ hematopoietic progenitors from human cord blood differentiate along two independent dendritic cell pathways in response to granulocyte-macrophage colony-stimulating factor plus tumor necrosis factor alpha: II. Functional analysis. *Blood* **90**, 1458–1470 (1997). doi: 10.1182/blood.V90.4.1458; pmid: 9269763
12. C. C. Chu *et al*., Resident CD141 $(BDCA3)^+$ dendritic cells in human skin produce IL-10 and induce regulatory T cells that suppress skin inflammation. *J. Exp. Med.* **209**, 935–945 (2012). doi: 10.1084/jem.20112583; pmid: 22547651
13. A. C. Villani *et al*., Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017). doi: 10.1126/science.aah4573; pmid: 28428369
14. D. Dudziak *et al*., Differential antigen processing by dendritic cell subsets in vivo. *Science* **315**, 107–111 (2007). doi: 10.1126/science.1136080; pmid: 17204652
15. A. D. Edwards *et al*., Relationships among murine $CD11c^{high}$ dendritic cell subsets as revealed by baseline gene expression patterns. *J. Immunol.* **171**, 47–60 (2003). doi: 10.4049/jimmunol.171.1.47; pmid: 12816982
16. Y. Kumamoto *et al*., $CD301b^+$ dermal dendritic cells drive T helper 2 cell-mediated immunity. *Immunity* **39**, 733–743 (2013). doi: 10.1016/j.immuni.2013.08.029; pmid: 24076051
17. K. L. Lewis *et al*., Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. *Immunity* **35**, 780–791 (2011). doi: 10.1016/j.immuni.2011.08.013; pmid: 22018469
18. D. Korenfeld *et al*., A type of human skin dendritic cell marked by CD5 is associated with the development of inflammatory skin disease. *JCI Insight* **2**, e96101 (2017). doi: 10.1172/jci.insight.96101; pmid: 28931765
19. X. Yin *et al*., Human blood $CD1c^+$ dendritic cells encompass $CD5^{high}$ and $CD5^{low}$ subsets that differ significantly in phenotype, gene expression, and functions. *J. Immunol.* **198**, 1553–1564 (2017). doi: 10.4049/jimmunol.1600193; pmid: 28087664
20. J. L. Hope, B. Pulendran, S. P. Schoenberger, P. D. Katsikis, 1st International Conference on Human & Translational Immunology. *Nat. Immunol.* **18**, 1–4 (2017). doi: 10.1038/ni.3635; pmid: 27984565
21. U. Cytlak *et al*., Differential IRF8 transcription factor requirement defines two pathways of dendritic cell development in humans. *Immunity* **53**, 353–370.e8 (2020). doi: 10.1016/j.immuni.2020.07.003; pmid: 32735845
22. M. Crowley, K. Inaba, M. Witmer-Pack, R. M. Steinman, The cell surface of mouse dendritic cells: FACS analyses of dendritic cells from different tissues including thymus. *Cell. Immunol.* **118**, 108–125 (1989). doi: 10.1016/0008-8749(89)90361-4; pmid: 2910499
23. M. O'Keeffe *et al*., Mouse plasmacytoid cells: Long-lived cells, heterogeneous in surface phenotype and function, that differentiate into $CD8^+$ dendritic cells only after microbial stimulus. *J. Exp. Med.* **196**, 1307–1319 (2002). doi: 10.1084/jem.20021031; pmid: 12438422
24. C. Ardavin, K. Shortman, Cell surface marker analysis of mouse thymic dendritic cells. *Eur. J. Immunol.* **22**, 859–862 (1992). doi: 10.1002/eji.1830220334; pmid: 1347747
25. A. Tarakhovsky *et al*., A role for CD5 in TCR-mediated signal transduction and thymocyte selection. *Science* **269**, 535–537 (1995). doi: 10.1126/science.7542801; pmid: 7542801
26. G. Bikah, J. Carey, J. R. Ciallella, A. Tarakhovsky, S. Bondada, CD5-mediated negative regulation of antigen receptor-induced growth signals in B-1 B cells. *Science* **274**, 1906–1909 (1996). doi: 10.1126/science.274.5294.1906; pmid: 8943203
27. H. S. Azzam *et al*., CD5 expression is developmentally regulated by T cell receptor (TCR) signals and TCR avidity. *J. Exp. Med.* **188**, 2301–2311 (1998). doi: 10.1084/jem.188.12.2301; pmid: 9858516
28. D. Hawiger, R. F. Masilamani, E. Bettelli, V. K. Kuchroo, M. C. Nussenzweig, Immunological unresponsiveness characterized by increased expression of CD5 on peripheral T cells induced by dendritic cells in vivo. *Immunity* **20**, 695–705 (2004). doi: 10.1016/j.immuni.2004.05.002; pmid: 15189735
29. H. Li *et al*., CD5 on dendritic cells regulates $CD4^+$ and $CD8^+$ T cell activation and induction of immune responses. *PLOS ONE* **14**, e0222301 (2019). doi: 10.1371/journal.pone.0222301; pmid: 31491023
30. J. L. Ceuppens, M. L. Baroja, Monoclonal antibodies to the CD5 antigen can provide the necessary second signal for activation of isolated resting T cells by solid-phase-bound OKT3. *J. Immunol.* **137**, 1816–1821 (1986). doi: 10.4049/jimmunol.137.6.1816; pmid: 3091691
31. F. Spertini, W. Stohl, N. Ramesh, C. Moody, R. S. Geha, Induction of human T cell proliferation by a monoclonal antibody to CD5. *J. Immunol.* **146**, 47–52 (1991). doi: 10.4049/jimmunol.146.1.47; pmid: 1701800
32. P. Vandenberghe, J. L. Ceuppens, Immobilized anti-CD5 together with prolonged activation of protein kinase C induce interleukin 2-dependent T cell growth: Evidence for signal transduction through CD5. *Eur. J. Immunol.* **21**, 251–259 (1991). doi: 10.1002/eji.1830210203; pmid: 1705509
33. T. Stanton, T. L. Stevens, J. A. Ledbetter, D. Wofsy, Anti-Ly-1 antibody induces interleukin 2 release from T cells. *J. Immunol.* **136**, 1734–1737 (1986). doi: 10.4049/jimmunol.136.5.1734; pmid: 3081630
34. C. Wang *et al*., CD5L/AIM regulates lipid biosynthesis and restrains Th17 cell pathogenicity. *Cell* **163**, 1413–1427 (2015). doi: 10.1016/j.cell.2015.10.068; pmid: 26607793
35. J. de Wit *et al*., CD5 costimulation induces stable Th17 development by promoting IL-23R expression and sustained STAT3 activation. *Blood* **118**, 6107–6114 (2011). doi: 10.1182/blood-2011-05-352682; pmid: 21926348
36. C. A. Matson *et al*., CD5 dynamically calibrates basal NF-κB signaling in T cells during thymic development and peripheral activation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14342–14353 (2020). doi: 10.1073/pnas.1922525117; pmid: 32513716
37. G. Blaize *et al*., CD5 signalosome coordinates antagonist TCR signals to control the generation of Treg cells induced by foreign antigens. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12969–12979 (2020). doi: 10.1073/pnas.1917182117; pmid: 32434911
38. Q. Zhang *et al*., Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* **179**, 829–845.e20 (2019). doi: 10.1016/j.cell.2019.10.003; pmid: 31675496
39. B. Maier *et al*., A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* **580**, 257–262 (2020). doi: 10.1038/s41586-020-2134-y; pmid: 32269339
40. M. N. Artyomov *et al*., Modular expression analysis reveals functional conservation between human Langerhans cells and mouse cross-priming dendritic cells. *J. Exp. Med.* **212**, 743–757 (2015). doi: 10.1084/jem.20131675; pmid: 25918340
41. R. Cabrita *et al*., Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020). doi: 10.1038/s41586-019-1914-8; pmid: 31942071
42. J. Banchereau, R. M. Steinman, Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998). doi: 10.1038/32588; pmid: 9521319
43. T. I. Jensen *et al*., Targeted regulation of transcription in primary cells using CRISPRa and CRISPRi. *Genome Res.* **31**, 2120–2130 (2021). doi: 10.1101/gr.275607.121; pmid: 34407984
44. See supplementary materials.
45. S. T. Ferris *et al*., cDC1 prime and are licensed by $CD4^+$ T cells to induce anti-tumour immunity. *Nature* **584**, 624–629 (2020). doi: 10.1038/s41586-020-2611-3; pmid: 32788723
46. E. Alspach *et al*., MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019). doi: 10.1038/s41586-019-1671-8; pmid: 31645760

47. R. O. Bak, D. P. Dever, M. H. Porteus, CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat. Protoc.* **13**, 358–376 (2018). doi: 10.1038/nprot.2017.143; pmid: 29370156
48. O. Padilla *et al*., Genomic organization of the human *CD5* gene. *Immunogenetics* **51**, 993–1001 (2000). doi: 10.1007/s002510000235; pmid: 11061284
49. O. Lecomte, J. B. Bock, B. W. Birren, D. Vollrath, J. R. Parnes, Molecular linkage of the mouse CD5 and CD6 genes. *Immunogenetics* **44**, 385–390 (1996). doi: 10.1007/BF02602784; pmid: 8781125
50. M. Ogishi *et al*., Inherited PD-1 deficiency underlies tuberculosis and autoimmunity in a child. *Nat. Med.* **27**, 1646–1654 (2021). doi: 10.1038/s41591-021-01388-5; pmid: 34183838
51. D. Korenfeld *et al*., STAT3 gain-of-function mutations underlie deficiency in human nonclassical CD16$^+$ monocytes and CD141$^+$ dendritic cells. *J. Immunol.* **207**, 2423–2432 (2021). doi: 10.4049/jimmunol.2000841; pmid: 34654687
52. P. Chomarat, J. Banchereau, J. Davoust, A. K. Palucka, IL-6 switches the differentiation of monocytes from dendritic cells to macrophages. *Nat. Immunol.* **1**, 510–514 (2000). doi: 10.1038/82763; pmid: 11101873
53. S. Wang, L. Pang, Z. Liu, X. Meng, SERPINE1 associated with remodeling of the tumor microenvironment in colon cancer progression: A novel therapeutic target. *BMC Cancer* **21**, 767 (2021). doi: 10.1186/s12885-021-08536-7; pmid: 34215248
54. P. Tan *et al*., MMP25-AS1/hsa-miR-10a-5p/SERPINE1 axis as a novel prognostic biomarker associated with immune cell infiltration in KIRC. *Mol. Ther. Oncolytics* **22**, 307–325 (2021). doi: 10.1016/j.omto.2021.07.008; pmid: 34553021
55. A. Moreno-Manuel *et al*., CD5 and CD6 as immunoregulatory biomarkers in non-small cell lung cancer. *Transl. Lung Cancer Res.* **9**, 1074–1083 (2020). doi: 10.21037/tlcr-19-445; pmid: 32953486
56. M. C. Cenit *et al*., Analysis of ancestral and functionally relevant CD5 variants in systemic lupus erythematosus patients. *PLOS ONE* **9**, e113090 (2014). doi: 10.1371/journal.pone.0113090; pmid: 25402503
57. S. Casadó-Llombart *et al*., Gene variation impact on prostate cancer progression: Lymphocyte modulator, activation, and cell adhesion gene variant contribution. *Prostate* **82**, 1331–1337 (2022). doi: 10.1002/pros.24407; pmid: 35767366
58. M. Potrony *et al*., Inherited functional variants of the lymphocyte receptor CD5 influence melanoma survival. *Int. J. Cancer* **139**, 1297–1302 (2016). doi: 10.1002/ijc.30184; pmid: 27169428
59. K. E. Pauken *et al*., Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016). doi: 10.1126/science.aaf2807; pmid: 27789795
60. C. Peña-Rossi *et al*., Negative regulation of CD4 lineage development and responses by CD5. *J. Immunol.* **163**, 6494–6501 (1999). doi: 10.4049/jimmunol.163.12.6494; pmid: 10586041
61. K. Inaba *et al*., Efficient presentation of phagocytosed cellular fragments on the major histocompatibility complex class II products of dendritic cells. *J. Exp. Med.* **188**, 2163–2173 (1998). doi: 10.1084/jem.188.11.2163; pmid: 9841929
62. L. M. Wakim, M. J. Bevan, Cross-dressed dendritic cells drive memory CD8$^+$ T-cell activation after viral infection. *Nature* **471**, 629–632 (2011). doi: 10.1038/nature09863; pmid: 21455179
63. L. A. Smyth *et al*., Acquisition of MHC:peptide complexes by dendritic cells contributes to the generation of antiviral CD8$^+$ T cell immunity in vivo. *J. Immunol.* **189**, 2274–2282 (2012). doi: 10.4049/jimmunol.1200664; pmid: 22821960
64. L. Biancone *et al*., Identification of a novel inducible cell-surface ligand of CD5 on activated lymphocytes. *J. Exp. Med.* **184**, 811–819 (1996). doi: 10.1084/jem.184.3.811; pmid: 9064341
65. J. N. Mandl, J. P. Monteiro, N. Vrisekoop, R. N. Germain, T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. *Immunity* **38**, 263–274 (2013). doi: 10.1016/j.immuni.2012.09.011; pmid: 23290521
66. S. P. Persaud, C. R. Parker, W. L. Lo, K. S. Weber, P. M. Allen, Intrinsic CD4$^+$ T cell sensitivity and response to a pathogen are set and sustained by avidity for thymic and peripheral complexes of self peptide and MHC. *Nat. Immunol.* **15**, 266–274 (2014). doi: 10.1038/ni.2822; pmid: 24487322
67. R. B. Fulton *et al*., The TCR's sensitivity to self peptide-MHC dictates the ability of naive CD8$^+$ T cells to respond to foreign antigens. *Nat. Immunol.* **16**, 107–117 (2015). doi: 10.1038/ni.3043; pmid: 25419629
68. K. S. Weber *et al*., Distinct CD4$^+$ helper T cells involved in primary and secondary responses to infection. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9511–9516 (2012). doi: 10.1073/pnas.1202408109; pmid: 22645349
69. W. C. Kieper, J. T. Burghardt, C. D. Surh, A role for TCR affinity in regulating naive T cell homeostasis. *J. Immunol.* **172**, 40–44 (2004). doi: 10.4049/jimmunol.172.1.40; pmid: 14688307
70. D. Rogers *et al*., Pre-existing chromatin accessibility and gene expression differences among naive CD4$^+$ T cells influence effector potential. *Cell Rep.* **37**, 110064 (2021). doi: 10.1016/j.celrep.2021.110064; pmid: 34852223
71. C. E. Nelson *et al*., Reprogramming responsiveness to checkpoint blockade in dysfunctional CD8 T cells. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2640–2645 (2019). doi: 10.1073/pnas.1810326116; pmid: 30679280
72. G. Dorothée *et al*., In situ sensory adaptation of tumor-infiltrating T lymphocytes to peptide-MHC levels elicits strong antitumor reactivity. *J. Immunol.* **174**, 6888–6897 (2005). doi: 10.4049/jimmunol.174.11.6888; pmid: 15905531
73. M. M. Meredith *et al*., Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. *J. Exp. Med.* **209**, 1153–1165 (2012). doi: 10.1084/jem.20112675; pmid: 22615130
74. H. Van de Velde, I. von Hoegen, W. Luo, J. R. Parnes, K. Thielemans, The B-cell surface protein CD72/Lyb-2 is the ligand for CD5. *Nature* **351**, 662–665 (1991). doi: 10.1038/351662a0; pmid: 1711157
75. M. H. Brown, E. Lacey, A ligand for CD5 is CD5. *J. Immunol.* **185**, 6068–6074 (2010). doi: 10.4049/jimmunol.0903823; pmid: 20952682
76. C. Zhang *et al*., CD5 binds to interleukin-6 and induces a feed-forward loop with the transcription factor STAT3 in B cells to promote cancer. *Immunity* **44**, 913–923 (2016). doi: 10.1016/j.immuni.2016.04.003; pmid: 27096320
77. J. Vera *et al*., The CD5 ectodomain interacts with conserved fungal cell wall components and protects from zymosan-induced septic shock-like syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1506–1511 (2009). doi: 10.1073/pnas.0805846106; pmid: 19141631
78. M. Binnewies *et al*., Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018). doi: 10.1038/s41591-018-0014-x; pmid: 29686425
79. Y. P. Chen *et al*., Single-cell transcriptomics reveals regulators underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. *Cell Res.* **30**, 1024–1042 (2020). doi: 10.1038/s41422-020-0374-x; pmid: 32686767
80. A. J. Combes *et al*., Discovering dominant tumor immune archetypes in a pan-cancer census. *Cell* **185**, 184–203.e19 (2022). doi: 10.1016/j.cell.2021.12.004; pmid: 34963056
81. H. Matsushita *et al*., Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* **482**, 400–404 (2012). doi: 10.1038/nature10755; pmid: 22318521
82. T. H. Corbett, D. P. Griswold Jr., B. J. Roberts, J. C. Peckham, F. M. Schabel Jr., Tumor induction relationships in development of transplantable cancers of the colon in mice for chemotherapy assays, with a note on carcinogen structure. *Cancer Res.* **35**, 2434–2439 (1975). pmid: 1149045

ACKNOWLEDGMENTS

We thank the McDonnell Genome Institute at Washington University for sequencing; E. M. Lantelme, D. Brinja, and the Flow Cytometry Core at Washington University for cell sorting; X. Cui, M. Sentmamat, and the Genome Engineering and iPSC Center (GEiC) for human and mouse sgRNA design and validation; A. Burroughs for technical help; Y. Tao for biostatistical analysis; M. Diamond, J. Kipnis, G. Amarashinghe, and S. Stewart for helpful discussions; K. Ravichandran for critical reading of the manuscript; M. Bambouskova and N. Wiggie for help with experiments; D. Kreamalmeyer and R. Meade for mouse breeding and care; M. Colonna, S. Gilfillan, and M. Molgora for OT-I mice and the MC-38 cell line; M. Cella for providing access to tissue specimens and guidance; C.-S. Hsieh and J. Yi for OT-II mice; K. Murphy for providing access to the MCA1956-mOVA line; T. Saito from RIKEN (the Institute of Physical and Chemical Research, Japan) for providing the NFAT-expressing 2B4 cell line (43-1); T. Keller, H. Marsh, and M. Yellin from Celldex Therapeutics for providing CDX-301; and the surgeons, residents, nurses, and staff at Barnes-Jewish Hospital and the Washington University School of Medicine in St. Louis Department of Surgery for providing access to skin and melanoma specimens. Aspects of the studies, including tetramer production and flow cytometry analysis, were performed with the assistance of the Immunomonitoring Laboratory (IML), which is supported by the Andrew M. and Jane M. Bursky Center for Human Immunology and Immunotherapy Programs and the Alvin J. Siteman Comprehensive Cancer Center. We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, MO, for the use of the GEiC and the IML. **Funding:** This work was funded by National Institutes of Health (NIH) grants 1R01CA245277-01A1 (E.K.), 5R01AR075959-02 (E.K.), and 5R21EB024767-03 (E.K.); and a National Psoriasis Foundation Translational Research Grant (E.K.). C.S. was supported in part by a postdoctoral training grant from NIH T32 CA009547-34 (R.D.S.). The Siteman Cancer Center is supported in part by National Cancer Institute (NCI) Cancer Center Support Grant #P30 CA091842. T.I.J. was funded by a grant from Novo Nordisk (NNF17OC0028894). Partial salary support for R.L.M. was provided by NIH grants R01 AI022553, R01 AR040312, R01 AR073252, R01 AR074302, R01 AI166313, and P50 AR080594. **Author contributions:** M.H., K.R., C.S., V.G., Y.Z., H.B., and L.G. performed experiments and analyzed the data; C.A.I., K.S., D.K., A.U.A., S.D., and D.D. provided technical help and performed experiments; F.M. and N.B. performed in silico analysis; M.W., J.G., T.I.J., R.B., C.L.A., and A.S. generated reagents; S.F. provided the MCA1956-mOVA cell line; J.L. performed biostatistical analysis; P.M.A. provided mice; R.D.S. provided the MCA1956 cell line; R.C.F. and T.T. provided patient specimens; M.W., M.M.G., M.P., R.L.M., R.B., R.D.S., and P.M.A. provided expertise; and E.K. conceived and performed experiments, analyzed and interpreted data, and wrote the manuscript. **Competing interests:** R.D.S. is a cofounder, scientific advisory board member, stockholder, and royalty recipient of Jounce Therapeutics and Neon Therapeutics and is a scientific advisory board member for A2 Biotherapeutics, BioLegend, Codiak Biosciences, Constellation Pharmaceuticals, NGM Biopharmaceuticals, and Sensei Biotherapeutics. R.B. is a cofounder, stockholder, and part-time employee of UNIKUM Therapeutics. All other authors declare that they have no competing interests. **Data and materials availability:** Single-cell data are available at Gene Expression Omnibus (GEO) under accession number (GSE219197). All other data are available in the manuscript or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg2752
Materials and Methods
Fig. S1 to S14
Tables S1 to S8
References (*83*–*101*)
MDAR Reproducibility Checklist

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

Submitted 2 February 2021; resubmitted 14 September 2022
Accepted 17 January 2023
10.1126/science.abg2752

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

A volatile sex attractant of tsetse flies

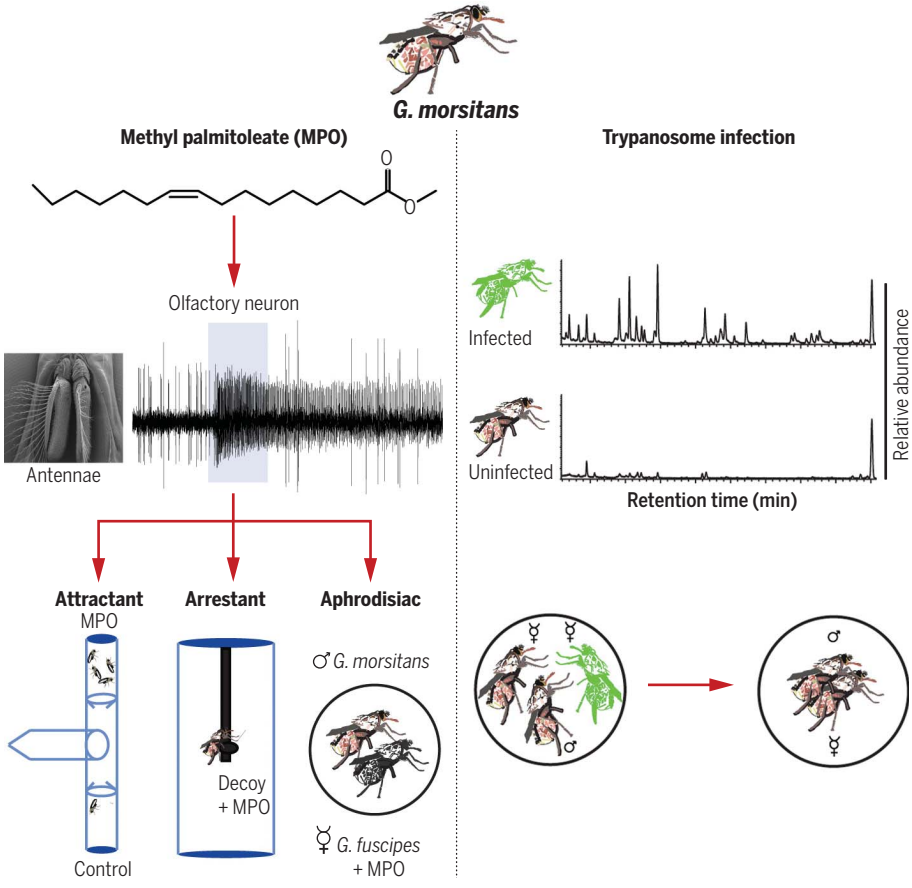
Shimaa A. M. Ebrahim*, Hany K. M. Dweck, Brian L. Weiss, John R. Carlson*

INTRODUCTION: Tsetse flies are vectors of trypanosomes that cause human and animal diseases in sub-Saharan Africa. Historically these flies and the trypanosomes they transmit have had extremely detrimental effects on the health and development of this region. Tsetse continues to be a major cause of rural poverty in this area because of the impact of trypanosomiasis on people and domestic farm animals. Moreover, the geographic range of tsetse is expected to extend to new regions as a result of climate change, thus placing more people and domestic animals at risk. Despite more than a century of tsetse research, volatile sex pheromones have not been identified. Because pheromones have been successfully used in the control of a wide variety of other insects, their identification in tsetse could be useful in controlling these flies and trypanosomiasis.

RATIONALE: To identify volatile pheromones of tsetse, we used gas chromatography-mass spectrometry (GC-MS). We studied the behavioral responses of tsetse flies to pheromones and to other flies perfumed with them. To elucidate the cellular effects of pheromones we analyzed the electrophysiological responses they elicited from olfactory neurons in the antenna. We analyzed both males and females. Although we focused on *Glossina morsitans,* we also used the related species *Glossina fuscipes*, which accounts for the greatest number of cases of human African trypanosomiasis. We also investigated the effects of trypanosome infection on the chemical profiles and sexual behavior of flies.

RESULTS: When introduced into a simple mating paradigm, pairs of tsetse began copulating within seconds. A single male placed in a chamber with a tsetse-sized decoy landed on it and stayed attached for minutes if the decoy was dosed with female extract, but not if dosed with male extract. Using GC-MS we identified methyl palmitoleate (MPO), methyl oleate (MO), and methyl palmitate (MP) as volatile compounds produced by *G. morsitans* that elicited strong behavioral responses. All three compounds have a 16-carbon unbranched backbone. We focused on MPO, which elicited a strong response from males—but not virgin females—in the decoy assay. This response depended on the antenna. MPO also elicited an attractive response from males in a T-maze paradigm. Male *G. morsitans* mounted *G. fuscipes* females perfumed with MPO, but not control females. We characterized olfactory neurons of the *G. morsitans* and *G. fuscipes* antennae. We identified a class of neurons on the *G. morsitans* antenna that responds to MPO. These responses are sexually dimorphic, and this class of neurons also responded to certain odorants that attract tsetse in the field. MPO had little if any effect on *G. fuscipes* in our physiological or behavioral testing. The pheromone profiles of the two species are different. Notably, infection of mated *G. morsitans* with trypanosomes resulted in the appearance of 21 small volatile compounds in their chemical profiles. Infection also reduced the sexual receptivity of females.

CONCLUSION: We have identified a volatile sex attractant of the tsetse fly. MPO acts as an aphrodisiac, elicits sex-specific behavioral effects, and evokes sexually dimorphic physiological responses. It also elicits species-specific behavioral and physiological effects. Our work invites exploration of whether MPO may be useful in tsetse control. We also found that infection with trypanosomes affects both the chemical profile and sexual behavior of tsetse. ■



Volatile chemicals of tsetse flies. Methyl palmitoleate (MPO), produced by the tsetse fly *Glossina morsitans*, elicits physiological responses from olfactory neurons. MPO also elicits behavioral responses, acting as an attractant and an arrestant, and as an aphrodisiac when used to perfume females of another species. Trypanosome infection alters the tsetse chemical profile. When placed with two females, one uninfected and one infected, males mated with the uninfected female.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: john.carlson@yale.edu
Cite this article as S. A. M. Ebrahim *et al.*, *Science* **379**, eade1877 (2023). DOI: 10.1126/science.ade1877

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.ade1877

CREDIT: TSE TSE FLY ILLUSTRATION: ESTEFANIA ALONSO GÓMEZ, CC BY-SA 4.0

RESEARCH ARTICLE

NEUROSCIENCE

A volatile sex attractant of tsetse flies

Shimaa A. M. Ebrahim[1]*, Hany K. M. Dweck[1], Brian L. Weiss[2], John R. Carlson[1]*

Tsetse flies transmit trypanosomes—parasites that cause devastating diseases in humans and livestock—across much of sub-Saharan Africa. Chemical communication through volatile pheromones is common among insects; however, it remains unknown if and how such chemical communication occurs in tsetse flies. We identified methyl palmitoleate (MPO), methyl oleate, and methyl palmitate as compounds that are produced by the tsetse fly *Glossina morsitans* and elicit strong behavioral responses. MPO evoked a behavioral response in male—but not virgin female—*G. morsitans*. *G. morsitans* males mounted females of another species, *Glossina fuscipes*, when they were treated with MPO. We further identified a subpopulation of olfactory neurons in *G. morsitans* that increase their firing rate in response to MPO and showed that infecting flies with African trypanosomes alters the flies' chemical profile and mating behavior. The identification of volatile attractants in tsetse flies may be useful for reducing disease spread.

Chemical communication among insects is used to identify and locate suitable mating partners. Pheromones are used to recognize a conspecific in a habitat that may contain thousands of the world's millions of insect species. Volatile pheromones have been identified and their mechanisms of action elucidated in a wide diversity of species (*1*). However, little is known about chemical communication among tsetse flies. One pheromone has been identified in the tsetse fly *Glossina morsitans morsitans*, but this molecule [(15,19,23-trimethylheptatriacontane (morsilure)] consists of a chain of 37 carbon atoms and its relative vapor pressure is vanishingly low (*2*–*7*). Tsetse flies spread disease across much of sub-Saharan Africa. They can carry African trypanosomes, which are transmitted when they bite humans or animals. In humans these parasites cause African sleeping sickness whereas in livestock they cause a disease called nagana, which has had a devastating effect on agricultural and economic development in Africa (*8*, *9*). The primary means of controlling these diseases is to control the tsetse flies that spread them, including the use of traps and targets containing attractive odorants derived from animal hosts, which have been shown as effective approaches for controlling disease spread (*10*). Because pheromones have been successfully used in the control of many other insects (*11*), the identification of airborne chemical communication among tsetse flies could be useful in further enhancing their control. Tsetse flies are also of substantial intrinsic biological interest: Rather than laying eggs as with most other insects, females give birth to larvae (*12*) and all embryonic and larval stages occur within the female's uterus, where progeny are nourished by maternal milk. Once laid, larvae burrow into the substrate and pupate within 30 min. Following eclosion, both male and female flies feed exclusively on vertebrate blood. It is an open question whether their chemical communication is also unusual. The antenna of tsetse contains several classes of olfactory sensilla, including trichoid sensilla (*13*). Sensilla in the medial portion of the *G. morsitans* antenna respond to a variety of monomolecular odorants (*14*). The antennae of other tsetse species have also been shown to respond to natural mixtures of plant or host animal volatiles (*15*, *16*).

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA. [2]Department of Epidemiology of Microbial Disease, Yale School of Public Health, New Haven, CT, USA.
*Corresponding author. Email: john.carlson@yale.edu (J.R.C.), shimaa.ebrahim@yale.edu (S.A.M.E.)

Mating in *G. morsitans* is initiated rapidly

To identify chemical factors affecting tsetse mating behavior we first placed a single male and a single virgin *G. morsitans* female together in a tube and found that the male quickly mounted the female (Fig. 1, A and B). The half time to mounting was on the order of 5 s. Out of 20 pairs, nearly all began copulating within 15 s. In a second experiment, we directly compared the copulation of *G. morsitans* to that of *Drosophila melanogaster* in the same assay. The mean latency of *G. morsitans* was two orders of magnitude faster than that of *D. melanogaster*: 0.15 min ± 0.02 min versus 22.14 min ± 4.9 min (Fig. 1C; $P < 0.0001$, Mann-Whitney test). *G. morsitans* also showed a much longer duration of copulation: 58.5 min ± 5.3 min, compared with only 20.3 ± 1.0 min for *D. melanogaster* (Fig. 1D, $P < 0.0001$, Mann-Whitney test). These latencies and durations were calculated from the pairs that had initiated mating within one hour, i.e., 18 of 20 pairs for *G. morsitans* and 14 of 20 pairs for *D. melanogaster*.

In contrast to the rapid copulation that ensued when a male *G. morsitans* encountered a virgin female, no copulation was observed when a male was paired with a non-virgin female (0%, $n = 20$) (Fig. 1E). Thus, tsetse showed copious mating behavior in this paradigm, with males mounting virgin females quickly and copulating for approximately an hour. Based on these data, we hypothesized that multiple sensory modalities could contribute to tsetse mating behavior, and we investigated whether olfaction played a role.

Female but not male hexane extracts attract *G. morsitans* males

We established a behavioral paradigm to measure the attraction of *G. morsitans* males to a tsetse-sized decoy (Fig. 2A) (*5*). We dosed the decoy with a chemosensory stimulus of interest, placed it in a chamber with a single male, and observed for one hour. In several instances males landed on the decoy and stayed attached, often for the remainder of the experiment. We scored this response as the percentage of male flies that landed on the decoy and remained attached for at least 5 min. We prepared extracts from virgin females, mated females, virgin males, and mated males by soaking flies in hexane with gentle shaking for 10 min. This procedure has been used previously to isolate pheromones from the cuticles of other insects (*17*–*20*). None of these extracts elicited attraction in this paradigm (fig. S1A). We then considered the possibility that tsetse flies might contain attractive compounds that were temporarily sequestered beneath the surface of the fly, either in internal oenocytes or in internal glands such as those of other insects that synthesize and eventually secrete attractive pheromones (*21*–*23*). Accordingly, we prepared hexane extracts by soaking flies in hexane with gentle shaking for 24 hours, a procedure that has been successfully used to isolate pheromones from *Drosophila* and other insects (*24*–*27*). Extracts from virgin females elicited a behavioral response in 9 of 15 (60%) of cases (Fig. 2B). Mated female extracts elicited a response in 4 of 15 (27%) of cases. Neither male extract evoked any response in male tsetse flies. This sex specificity in the effect of extracts suggested a sexually dimorphic composition of effective compounds. To identify these compounds we carried out a gas chromatography-mass spectrometry (GC-MS) analysis of the extracts.

G. morsitans hexane extracts contain candidate attractants

We injected *G. morsitans* extracts into a GC-MS instrument; among the compounds identified were three fatty acids [palmitoleic acid (POA), palmitic acid (PA), and oleic acid (OA)] and three fatty acid methyl esters [methyl palmitoleate (MPO), methyl palmitate (MP),

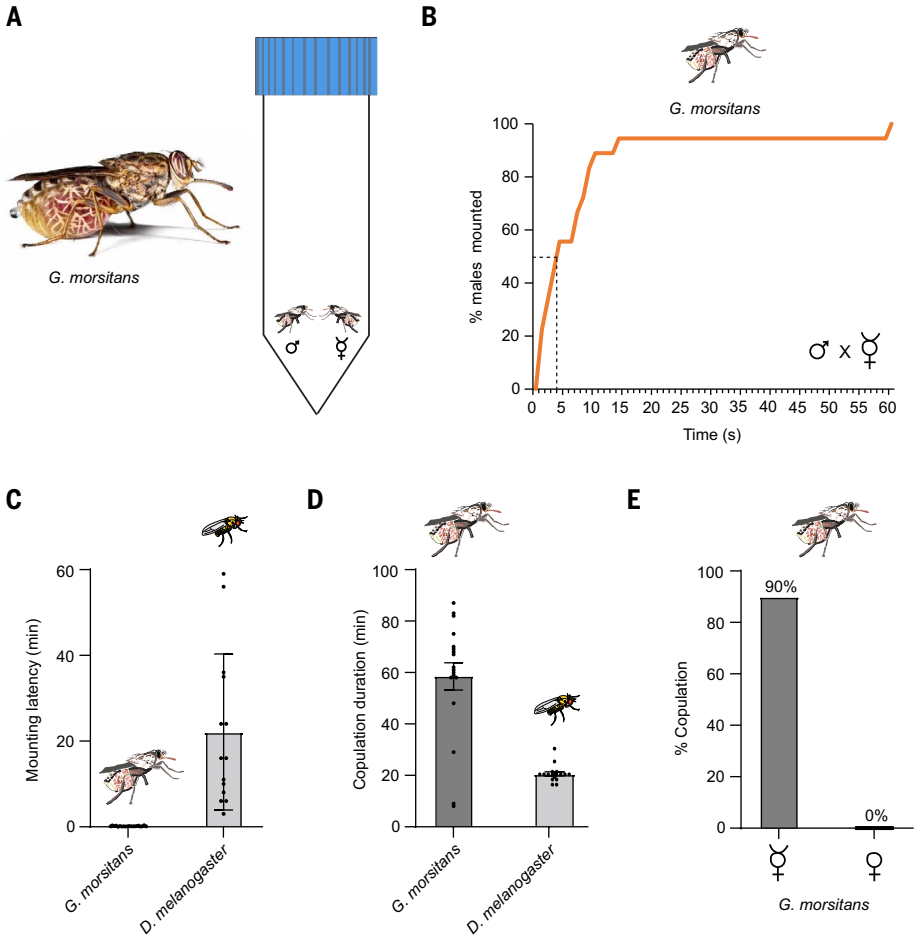


Fig. 1. Mating of *G. morsitans*. (**A**) The mating assay, in which a male and a female are placed together in a 50-ml tube, 11.5 cm in height. (**B**) Percentage of *G. morsitans* males that had mounted a virgin female in the mating assay as a function of time; n = 20. (**C**) Copulation latencies of *G. morsitans* and *D. melanogaster* males with a virgin female. Out of 20 pairs each for *D. melanogaster* and *G. morsitans*, 14 (n = 14) and 18 (n = 18) pairs mated, respectively. Error bars are SEM. (**D**) Copulation duration; n = 14 for *D. melanogaster* and n = 18 for *G. morsitans*. Error bars are SEM. (**E**) Percentage of copulation of *G. morsitans* males with either virgin females or mated females; n = 20.

and methyl oleate (MO)], which are pheromones in other insects (*4*, *22*, *26*, *28*, *29*, *30*–*32*) (Fig. 2C). The relative abundance of these six compounds in the 24-hour extracts depended on sex and mating status (Fig. 2D); POA was more abundant in females than in males whereas PA, MO, and OA were more abundant in females than in virgin males. We did not detect the six compounds from an extract prepared by brief (10 min) immersion of flies in hexane (fig. S1B). We note that the abundance of a certain methyl branched alkane was higher in mated females than in virgin females (Fig. 2E; P = 0.029, Mann-Whitney test; n = 4), suggesting that this compound—which was present at relatively high levels in males—was transferred from males to females during copulation. We speculated that this compound could reduce the attractiveness of mated females, thereby contributing to the lack of copulation observed among previously mated females (Fig. 1E). It could also explain why male extracts do not elicit a response from males in our decoy paradigm (Fig. 2B). Transfer of other compounds, some inhibitory, during mating has previously been shown or suggested in species of *Drosophila* (*33*–*36*). Taken together, these results suggested the possibility that some of the six compounds might have a behavioral effect on tsetse flies.

MPO, MP, and MO attract *G. morsitans* males

We measured the response of *G. morsitans* males to the tsetse-sized decoy perfumed with each of the six compounds (Fig. 3A). MO, MPO, and MP all elicited a strong response (Fig. 3B). MPO elicited a response in 13 of 15 (87%) of cases when tested at a 10^{-3} dilution. MO elicited a response at concentrations spanning three orders of magnitude. The other three compounds did not elicit a response. We also tested nine other compounds, including three known attractants of *G. morsitans*—1-octen-3-ol, 4-methylphenol, and acetone (*10*)—and found that none of these compounds elicited a response in this paradigm at a 10^{-2} dilution (fig. S2). We further investigated MPO, which elicited the greatest response in the decoy paradigm. We tested virgin females with MPO and found no response at any concentration (Fig. 3C). Thus, MPO elicited a male-specific response in this paradigm. Surgical removal of the antennae eliminated the response, suggesting that the response relied on the olfactory system (Fig. 3D). In insect chemical ecology, compounds are described as arrestants, aphrodisiacs, and/or attractants (*37*), based on the elicited behavior. Compounds can belong to more than one category. The long periods of time that males spend on a decoy dosed with MPO are reminiscent of the behavior elicited by an arrestant. However, we tested the possibility that MPO also acted as an aphrodisiac and attractant. We paired a *G. morsitans* male with a *G. fuscipes* female perfumed with MPO diluted 10^{-3} in hexane (Fig. 3E). *G. fuscipes* is a close relative of *G. morsitans* and accounts for the greatest number of cases of human African trypanosomiasis (*38*). In control experiments using hexane solvent alone, *G. morsitans* males made no attempts to mate with *G. fuscipes* females during a 1-hour observation period, i.e., 0% (Fig. 3F; n = 20). By contrast, when paired with a *G. fuscipes* female perfumed with MPO, in 9 of 15 (60%) of cases the *G. morsitans* male mounted the *G. fuscipes* females in an apparent attempt to copulate (Fig. 3F). The coupling did not persist and in all cases the female separated from the male shortly after mounting. These data suggest that MPO acts as an aphrodisiac for *G. morsitans* males. To test whether the compounds identified acted as attractants, we measured olfactory attraction in a T-maze assay (Fig. 3G). In this paradigm flies made a choice between two tubes, one containing a volatile odor and the other containing a diluent control. Flies that entered either tube rarely left. We tested all six compounds in this assay at concentrations spanning five orders of magnitude. MO, MPO, and MP all elicited an attractive response at one or more concentrations (Fig. 3H). In summary, MPO elicited behavior expected of an arrestant, an aphrodisiac, and an attractant. MO and MP also elicited arrestant and attractive responses from males. Next, we asked whether there are olfactory neurons in *G. morsitans* antennae that might mediate these responses.

G. morsitans olfactory neurons respond to extracts and to MPO

We first asked whether the extracts themselves elicited electrophysiological responses from ORNs of the tsetse antenna (Fig. 4A). We focused on trichoid sensilla because volatile pheromones primarily elicit responses

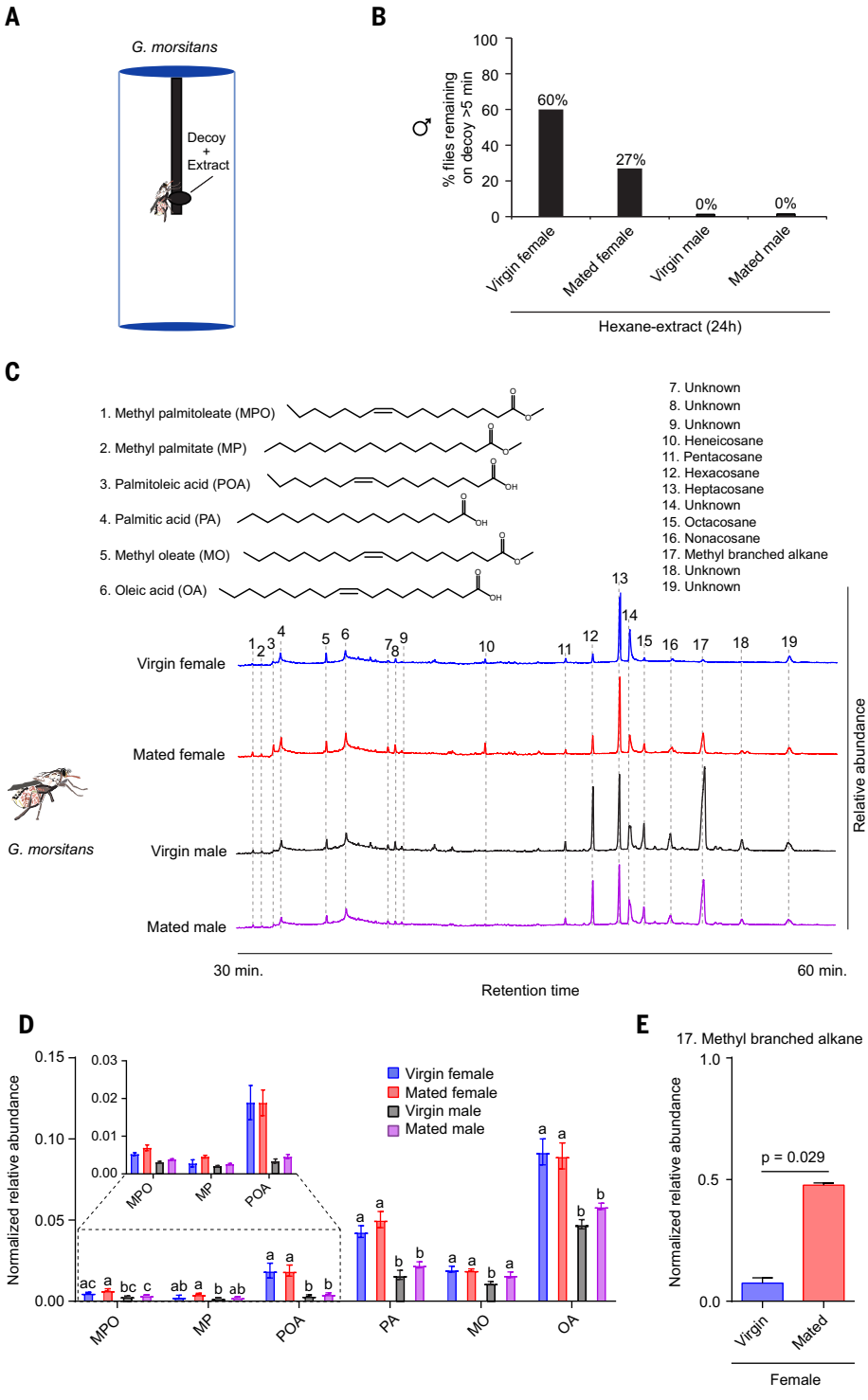


Fig. 2. *G. morsitans* hexane extracts contain candidate attractants. (**A**) The decoy paradigm, in which the decoy consists of a knot in a length of black yarn dosed with extract. (**B**) Percentage of males that landed on the decoy when dosed with a 24-hour extract and remained on it continuously for more than 5 min; $n = 15$. (**C**) Examples of total ion current chromatograms of 24-hour hexane extracts. Peak numbers correspond to the identified compounds. Additional data are shown in fig. S1. (**D**) Normalized relative abundance of compounds identified in 24-hour hexane extracts. Normalization is to the sum of the areas of all peaks. One-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test; $n = 4$. Values indicated with different letters are significantly different. Error bars are SEM. (**E**) Normalized relative abundance of methyl branched alkane (compound 17) in virgin and mated females. Mann-Whitney test; $n = 4$.

from them in *Drosophila* (*26*). Using single-sensillum electrophysiology and a modified stimulus delivery system in which an airflow carried odors to the antenna from a source 24 cm away (Fig. 4B), we recorded from trichoid sensilla located on the lateral face of the antenna (Fig. 4A). We tested hexane extracts from virgin females, mated females, virgin males, and mated males. Each of the four extracts elicited responses from a number of sensilla, and each of the four groups of flies (virgin female, mated female, virgin male, mated male) contained sensilla that responded to extracts (fig. S3). Overall, 12 of 82 (15%) of the tested sensilla responded to a least one extract with a response of ≥10 spikes per s. We next tested the six compounds against trichoid sensilla of the antenna. We tested trichoid sensilla from males ($n = 56$) and virgin females ($n = 50$). Delivering the compounds through an airflow using the system shown in Fig. 4B, we found that MPO elicited excitatory responses from a subset of sensilla (Fig. 4C). The responding sensilla have two ORNs, one designated the "A" neuron, which produces spikes of large amplitude and the other designated the "B" neuron, which produces spikes of small amplitude (Fig. 4D). The responses to MPO were from the B neuron (Fig. 4, C and E). Responses were observed in the sensilla of both males and virgin females (Fig. 4, C and E); 20% of tested male sensilla (11 of 56) and 26% of female sensilla (13 of 50) responded with >20 spikes per s at the tested dose. We observed few if any responses from the tested trichoid sensilla with the other five compounds using this delivery system (Fig. 4E). The response magnitudes to MPO were greater in males than in virgin females (Fig. 4F). To determine whether responses were dose-dependent we tested a range of concentrations. Because trichoid sensilla were heterogeneous in their responses to MPO we carried out this analysis on two individual sensilla, one in males and one in virgin females, which are located on the proximal portion of the antenna and produced the highest responses in our survey. Increasing doses produced increasing responses in both cases (Fig. 4, G and H). To identify and classify the ORNs that respond to MPO, we tested the sensilla with a panel of volatile compounds. Trichoid sensilla in *G. morsitans* on the opposite side of the antenna (the medial side) respond to odorants (*14*). Therefore, in our analysis of the trichoid sensilla we included nine odorants in addition to the six compounds from the tsetse extract. The nine odorants included three that are known to attract *G. morsitans*: 1-octen-3-ol, acetone, and 4-methyl phenol (also known as p-cresol) (*10*). We then classified the ORNs based on their responses to the 15 compounds, using a hierarchical cluster analysis. Neurons classified as "A" fell into three functional classes, whereas neurons classified as "B" fell

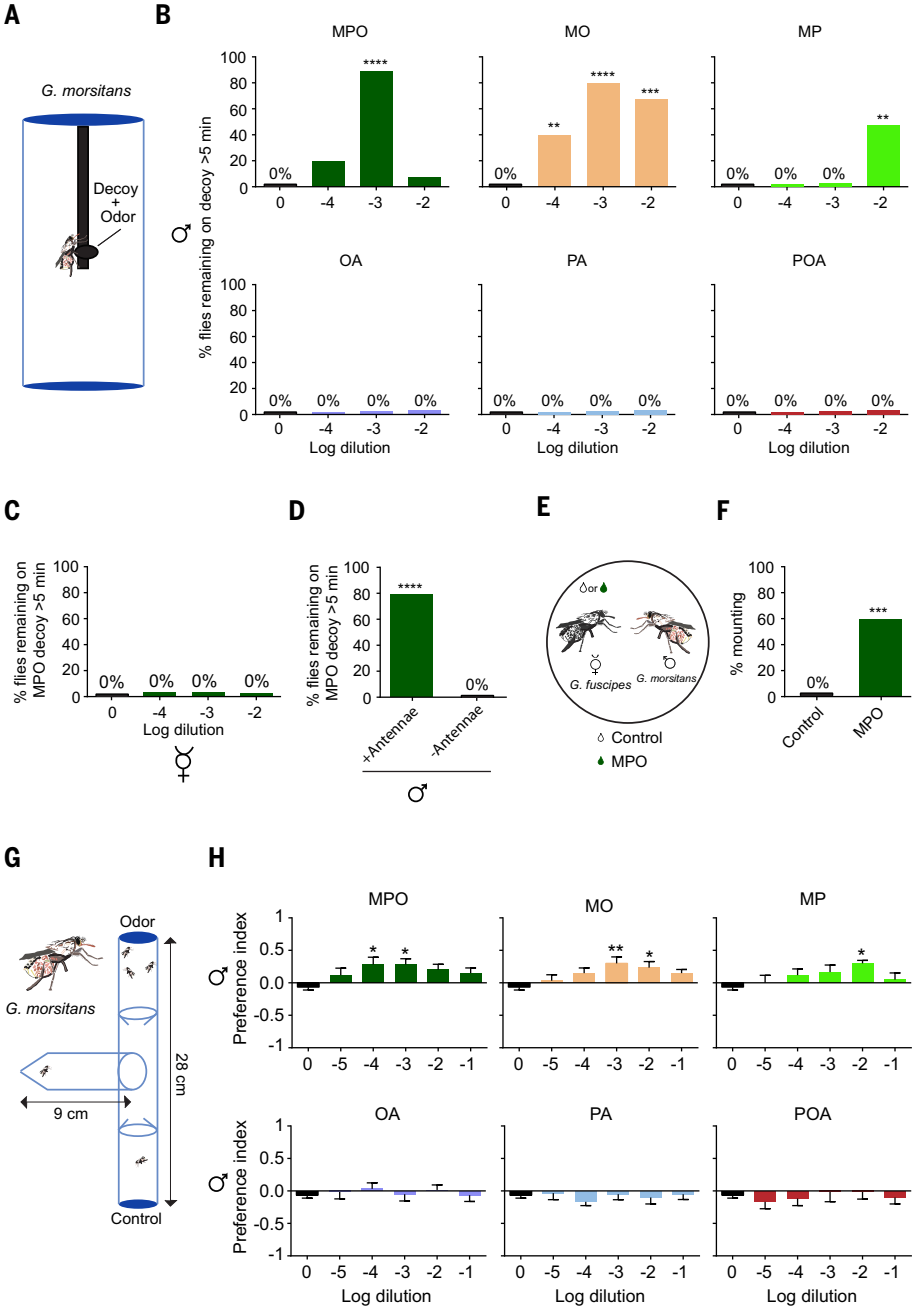


Fig. 3. MPO acts as an aphrodisiac pheromone in *G. morsitans* males. (**A**) The decoy paradigm. The decoy is dosed with a candidate pheromone or odor. (**B**) Percentage of males that landed on the decoy and remained on it continuously for more than 5 min; n = 15. Chi-squared test. (**C**) Percentage of *G. morsitans* virgin females that stayed for more than 5 min on a decoy dosed with a range of dilutions of methyl palmitoleate; n = 15. (**D**) Dependence of *G. morsitans* response to decoy on antennae. The decoy was dosed with a 10^{-3} dilution of MPO in paraffin oil; n = 15. Chi-squared test. (**E**) Schematic of a test with a perfumed *G. fuscipes* virgin female and a *G. morsitans* male. (**F**) Percentage of mounting when a *G. fuscipes* virgin female is perfumed with a 10^{-3} dilution of MPO, compared with a female perfumed with diluent alone; n = 15. Chi-squared test. (**G**) The T-maze paradigm. Flies are initially in the tube shown on the left. The apparatus is placed horizontally on a benchtop. (**H**) Behavioral responses of *G. morsitans* males in the T-maze paradigm. *P < 0.05, **P < 0.01, Wilcoxon Signed Ranked Test, n =12. Error bars are SEM.

into four functional classes, in both males (Fig. 5A) and females (fig. S4). In some cases the classes appeared qualitatively similar but with different response magnitudes (e.g., B2 and B3 in Fig. 5B). The response profiles of most male ORNs, e.g., B1, have female counterparts that appear similar. [In the case of B1, the male and female counterparts were indistinguishable; $P > 0.05$, Analysis of Similarities (ANOSIM) test]. Neurons that responded to MPO, the B1 neurons, also responded to known *G. morsitans* attractants: two of the strongest responses were to the attractants 4-methylphenol and 1-octen-3-ol (Fig. 5B). These results suggest that MPO may mediate attractive behaviors at least in part through these ORNs.

MO and MP activate trichoid neurons at close range

MO and MP elicited behavioral responses from *G. morsitans* (Fig. 3, B and H), but not electrophysiological responses in our initial analysis (Fig. 4E). We reasoned that MO and MP might elicit physiological responses from sensilla on untested regions of the large and complex tsetse antenna (*13*). Alternatively, it is possible that given the low volatility of these long-chain compounds, the dosage or odorant dynamics used in the preceding survey were not adequate to elicit responses. To address this latter possibility, we modified the odor delivery system (Fig. 5C). We reduced the distance from the odor source to the antennal preparation (to ~1 mm) and eliminated the airflow, analogous to an approach successfully used with pheromones in *Drosophila* (*39*). This close proximity of odor source to antenna is reminiscent of the proximity of two mating flies in a natural context. The diluent control produced no firing among the tested male trichoid sensilla, and MPO elicited responses from B neurons but not A neurons (Fig. 5, D and E), consistent with our earlier results (Fig. 4C). However, using this new delivery system, MP elicited responses from B neurons and MO activated both A and B neurons. These results indicate that MP and MO, as well as MPO, activate olfactory neurons, consistent with their activity in attracting and arresting *G. morsitans* males.

G. fuscipes differs from *G. morsitans* in responses to MPO, MP, MO, and other odorants

We next examined the response of *G. fuscipes* trichoid sensilla to MPO, MP, and MO, using the same close-range delivery system used for *G. morsitans*. We again focused on trichoid sensilla that cover the lateral side of the male antenna. In contrast to our results with *G. morsitans*, the tested methyl esters elicited little if any response from any tested sensilla when used at the same concentrations (fig. S5). Consistent with these physiological results, behavioral testing did not reveal a response to these compounds in either the decoy paradigm (fig. S5B) or the T-maze preference paradigm (fig. S5C). We also measured the abundance of MPO, MP, MO, and the other three compounds in virgin female, mated female, virgin male, and mated male *G. fuscipes*. In almost all cases the relative abundance of these compounds in *G. fuscipes* was lower

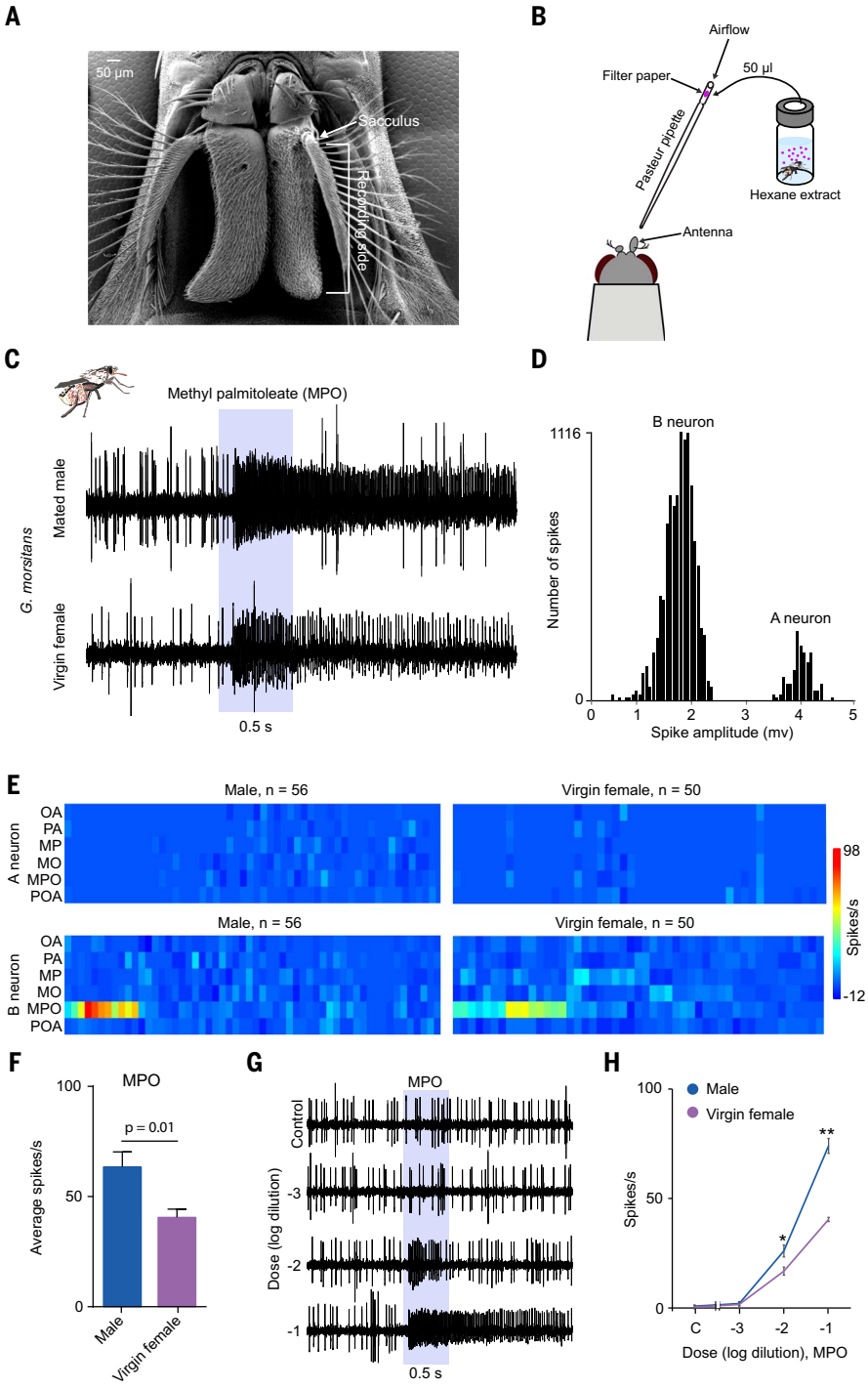


Fig. 4. MPO activates ORNs in certain antennal sensilla. (**A**) Scanning electron micrograph of antennae showing the region from which electrophysiological recordings were taken from trichoid sensilla. (**B**) The stimulus delivery system. (**C**) Example traces of electrophysiological responses of trichoid sensilla in males and virgin females to a 10^{-1} dilution of methyl palmitoleate. The shaded area indicates the 0.5-s odor stimulus. (**D**) The bimodal distribution of spike amplitudes in trichoid sensilla that respond to MPO in mated males. "A" and "B" indicate subpopulations of spikes attributed to neurons A and B. (**E**) Heatmap of electrophysiological responses of trichoid sensilla to compounds identified in body-wash extracts. Each rectangle shows the response magnitude of one neuron in spikes per s, each from a different trichoid sensillum, when tested with a 10^{-1} dilution of each compound. (**F**) Mean responses to a 10^{-1} dilution of MPO in males and virgin females. Mann-Whitney test, $n = 11$ for males and $n = 15$ for virgin females. Error bars are SEM. (**G**) Example traces of electrophysiological responses of trichoid sensilla in virgin females to a range of dilutions of methyl palmitoleate. In the control trace the stimulus was the diluent control. (**H**) Dose-response curves of trichoid sensilla to a range of dilutions of methyl palmitoleate. *$P < 0.05$; **$P < 0.01$, Mann-Whitney test; $n = 5$. Mean ± SEM; "C" = control diluent.

than that in *G. morsitans* (fig. S6). The lack of responses to MPO, MP, and MO motivated us to ask whether these *G. fuscipes* sensilla respond to other odorants, and if so whether the response profiles to these odorants differ from those of *G. morsitans*. Accordingly, we surveyed the trichoid sensilla on the lateral face of the male antenna with the same nine odorants tested against *G. morsitans*. Strong olfactory responses were recorded in *G. fuscipes* (fig. S5D). The A neurons could be divided through cluster analysis into four distinguishable classes, in contrast to three in *G. morsitans*; the B neurons fell into four classes in both species. In both species there are classes that responded to none of the tested odorants. Testing with more odorants might allow more detailed classification. Inhibitory responses were observed in *G. fuscipes* but not in *G. morsitans*. One class of B neurons, B4, was strongly inhibited by several odorants, including 6-methyl-5-hepten-2-one, which excited another class of B neurons, B3.

Trypanosome infection affects chemical profiles and mating behavior

We next investigated the impact of trypanosome infection on olfactory physiology, profiles of body-wash extracts, and mating. All previous experiments in this study were conducted with uninfected tsetse flies. We infected *G. morsitans* with *Trypanosoma brucei brucei* (strain RUMP 503) under laboratory conditions and compared them to uninfected controls. We first asked whether olfactory responses underwent major alterations after infection. From infected flies we measured the responses of 27 trichoid sensilla on the lateral surface of the antenna to a panel of 10 compounds, including MPO (fig. S7A). We then compared neuronal spaces constructed from the neuronal responses of infected (fig. S7B) and uninfected *G. morsitans* (Fig. 5). We found many common features and much overall similarity ($P = 0.08$, R = 0.0103 for A neurons; $P = 0.04$, R = 0.8 for B neurons; ANOSIM based on Bray-Curtis similarity; fig. S7B). We then compared chemical profiles of infected and uninfected flies. Hexane extracts of infected mated males and females both contained 21 small volatile compounds, including α-pinene, that were not observed in uninfected control flies or in the solvent control (Fig. 6A). These 21 compounds were specific not only to infected flies but also to mated flies: we did not observe them in either infected or uninfected virgin males or females (fig. S7C). To test the robustness of these results we individually examined 19 to 29 flies of each of the eight types (infected and uninfected mated males, infected and uninfected mated females, infected and uninfected virgin males, infected and uninfected virgin females) and found a high degree of consistency: the appearance of these compounds depended on both infection and

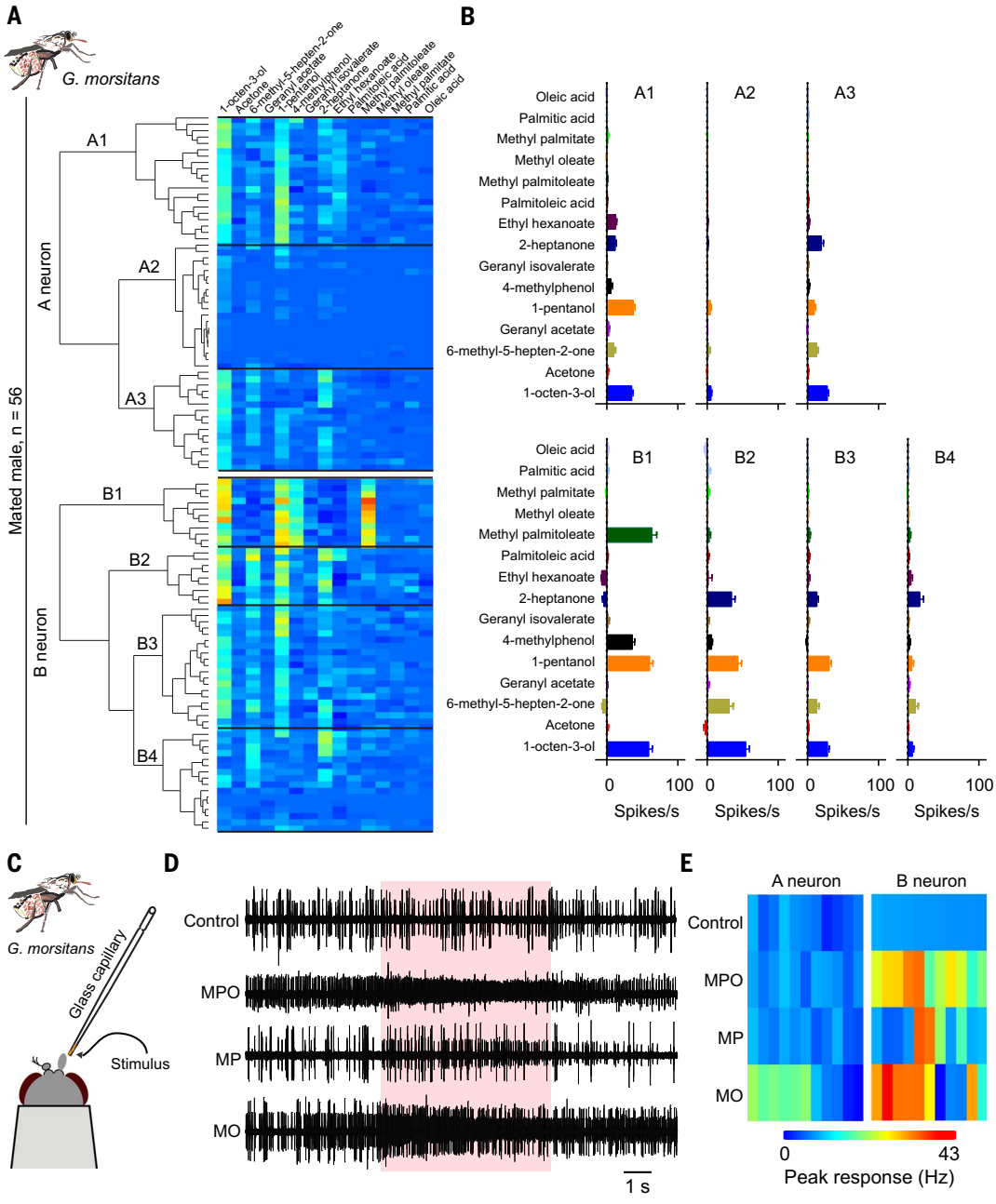


Fig. 5. ORNs that respond to MPO also respond to known *G. morsitans* attractants. (**A**) Heatmap based on hierarchical cluster analysis of responses of male trichoid sensilla to a panel of nine odorants and the six newly identified compounds. In the heatmap each horizontal row represents one trichoid sensillum and each vertical column represents one of the olfactory stimuli. The classification was carried out with Ward's method. The 9 odorants were diluted 10^{-2} in paraffin oil; the six compounds were diluted 10^{-1} in paraffin oil. (**B**) Response profiles of neuronal classes in trichoid sensilla; means ± SEM. n values range from 9 to 20, as shown in panel A. (**C**) Close-range stimulus delivery system. The stimulus is placed ~1 mm from the antenna, at the end of a glass capillary. (**D**) Example traces of electrophysiological responses of male trichoid sensilla to MPO (neat), MP (neat), and MO (10^{-1} dilution). MPO and MP elicit an increase in the frequency of the small spikes (B neurons); MO elicits an increase in the frequency of both small and large spikes (B and A neurons, respectively). (**E**) Heatmaps of electrophysiological responses of A and B neurons to MPO (neat), MP (neat), and MO (10^{-1} dilution) in males. Responses are peak responses.

mating. Finally we investigated whether infection had any effects on mating behavior. We placed a virgin female in a tube with two males, one infected and one uninfected. The female mated with the two kinds of males with equal frequency (Fig. 6B, $P > 0.05$, Chi-squared test, $n = 20$). However, when an uninfected virgin male was placed with two females, one infected and one uninfected, in all of 20 trials the male mated with the uninfected female (Fig. 6C); the infected females were much less receptive to the males.

Discussion

G. morsitans males mounted females very quickly upon coming into each other's vicinity. Although behaviors in laboratory settings are imperfect representations of behaviors in natural settings, in the same laboratory paradigm, pairs of *G. morsitans* had mounting latencies that were 1/100th those of *D. melanogaster*. Whereas *D. melanogaster* males engage in a prolonged and elaborate series of courtship behaviors prior to copulation (*40*, *41*), *G. morsitans* males were quick to mount. Once initiated, copulation lasted much longer in *G. morsitans* than in *D. melanogaster*. Moreover, the copulation time of *G. morsitans*, 58 min, exceeded that of all but one of 82 species of *Drosophila* considered in an analysis (*42*). Our measurement of mean copulation time is consistent with the observation of long-lasting copulation in studies of *G. morsitans* insemination and ovulation (*43*), and of interactions between males and decoys (*44*). These differences between the two fly species raise questions about whether the molecular, cellular, and circuit basis of sexual behavior differs markedly between tsetse flies and *Drosophila*, whose ecology, physiology, and behavior are distinctly different (*40*, *41*).

We have carried out behavioral, chemical, and electrophysiological analysis of tsetse to gain insight into the underpinnings of tsetse fly sexual behavior, and found evidence that volatile compounds can affect mating behavior in tsetse flies, despite little prior evidence supporting a role for volatile pheromones in these animals. The compound

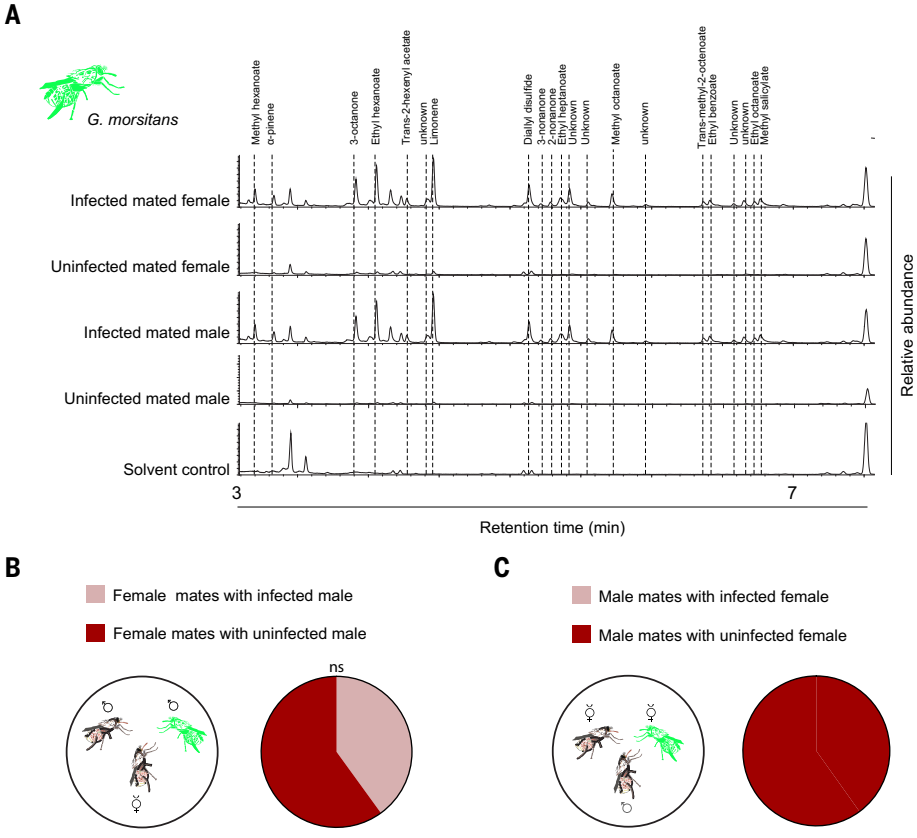


Fig. 6. Trypanosome infection changes chemical profile and behavior. (**A**) Total ion current chromatograms of 24-hour hexane extracts of infected and uninfected mated *G. morsitans*. All flies were 14 days old. (**B**) Healthy females mated with both infected and uninfected males. ($P > 0.05$, Chi-squared test, $n = 20$). All flies were virgins and all pairs copulated. (**C**) Healthy males mated with uninfected but not infected females. All flies were virgins and all pairs copulated; $n = 20$.

15,19,23-trimethylheptatriacontane has previously been identified as a contact sex pheromone in *G. morsitans* (*5*, *45*). Five-day-old females contain more than 4 mg of 15,19,23-trimethylheptatriacontane, and when microgram quantities were placed on dead males they elicited copulatory responses from other males (*7*). Although widely considered a contact pheromone, this compound was reported in one study to elicit an olfactory response from the antenna (*46*); however, this finding has been controversial (*2*, *3*), in part because the compound consists of a chain of 37 carbon atoms and its relative vapor pressure has been calculated to be 12 orders of magnitude lower than that of (Z)-9-tricosene, a volatile sex pheromone of houseflies (*47*).

We have found that *G. morsitans* flies produce MPO, MO, and MP; these volatile compounds have chain lengths of only 16 carbons, much shorter than the 37-carbon chain length of 5,19,23-trimethylheptatriacontane or the 23 carbons of (Z)-9-tricosene. MPO, MO, and MP had an arrestant effect in a decoy assay and an attractive effect in a T-maze assay. MPO is the compound for which we found the most evidence supporting a potential role as a volatile pheromone in *G. morsitans*. There are several arguments in favor of such a role: (i) MPO acts as a pheromone in other species (*22*, *28*). (ii) It elicited a response from males but not virgin females in the decoy assay. (iii) MPO showed characteristics of an aphrodisiac: *G. morsitans* males mounted *G. fuscipes* virgin females that were perfumed with MPO but not control *G. fuscipes* females, consistent with a role for MPO in driving male *G. morsitans* sexual behavior. (iv) MPO elicited greater responses from ORNs of males than virgin females. (v) It elicited little if any response from any tested sensilla in *G. fuscipes* (fig. S5A). Although we cannot exclude the possibility of responses from untested sensilla, we observed no behavioral responses to MPO in *G. fuscipes*, in either of two paradigms. MPO was recovered from a 24-hour hexane extract but was not detectable in a 10-min extract as one might expect of a cuticular pheromone. The 24-hour extracts have two salient properties: They showed sexual dimorphism in their effects in the decoy paradigm, and the number of peaks in the 24-hour extract did not greatly exceed the number in the 10 min extracts. Thus, of the vast number of molecules in the fly, only a few were detected in the 24-hour extract, MPO among them.

Although we do not know the typical distribution of MPO within flies, MPO might be synthesized and stored in an internal gland, such as those of Tephritid fruit flies, cockroaches, and a variety of hymenopterans and lepidopterans, which produce pheromones and release them during certain behaviors, often through a duct to a pore in the cuticle (*48*–*51*); in some species a pheromone is released only when needed (*48*, *52*). Alternatively, MPO might reside largely in internal oenocytes that produce pheromones and deliver them to the cuticle only when the fly is in certain nutritional or behavioral states (*53*–*57*). However, the absence of detectable MPO in the cuticular layer of flies fed and maintained under standard laboratory conditions may explain why it has not been identified previously as a pheromone in tsetse.

Insect pheromones are extremely diverse in many ways. Some, such as the classic example of bombykol in moths, act in long-distance attraction, are secreted specifically by females, elicit electrophysiological responses from narrowly tuned male-specific ORNs, and elicit male-specific behavioral responses (*58*, *59*). By contrast, other insect pheromones (i) elicit other behaviors (*60*); (ii) are secreted at comparable levels by both sexes (*26*, *39*); (iii) elicit electrophysiological responses from broadly tuned ORNs in both sexes (*5*, *26*, *61*); or (iv) elicit behavioral responses from both sexes (*26*). The sensitivity of ORNs to their cognate pheromones also varies greatly: one molecule of bombykol is sufficient to elicit a nervous impulse from a male moth ORN (*62*, *63*); by contrast, a 10^{-2} dilution of *cis*-vaccenyl acetate was required to elicit responses from its cognate at1 neuron of *Drosophila* (*64*), similar to the sensitivity of tsetse trichoid sensilla to MPO that we have observed. MPO has shown characteristics of both an attractant and an arrestant in laboratory tests. We speculate that as an attractant produced by females it might contribute to initiation of mating whereas as an arrestant it might contribute to the maintenance of mating, preventing premature termination. MPO activates neurons that also respond to the olfactory attractants 4-methylphenol and 1-octen-3-ol. These results support the interpretation that MPO activates a circuit that mediates olfactory attraction. Mating encounters of *Glossina* generally occur on or near hosts (*9*), and 4-methylphenol and 1-octen-3-ol are both host odors (*10*). The circuit might be activated more strongly when both fly and host cues are presented together. We note precedent for the dual response of neurons to pheromones and odorants: in *Drosophila* the ab9A ORNs respond both to the pheromone (Z)-4-undecenal and to food odors (*61*). These ORNs express two

receptors, one of which responds to (Z)-4-undecenal and the other to food odors.

In addition to promoting or maintaining mounting, MPO could have roles in any of a series of stereotyped sexual behaviors that occur after mounting (*65*). MPO could also affect female receptivity. Our results support a role for olfactory cues in tsetse mating behavior. The overall role of olfactory cues in tsetse mating, however, remains to be determined. Multiple sensory modalities are likely to influence mating behavior in tsetse, as in many other animal species (*3*). Moreover, there is evidence that *G. morsitans* males from which antennae have been surgically removed are able to mount conspecific females (*2*). These results are consistent with those from *Drosophila*, in which cues of several modalities act in mating behavior, but mating does not require the presence of all modalities (*66*).

We found that infection with trypanosomes changed the chemical profile of mated males and mated females: 21 volatile compounds were identified in the body-wash extracts of infected flies but not in their uninfected counterparts. This finding raises a series of questions, including whether these compounds are synthesized by the fly or by the parasites. Malaria parasites that infect mosquitoes produce a variety of volatile compounds, including α-pinene, which is one of the 21 compounds we identified (*67*). Although trypanosomes could synthesize the 21 identified compounds directly, infection also causes a wide variety of effects on the gene expression of tsetse flies (*68*, *69*), some of which could promote the production of volatile compounds by the fly. Perhaps the appearance of these compounds requires mating because it allows their transport from the interior of the fly to the external surface. In support of this possibility, evidence indicates that the act of mating exposes pores in the ninth tergite of the male, through which compounds could be released (*70*). Previous data have shown that a *G. morsitans* odor receptor, GmmOr35, responds to α-pinene (*13*), supporting the possibility that the infection or mating status of a fly could be discerned by another fly through olfaction; α-pinene also elicited antennal responses from three other species of *Glossina* (*16*). Infection with trypanosomes reduced mating receptivity in females. Animals fighting infections may invest less in reproduction as there is a trade-off between these processes due to limited energy resources (*71*). Consistent with this argument, fecundity of infected female tsetse flies is reduced (*72*), and mosquitoes infected with malaria parasites produce fewer eggs (*73*, *74*).

The identification of volatile pheromones in tsetse flies may in the long term have important implications for disease control. One of the most effective means of controlling tsetse flies is with traps and targets, which historically have included attractive host odors. Our results suggest the possibility of using tsetse odors, in combination with long-range attraction to host odors, for controlling tsetse flies and disease spread.

REFERENCES AND NOTES

1. C. Mucignat-Caretta, *Neurobiology of Chemical Communication*. C. Mucignat-Caretta, Ed. (CRC Press, 2014).
2. P. A. Lan gley, P. M. Huyton, D. A. Carlson, Sex pheromone perception by males of the tsetse fly, Glossina morsitans morsitans. *Physiol. Entomol.* **12**, 425–433 (1987). doi: 10.1111/j.1365-3032.1987.tb00769.x
3. R. Wall, P. A. Langley, The mating behaviour of tsetse flies (Glossina): A review. *Physiol. Entomol.* **18**, 211–218 (1993). doi: 10.1111/j.1365-3032.1993.tb00470.x
4. T. Engl *et al*., Effect of antibiotic treatment and gamma-irradiation on cuticular hydrocarbon profiles and mate choice in tsetse flies (Glossina m. morsitans). *BMC Microbiol.* **18**, 145 (2018). doi: 10.1186/s12866-018-1292-7; pmid: 30470188
5. D. A. Carlson, P. A. Langley, P. Huyton, Sex pheromone of the tsetse fly: Isolation, identification, and synthesis of contact aphrodisiacs. *Science* **201**, 750–753 (1978). doi: 10.1126/science.675256; pmid: 675256
6. M. J. R. Hall, P. A. Langley, The responses of individual males in an isolated population of Glossina morsitans morsitans Westwood (Diptera: Glossinidae) to pheromone-baited decoy 'females'. *Bull. Entomol. Res.* **79**, 319–334 (1989). doi: 10.1017/S0007485300018307
7. P. M. Huyton, P. A. Langley, D. A. Carlson, T. W. Coates, The role of sex pheromones in initiation of copulatory behaviour by male tsetse flies, Glossina morsitans morsitans. *Physiol. Entomol.* **5**, 243–252 (1980). doi: 10.1111/j.1365-3032.1980.tb00232.x
8. P. P. Simarro *et al*., Estimating and mapping the population at risk of sleeping sickness. *PLOS Negl. Trop. Dis.* **6**, e1859 (2012). doi: 10.1371/journal.pntd.0001859; pmid: 23145192
9. T. A. M. Nash, *Africa's Bane: The Tsetse Fly*. (Collins, 1969).
10. S. J. Torr, G. A. Vale, Know your foe: Lessons from the analysis of tsetse fly behaviour. *Trends Parasitol.* **31**, 95–99 (2015). doi: 10.1016/j.pt.2014.12.010; pmid: 25599585
11. G. V. Reddy, A. Guerrero, New pheromones and insect control strategies. *Vitam. Horm.* **83**, 493–519 (2010). doi: 10.1016/S0083-6729(10)83020-1; pmid: 20831959
12. J. B. Benoit, G. M. Attardo, A. A. Baumann, V. Michalkova, S. Aksoy, Adenotrophic viviparity in tsetse flies: Potential for population control and as an insect model for lactation. *Annu. Rev. Entomol.* **60**, 351–371 (2015). doi: 10.1146/annurev-ento-010814-020834; pmid: 25341093
13. J. S. Chahda *et al*., The molecular and cellular basis of olfactory response to tsetse fly attractants. *PLOS Genet.* **15**, e1008005 (2019). doi: 10.1371/journal.pgen.1008005; pmid: 30875383
14. N. Soni, J. S. Chahda, J. R. Carlson, Odor coding in the antenna of the tsetse fly *Glossina morsitans*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14300–14308 (2019). doi: 10.1073/pnas.1907075116; pmid: 31221757
15. V. Harraca, Z. Syed, P. M. Guerin, Olfactory and behavioural responses of tsetse flies, Glossina spp., to rumen metabolites. *J. Comp. Physiol. A* **195**, 815–824 (2009). doi: 10.1007/s00359-009-0459-y; pmid: 19626328
16. Z. Syed, P. M. Guerin, Tsetse flies are attracted to the invasive plant Lantana camara. *J. Insect Physiol.* **50**, 43–50 (2004). doi: 10.1016/j.jinsphys.2003.09.007; pmid: 15037092
17. Y. Snellings *et al*., The role of cuticular hydrocarbons in mate recognition in Drosophila suzukii. *Sci. Rep.* **8**, 4996 (2018). doi: 10.1038/s41598-018-23189-6; pmid: 29567945
18. B. Foley, S. F. Chenoweth, S. V. Nuzhdin, M. W. Blows, Natural genetic variation in cuticular hydrocarbon expression in male and female Drosophila melanogaster. *Genetics* **175**, 1465–1477 (2007). doi: 10.1534/genetics.106.065771; pmid: 17194783
19. L. Vaníčková *et al*., Cuticular hydrocarbons of the South American fruit fly Anastrepha fraterculus: Variability with sex and age. *J. Chem. Ecol.* **38**, 1133–1142 (2012). doi: 10.1007/s10886-012-0177-8; pmid: 22948785
20. C. L. Vernier *et al*., The cuticular hydrocarbon profiles of honey bee workers develop via a socially-modulated innate process. *eLife* **8**, e41855 (2019). doi: 10.7554/eLife.41855; pmid: 30720428
21. A. Mitra, Function of the Dufour's gland in solitary and social Hymenoptera. *J. Hymenopt. Res.* **35**, 33–58 (2013). doi: 10.3897/jhr.35.4783
22. S. Noushini *et al*., Attraction and Electrophysiological Response to Identified Rectal Gland Volatiles in *Bactrocera frauenfeldi* (Schiner). *Molecules* **25**, 1275 (2020). doi: 10.3390/molecules25061275; pmid: 32168881
23. V. B. Wigglesworth, The source of lipids and polyphenols for the insect cuticle: The role of fat body, oenocytes and oenocytoids. *Tissue Cell* **20**, 919–932 (1988). doi: 10.1016/0040-8166(88)90033-X; pmid: 18620248
24. S. Abril *et al*., Cuticular hydrocarbons correlate with queen reproductive status in native and invasive Argentine ants (Linepithema humile, Mayr). *PLOS ONE* **13**, e0193115 (2018). doi: 10.1371/journal.pone.0193115; pmid: 29470506
25. L. G. Batista-Pereira *et al*., Cuticular hydrocarbons of Heterotermes tenuis (Isoptera: Rhinotermitidae): analyses and electrophysiological studies. *Z. Naturforsch. C J. Biosci.* **59**, 135–139 (2004). doi: 10.1515/znc-2004-1-226; pmid: 15018067
26. H. K. Dweck *et al*., Pheromones mediating copulation and attraction in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2829–E2835 (2015). doi: 10.1073/pnas.1504527112; pmid: 25964351
27. R. A. Jurenka, M. Subchev, Identification of cuticular hydrocarbons and the alkene precursor to the pheromone in hemolymph of the female gypsy moth, Lymantria dispar. *Arch. Insect Biochem. Physiol.* **43**, 108–115 (2000). doi: 10.1002/(SICI)1520-6327(200003)43:3<108::AID-ARCH2>3.0.CO;2-V; pmid: 10685098
28. M. Ayasse, T. Marlovits, J. Tengö, T. Taghizadeh, W. Francke, Are there pheromonal dominance signals in the bumblebee Bombus hypnorum L (Hymenoptera, Apidae)? *Apidologie (Celle)* **26**, 163–180 (1995). doi: 10.1051/apido:19950301
29. C. I. Keeling, K. N. Slessor, H. A. Higo, M. L. Winston, New components of the honey bee (Apis mellifera L.) queen retinue pheromone. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4486–4491 (2003). doi: 10.1073/pnas.0836984100; pmid: 12676987
30. A. M. El-Sayed, "The Pherobase: database of pheromones and semiochemicals" (Pherobase, 2022); www.pherobase.com.
31. J. M. González, A. Cusumano, H. J. Williams, S. Colazza, S. B. Vinson, Behavioral and chemical investigations of contact kairomones released by the mud dauber wasp Trypoxylon politum, a host of the parasitoid Melittobia digitata. *J. Chem. Ecol.* **37**, 629–639 (2011). doi: 10.1007/s10886-011-9960-1; pmid: 21552995
32. H. A. Lloyd, N. R. Schmuff, A. Hefetz, Chemistry of the anal glands of Bothriomyrmex syrius forel. Olfactory mimetism and temporary social parasitism. *Comp. Biochem. Physiol. B* **83**, 71–73 (1986). doi: 10.1016/0305-0491(86)90333-0; pmid: 3943313
33. G. Guiraudie-Capraz, D. B. Pho, J. M. Jallon, Role of the ejaculatory bulb in biosynthesis of the male pheromone cis-vaccenyl acetate in Drosophila melanogaster. *Integr. Zool.* **2**, 89–99 (2007). doi: 10.1111/j.1749-4877.2007.00047.x; pmid: 21396023
34. M. A. Khallaf *et al*., Mate discrimination among subspecies through a conserved olfactory pathway. *Sci. Adv.* **6**, eaba5279 (2020). doi: 10.1126/sciadv.aba5279; pmid: 32704542
35. M. A. Khallaf *et al*., Large-scale characterization of sex pheromone communication systems in Drosophila. *Nat. Commun.* **12**, 4165 (2021). doi: 10.1038/s41467-021-24395-z; pmid: 34230464
36. J. Y. Yew, K. Dreisewerd, C. C. de Oliveira, W. J. Etges, Male-specific transfer and fine scale spatial differences of newly identified cuticular hydrocarbons and triacylglycerides in a Drosophila species pair. *PLOS ONE* **6**, e16898 (2011). doi: 10.1371/journal.pone.0016898; pmid: 21369358
37. M. F. Ali, E. D. Morgan, Chemical communication in insect communities: A guide to insect pheromones with special emphasis on social insects. *Biol. Rev.* **65**, 227–247 (1990). doi: 10.1111/j.1469-185X.1990.tb01425.x
38. E. S. Krafsur, Tsetse flies: Genetics, evolution, and role as vectors. *Infect. Genet. Evol.* **9**, 124–141 (2009). doi: 10.1016/j.meegid.2008.09.010; pmid: 18992846
39. W. van der Goes van Naters, J. R. Carlson, Receptors and neurons for fly odors in Drosophila. *Curr. Biol.* **17**, 606–612 (2007). doi: 10.1016/j.cub.2007.02.043; pmid: 17363256
40. H. J. Pavlou, S. F. Goodwin, Courtship behavior in Drosophila melanogaster: Towards a 'courtship connectome'. *Curr. Opin. Neurobiol.* **23**, 76–83 (2013). doi: 10.1016/j.conb.2012.09.002; pmid: 23021897
41. B. J. Dickson, Wired for sex: The neurobiology of Drosophila mating decisions. *Science* **322**, 904–909 (2008). doi: 10.1126/science.1159276; pmid: 18988843
42. T. Markow, "Evolution of Drosophila mating systems" in Evolutionary Biology, M. K. Hecht, ed. (Plenum press, 1996), pp. 73–106.

43. D. S. Saunders, C. W. H. Dodd, Mating, insemination, and ovulation in the tsetse fly, Glossina morsitans. *J. Insect Physiol.* **18**, 187–198 (1972). doi: 10.1016/0022-1910(72)90119-9
44. R. Wall, Tsetse mating behaviour: Effects of age and hunger in Glossina morsitans morsitans and G.pallidipes. *Physiol. Entomol.* **13**, 479–486 (1988). doi: 10.1111/j.1365-3032.1988.tb01132.x
45. P. M. Huyton, P. A. Langley, D. A. Carlson, M. Schwarz, Specificity of contact sex pheromones in tsetse flies, Glossina spp. *Physiol. Entomol.* **5**, 253–264 (1980). doi: 10.1111/j.1365-3032.1980.tb00233.x
46. C. J. den Otter, R. K. Saini, Pheromone perception in the tsetse fly Glossina morsitans morsitans. *Entomol. Exp. Appl.* **39**, 155–161 (1985). doi: 10.1111/j.1570-7458.1985.tb03557.x
47. D. A. Carlson *et al*., Sex attractant pheromone of the house fly: Isolation, identification and synthesis. *Science* **174**, 76–78 (1971). doi: 10.1126/science.174.4004.76; pmid: 5120874
48. C. Schal, Y. Fan, G. J. Blomquist, in *Insect Pheromone Biochemistry and Molecular* Biology, G. Blomquist, R. Vogt, Eds. (Academic Press, 2003), pp. 283–322.
49. F. Scolari, F. Valerio, G. Benelli, N. T. Papadopoulos, L. Vaníčková, Tephritid Fruit Fly Semiochemicals: Current Knowledge and Future Perspectives. *Insects* **12**, 408 (2021). doi: 10.3390/insects12050408; pmid: 33946603
50. M. Ayasse, R. J. Paxton, J. Tengö, Mating behavior and chemical communication in the order Hymenoptera. *Annu. Rev. Entomol.* **46**, 31–78 (2001). doi: 10.1146/annurev.ento.46.1.31; pmid: 11112163
51. Z. H. K. M. A. D. Mushtaq A Ganai, Pheromones in lepidopteran insects: Types, production, reception and its application. *J. Pharmacogn. Phytochem.* **6**, 2552–2558 (2017).
52. J. E. Percy-Cunningham, J. A. MacDonald, in *Pheromone Biochemistry*, G.D. Prestwich, G.J. Blomquist, Eds. (Academic Press, 1987), pp. 27–75.
53. K. Jensen *et al*., Change in sex pheromone expression by nutritional shift in male cockroaches. *Behav. Ecol.* **28**, 1393–1401 (2017). doi: 10.1093/beheco/arx120
54. F. C. Ingleby *et al*., Genotype-by-environment interactions for cuticular hydrocarbon expression in Drosophila simulans. *J. Evol. Biol.* **26**, 94–107 (2013). doi: 10.1111/jeb.12030; pmid: 23163514
55. Q. L. Ming, S. M. Lewis, Pheromone production by male Tribolium castaneum (Coleoptera: Tenebrionidae) is influenced by diet quality. *J. Econ. Entomol.* **103**, 1915–1919 (2010). doi: 10.1603/EC10110; pmid: 21061996
56. M. J. Rantala, R. Kortet, J. S. Kotiaho, A. Vainikka, J. Suhonen, Condition dependence of pheromones and immune function in the grain beetle Tenebrio molitor. *Funct. Ecol.* **17**, 534–540 (2003). doi: 10.1046/j.1365-2435.2003.00764.x
57. C. B. Weddle, C. Mitchell, S. K. Bay, S. K. Sakaluk, J. Hunt, Sex-specific genotype-by-environment interactions for cuticular hydrocarbon expression in decorated crickets, Gryllodes sigillatus: Implications for the evolution of signal reliability. *J. Evol. Biol.* **25**, 2112–2125 (2012). doi: 10.1111/j.1420-9101.2012.02593.x; pmid: 22900500
58. K.-E. Kaissling, G. Kasang, H. J. Bestmann, W. D. D.-C. Stransky, O. Vostrowsky, A new pheromone of the silkworm moth Bombyx mori. *Naturwissenschaften* **65**, 382–384 (2004). doi: 10.1007/BF00439702
59. T. Daimon *et al*., Female sex pheromone and male behavioral responses of the bombycid moth Trilocha varians: Comparison with those of the domesticated silkmoth Bombyx mori. *Naturwissenschaften* **99**, 207–215 (2012). doi: 10.1007/s00114-012-0887-3; pmid: 22307535
60. G. J. Blomquist, R. G. Vogt,. *Insect Pheromone Biochemistry and Molecular Biology: The Biosynthesis and Detection of Pheromones and Plant Volatiles* (Academic Press, 2003).
61. S. Lebreton *et al*., A Drosophila female pheromone elicits species-specific long-range attraction via an olfactory channel with dual specificity for sex and food. *BMC Biol.* **15**, 88 (2017). doi: 10.1186/s12915-017-0427-x; pmid: 28962619
62. K. E. Kaissling, C. Mucignat-Caretta, in *Neurobiology of Chemical Communication*, K. E. Kaissling, C. Mucignat-Caretta, Eds. (CRC Press, 2014), Ch. 4
63. K. E. Kaissling, E. Priesner, [Smell threshold of the silkworm]. *Naturwissenschaften* **57**, 23–28 (1970). doi: 10.1007/BF00593550; pmid: 5417282
64. P. Clyne, A. Grant, R. O'Connell, J. R. Carlson, Odorant response of individual sensilla on the Drosophila antenna. *Invert. Neurosci.* **3**, 127–135 (1997). doi: 10.1007/BF02480367; pmid: 9783438
65. P. M. Huyton, P. A. Langley, Copulatory behaviour of the tsetse flies Glossina morsitans and G.austeni. *Physiol. Entomol.* **7**, 167–174 (1982). doi: 10.1111/j.1365-3032.1982.tb00286.x
66. D. Yamamoto, M. Koganezawa, Genes and circuits of courtship behaviour in Drosophila males. *Nat. Rev. Neurosci.* **14**, 681–692 (2013). doi: 10.1038/nrn3567; pmid: 24052176
67. M. Kelly *et al*., Malaria parasites produce volatile mosquito attractants. *mBio* **6**, e00235-15 (2015). doi: 10.1128/mBio.00235-15; pmid: 25805727
68. E. Aksoy *et al*., Mammalian African trypanosome VSG coat enhances tsetse's vector competence. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6961–6966 (2016). doi: 10.1073/pnas.1600304113; pmid: 27185908
69. Z. Hao *et al*., Tsetse immune responses and trypanosome transmission: Implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12648–12653 (2001). doi: 10.1073/pnas.221363798; pmid: 11592981
70. A. Vigneron *et al*., A fine-tuned vector-parasite dialogue in tsetse's cardia determines peritrophic matrix integrity and trypanosome transmission success. *PLOS Pathog.* **14**, e1006972 (2018). doi: 10.1371/journal.ppat.1006972; pmid: 29614112
71. H. Hurd, Manipulation of medically important insect vectors by their parasites. *Annu. Rev. Entomol.* **48**, 141–161 (2003). doi: 10.1146/annurev.ento.48.091801.112722; pmid: 12414739
72. C. Hu *et al*., Infections with immunogenic trypanosomes reduce tsetse reproductive fitness: Potential impact of different parasite strains on vector population structure. *PLOS Negl. Trop. Dis.* **2**, e192 (2008). doi: 10.1371/journal.pntd.0000192; pmid: 18335067
73. A. M. Ahmed, H. Hurd, Immune stimulation and malaria infection impose reproductive costs in Anopheles gambiae via follicular apoptosis. *Microbes Infect.* **8**, 308–315 (2006). doi: 10.1016/j.micinf.2005.06.026; pmid: 16213176
74. C. S. Hacker, The differential effect of Plasmodium gallinaceum on the fecundity of several strains of Aedes aegypti. *J. Invertebr. Pathol.* **18**, 373–377 (1971). doi: 10.1016/0022-2011(71)90040-1; pmid: 5151912

ACKNOWLEDGMENTS

We are very grateful to S. Aksoy for the use of equipment in her laboratory and for providing financial support required to rear the tsetse flies and African trypanosomes used in this study. We thank A.M.M. Abd-Alla (International Atomic Energy Agency, Vienna, Austria) for providing *G. fuscipes fuscipes* pupae, and S. Chahda for the micrograph in Fig. 4A. Several figures make use of a Wikicommons image created by E. A. Gómez, which was modified as needed: https://en.wikipedia.org/wiki/Glossina_morsitans#/media/File:Glossina_morsitans_morsitans_(white_paper).jpeg. **Author contributions:** Conceptualization: S.A.M.E., H.K.M.D., B.L.W., and J.R.C. Investigation: S.A.M.E. and H.K.M.D. Methodology: S.A.M.E. and H.K.M.D. Visualization: S.A.M.E., H.K.M.D., and J.R.C. Funding acquisition: S.A.M.E., H.K.M.D., and J.R.C. Project administration: S.A.M.E. Supervision: J.R.C. Writing: S.A.M.E. and J.R.C. Writing – review and editing: S.A.M.E., H.K.M.D., B.L.W., and J.R.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main paper or supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade1877
Materials and Methods
Figs. S1 to S7
MDAR Reproducibility Checklist

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

Submitted 29 July 2022; accepted 12 December 2022
10.1126/science.ade1877

RESEARCH ARTICLE

ORGANIC CHEMISTRY

Dynamic kinetic asymmetric arylation and alkenylation of ketones

Lin-Xin Ruan, Bo Sun, Jia-Ming Liu, Shi-Liang Shi*

Despite the importance of enantioenriched alcohols in medicinal chemistry, total synthesis, and materials science, the efficient and selective construction of enantioenriched tertiary alcohols bearing two contiguous stereocenters has remained a substantial challenge. We report a platform for their preparation through the enantioconvergent, nickel-catalyzed addition of organoboronates to racemic, nonactivated ketones. We prepared several important classes of α,β-chiral tertiary alcohols in a single step with high levels of diastereo- and enantioselectivity through a dynamic kinetic asymmetric addition of aryl and alkenyl nucleophiles. We applied this protocol to modify several profen drugs and to rapidly synthesize biologically relevant molecules. We expect this nickel-catalyzed, base-free ketone racemization process to be a widely applicable strategy for the development of dynamic kinetic processes.

The development of general stereoselective methods with simultaneous control of contiguous stereocenters is critical to the efficient discovery and manufacture of new drugs. Enantioenriched alcohols constitute an important class of compounds that are found in a myriad of pharmaceutical agents and bioactive natural products (*1*, *2*). A variety of protocols for the asymmetric addition of organometallic nucleophiles or hydride reagents to carbonyls has enabled the efficient and stereoselective synthesis of alcohols with a single stereocenter (*3*–*5*). However, efficient methods for the preparation of alcohols bearing two adjacent stereocenters remain relatively underdeveloped. In particular, synthetic approaches to chiral tertiary alcohols with stereocenters at the α and β positions from simple and easily available precursors are rare, despite the frequent presence of this structural motif in bioactive molecules and drugs (Fig. 1A and fig. S1) (*6*–*8*). A typical method for the synthesis of α,β-chiral tertiary alcohols is the diastereoselective nucleophilic addition to enantiopure α-stereogenic ketones (Fig. 1B) (*9*, *10*). However, these precursors are difficult to prepare and potentially suffer from racemization during carbonyl addition reactions or during storage. Moreover, the stereocontrolled addition to the carbonyl group is generally nontrivial, and the results may be difficult to predict in many cases (*10*). In addition, to obtain reasonable levels of diastereoselectivity, reactions need to be conducted under low-temperature conditions with highly basic and nucleophilic organometallic reagents (for example, Mg or Li), which limit the type of compatible substrates and functional groups. In this context, it would be more desirable to use readily available racemic ketones and air- and moisture-stable nucleophiles for the general construction of stereodefined α,β-stereogenic tertiary alcohols in an enantioconvergent manner (Fig. 1C).

Among general approaches for achieving an enantioconvergent synthesis of tertiary alcohols with contiguous stereocenters, the dynamic kinetic asymmetric transformation (DyKAT) of racemic α-chiral ketones through the addition of air- and moisture-stable organoboronate reagents provides an elegant and efficient solution that is applicable to synthesize complex and functionalized targets (*11*–*18*). In general, to achieve a selective and efficient DyKAT, the first requirement is that the substrate racemization must be rapid [with a high k_{rac} (rate of racemization) in the equilibration between enantiomeric starting materials (*S*)-SM and (*R*)-SM] (Fig. 1D). In addition, there is a second prerequisite: the reaction of the chiral catalyst with one enantiomer of the substrate must occur with high stereoselectivity and at a substantially higher rate than the other enantiomer ($k_S \gg k_R$) to give the enantioenriched product. As one enantiomer of substrate is selectively consumed, the equilibrium shifts to allow both enantiomers of the racemic starting material to eventually convert into the chiral product. Thus, in principle, the maximum theoretical yield of a DyKAT process is 100%, which is superior to processes that are based on kinetic or classical resolutions with yields limited to 50%. The dynamic kinetic asymmetric hydrogenation of carbonyls has been well studied and even performed on more than a 100-ton scale annually to provide an ideal asymmetric synthesis of secondary alcohols, and important contributions have been made by the groups of Noyori (*19*), Zhou (*20*), Johnson (*21*), Zhang (*22*), and others (*23*). By contrast, dynamic kinetic asymmetric ketone additions to furnish chiral tertiary alcohols bearing adjacent stereocenters have been less explored. Currently, the DyKAT chemistry of ketones is limited to the addition of activated ketones (α-keto esters) to form complex glycolates, a technique developed by the Johnson group (*24*, *25*). However, to date, dynamic kinetic asymmetric additions to nonactivated ketones for general synthesis of chiral tertiary alcohols remain unreported.

Several factors have impeded the development of a protocol for enantioconvergent addition of nonactivated ketones (Fig. 1D). First, the racemization of nonactivated substrates through enolization is sluggish and requires strong bases, making it difficult to develop functional group–tolerant conditions that meet the kinetic requirements for a DyKAT. Second, the attenuated electrophilicity and increased steric hindrance of simple ketones lead to diminished reactivity for addition reactions in general. Third, the control over the diastereoselectivity and enantioselectivity for unactivated ketones is more challenging. For these reasons, almost all DyKATs of ketones use substrates bearing electron-withdrawing functional groups (such as ester or halogen groups) at the α-position of the ketone, and simple ketones are seldom used. However, we considered that developing a distinct reactivity paradigm for racemization and using a nontraditional carbonyl activation mode with a judicious choice of chiral ligand would address these longstanding critical challenges.

Toward this goal, we recently developed a series of bulky yet flexible C_2-symmetric chiral *N*-heterocyclic carbenes (NHC) (*26*), namely 1,3-bis(4-methyl-2,6-bis((*R*)-1-phenylethyl)phenyl)-4,5-dihydro-1*H*-imidazol-3-ium-2-ide (SIPE)– and 7,9-bis(4-methyl-2,6-bis((*R*)-1-phenylethyl)phenyl)-7*H*-acenaphtho[1,2-*d*]imidazol-9-ium-8-ide (ANIPE)–type ligands (*27*–*29*). We successfully applied the ligands to several challenging asymmetric transformations, including a highly enantioselective nickel-catalyzed arylation of ketones with organoboronates (*30*). A rare enantioselective η^2-coordinating activation of ketone carbonyls and the subsequent cross-coupling–like mechanism is the key to enable a general synthesis of chiral tertiary alcohols (*31*). Recently, we made the discovery that under base-free conditions, Ni-ANIPE complex can rapidly catalyze the racemization process of a chiral nonactivated ketone [(*R*)-**1b**], with the enantiomeric excess (ee) of (*R*)-**1b** decreasing from 98 to 0% in 15 minutes in the presence of 5 mol % Ni-**L1** at 30°C (Fig. 1E, reaction 1). We reasoned that electron-donating NHC-ligated Ni(0) species facilitate oxidative cyclization to form an η^2 Ni-ketone complex, which could undergo a reversible β-H elimination

State Key Laboratory of Organometallic Chemistry, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200032, China.
*Corresponding author. Email: shiliangshi@sioc.ac.cn

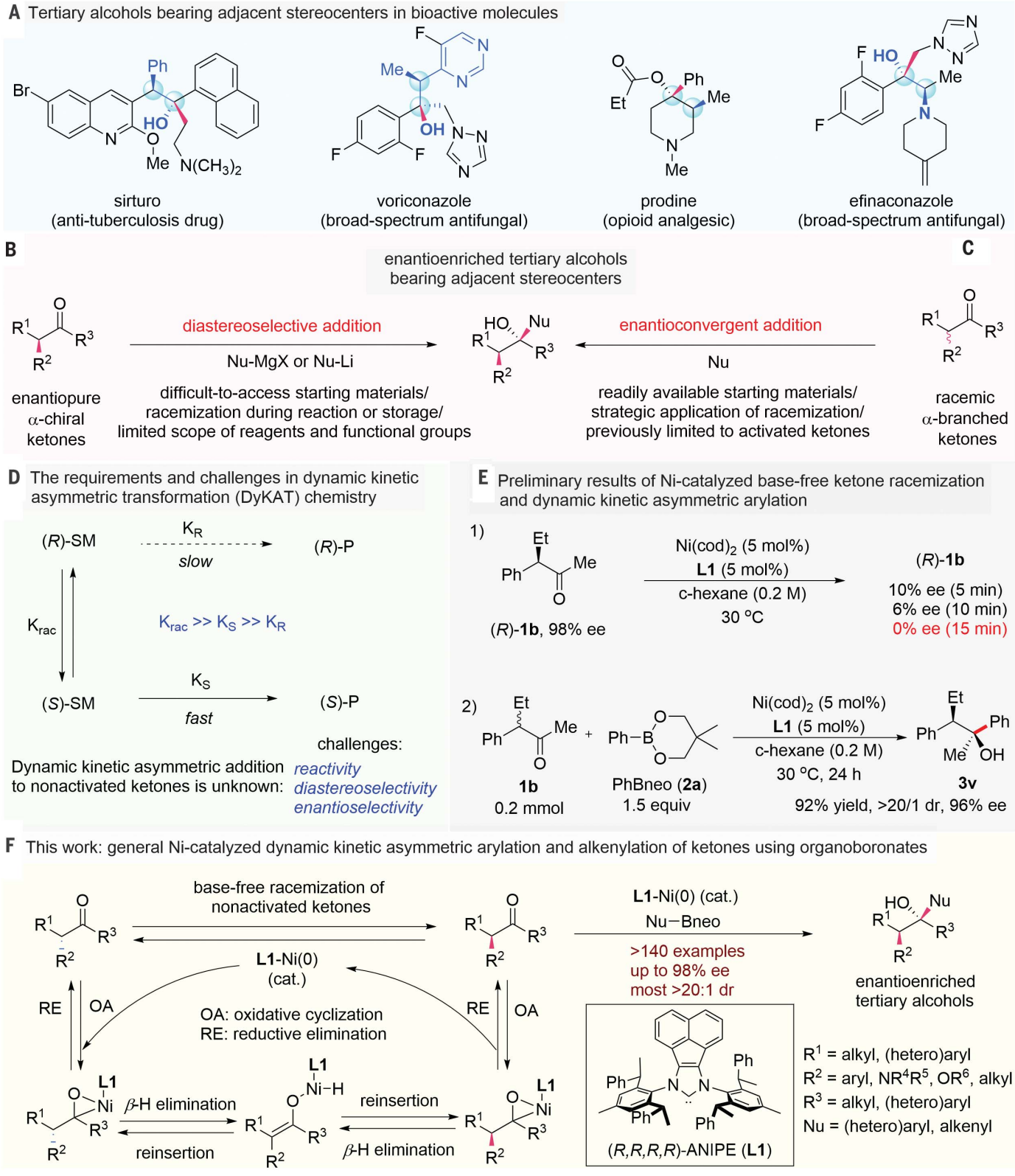


Fig. 1. Bioactive tertiary alcohols, their synthesis by traditional diastereoselective addition, and this work's DyKAT strategy. (**A**) Representative drugs demonstrating the ubiquitous nature of chiral tertiary alcohols in bioactive organic molecules. (**B** and **C**) Contrasting traditional methods and our proposed enantioconvergent approaches to enantioenriched tertiary alcohols bearing adjacent stereocenters. (**D**) The requirements and challenges of dynamic kinetic asymmetric addition to nonactivated ketones. (**E**) The discovery of a rapid nickel-catalyzed racemization of unactivated ketones and its application in DyKAT chemistry. (**F**) A nickel catalyst for general enantioconvergent addition of racemic nonactivated ketones to generate various chiral tertiary alcohols bearing two contiguous stereocenters.

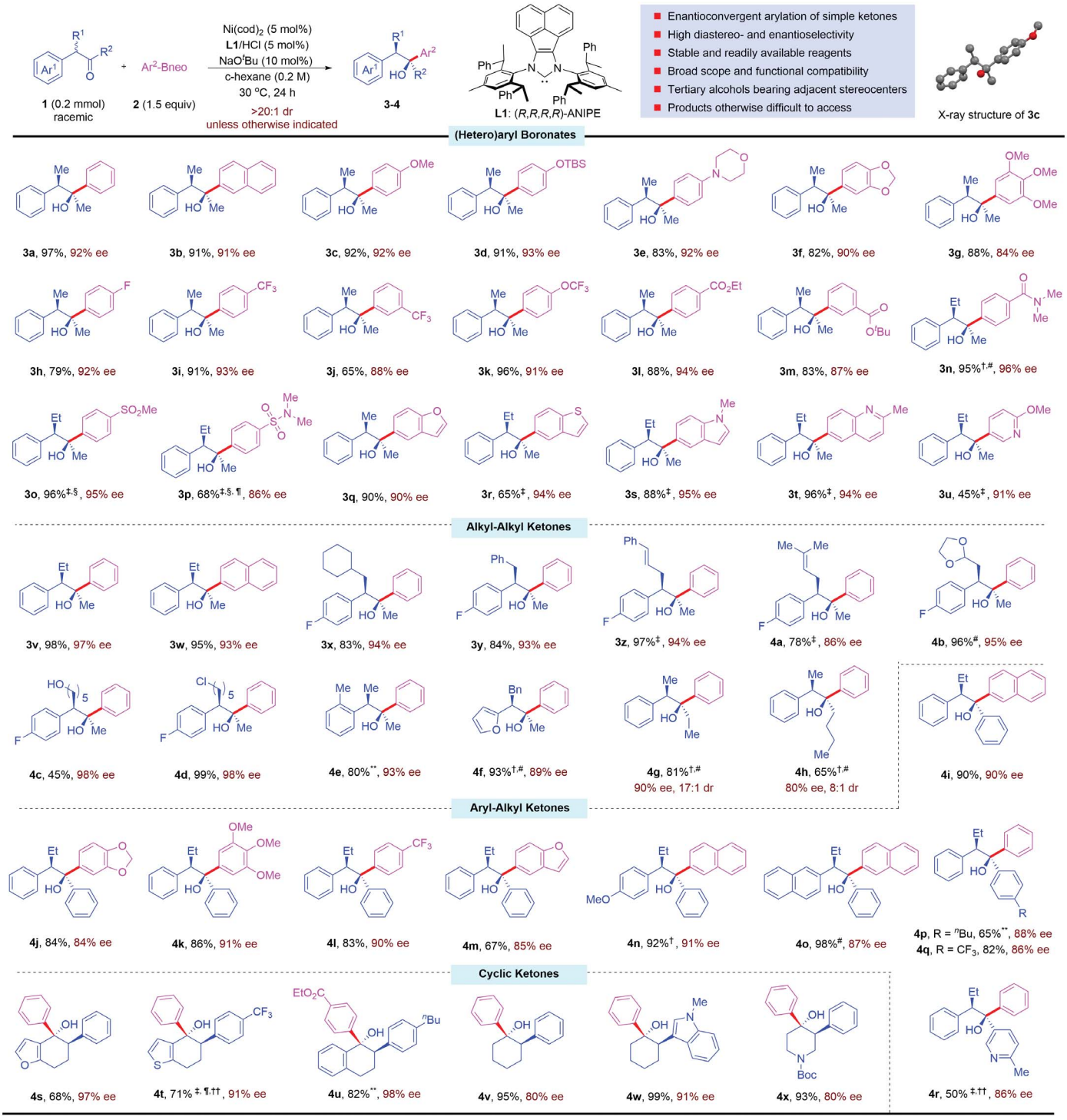


Fig. 2. Scope of nickel-catalyzed enantioconvergent arylation of α-aryl ketones. The yields of isolated products on a 0.2 mmol scale are shown. Values ee and dr were determined with HPLC and NMR analysis. The relative and absolute stereochemistries of products are shown according to the x-ray structure of **3c**. †10 mol % catalyst, 20 mol % NaO^{t}Bu; ‡10 mol % catalyst, 20 mol % NaO^{t}Bu, 50°C, 48 hours; §toluene as the solvent; ¶80°C. #50°C; **48 hours; ††Ti(O^{i}Pr)$_4$ was added (supplementary materials).

and reinsertion to rapidly racemize the ketone (Fig. 1F). We considered that these exceptionally rapid and mild racemization conditions would constitute a promising platform for the development of DyKATs that use α-chiral ketones as substrates. Moreover, given the excellent levels of stereocontrol that have previously been achieved with ligands in the ANIPE family, we speculated that Ni-ANIPE would be a general and highly selective catalyst for the enantioconvergent addition of sp^2 carbon nucleophiles into racemic, non-activated ketones. A preliminary test of the feasibility with racemic ketone **1b** with phenylboronic acid neopentylglycol ester (PhBneo;

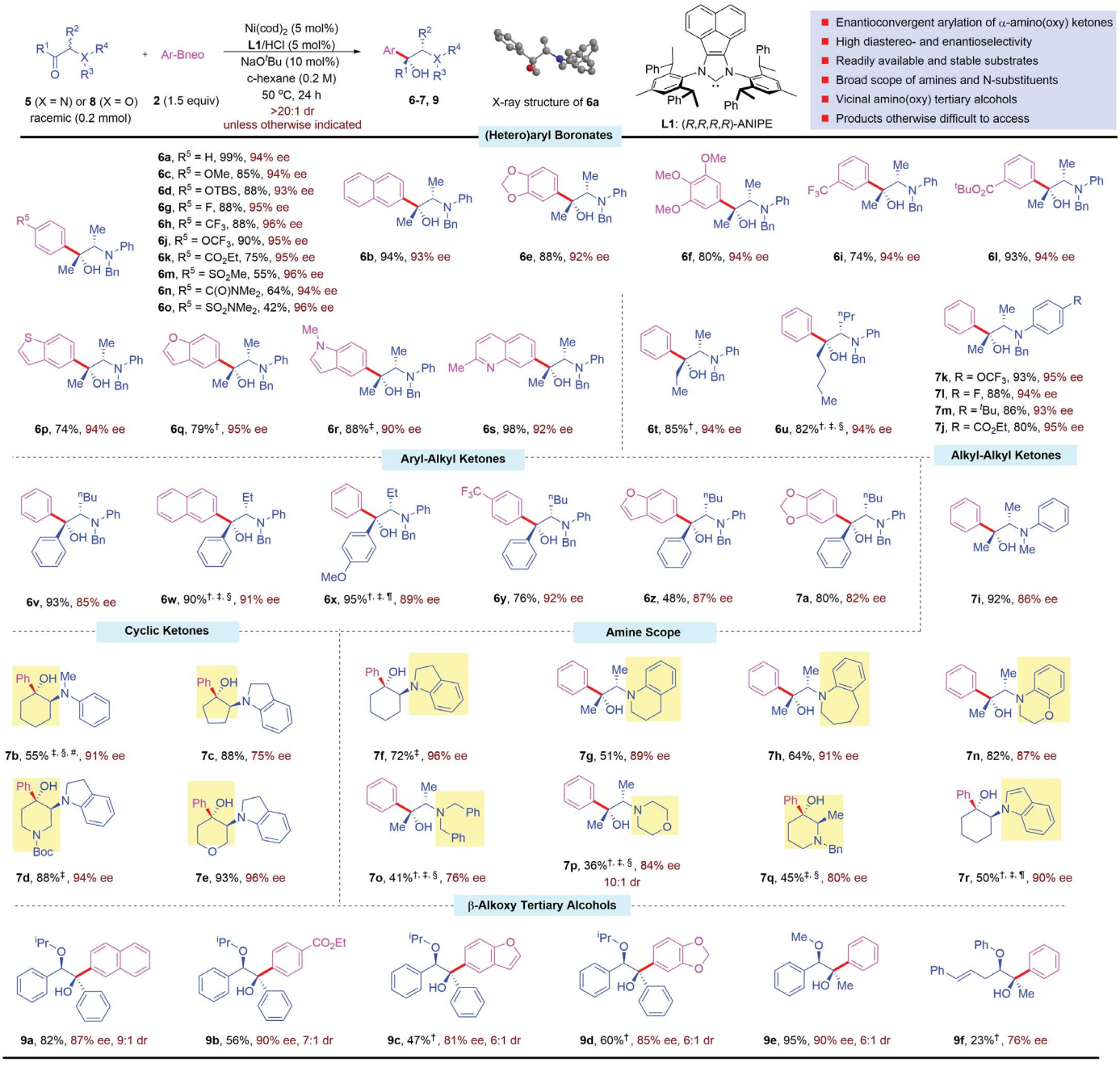


Fig. 3. Scope of nickel-catalyzed enantioconvergent arylation of α-amino or α-oxy ketones. The yields of isolated products on a 0.2 mmol scale are shown. Values ee and dr were determined with HPLC and NMR analysis. The relative and absolute stereochemistries of products are shown according to the x-ray structure of **6a**. †10 mol % catalyst, 20 mol % NaO^{t}Bu; ‡Ti(OiPr)$_4$ was added; §60°C; ¶80°C; #1.0 equiv NaO^{t}Bu (supplementary materials).

2a) delivered product **3v** in 92% yield with 96% ee and complete [>20:1 diastereomeric ratio (dr)] diastereocontrol (Fig. 1E, reaction 2). On the basis of these key preliminary findings, we report enantioconvergent addition (including arylation and alkenylation) of boronate esters into nonactivated ketones to deliver enantioenriched tertiary alcohols with adjacent stereocenters in high diastereo- and enantioselectivities and chemical yields (Fig. 1F). In a single operation, multiple important classes of challenging and value-added tertiary alcohols—including α-aryl tertiary alcohols, tertiary allylic alcohols, α,β-amino alcohols, β-alkoxy alcohols, and β,β-dialkyl alcohols, compounds that previously required multistep synthetic procedures to access—were prepared directly from readily available and inexpensive organoboronates and racemic ketones.

Reaction optimization

To develop the enantioconvergent arylation of ketone, we selected **2a** and ketone **1a** as the model substrates using imidazolium chloride and base as the ligand for nickel catalyst, instead of free NHC (supplementary materials). After extensive investigation of reaction parameters, we identified the optimized reaction condition as **1a** (1.0 equiv), **2a** (1.5 equiv), 5 mol % of Ni(cod)$_2$ and NHC (**L1**/HCl, readily prepared on 50 g scale) (supplementary materials), and 10 mol % of NaO^{t}Bu (sufficient for in situ deprotonation of imidazolium chloride, but insufficient to promote racemization) in cyclohexane at ambient temperature (30°C)

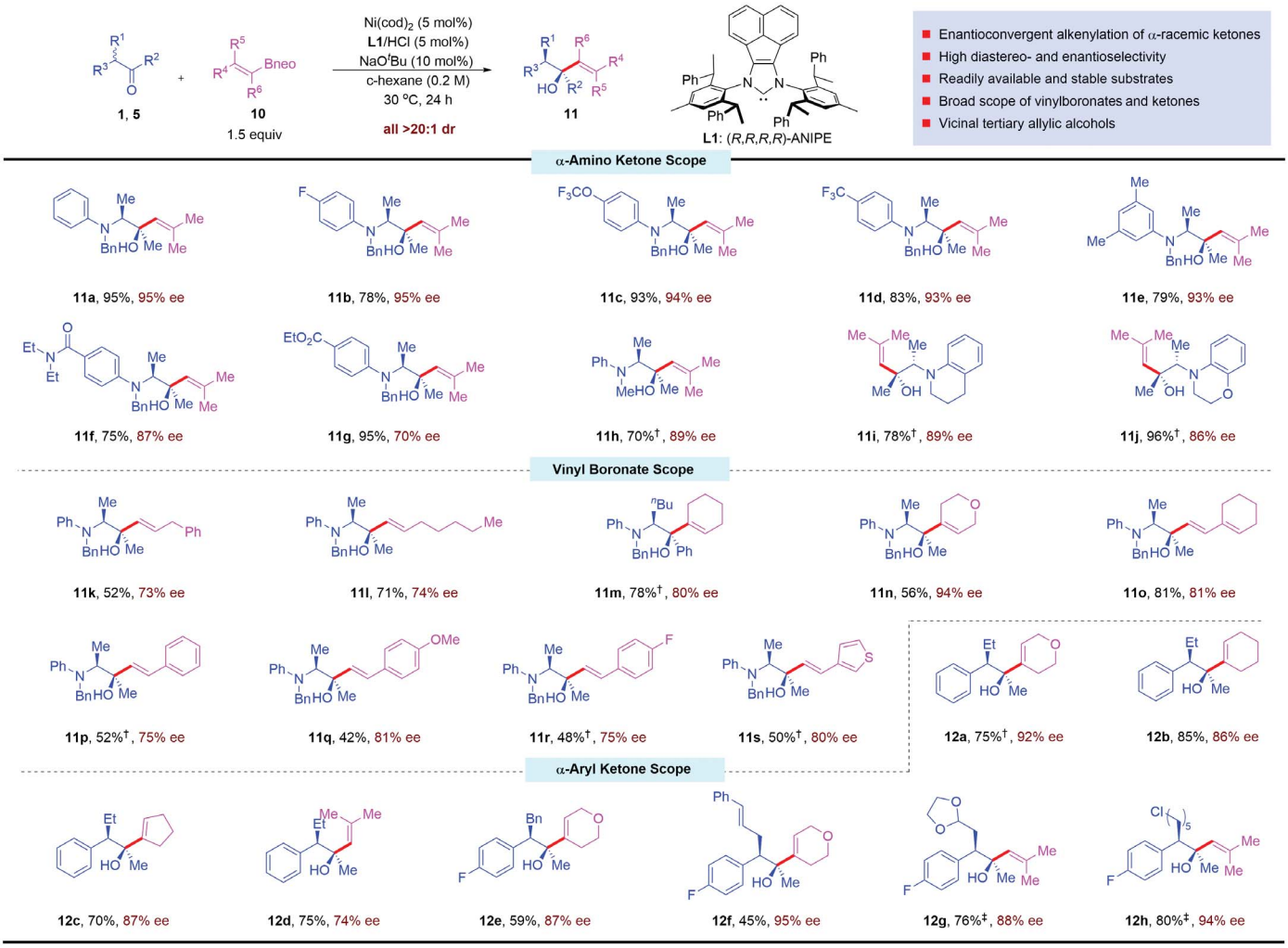


Fig. 4. Scope of nickel-catalyzed enantioconvergent alkenylation of racemic ketones. The yields of isolated products on a 0.2 mmol scale are shown. †50 mol % $Ti(OiPr)_4$ was added; ‡10 mol % catalyst, 20 mol % NaO^tBu, 50°C.

for 24 hours (Fig. 2). The tertiary alcohol **3a** (Fig. 2) was obtained in quantitative yield with 92% ee and >20:1 dr. A completely base-free condition with free NHC (**L1**) gave comparable results (supplementary materials). We found a series of commonly used chiral phosphine and NHC ligands to be ineffective for this arylation reaction, probably because of the low reactivity of the bulky ketone substrate and difficult enantiodiscrimination of the dialkyl substituents. Increasing the amount of NaO^tBu substantially lowered the chemical yield, and three equivalents of base completely inhibited the reaction (supplementary materials). We reasoned that a significant excess of the base could build up a high concentration of the enolate form of the ketone substrate, resulting in no desired arylation but an aldol side reaction. Control experiments without ligand or nickel catalyst led to no reaction, suggesting the critical role of the Ni-NHC catalyst (supplementary materials).

Arylation substrate scope

With the optimized reaction conditions, we first explored the scope of arylboron components for this DyKAT protocol (Fig. 2, **3a** to **3u**). Both electron-rich and electron-poor arylboronates readily underwent arylation, giving rise to tertiary alcohol products in high yields and diastereo- and enantioselectivities (**3a** to **3p**, 84 to 96% ee, >20:1 dr). Many functional groups such as ether (**3c**), silyl ether (**3d**), tertiary amine (**3e**), fluorine (**3h**), trifluoromethyl (**3i** and **3j**), trifluoromethoxy (**3k**), ester (**3l** and **3m**), amide (**3n**), sulfone (**3o**), and sulfonamide (**3p**) were compatible with the arylation reaction. Arylboronic esters bearing pharmaceutically important heterocycles such as benzofuran (**3q**), benzothiophene (**3r**), indole (**3s**), quinolone (**3t**), morpholine (**3e**), and pyridine (**3u**) all served as competent substrates that create products in good to excellent yields with high diastereo- and enantioselectivities (88 to 94% ee, >20:1 dr).

Subsequently, we surveyed the scope of ketone substrates (Fig. 2, **3v** to **4x**). Ketones with different α-substituents—including methyl, ethyl, bulky cyclohexylmethyl, benzyl, cinnamyl, and prenyl groups—could all be converted to desired tertiary alcohols in high yield and with high ee (**3v** to **4a**, 86 to 97% ee, >20:1 dr). Substrates containing an acetal, a free hydroxyl group, and an alkyl chloride underwent an arylation reaction to afford complex alcohol products in good yields and with excellent ee (**4b** to **4d**, 95 to 98% ee, >20:1 dr). Then, we investigated the ketone substrates substituted with different α-aryl groups and found that electron-rich, -neutral, and -poor substituents (**3x** to **4e**, **4n**, **4t**, **4o**, and **4u**) were all compatible with this transformation. Moreover, substrates bearing furan (**4f**) and *N*-methylindole (**4w**) also worked well. Dialkyl ketones with different chain lengths were also suitable partners for this reaction, albeit with a slight loss of diastereo- and enantioselectivity as the chain grew (**4g** to **4h**, 80 to 90% ee, 8:1 to 17:1 dr).

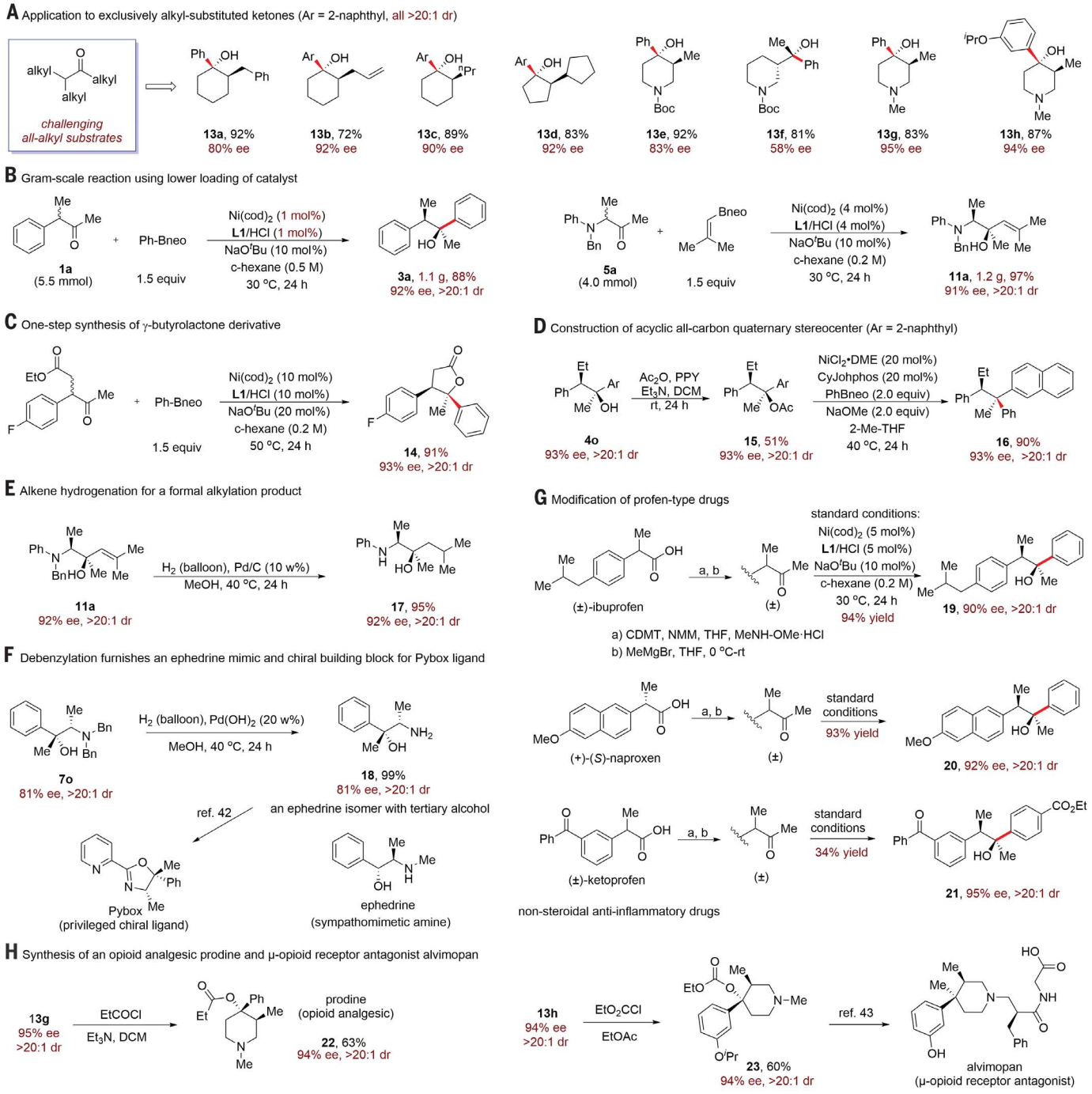


Fig. 5. Application of the DyKAT arylation to exclusively alkyl-substituted ketones, synthetic elaboration, and drug modification. (**A**) Reactions with challenging α,α-dialkyl ketones. (**B**) Gram-scale arylation and alkenylation reaction. (**C**) One-step synthesis of a γ-butyrolactone derivative. (**D**) Stereoretentive construction of an acyclic all-carbon quaternary stereocenter. (**E**) Combination of vinylation and hydrogenation for a formal alkylation reaction. (**F**) Synthesis of an ephedrine mimic. (**G** and **H**) Application of the enantioconvergent arylation strategy in modifying profen-type drugs and synthesizing prodine and alvimopan (supplementary materials).

We also examined the reactivity of aryl ketone substrates (**4i** to **4r**). Commercially available 1,2-diphenylbutan-1-one effectively underwent enantioconvergent arylation with different arylboronates to give diaryl-substituted tertiary alcohols in moderate to excellent yield with high levels of diastereo- and enantioselectivity, despite the potential low reactivity caused by extremely hindered substrates (**4i** to **4r**, 84 to 91% ee, > 20:1 dr). Especially aryl ketones bearing heterocycles—such as pyridine (**4r**), furan (**4s**), and thiophene (**4t**)—were all suitable substrates (86 to 97% ee, >20:1 dr). Aside from acyclic ketones, various α-stereogenic cyclic ketones also performed well to afford highly complex tertiary alcohols bearing adjacent stereocenters (**4s** to **4x**, 80 to 98% ee, >20:1 dr). Last, the absolute and relative configuration of **3c** was determined by means of x-ray crystallographic diffraction analysis. Ketones with α-aryl α-alkyl substituents

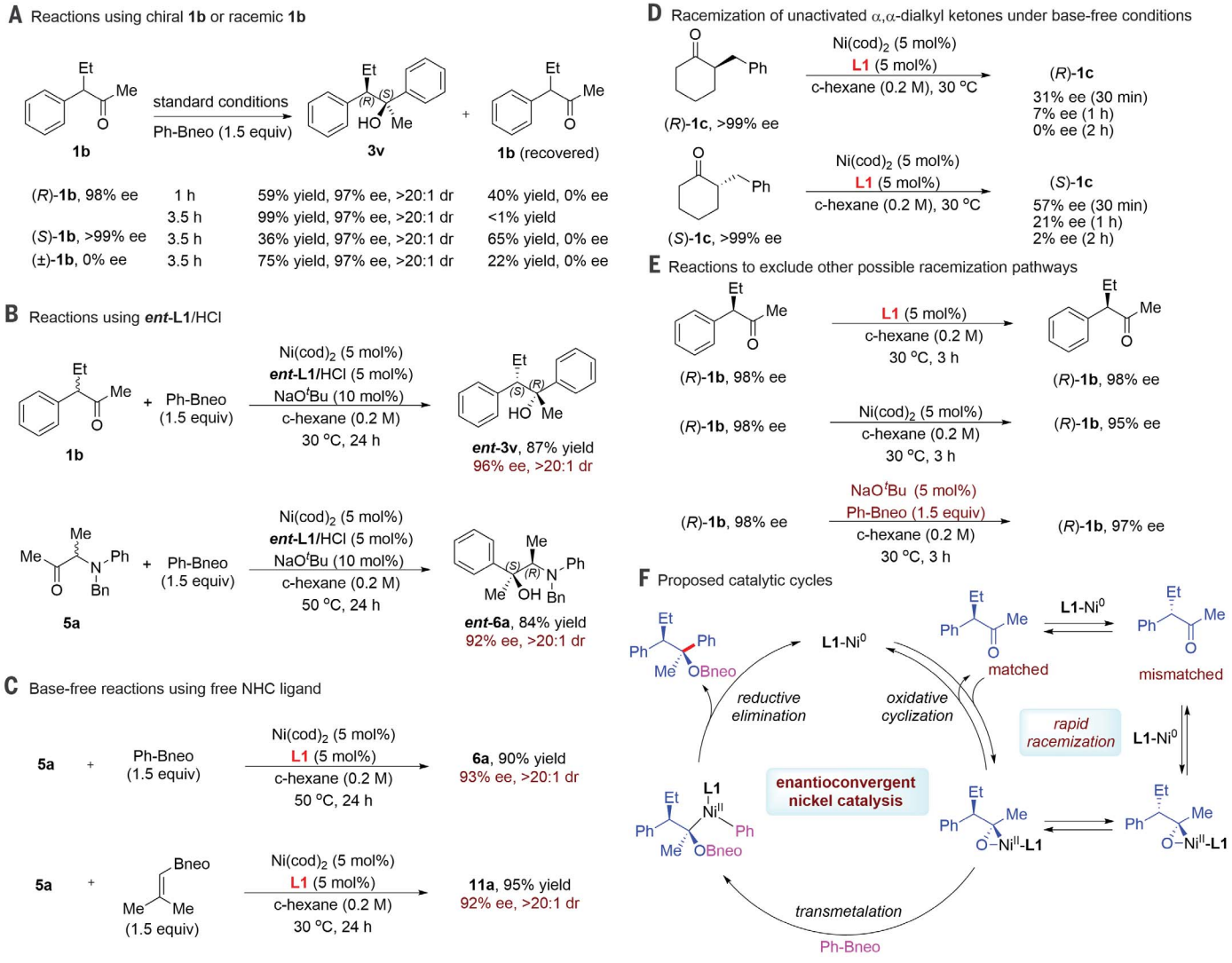


Fig. 6. Control experiments and proposed catalytic cycles. (**A**) Reactions with chiral and racemic ketone substrates to support the rapid racemization process and identify match or mismatch substrates. (**B**) Reactions with enantiomeric ligand to obtain the enantiomeric products. (**C**) DyKAT arylation and vinylation of ketones under base-free conditions. (**D**) Fast racemization of unactivated α,α-dialkyl ketones under base-free conditions. (**E**) Reactions to exclude other possible racemization pathways. (**F**) Proposed catalytic cycles.

are rarely used in DyKATs because of their low acidity. However, our use of these less activated substrates in this method allowed the expedient synthesis of complex compounds that could be difficult to access by other approaches.

Extension to α-amino or α-oxy ketones

Optically pure amino alcohols are prominent structural motifs prevalent in pharmaceuticals and bioactive natural products (Fig. 1A). Furthermore, these compounds are also used as chiral auxiliaries, ligands, and catalysts in asymmetric synthesis. The asymmetric hydrogenation of amino ketones is the most efficient for synthesizing chiral amino alcohols. Dynamic kinetic asymmetric hydrogenation of racemic α-amino ketones to vicinal amino secondary alcohols has also been developed (*32*). However, asymmetric catalytic construction of vicinal amino tertiary alcohols is rare, despite this functionality being an essential building block for synthesis and privileged substructure in marketed antifungal agents (Fig. 1A). Typically, this moiety is synthesized through diastereoselective organometallic addition to laboriously prepared chiral α-amino ketones (*33*). Enantio- and diastereoselective catalytic nitroaldol reactions of ketones are few (*34*), even though the subsequent reduction of the nitro group to amine is required. Moreover, the direct catalytic asymmetric synthesis of 1,2-amino tertiary alcohols is limited in the scope of coupling partners (*35*). To date, the only general enantioselective method is the 2-azadiene-ketone reductive coupling recently reported by Li *et al.* (*36*). However, these methods require substrates with *N*-protecting groups that inherently affect the step or redox economy. Therefore, a straightforward and general approach that enantioselectively assembles tertiary amino alcohols with various desired amine substituents is highly sought after.

We therefore focused on the enantioconvergent synthesis of tertiary amino alcohols using a racemic α-benzyl(phenyl)amino ketone (**5a**) and PhBneo (**2a**) as the model substrate. After a quick screening of reaction parameters (supplementary materials), we identified the optimal reaction conditions as otherwise conserved except for the temperature, which we simply elevated to 50°C. With the conditions established, we first investigated the scope of arylboron coupling partners for this enantioconvergent arylation reaction. As shown in Figure 3 (**6a** to

6s), a wide variety of arylboronates with both electron-donating and electron-withdrawing substituents readily participated in the reaction, affording vicinal amino tertiary alcohols in high yelds and enantioselectivities with complete diastereocontrol (**6a** to **6o**, 92 to 96% ee, >20:1 dr). Moreover, functional groups including ether (**6c**, **6e**, and **6f**), silyl ether (**6d**), fluoride (**6g**), trifluoromethyl (**6h** and **6i**), trifluoromethoxy (**6j**), ester (**6k** and **6l**), sulfone (**6m**), amide (**6n**), and sulfonamide (**6o**) were well compatible under the arylation conditions. In addition to arylboronates, heteroarylboronates bearing benzothiophene (**6p**), benzofuran (**6q**), indole (**6r**), and quinolone (**6s**) were all viable substrates giving products in good yield with excellent enantio- and diastereoselectivities (90 to 95% ee, >20:1 dr).

We next examined the scope of racemic α-amino ketone substrates (Fig. 3, **6t** to **7r**). Dialkyl ketones with longer alkyl chains, such as ethyl and butyl groups, delivered tertiary amino alcohols in similarly high yields and selectivities (**6t** to **6u**, 94% ee, >20:1 dr). A series of exceptionally bulky aryl-alkyl ketones proceeded effectively, with heteroaryl and arylboronates providing complex vicinal amino diaryl tertiary alcohols in 82 to 92% ee and excellent diastereoselectivities (**6v** to **7a**, >20:1 dr). Cyclic α-amino ketones such as cyclohexanone, cyclopentanone, tetrahydropyranone, and piperidinone were all feasible coupling components affording cyclic amino tertiary alcohols in a single operation with high yield and selectivities (**7b** to **7f** and **7q**, 75 to 96% ee, >20:1 dr). Subsequently, we assessed the scope and type of the α-amino substituent. As we expected, various substrates containing alkyl-aryl amines with different substituents (**7i** to **7m**, 86 to 95% ee, >20:1 dr) and dialkyl amino groups successfully delivered products (**7o** to **7q**, 76 to 84% ee, 10:1 to >20:1 dr), albeit with diminished reactivity and selectivity in some cases. Substrates with medicinally important saturated or unsaturated *N*-heterocycles such as indoline (**7c** to **7f**), dihydroquinoline (**7g**), tetrahydrobenzoazepine (**7h**), dihydrobenzooxazine (**7n**), morpholine (**7p**), piperidine (**7q**), and indole (**7r**) all effectively underwent the enantioconvergent arylation reaction furnishing cyclic amino tertiary alcohols. These products, especially those containing cyclic amines, are very challenging to synthesize by methods with substrates with *N*-protecting groups. Additionally, the relative and absolute configuration of the product (**6a**) was confirmed with x-ray crystallography.

Chiral β-alkoxy alcohols are essential chiral building blocks in asymmetric synthesis. Asymmetric ketone hydrogenation has been reported to prepare β-alkoxy secondary alcohols (*20*). However, enantioselective catalytic synthesis of β-alkoxy tertiary alcohols remains elusive. We found that the coupling of commercially available racemic benzoin isopropyl ether and several functionalized aryl or heteroaryl boronates in our standard nickel-catalyzed conditions proceeded effectively, producing complex β-alkoxy tertiary alcohols in good yields with synthetically useful diastereo- and enantioselectivities [**9a** to **9d**, 81 to 90% ee, 6:1 to 9:1 dr (Fig. 3)]. Less hindered dialkyl α-oxy-ketones also performed well, although a homoallylic ether substrate showed lower reactivity when we used a bulkier ANIPE-type ligand (**9e** to **9f**, 76 to 90% ee, 6:1 to >20:1 dr) (supplementary materials).

Asymmetric alkenylation

Enantioenriched allylic alcohols are among the most important structural units in organic synthesis. They are found in a myriad of bioactive molecules and drugs; they are vital chiral synthons that undergo diverse transformations to afford various stereodefined targets of value in chemistry. Consequently, the asymmetric synthesis of allylic alcohols has been a critical topic for methodology development for decades (*37*). However, the enantioselective synthesis of tertiary allylic alcohols is still challenging. Recent progress on asymmetric vinylation of ketones, arguably the most efficient method, has been achieved with stepwise-preformed vinyl zinc (*38*), enyne (*39*), or vinyl boronic acid (*40*). Nonetheless, these reactions are limited to sensitive organometallic vinyl sources, activated ketones, or the preparation of dienyl allylic alcohols. The established systems are not amenable to synthesizing chiral vicinal tertiary allylic alcohols. Thus, a general and practical asymmetric vinylation of ketones, especially one with readily available reagents and capable of simultaneous construction of two stereocenters, is in high demand.

To this end, we focused on expanding our arylation reaction to enantioconvergent vinylation of racemic ketones (Fig. 4). We found that the vinylation reaction between α-amino ketones (**5**) and a trisubstituted alkenylboronate performed well to give vicinal amino tertiary allylic alcohol (**11a**) in nearly quantitative yield with 95% ee and >20:1 dr. Regarding the scope of α-amino substituents, a range of alkyl aniline containing functional groups—such as fluorine (**11b**), trifluoromethoxy (**11c**), trifluoromethyl (**11d**), amide (**11f**), and ester (**11g**), as well as *N*-heterocycles including dihydroquinoline (**11i**) and dihydrobenzooxazine (**11j**)—were all readily incorporated into the amino allylic alcohol products in high yields with 70 to 95% ee and >20:1 dr. As for the scope of vinylboronates, both acyclic or cyclic substrates and alkyl-, alkenyl-, and aryl-substituted vinyl substrates all successfully coupled with α-amino ketones to deliver products in 73 to 94% ee and >20:1 dr (**11k** to **11s**). As we expected, this dynamic kinetic asymmetric vinylation strategy is also suitable for α-aryl ketones, giving rise to many chiral tertiary allylic alcohols in 74 to 95% ee with all >20:1 dr (**12a** to **12h**).

Further applications

Probably because of the formidable difficulty in simultaneously achieving facile racemization and stereocontrol, α,α-dialkyl ketones are particularly challenging substrates with lower acidity and have not been used in the DyKAT addition chemistry. Encouraged by the racemization reactivity and the excellent levels of stereocontrol in our protocol, we thought that this strategy would also be applicable to these most challenging substrates. We found that a wide range of commercially available cyclic or acyclic α,α-dialkyl ketones (with benzyl, allyl, linear alkyl, and branched alkyl substituents) successfully participated in the DyKAT delivery of tertiary alcohol products with high yield and stereoselectivities [**13a** to **13h**, 58 to 95% ee, >20:1 dr (Fig. 5A)], further highlighting the generality of our method. To demonstrate the scalability of the reaction, we performed gram-scale reactions (Fig. 5B). Arylation of α-aryl ketone 1a (5.5 mmol) or alkenylation of α-amino ketone **5a** (4.0 mmol) in the presence of lower loading of catalyst (1 or 4 mol %) provided the corresponding tertiary alcohols in high yield (>1 g obtained) with similar levels of enantio- and diastereoselectivity as before. Moreover, our use of an air-stable Ni(0)-precatalyst allowed us to perform the reaction under glovebox-free conditions [using Schlenk techniques (supplementary materials)]. Considering the sterically encumbered nature of the tertiary alcohol installed in the arylation reaction, we considered whether this functionality could be applied in downstream transformations. The enantioconvergent arylation of a racemic γ-ketoester and the subsequent lactonization occurred in a single operation providing access to a γ-butyrolactone derivative (**14**) featuring two adjacent stereocenters in 91% yield with 93% ee and >20:1 dr (Fig. 5C). This result prompted us to explore more general transformations. By using Watson's stereoretentive nickel-catalyzed Suzuki-Miyaura coupling protocol (*41*), we successfully converted the acylated tertiary alcohol to the complex molecule **16**, which contained two contiguous chiral centers with an acyclic all-carbon quaternary center in high efficiency (Fig. 5D). Furthermore, Pd-catalyzed hydrogenation of amino allylic alcohol **11a** gave the corresponding alkylated amino alcohol **17** in 93% yield (Fig. 5E). Therefore, the combination of vinylation and reduction provides a formal DyKAT alkylation of racemic ketones. Another hydrogenation reaction through Pd catalysis afforded debenzylated product **18**, a tertiary amino alcohol isomer of sympathomimetic amine ephedrine (Fig. 5F). In addition, **18** is

the key precursor for synthesizing privileged chiral Pybox ligand (*42*). To further showcase the applicability of the DyKAT strategy in drug modification, we converted the commonly used, nonsteroidal, and anti-inflammatory medicines ibuprofen, naproxen, and ketoprofen to their ketone derivatives, which can all undergo the enantioconvergent arylation with excellent diastereo- and enantiocontrol regardless of whether racemic or enantiopure precursors were used (Fig. 5G). Drug mimics **19** and **20** were obtained in high yields (93 to 94%), whereas the ketoprofen mimic **21** was prepared in a lower yield because of the competitive reaction of two ketone groups in the substrate. The arylation products (**13g** and **13h**) were subjected to esterification conditions to give an opioid analgesic, prodine (**22**), and the precursor to μ-opioid receptor antagonist alvimopan (**23**) (*43*), respectively, further demonstrating the utility of the current protocol (Fig. 5H).

Mechanistic studies

We next conducted several control experiments to gain insight into the DyKAT process. First, we separated the two enantiomers of α-branched ketone (*R*)-**1b** and (*S*)-**1b** by using preparative high-performance liquid chromatography (HPLC) and subjected them and racemic-**1b** to the standard reaction conditions, respectively (Fig. 6A). Both enantiomers and racemate were transformed to product **3v** in 97% ee with the same absolute and relative configuration, which indicates a complete catalyst-controlled process. When we used (*R*)-**1b**, the product **3v** was obtained with 59% yield in 1 hour (with 40% **1b** recovered) and with 99% yield in 3.5 hours. However, when we used (*S*)-**1b** as substrate, only 36% **3v** was obtained in 3.5 hours (with 65% **1b** recovered). The reaction with racemic **1b** gave a middle yield of 75% in 3.5 hours (with 22% **1b** recovered). These results reveal that (*R*)-**1b** is the stereochemically matched substrate with the Ni-**L1** catalyst, giving a faster reaction than that of (*S*)-**1b**. The recovered substrates were all racemic, which suggests that racemization is far more rapid than the arylation step, which is a requirement for a DyKAT to occur. Moreover, when we used the enantiomer of **L1**/HCl [(*S*,*S*,*S*,*S*)-ANIPE/HCl] instead of **L1**/HCl, enantiomers of **3v** and **6a** were generated efficiently, again indicating the excellent catalyst-controlled course (Fig. 6B).

The reaction performed well under base-free conditions. We then subjected free NHC (**L1**) to the DyKAT arylation and vinylation reactions of α-amino ketone **5a** (Fig. 6C). These reactions also proceeded effectively, giving the corresponding products in similar levels of yield, enantio-, and diastereoselectivity with the in situ NHC generation method. We also observed fast racemization of the less acidic α,α-dialkyl ketones [(*R*)-**1c** and (*S*)-**1c**] under base-free conditions (Fig. 6D). We next performed several control experiments to rule out other possible racemization pathways. The reactions with free carbene and $Ni(cod)_2$, respectively, did not promote ketone racemization (Fig. 6E). In another experiment, we found that the combination of base and boronic ester (5 mol % of base and 1.5 equiv of PhBneo for mimicking the reaction conditions) also did not racemize the ketone. These results suggest that both the racemization and C–C bond-forming processes occur in the absence of a base, and that the ketone racemization may not happen through a traditional base-deprotonating enolate-formation process. Thus, on the basis of our observations and previous reports (*30*, *31*), we propose a catalytic cycle (Fig. 6F). A chiral **L1**-Ni(0) complex and the matched ketone substrate [(*R*)-**1b**] undergo oxidative cyclization to afford an oxanickelacycle. The oxanickelacycle can be opened by PhBneo through transmetalation to form an aryl-alkyl nickel species, which undergoes reductive elimination to give product and regenerate nickel catalyst. The rapid racemization under the Ni-NHC complex through a reversible β-H elimination and reinsertion efficiently converts (*S*)-**1b** to (*R*)-**1b**, allowing complete consumption of ketone substrate.

Beyond the immediate utility of this method, we anticipate that the Ni-NHC–enabled enantioconvergent strategy will spur the development of other DyKAT processes for the rapid synthesis of valuable complex targets.

REFERENCES AND NOTES

1. E. J. Corey, L. Kürti, *Enantioselective Chemical Synthesis: Methods, Logic and Practice* (Direct Book Publishing, 2010).
2. E. N. Jacobsen, A. Pfaltz, H. Yamamoto, *Comprehensive Asymmetric Catalysis I–III* (Springer-Verlag Berlin Heidelberg, 1999).
3. M. Shibasaki, M. Kanai, *Chem. Rev.* **108**, 2853–2873 (2008).
4. J. L. Stymiest, V. Bagutski, R. M. French, V. K. Aggarwal, *Nature* **456**, 778–782 (2008).
5. H. Zhou *et al*., *Nature* **605**, 84–89 (2022).
6. R. Y. Liu, S. L. Buchwald, *Acc. Chem. Res.* **53**, 1229–1243 (2020).
7. Y. Yang, I. B. Perry, G. Lu, P. Liu, S. L. Buchwald, *Science* **353**, 144–150 (2016).
8. A. Saxena, B. Choi, H. W. Lam, *J. Am. Chem. Soc.* **134**, 8428–8431 (2012).
9. D. A. Evans, *Science* **240**, 420–426 (1988).
10. N. D. Bartolo, J. A. Read, E. M. Valentín, K. A. Woerpel, *Chem. Rev.* **120**, 1513–1619 (2020).
11. B. M. Trost, D. R. Fandrick, *Aldrichim Acta* **40**, 59–72 (2007).
12. V. Bhat, E. R. Welin, X. Guo, B. M. Stoltz, *Chem. Rev.* **117**, 4528–4561 (2017).
13. S. Caddick, K. Jenkins, *Chem. Soc. Rev.* **25**, 447–456 (1996).
14. F. F. Huerta, A. B. E. Minidis, J. E. Bäckvall, *Chem. Soc. Rev.* **30**, 321–331 (2001).
15. J. Choi, G. C. Fu, *Science* **356**, eaaf7230 (2017).
16. J. S. DeHovitz *et al*., *Science* **369**, 1113–1118 (2020).
17. H. You, E. Rideau, M. Sidera, S. P. Fletcher, *Nature* **517**, 351–355 (2015).
18. M. Huang, L. Zhang, T. Pan, S. Luo, *Science* **375**, 869–874 (2022).
19. R. Noyori *et al*., *J. Am. Chem. Soc.* **111**, 9134–9135 (1989).
20. J.-H. Xie, Q.-L. Zhou, *Aldrichim Acta* **48**, 33–40 (2015).
21. K. M. Steward, E. C. Gentry, J. S. Johnson, *J. Am. Chem. Soc.* **134**, 7329–7332 (2012).
22. F. Wang *et al*., *J. Am. Chem. Soc.* **143**, 2477–2483 (2021).
23. J. Liu, S. Krajangsri, J. Yang, J.-Q. Li, P. G. Andersson, *Nat. Catal.* **1**, 438–443 (2018).
24. S. L. Bartlett, J. S. Johnson, *Acc. Chem. Res.* **50**, 2284–2296 (2017).
25. S. L. Bartlett, K. M. Keiter, J. S. Johnson, *J. Am. Chem. Soc.* **139**, 3911–3916 (2017).
26. D. Janssen-Müller, C. Schlepphorst, F. Glorius, *Chem. Soc. Rev.* **46**, 4845–4854 (2017).
27. Y. Cai *et al*., *Angew. Chem. Int. Ed.* **57**, 1376–1380 (2018).
28. W.-B. Zhang, X.-T. Yang, J.-B. Ma, Z.-M. Su, S.-L. Shi, *J. Am. Chem. Soc.* **141**, 5628–5634 (2019).
29. D. Shen, Y. Xu, S.-L. Shi, *J. Am. Chem. Soc.* **141**, 14938–14945 (2019).
30. Y. Cai, L.-X. Ruan, A. Rahman, S.-L. Shi, *Angew. Chem. Int. Ed.* **60**, 5262–5267 (2021).
31. Y. Hoshimoto, M. Ohashi, S. Ogoshi, *Acc. Chem. Res.* **48**, 1746–1755 (2015).
32. J.-H. Xie *et al*., *J. Am. Chem. Soc.* **131**, 4222–4223 (2009).
33. T. Ooi *et al*., *J. Am. Chem. Soc.* **127**, 5073–5083 (2005).
34. T. Karasawa, R. Oriez, N. Kumagai, M. Shibasaki, *J. Am. Chem. Soc.* **140**, 12290–12295 (2018).
35. C. Wang *et al*., *Angew. Chem. Int. Ed.* **55**, 685–688 (2016).
36. K. Li, X. Shao, L. Tseng, S. J. Malcolmson, *J. Am. Chem. Soc.* **140**, 598–601 (2018).
37. A. Lumbroso, M. L. Cooke, B. Breit, *Angew. Chem. Int. Ed.* **52**, 1890–1932 (2013).
38. H. Li, P. J. Walsh, *J. Am. Chem. Soc.* **126**, 6538–6539 (2004).
39. V. Komanduri, M. J. Krische, *J. Am. Chem. Soc.* **128**, 16448–16449 (2006).
40. Y. Huang, R.-Z. Huang, Y. Zhao, *J. Am. Chem. Soc.* **138**, 6571–6576 (2016).
41. J. Xu, O. P. Bercher, M. P. Watson, *J. Am. Chem. Soc.* **143**, 8608–8613 (2021).
42. Y.-F. Zhang *et al*., *J. Am. Chem. Soc.* **143**, 15413–15419 (2021).
43. J. A. Werner *et al*., *J. Org. Chem.* **61**, 587–597 (1996).

ACKNOWLEDGMENTS

We thank Y. Wang (University of Pittsburgh) for helpful discussion on the manuscript. **Funding:** This work is supported by the National Key R&D Program of China (2022YFA1503702 and 2021YFF0701600), the National Natural Science Foundation of China (92256303, 21871288, 21821002, and 22171280), the Ningbo Natural Science Foundation (2022J017), the Program of Shanghai Academic Research Leader (22XD1424900), and the CAS Youth Interdisciplinary Team (JCTD-2021-11). **Author contributions:** S.-L.S. conceived and directed the projects and wrote the manuscript. L.-X.R. performed most of the experiments including reaction development and application, substrate scope investigation, and mechanistic study. B.S. and J.-M.L. performed part of the substrate scope study. All authors analyzed the data. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Crystallographic data for compounds 3c and 6a are available from the Cambridge Crystallographic Data Center under reference numbers CCDC 2052155 and 2184616, respectively. All other data to support the conclusions are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade0760
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
HPLC Traces
NMR Spectra
References (*44*–*70*)

Submitted 22 July 2022; accepted 18 January 2023
10.1126/science.ade0760

ELECTRON FLUIDS

Imaging the breaking of electrostatic dams in graphene for ballistic and viscous fluids

Zachary J. Krebs[1]†, Wyatt A. Behn[1]†, Songci Li[1], Keenan J. Smith[1], Kenji Watanabe[2], Takashi Taniguchi[3], Alex Levchenko[1], Victor W. Brar[1]*

The charge carriers in a material can, under special circumstances, behave as a viscous fluid. In this work, we investigated such behavior by using scanning tunneling potentiometry to probe the nanometer-scale flow of electron fluids in graphene as they pass through channels defined by smooth and tunable in-plane p-n junction barriers. We observed that as the sample temperature and channel widths are increased, the electron fluid flow undergoes a Knudsen-to-Gurzhi transition from the ballistic to the viscous regime characterized by a channel conductance that exceeds the ballistic limit, as well as suppressed charge accumulation against the barriers. Our results are well modeled by finite element simulations of two-dimensional viscous current flow, and they illustrate how Fermi liquid flow evolves with carrier density, channel width, and temperature.

The interactions between particles that make up a fluid play a critical role in how the fluid flows. At low densities, where the particles are free to move ballistically, such as in gases, the conductance of a constriction is dependent on only the channel width and particle scattering from the walls, which leads to momentum loss. At higher densities, interactions between particles—which preserve momentum—become more frequent and can lead to collective flows that enhance conductance through the constriction beyond the ballistic limit (*1*, *2*). This phenomenon, which is exhibited by viscous fluids with laminar flow, was predicted by Gurzhi to also occur in electronic systems when the electron-electron (e-e) scattering length l_{ee} becomes shorter than the momentum-relaxing scattering length l_{mr} stemming from electron-impurity and electron-phonon scattering (*3*, *4*). Initially, signatures of this effect were observed in high-mobility gallium arsenide quantum wells and wires, manifested by a decrease of resistance with an increase of temperature (*5*–*7*) as well as other electron fluid–like behaviors (*8*–*11*).

The focus of recent attention on hydrodynamic phenomena largely concerns graphene, other two-dimensional (2D) van der Waals materials, and topological semimetals, where short e-e mean free paths and low Umklapp scattering rates allow the quasiparticles to form strongly correlated viscous Fermi or Dirac fluids (*12*–*31*). The range of hydrodynamic behaviors that are predicted to occur in these viscous fluid states include current vortex formation (*15*, *16*), higher-than-ballistic conduction (*17*, *18*), strong nonlocality (*22*–*24*), and anomalous thermoelectric responses (*25*–*28*). To observe these effects, several experimental methods have been implemented. Transport measurements through a series of lithographed constrictions and strategically contacted samples have recently observed signatures of superballistic conductance as well as negative nonlocal resistance (*32*–*34*). In a parallel vein, a departure from the Mott relations, a giant violation of the Wiedemann-Franz law, and breakdown of Matthiessen's rule were observed in the hydrodynamic regime in graphene (*35*, *36*). Meanwhile, scanned single electron transistor and nitrogen vacancy measurements of etched, encapsulated graphene devices have imaged Poiseuille current density profiles—a hallmark of viscous flow—in Hall bar and Corbino geometries (*37*–*41*). And, very recently, a scanned SQUID (superconducting quantum interference device) imaging technique has enabled visualization of current whirlpools in tungsten ditelluride (*42*). Lastly, Kelvin probe force microscopy has been used to image, with 60-nm resolution, regions

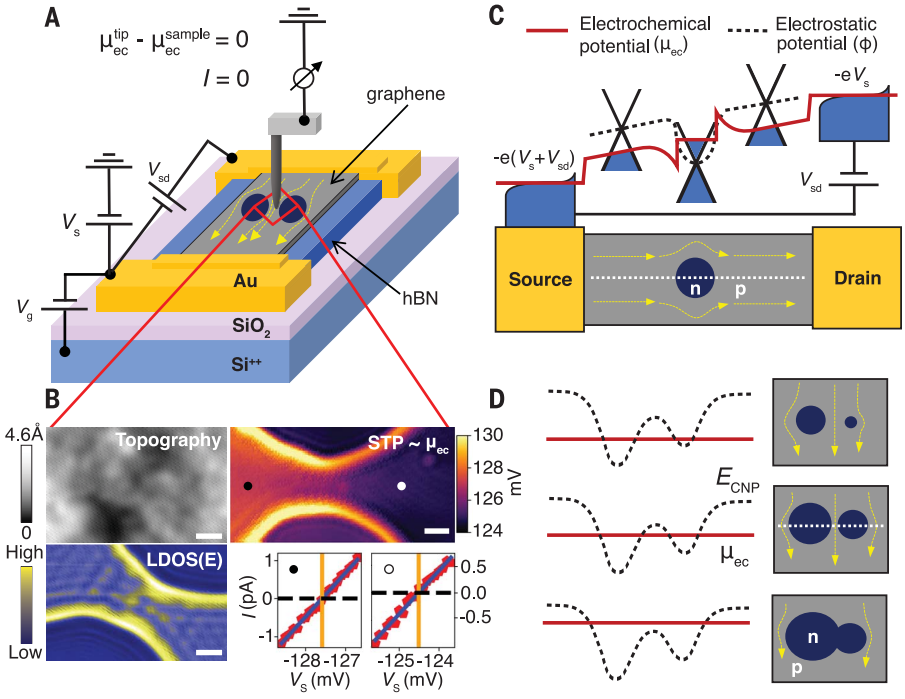


Fig. 1. Experimental method for imaging charge flow through variable-width channels. (**A**) Schematic of the STP experimental setup. V_{sd} drives current through the sample, and V_s (I = 0) determines the difference between the sample and tip electrochemical potentials. The carrier density (and E_F) is globally modified with an electrostatic gate electrode V_g. (**B**) Simultaneously acquired topography (top left) and electronic local density of states (LDOS) (bottom left) (V_s = −10 mV, V_g = −2 V) of the electrostatic dam. (Top right) Example STP map (V_g = −2 V) of the same area acquired under transport. All scale bars are 100 nm. (Bottom right) Example tunneling I-V curves recorded at two points shown on the STP map. I = 0 (black dashed line) determines the local electrochemical potential $\mu_{ec} = -V_s$ (orange vertical line) of the sample; by fitting many such curves (blue solid line), μ_{ec} is mapped spatially. (**C**) An energy diagram depicting how the electrostatic and electrochemical potentials vary along the flow direction across a potential well (white dashed line). The variations in μ_{ec} near the scatterer are exaggerated for clarity. (**D**) Schematic of an electrostatic dam at several gating conditions. The size of two p-n junction barriers increases from high to low hole doping (top to bottom), closing off the channel between them and blocking the flow of current. The leftmost panels show the local charge neutrality point (E_{CNP}) and μ_{ec} along the white dashed line. E_{CNP} follows the in-plane electrostatic potential defining the dam.

[1]Department of Physics, University of Wisconsin–Madison, Madison, WI 53706, USA. [2]Research Center for Functional Materials, National Institute for Materials Science, Tsukuba 305-0044, Japan. [3]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, Tsukuba 305-0044, Japan.
*Corresponding author. Email: vbrar@wisc.edu
†These authors contributed equally to this work.

of negative local resistance in chemical vapor–deposited graphene on silicon dioxide that were attributed to viscous flow behavior around intrinsic defects (*43*).

In this work, we used scanning tunneling potentiometry (STP) to image, with nanometer-scale spatial resolution, the electrochemical potential profile associated with quasiparticle flow in graphene around electrostatic barriers that are "drawn" using voltage pulses from the tip of a scanning tunneling microscope (STM) (*44*). This methodology allows for the creation of smooth barriers defined by in-plane p-n junctions, which confine the particle flow without introducing diffusive scattering or other momentum-relaxing processes that occur at the edges of lithographed samples (*45*, *46*). Moreover, we were able to vary the width of the conduction channels from micrometer scale to pinch-off (where the barriers form "electrostatic dams" that suppress flow) by tuning the carrier density. Another distinguishing feature of our experiment is that the graphene/hexagonal boron nitride (hBN) samples that we probed were nonencapsulated, which serves to reduce dielectric screening and therefore enhance e-e interactions, allowing viscous flow to be observed at shorter length scales.

Imaging and quantifying charge flow using STP

A schematic of our STP measurement geometry is shown in Fig. 1A. In STP, a source-drain bias (V_{sd}) is used to drive current through a thin sample, and the subsequent spatially varying electrochemical potential (μ_{ec}) is measured locally using an STM tip (*47*–*51*). To measure μ_{ec}, the feedback of the STM tip is turned off, and the tip bias needed to zero the tunneling current (I) is determined by performing a linear fit of the tunneling current-voltage (I-V) curve (Fig. 1B). This allows μ_{ec} to be determined to within ~10 μV and at the angstrom-scale spatial resolution of standard STM. Figure 1C illustrates how μ_{ec} varies across the sample under transport conditions and in the presence of a p-n junction, which scatters incident carriers. When $V_{sd} = 0$, all local accumulations of charge that affect the chemical potential, or Fermi level (E_F), are offset by changes in electrostatic potential ϕ, such that $\mu_{ec} = \phi + E_F$ is constant across the surface. For $V_{sd} \neq 0$, however, μ_{ec} will change continuously across the sample, with a spatially varying slope that depends on the local conductivity; meanwhile, any changes in local charge accumulation that are caused by active carrier scattering or ballistic transport will also affect μ_{ec} and be detectable with STP (*51*–*54*).

Previous STP measurements of graphene devices on silicon carbide (SiC) substrates have revealed sharp drops in potential associated with monolayer-bilayer boundaries, as well as subsurface crystal steps. In some cases, Landauer residual resistivity dipoles (LRRDs) were observed near defect features, which could be used to model the electron-barrier scattering mechanisms (*55*–*61*). STP measurements have also been performed on graphene nanoribbons on SiC in the ballistic transport regime (*62*). The use of SiC as a substrate, with its large dielectric constant, strongly screens e-e interactions and therefore suppresses hydrodynamic effects. Those measurements also used topographic features to act as scattering barriers, which are known to introduce artifacts into STP measurements owing to tip convolution (*63*).

In this work, we probed monolayer graphene/hBN samples with electrostatic barriers that were introduced by "drawing" them with the STM tip using a previously developed methodology (*44*, *64*, *65*). Each barrier was created by introducing subsurface charges in the underlying hBN by applying a 1- to 2-min, 5-V pulse with the STM tip. This technique creates an electrostatic potential well in the plane of the graphene sheet that scatters incident holes and electrons. Within a suitable range of negative gate voltages, a circular p-n (outside-inside) junction forms on the periphery of the potential well, which acts as a reflective boundary for graphene quasiparticles (*66*). By placing two of these p-n junctions in proximity, we built a small channel that current can flow through when a source-drain bias is applied. Moreover, as shown in Fig. 1D, the width of this current-carrying channel can be tuned by using an electrostatic back gate to adjust E_F, which alters the radii of the p-n junction barriers; the p-n junctions considered here decrease in radius with higher hole concentrations, which leads to an increased channel width. Prior to introducing the barriers, STP measurements of transport

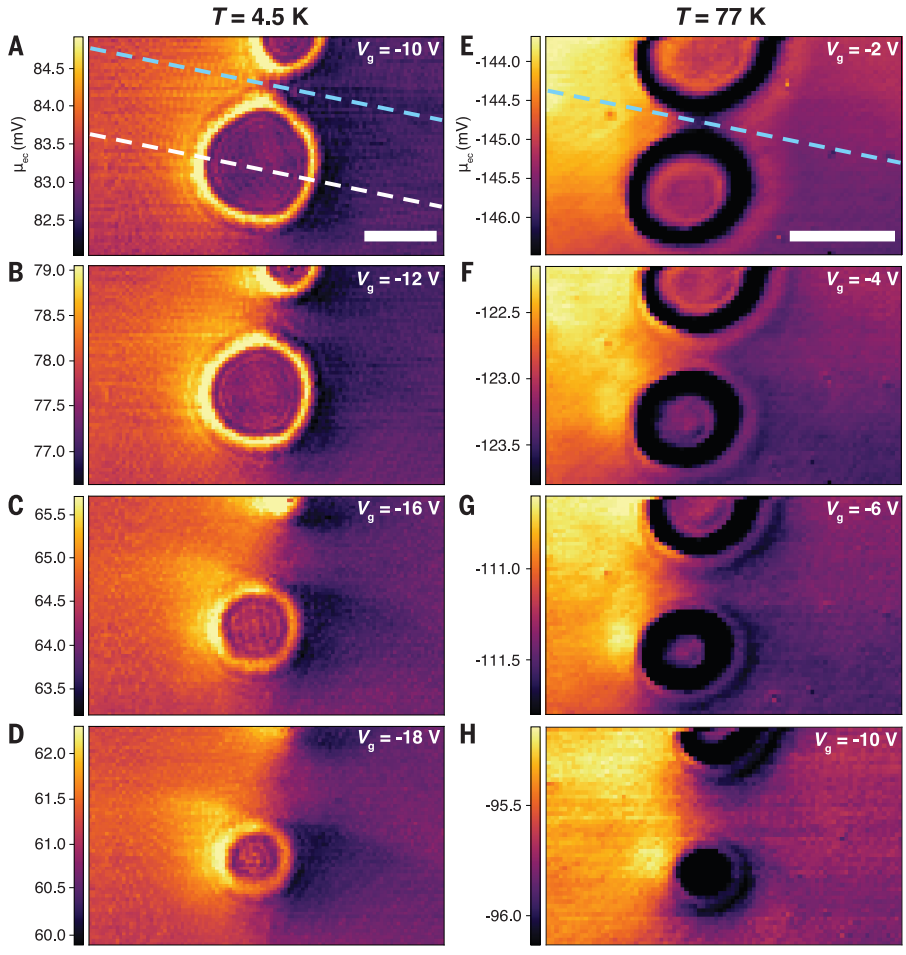


Fig. 2. Mapping the electrochemical potential near electrostatic dams at low and high temperature. (**A** to **D**) STP maps of an electrostatic dam at T = 4.5 K, V_{sd} = 0.4 V, and four selected gate voltages: −10, −12, −16, and −18 V, in order of increasing channel width. (**E** to **H**) STP maps of a second electrostatic dam at T = 77 K, V_{sd} = −0.4 V, and four gate voltages: −2, −4, −6, and −10 V, in order of increasing channel width. The scale bars are 250 nm. The direction of incident current flow for all images is from left to right along the dashed lines in (A) and (E), which are used later to plot linecuts of μ_{ec} in Fig. 3.

across the bare surface were performed to obtain the potential drop caused solely by momentum-relaxing scattering sources, which was then used to calculate l_{mr} = 5 μm at 4.5 K (*67*).

Mapping the electrochemical potential drop near barriers

STP images acquired after the formation of the potential wells at 4.5 and 77 K are shown in Fig. 2. At both 4.5 and 77 K, μ_{ec} is observed to increase (decrease) on the upstream (downstream) side of the potential wells, creating in-plane dipoles. We identify these features as LRRDs, which are known to occur in ballistic or near-ballistic transport conditions when scattered charge carriers accumulate (deplete) on the upstream (downstream) side of a barrier (*47*). To better visualize this behavior, cross-sectional cuts across the barriers are shown in Fig. 3, A and B, which capture the effect of the charge accumulation on μ_{ec}. At 4.5 K, we find that the changes in μ_{ec} in both the accumulation and depletion zones can be fit well with a R^{-1} dependence, where R is the distance from the center of the barrier, in good agreement with LRRD theory for ballistic transport (*68*, *69*). A more detailed, full image of the theoretical LRRD for ballistic transport in a 2D system is shown in Fig. 3C, which was calculated using a framework introduced in (*70*) and described in (*67*). The dipole profile at 77 K shows a notable difference (Fig. 3B): The accumulation and depletion profiles are decreased in magnitude by a factor of ~2.5 from those at 4.5 K. The decreased dipole strength at high temperatures can be explained in terms of viscous corrections to the electronic flow. Guo *et al.* showed that frequent e-e interactions reduce the effective scattering strength of obstacles seen by incident charge carriers (*70*). This is because electrons reflected by an obstacle can receive momentum from incoming electrons back in the direction of current flow, keeping them from fully backscattering. The magnitude of this dipole reduction can be modeled as a function of the quantity r/l_{ee}, where r is the scattering length of a barrier and is approximately equal to the barrier radius. Figure 3D shows the predicted dipole potential profile on one side of the barrier as a function of l_{ee}, plotted alongside the experimental data. The observed reduction in dipole strength at 77 K is consistent with $r/l_{ee} \sim 1$, which corresponds to $\tilde{l}_{ee}$ = 150 nm. We denote this experimentally derived value of l_{ee} with a tilde because this value is only accurate up to a multiplicative factor of order one owing to uncertainty in the scattering length r.

Within the quantum wells, meanwhile, the STP images reveal standing waves associated with circular quasibound states trapped in the well. These states are excited by the incident current-carrying carriers, creating changes to the local charge density that are visible in STP (*53*). The p-n barrier can be observed as the bright (dark) ring in the 4.5 K (77 K) measurements. The same ring feature has been determined in previous scanning tunneling spectroscopy and Kelvin probe force microscopy measurements to indicate the position of the classical turning point of the quasi-bound states, where there is an accumulation

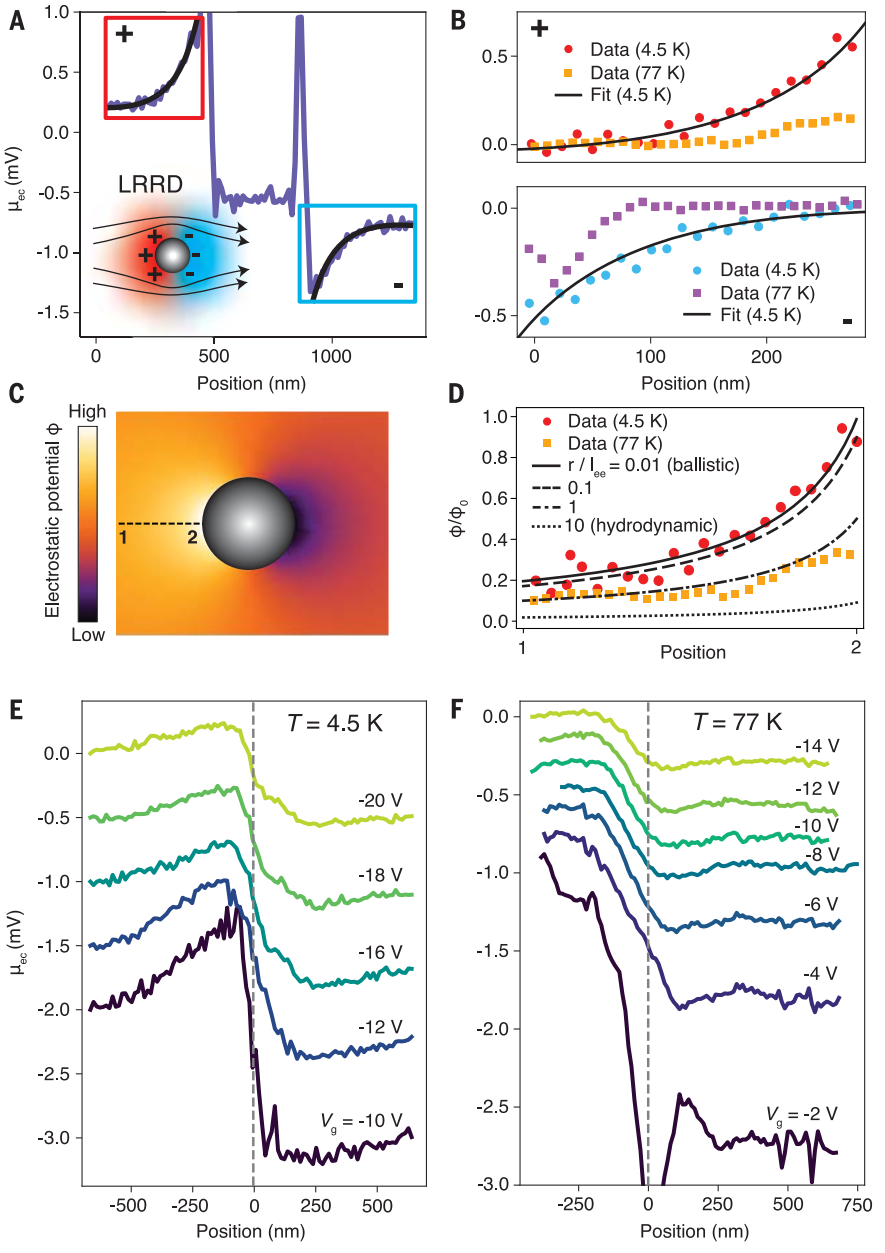


Fig. 3. Tracking the gate-dependent voltage drop through the channels and formation of LRRDs across the electrostatic barriers. (**A**) Linecut along the white dashed line in Fig. 2A revealing the LRRD structure at T = 4.5 K. Data within the red (blue) box corresponds to an accumulation (depletion) of holes. (**B**) (Top) Zoom-in of the accumulation zone in (A). (Bottom) Zoom-in of the depletion zone in (A). Both are plotted alongside the equivalent linecuts at 77 K. Solid lines show R^{-1} fits to the 4.5 K data. (**C**) Predicted electrostatic potential of an LRRD around a circular obstacle when current is flowing from left to right (*67*). (**D**) Normalized linecuts of the electrostatic potential in (C) for varying e-e scattering lengths, parameterized by r/l_{ee} and compared with the data from (B). (**E**) Linecuts along the blue dashed line in Fig. 2A at T = 4.5 K. (**F**) Linecuts through the blue dashed line in Fig. 2E at T = 77 K.

of quasiparticle density (*44*, *71*–*73*). It is not fully understood why the p-n boundary appears bright for measurements at 4.5 K but dark at 77 K. This effect was observed in multiple separate measurements, and we speculate that STP is sensitive to thermovoltages generated on the p-n boundary because of strong tip-induced resonances in the local density of states (*72*, *74*, *75*). Figure S3 shows how the p-n junctions change appearance in response to heating.

In addition to the features described above, we also observe a drop in μ_{ec} along a path through the channel between the wells in the direction of current flow, which represents a central focus of this work. This change in μ_{ec} is associated with current that flows through the channel, the width of which can be tuned via electrostatic gating (as illustrated in Fig. 1D). Figure 2 shows how μ_{ec} evolves in the vicinity of the channels as their width is varied from as wide as 400 nm to pinch-off, where it forms an electrostatic dam that blocks the incident current. At 4.5 K, raylike "streams" of current are visible emerging from the downstream side of the channel, a property that is consistent with ballistic carriers passing through the gap and locally increasing μ_{ec} (*54*, *76*). Such qualitative streams are not as apparent for data obtained at 77 K. To quantify the potential drop induced by the channels, we recorded μ_{ec} along linecuts through the channels along the direction of current flow (Fig. 3, E and F). At both low and high measurement temperatures, an increase in hole carrier density at lower gate voltages is associated with a smaller potential drop through the channels. This trend is expected, because a larger carrier density both widens the channel and increases the conductivity of graphene, creating a more conductive channel overall. We also note that the potential drop in the 4.5 K data is highly symmetric about the channel midpoint, whereas in the 77 K data, the potential drop seems to occur mostly on one side (to the left of the vertical dashed line in Fig. 3F). We attribute this to a slight misalignment of the incident current direction with the channel axis.

Channel conductances

To quantitatively characterize the carrier flow, we used the measured electrochemical drop ($\Delta\mu_{ec}$) through the channels (Fig. 3, E and F) to calculate the conductance of each channel, $G_{data} = I_{channel}/\Delta\mu_{ec}$. The current flowing through each channel ($I_{channel}$) is first estimated using the assumption that the ratio between the channel width (w) and the width of the graphene flake (W = 15 μm) is proportional to the ratio between $I_{channel}$ and the current passing through the whole flake (J), measured using an ammeter—that is, $w/W = \alpha I_{channel}/J$. The proportionality constant α depends on the precise boundary conditions and geometry of our sample, as well as on any deflection of current caused by the perturbing potential of the STM tip (*66*). Although α cannot be calculated a priori, it can be estimated by comparing our data at 4.5 K to the Sharvin formula for ballistic transmission through a channel

$$G_{Sh} = G_Q \frac{\overline{E_F}}{\pi\hbar v_F} w \qquad (1)$$

where $\hbar$ is Planck's constant (h) divided by 2π, $G_Q = 4e^2/h$ is the conductance quantum, w is the minimum width inside the channel, v_F is the Fermi velocity, and $\overline{E_F}$ is the Fermi level (chemical potential) averaged over the narrowest cross section of the channel. In repeated measurements at 4.5 K using a new electrostatic dam each time, we find excellent agreement between G_{Sh} and G_{data} for channel widths up to 400 nm when α = 2.8, as shown in Fig. 4A. This agreement is expected, given that the only changes between subsequent measurements are slight variations in the tip potential near its apex and the positioning of our electrostatic dams within the center of the graphene flake. For w > 400 nm, the Sharvin prediction is consistently less than the measured conductance, because the geometry starts to resemble two separate barriers instead of a well-defined channel (*67*). In contrast to measurements at 4.5 K, data taken at 77 K demonstrate a channel conductance that is larger than the ballistic Sharvin prediction with α = 2.8, and this deviation increases as the channel is widened. Simply adjusting α as a fitting parameter does not produce good agreement between the data at 77 K and the ballistic theory across all channel widths (*67*). This suggests that a viscous Gurzhi contribution (G_G)—which is known to depend quadratically on channel width (*17*)—should be added to the channel conductance at elevated temperatures. Notably, modeling the viscous corrections to the channel conductance only requires the addition of a single term to the Sharvin formula. This additive correction can be qualitatively understood in the following way: When e-e interactions are prevalent, charge carriers that are backscattered by the channel edges now experience forward scattering processes that push them back through the channel, which in effect increases the number of conduction channels beyond the Sharvin limit (*18*). To account for this viscous contribution and provide another experimental estimate of the e-e scattering length l_{ee}, we turn to (*17*, *32*)

$$G = G_{Sh} + G_G, G_G = c_G G_Q \frac{\overline{E_F}}{\pi\hbar v_F} \frac{w^2}{l_{ee}} \qquad (2)$$

which holds in the absence of significant momentum-relaxing scattering processes (*76*). From Eq. 2, we can write

$$l_{ee} = c_G w \left(\frac{G_{data}}{G_{Sh}} - 1 \right)^{-1} \qquad (3)$$

where c_G is a nonuniversal numerical factor that is specific to viscous flow around our electrostatic dams. For example, $c_G = \pi^2/16 = 0.62$ for a perfect slit geometry (*17*); however, the value of c_G also depends on the boundary conditions and differs for flows with no-slip and no-stress conditions (*45*, *46*, *77*, *78*). For our geometry, we expect c_G = 0.75 on the basis of viscous fluid simulations described below [also, see (*67*)]. The resulting theoretical channel conductance is plotted in Fig. 4A, showing good agreement with 77 K measurements and providing strong evidence of viscous effects in the electron fluid increasing the channel conductance. Corresponding estimates of l_{ee} using Eq. 3 are shown in Fig. 4B, giving values that are

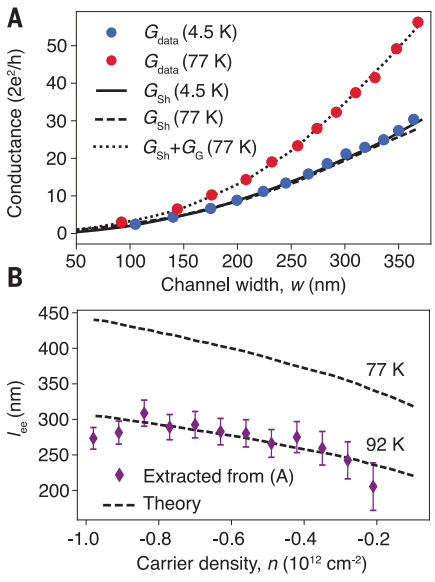


Fig. 4. Channel conductances and extracted electron-electron scattering lengths. (**A**) Channel conductance $G_{data} = I_{channel}/\Delta\mu_{ec}$ obtained from the STP data at both low (blue, 4.5 K) and high (red, 77 K) temperature. Theoretical curves (solid and dashed lines) are calculated using the Sharvin formula (Eq. 1). A viscous contribution G_G is included in the dotted line. (**B**) e-e scattering lengths extracted from the red (77 K) data points and the dotted fit in (A) using Eq. 3. Values for the dashed theory curves were taken from the microscopic calculations in (*14*). Error bars on the data points come from a 14-nm uncertainty in the channel widths, w. All data points presented in this figure were obtained from a repeated set of STP measurements distinct from those in Fig. 2. The results from a similar analysis of Fig. 2 are found in (*67*) and closely replicate the above results.

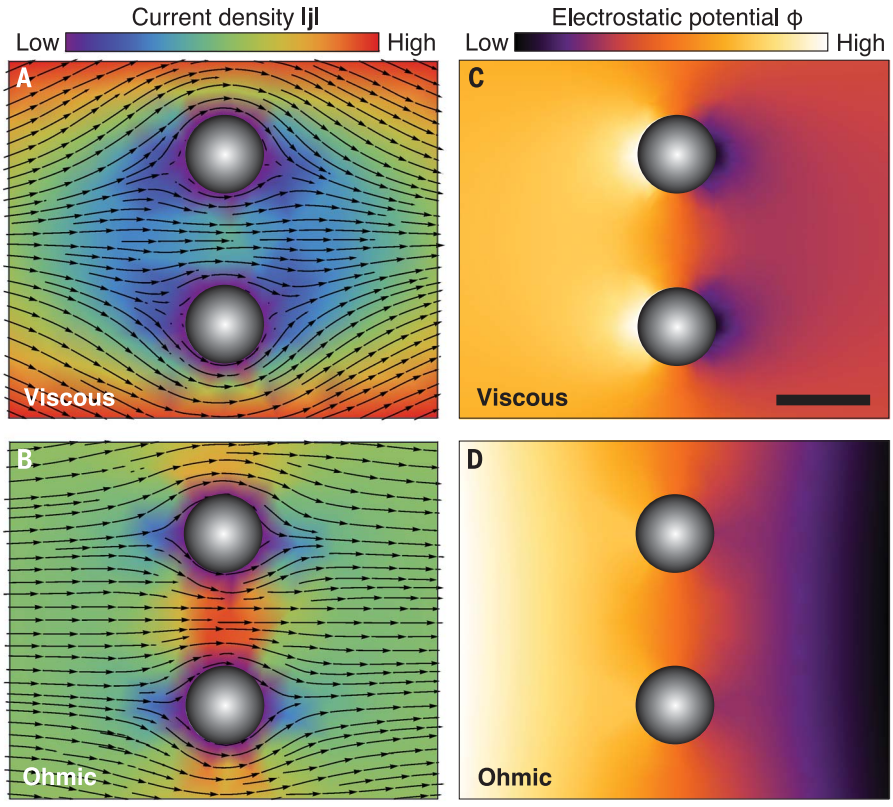


Fig. 5. Finite element models of carrier flow in the ohmic and viscous regimes. Numerical solutions for the current density and electric potential describing hydrodynamic flow through two circular barriers in the viscous regime (top panels) and ohmic regime (bottom panels). (**A** and **B**) The arrow plots show the streamline of the current density, and the color plots show the magnitude of the current density. (**C** and **D**) Electrostatic potential induced by current flow corresponding to the viscous and ohmic cases in (A) and (B), respectively. The scale bar is 250 nm.

shorter than previous measurements of hBN-encapsulated graphene largely because of the weaker dielectric environment ε of our samples (*32*). In Fig. 4B, we also plot theoretically predicted values of l_{ee} taken from (*14*) after multiplying by the factor $(\varepsilon/\varepsilon_{enc})^2 = (2.5/4)^2 = 0.39$ to account for the reduced dielectric screening of e-e interactions, where $\varepsilon_{enc} = 4$ is the dielectric constant for encapsulated samples (*14*). We note that for a temperature of $T = 77$ K, the measured scattering lengths l_{ee} are still less than the predicted value by ~30%, and we obtain a good fit using a temperature of 92 K (assuming $l_{ee} \sim T^{-2}$). This difference could be a result of multiple effects, such as Joule heating or a nonequilibrium energy distribution of electrons. Other factors include STM tip–induced gating effects and uncertainty in our estimate of c_G, which is known to depend sensitively on the channel boundary conditions.

Simulating carrier flow

These findings indicate that as the temperature of the graphene is increased from 4.5 K to 77 K, l_{ee} decreases until it is comparable to the width of the channel. Under these conditions, the carrier flow transitions from the Knudsen to the Gurzhi regime, where it behaves as a viscous Fermi liquid and exhibits superballistic conductance through a channel. To better understand the potential profiles measured using STP and how they relate to viscous flow, we compared our measurements with the following theoretical model. The motion of hydrodynamic electron flow under moderate external drive can be described by the linear Navier-Stokes equation in the following form

$$\nu\nabla^2\mathbf{u} - \frac{\mathbf{u}}{\tau_{mr}} = \frac{e}{m}\nabla\phi \qquad (4)$$

where $\mathbf{u}$ is the macroscopic flow velocity, τ_{mr} is the momentum relaxation time, ν is the kinematic viscosity of the electron fluid, e is the carrier charge, m is the carrier mass, and ϕ is the electric potential. The first term on the left-hand side of Eq. 4 describes viscous stress, while the second term accounts for ohmic loss. In addition, the continuity equation for current conservation is written as

$$\nabla\cdot\mathbf{j} = 0 \qquad (5)$$

Considering that $\mathbf{j} = ne\mathbf{u}$ and assuming constant electron density n, the linear Navier-Stokes Eq. 4 is recast in the form

$$l_G^2\nabla^2\mathbf{j} - \mathbf{j} = \sigma\nabla\phi \qquad (6)$$

where σ is the Drude conductivity (*67*). The interplay of viscous and momentum-relaxing terms introduces a natural length scale in the problem, namely the Gurzhi length, $l_G \equiv \sqrt{\nu\tau_{mr}} = \sqrt{l_{ee}l_{mr}}/2$. If $l_G \ll w$, where w is the typical size of the system, such as the width of the channel, the viscous stress in Eq. 6 can be neglected and the fluid enters the ohmic regime. In contrast, if $l_G \gg w$, the ohmic dissipation in Eq. 6 is small and the system is in the viscous regime.

In this framework, Eqs. 5 and 6 allow for calculations of the current density and electric potential profiles. Whereas analytical solutions are difficult to obtain for arbitrary geometries, numerical solutions using finite element methods can be obtained for hydrodynamic flow bypassing two circular barriers. Starting from the experimentally derived scattering lengths $l_{mr} = 5$ μm and $l_{ee} = 300$ nm, we simulated a viscous fluid with Gurzhi length of order $l_G = 1$ μm > w and compared these results with current flow in the ohmic regime, $l_G = 30$ nm < w. The distribution of current density and the electric potential from these two simulations are shown in Fig. 5. In these numerical simulations, we used no-slip boundary conditions for the flow velocity $\mathbf{u}$, and the flow is driven by the bias voltage V_{sd} applied to the left side of the sample (the right side is grounded to zero).

The results shown in Fig. 5C demonstrate good qualitative agreement with the 77 K STP measurements in Fig. 2, E to G. In particular, the streams of current that are anticipated for ballistic transport are not observed in the simulated viscous flow potentials; this observation is consistent with previous simulations of Boltzmann transport theory in graphene (*76*). Moreover, our numerical simulations predict the formation of cross-barrier dipoles in the electric potential profile in the viscous regime, which are observed in the measurements. The possibility of such dipole formation near channel edges in viscous electronic fluids was previously predicted (*17*) using analytical methods; this behavior can be attributed to the fact that the electric fields near the edges point against the current flow in order to push the electron liquid away from the boundary walls. The profiles of these regions of charge accumulation

or depletion depart in magnitude from what is expected for ballistic transport. This prediction compares well with our 77 K measurements and with our analytical modeling of the dipole described in Fig. 3, where we observe and calculate reduced dipole magnitudes.

Our numerical simulations also provide simultaneous measures of the cross-channel potential drop as well as the channel current, which, combined, allow for determination of the nonuniversal viscous contribution factor, c_G, used in Eqs. 2 and 3. To obtain c_G, we performed identical numerical simulations of a narrow-slit geometry, where c_G is known analytically to be $\pi^2/16$, and compared the observed channel conductance in that simulation to the results shown in Fig. 5, A and C, yielding $c_G = 0.75$ (*67*).

Outlook

We have shown that STP can be used to visualize hydrodynamic effects in graphene through direct imaging of the local electrochemical potential while a current is passed through the graphene sheet. This methodology offers a spatial resolution superior to that of other scanned probe measurements and allows for the creation and analysis of intricate flow geometries defined by smooth barriers created by in-plane p-n junctions. In this work, we used STP to reveal superballistic conductance through narrow channels in graphene, as well as local dipoles that form in both ballistic and viscous regimes owing to local carrier accumulation. These results provide insight into the electronic transport of hydrodynamic Fermi fluids. Specifically, the reduced dipoles that we observe indicate how the overall voltage drop generated by the summation of in-plane dipoles is naturally smaller in the viscous regime than in the ballistic regime, leading to a higher overall conductance in the viscous case. These measurements also provide a framework for analyzing more-complex flow patterns that are engineered to exhibit exotic effects, such as nonreciprocal flow (*79*). Angstrom-scale images, meanwhile, could be used to visualize atomistic transport features that are predicted to occur along grain boundaries and near defects (*53*).

REFERENCES AND NOTES

1. M. Knudsen, *Ann. Phys.* **28**, 75–130 (1909).
2. Y. V. Sharvin, *Sov. Phys. JETP* **21**, 655–656 (1965).
3. R. Gurzhi, *Sov. Phys. JETP* **17**, 521–522 (1963).
4. R. Gurzhi, *Sov. Phys. Usp.* **11**, 255–270 (1968).
5. L. W. Molenkamp, M. J. M. de Jong, *Phys. Rev. B* **49**, 5038–5041 (1994).
6. M. J. M. de Jong, L. W. Molenkamp, *Phys. Rev. B* **51**, 13389–13402 (1995).
7. X. P. A. Gao *et al*., *Phys. Rev. Lett.* **94**, 086402 (2005).
8. B. A. Braem *et al*., *Phys. Rev. B* **98**, 241304 (2018).
9. A. D. Levin, G. M. Gusev, E. V. Levinson, Z. D. Kvon, A. K. Bakarov, *Phys. Rev. B* **97**, 245308 (2018).
10. G. M. Gusev, A. S. Jaroshevich, A. D. Levin, Z. D. Kvon, A. K. Bakarov, *Sci. Rep.* **10**, 7860 (2020).
11. A. Gupta *et al*., *Phys. Rev. Lett.* **126**, 076803 (2021).
12. M. Müller, J. Schmalian, L. Fritz, *Phys. Rev. Lett.* **103**, 025301 (2009).
13. I. Torre, A. Tomadin, A. K. Geim, M. Polini, *Phys. Rev. B* **92**, 165433 (2015).
14. A. Principi, G. Vignale, M. Carrega, M. Polini, *Phys. Rev. B* **93**, 125410 (2016).
15. F. M. D. Pellegrino, I. Torre, A. K. Geim, M. Polini, *Phys. Rev. B* **94**, 155414 (2016).
16. L. Levitov, G. Falkovich, *Nat. Phys.* **12**, 672–676 (2016).
17. H. Guo, E. Ilseven, G. Falkovich, L. S. Levitov, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3068–3073 (2017).
18. A. Stern *et al*., *Phys. Rev. Lett.* **129**, 157701 (2022).
19. B. N. Narozhny, I. V. Gornyi, A. D. Mirlin, J. Schmalian, *Ann. Phys.* **529**, 1700043 (2017).
20. A. Lucas, K. C. Fong, *J. Phys. Condens. Matter* **30**, 053001 (2018).
21. D. Y. H. Ho, I. Yudhistira, N. Chakraborty, S. Adam, *Phys. Rev. B* **97**, 121404 (2018).
22. G. Falkovich, L. Levitov, *Phys. Rev. Lett.* **119**, 066601 (2017).
23. A. Shytov, J. F. Kong, G. Falkovich, L. Levitov, *Phys. Rev. Lett.* **121**, 176805 (2018).
24. S. Danz, M. Titov, B. N. Narozhny, *Phys. Rev. B* **102**, 081114 (2020).
25. M. S. Foster, I. L. Aleiner, *Phys. Rev. B* **79**, 085415 (2009).
26. A. Lucas, J. Crossno, K. C. Fong, P. Kim, S. Sachdev, *Phys. Rev. B* **93**, 075426 (2016).
27. M. Zarenia, A. Principi, G. Vignale, *2D Mater.* **6**, 035024 (2019).
28. S. Li, A. Levchenko, A. V. Andreev, *Phys. Rev. B* **102**, 075305 (2020).
29. A. Levchenko, J. Schmalian, *Ann. Phys.* **419**, 168218 (2020).
30. J. Gooth *et al*., *Nat. Commun.* **9**, 4093 (2018).
31. U. Vool *et al*., *Nat. Phys.* **17**, 1216–1220 (2021).
32. R. Krishna Kumar *et al*., *Nat. Phys.* **13**, 1182–1185 (2017).
33. D. A. Bandurin *et al*., *Science* **351**, 1055–1058 (2016).
34. D. A. Bandurin *et al*., *Nat. Commun.* **9**, 4533 (2018).
35. J. Crossno *et al*., *Science* **351**, 1058–1061 (2016).
36. F. Ghahari *et al*., *Phys. Rev. Lett.* **116**, 136802 (2016).
37. J. A. Sulpizio *et al*., *Nature* **576**, 75–79 (2019).
38. L. Ella *et al*., *Nat. Nanotechnol.* **14**, 480–487 (2019).
39. A. Jenkins *et al*., *Phys. Rev. Lett.* **129**, 087701 (2022).
40. M. J. H. Ku *et al*., *Nature* **583**, 537–541 (2020).
41. C. Kumar *et al*., *Nature* **609**, 276–281 (2022).
42. A. Aharon-Steinberg *et al*., *Nature* **607**, 74–80 (2022).
43. S. Samaddar *et al*., *Nano Lett.* **21**, 9365–9373 (2021).
44. J. Velasco Jr. *et al*., *Nano Lett.* **16**, 1620–1625 (2016).
45. E. I. Kiselev, J. Schmalian, *Phys. Rev. B* **99**, 035430 (2019).
46. R. Moessner, N. Morales-Durán, P. Surówka, P. Witkowski, *Phys. Rev. B* **100**, 155115 (2019).
47. P. Muralt, D. W. Pohl, in *Scanning Tunneling Microscopy* (Springer, 1986), pp. 252–254.
48. P. Muralt, H. Meier, D. W. Pohl, H. W. M. Salemink, *Appl. Phys. Lett.* **50**, 1352–1354 (1987).
49. C. S. Chu, R. S. Sorbello, *Phys. Rev. B* **40**, 5950–5955 (1989).
50. T. Druga, M. Wenderoth, J. Homoth, M. A. Schneider, R. G. Ulbrich, *Rev. Sci. Instrum.* **81**, 083704 (2010).
51. J. R. Kirtley, S. Washburn, M. J. Brady, *Phys. Rev. Lett.* **60**, 1546–1549 (1988).
52. K. H. Bevan, *Nanotechnology* **25**, 415701 (2014).
53. D. K. Morr, *Phys. Rev. B* **95**, 195162 (2017).
54. B. G. Briner, R. M. Feenstra, T. P. Chin, J. M. Woodall, *Phys. Rev. B* **54**, R5283–R5286 (1996).
55. R. Landauer, *IBM J. Res. Develop.* **1**, 223–231 (1957).
56. S.-H. Ji *et al*., *Nat. Mater.* **11**, 114–119 (2012).
57. F. Giannazzo, I. Deretzis, A. La Magna, F. Roccaforte, R. Yakimova, *Phys. Rev. B* **86**, 235422 (2012).
58. W. Wang, K. Munakata, M. Rozler, M. R. Beasley, *Phys. Rev. Lett.* **110**, 236802 (2013).
59. P. Willke, T. Druga, R. G. Ulbrich, M. A. Schneider, M. Wenderoth, *Nat. Commun.* **6**, 6399 (2015).
60. K. W. Clark *et al*., *ACS Nano* **7**, 7956–7966 (2013).
61. A. Sinterhauf *et al*., *Nat. Commun.* **11**, 555 (2020).
62. A. De Cecco *et al*., *Nano Lett.* **20**, 3786–3790 (2020).
63. J. P. Pelz, R. H. Koch, *Phys. Rev. B* **41**, 1212–1215 (1990).
64. J. Lee *et al*., *Nat. Phys.* **12**, 1032–1036 (2016).
65. J. Velasco Jr. *et al*., *Nano Lett.* **18**, 5104–5110 (2018).
66. B. Brun *et al*., *Phys. Rev. B* **100**, 041401 (2019).
67. See supplementary materials.
68. R. S. Sorbello, *Phys. Rev. B* **23**, 5119–5127 (1981).
69. R. S. Sorbello, C. S. Chu, *IBM J. Res. Develop.* **32**, 58–62 (1988).
70. H. Guo, E. Ilseven, G. Falkovich, L. Levitov, arXiv:1612.09239 [cond-mat.mes-hall] (2017).
71. C. Gutiérrez *et al*., *Science* **361**, 789–794 (2018).
72. E. A. Quezada-López *et al*., *Nanomaterials* **10**, 1154 (2020).
73. W. A. Behn *et al*., *Nano Lett.* **21**, 5013–5020 (2021).
74. J. A. Støvneng, P. Lipavský, *Phys. Rev. B* **42**, 9214–9216 (1990).
75. J. Park, G. He, R. M. Feenstra, A.-P. Li, *Nano Lett.* **13**, 3269–3273 (2013).
76. A. Lucas, *Phys. Rev. B* **95**, 115425 (2017).
77. S. S. Pershoguba, A. F. Young, L. I. Glazman, *Phys. Rev. B* **102**, 125404 (2020).
78. S. Li, M. Khodas, A. Levchenko, *Phys. Rev. B* **104**, 155305 (2021).
79. J. Geurs *et al*., arXiv:2008.04862 [cond-mat.mes-hall] (2020).
80. Z. J. Krebs *et al*., Imaging the breaking of electrostatic dams in graphene for ballistic and viscous fluids, version 1, Zenodo (2022); https://doi.org/10.5281/zenodo.7314592.

ACKNOWLEDGMENTS

The authors thank A. Marguerite for stimulating discussions. **Funding:** Funding was provided by US Department of Energy, Office of Science, Basic Energy Sciences (BES) Program for Materials and Chemistry Research in Quantum Information Science award DE-SC0020313 (Z.J.K., K.J.S., and A.L.); Office of Naval Research award N00014-20-1-2356 (W.A.B. and V.W.B.); National Science Foundation grant DMR-1653661 (S.L.); University of Wisconsin Materials Research Science and Engineering Center grant DMR1720415; Elemental Strategy Initiative, MEXT, Japan, grant JPMXP0112101001 (K.W. and T.T.); and JSPS KAKENHI grants JP19H05790 and JP20H00354 (K.W. and T.T.). **Author contributions:** Z.J.K. and W.A.B. performed all STP measurements and data analysis. K.J.S. fabricated the graphene devices. K.W. and T.T. grew the high-quality hBN crystals used in all graphene devices. S.L. and A.L. ran numerical simulations of electron fluid flow in graphene. V.W.B., Z.J.K., and W.A.B. conceived of the experiment and designed the measurement system. Z.J.K., S.L., and V.W.B. modeled the results using theoretical tools. Z.J.K., W.A.B., V.W.B., and S.L. wrote the manuscript. V.W.B. and A.L. supervised the experimental and theoretical aspects of the research, respectively. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and files used to generate the results of this article—experimental and theoretical—can be found in Zenodo (*80*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm6073
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (*81*–*84*)

Submitted 29 September 2021; accepted 24 January 2023
10.1126/science.abm6073

MOLECULAR BIOLOGY

Exon architecture controls mRNA m^{6}A suppression and gene expression

P. Cody He[1,2,3]†, Jiangbo Wei[1,3]†, Xiaoyang Dou[1,3]†, Bryan T. Harada[1,3]†‡, Zijie Zhang[1,3,4], Ruiqi Ge[1,3], Chang Liu[1,3]§, Li-Sheng Zhang[1,3], Xianbin Yu[1,3], Shuai Wang[5], Ruitu Lyu[1,3], Zhongyu Zou[1,3], Mengjie Chen[6,7], Chuan He[1,2,3]*

N^6-methyladenosine (m^{6}A) is the most abundant messenger RNA (mRNA) modification and plays crucial roles in diverse physiological processes. Using a massively parallel assay for m^{6}A (MPm^{6}A), we discover that m^{6}A specificity is globally regulated by suppressors that prevent m^{6}A deposition in unmethylated transcriptome regions. We identify exon junction complexes (EJCs) as m^{6}A suppressors that protect exon junction–proximal RNA within coding sequences from methylation and regulate mRNA stability through m^{6}A suppression. EJC suppression of m^{6}A underlies multiple global characteristics of mRNA m^{6}A specificity, with the local range of EJC protection sufficient to suppress m^{6}A deposition in average-length internal exons but not in long internal and terminal exons. EJC-suppressed methylation sites colocalize with EJC-suppressed splice sites, which suggests that exon architecture broadly determines local mRNA accessibility to regulatory complexes.

N^6-methyladenosine (m^{6}A), the most prevalent mRNA modification in mammals, influences wide-ranging aspects of gene expression in diverse physiological and pathophysiological processes (*1*–*3*). The METTL3-METTL14 methyltransferase complex installs m^{6}A methylation on mRNA in a common DRACH (where D = A, G, or U; R = A or G; and H = A, C, or U) sequence motif, but only a fraction of DRACH sequences (~5%) in a subset of cellular transcripts are selected for methylation (*4*). Additionally, m^{6}A exhibits a marked regional bias in its transcriptomic distribution, being strongly enriched in long internal exons and near stop codons (*5*, *6*). Despite the central importance of specific m^{6}A deposition in m^{6}A-mediated gene regulation, the mechanistic basis for m^{6}A specificity has remained poorly understood.

In this study, we discover the existence of prevalent regulatory mechanisms that restrict m^{6}A methylation to specific transcript regions through targeted suppression of m^{6}A in unmethylated regions. We find that pre-mRNA splicing selectively suppresses m^{6}A deposition in average-length internal exons but not in longer exons. We identify exon junction complexes (EJCs) as major m^{6}A suppressors that mediate this effect and control several key characteristics of global m^{6}A specificity. EJC depletion results in pervasive aberrant methylation of mRNAs and m^{6}A-mediated transcript destabilization. EJCs, together with interacting proteins, package and protect long stretches of proximal RNA from cellular methylation deposition, which may represent a general mechanism by which exon architecture and EJC positioning determine local mRNA accessibility to regulatory machineries.

[1]Department of Chemistry, Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA. [2]Committee on Immunology, The University of Chicago, Chicago, IL 60637, USA. [3]Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA. [4]State Key Laboratory for Conservation and Utilization of Bio-Resources, School of Life Sciences, Yunnan University, Kunming, Yunnan 650091, China. [5]Department of Neurobiology, The University of Chicago, Chicago, IL 60637, USA. [6]Department of Human Genetics, The University of Chicago, Chicago, IL 60637, USA. [7]Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA.
***Corresponding author. Email: chuanhe@uchicago.edu**
†These authors contributed equally to this work.
‡Present address: Cell Press, Cambridge, MA 02139, USA.
§Present address: Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA.

Massively parallel assay for m^{6}A

The extent to which global m^{6}A specificity is controlled has important implications for m^{6}A regulation but is poorly understood (*4*). We approached this problem by investigating whether the local sequence surrounding an m^{6}A methylated site, when uncoupled from its endogenous context, is sufficient to specify methylation at that site. Conversely, we also investigated whether the local sequence surrounding an unmethylated DRACH site, when uncoupled from its endogenous context, is sufficient to prevent methylation at that site.

To assess this systematically on an epitranscriptome-wide scale, we developed a massively parallel reporter assay (MPRA) that enables high-throughput assessment of the m^{6}A methylation status of thousands of designed sequences, which we named massively parallel assay for m^{6}A (MPm^{6}A) (Fig. 1A and fig. S1A). In the MPm^{6}A workflow, thousands of endogenously methylated m^{6}A sites and endogenously unmethylated DRACH sites, with 102 nucleotides (nt) of sequence surrounding each site, were synthesized and cloned into the 3′ untranslated region (3′UTR) of a plasmid-based, intronless green fluorescent protein (GFP) transgene. For each sequence, we also designed a corresponding negative control sequence in which all DRACH motifs were mutated to prevent methylation. The sequences were expressed and then m^{6}A methylated through transfection into cells or, when specified, through in vitro transcription and in vitro m^{6}A methylation. The methylation status of each individual sequence was assessed by its enrichment after m^{6}A–immunoprecipitation (IP) of mRNA. We selected 6897 HeLa m^{6}A sites and 3058 unmethylated DRACH sites to assay in HeLa cells and validated the assay's precision and accuracy (fig. S1, B to D).

Widespread mRNA m^{6}A suppression controls m^{6}A epitranscriptome specificity

When we compared the methylation levels of the endogenously methylated sequences with their negative control sequences, we found that 92.8% of the sequences exhibited significant methylation in this reporter assay (Fig. 1B and fig. S1E), which indicates that most endogenously methylated sites do not strictly require their larger surrounding native context for methylation. Unexpectedly, 90.2% of endogenously unmethylated sequences also exhibited significant methylation (Fig. 1B and fig. S1E). The MPm^{6}A enrichment scores of the endogenously unmethylated group were similar to those of the endogenously methylated group, despite their diverging endogenous methylation states (Fig. 1B). We observed similar results when the sequences were in vitro transcribed and methylated with recombinant METTL3-METTL14 (fig. S2). Thus, thousands of endogenously unmethylated DRACH sites became methylated when they were uncoupled from their endogenous contexts and expressed in an artificial reporter context. We call these identified sites suppressed m^{6}A sites. We validated these results for three selected sequences (fig. S1F) and confirmed that methylation was not notably influenced by the cytomegalovirus (CMV) promoter of the MPm^{6}A plasmid (fig. S3). We observed similar results when sequences were expressed within coding sequences (CDSs) or 5′UTRs, though m^{6}A enrichment was significantly lower for many sequences when placed in the CDSs or 5′UTRs versus in the 3′UTRs (fig. S4, A to D). This suggests that 5′UTRs are generally less conducive for m^{6}A methylation compared with 3′UTRs (fig. S4E). Collectively, these results reveal the existence of thousands of suppressed m^{6}A sites that are silenced by unknown mechanisms.

We noted that suppressed m^{6}A sites were enriched in the CDS and 3′UTR and were depleted near the stop codon, which is the reverse of endogenous m^{6}A site enrichment (Fig. 1C and fig. S5A). Further, suppressed m^{6}A sites in internal exons reside within much shorter exons (median = 167 nt) compared with endogenous m^{6}A sites (median = 915 nt) (Fig. 1D). These observations suggest that endogenous m^{6}A enrichment in long internal exons may

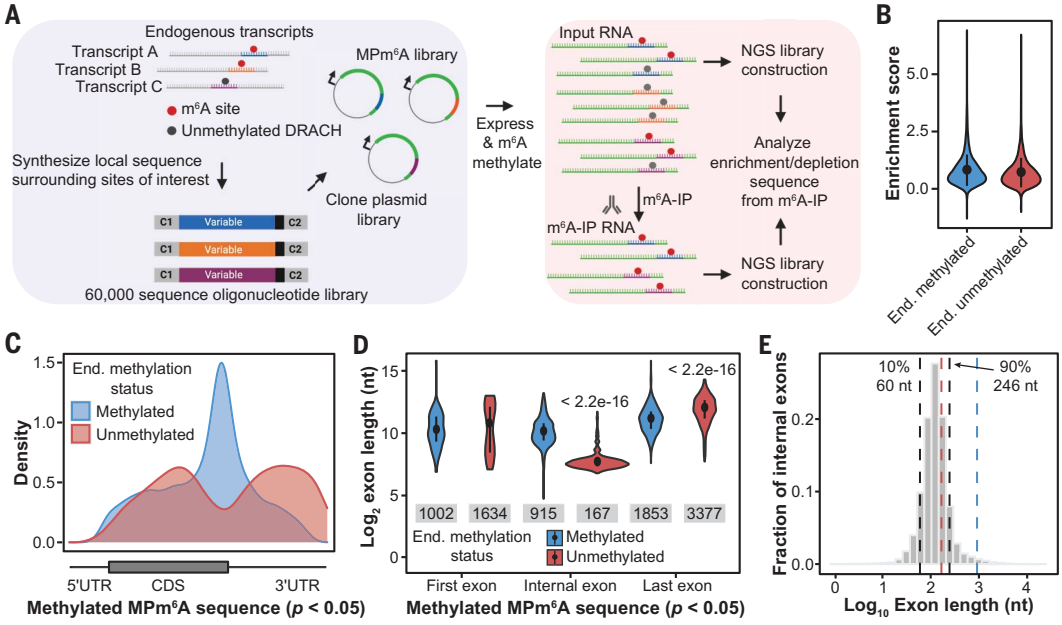


Fig. 1. MPm^{6}A reveals suppression of thousands of m^{6}A sites in unmethylated transcriptome regions. (**A**) Schematic of the MPm^{6}A workflow. NGS, next generation sequencing. (**B**) MPm^{6}A enrichment scores [(experimental IP/input) − (negative control IP/input)] for endogenously methylated (n = 6095) and unmethylated (n = 2716) sequences. Values are means ± SDs from four biological replicates. (**C**) Metagenes of endogenously methylated and unmethylated sequences that are significantly methylated in MPm^{6}A. (**D**) Exon lengths of endogenously methylated and unmethylated sequences that are significantly methylated in MPm^{6}A. Values are medians and interquartile ranges (IQRs). Statistics by Wilcoxon rank sum test. Sample sizes for each violin plot from left to right are: n = 175, n = 22, n = 696, n = 519, n = 3539, and n = 1328. (**E**) Distribution of internal exon lengths in the human genome. Black lines indicate 10th percentile (left; 60 nt) and 90th percentile (right; 246 nt). Blue and red lines indicate median internal exon length for MPm^{6}A endogenously methylated (915 nt) and unmethylated (167 nt) sequences, respectively.

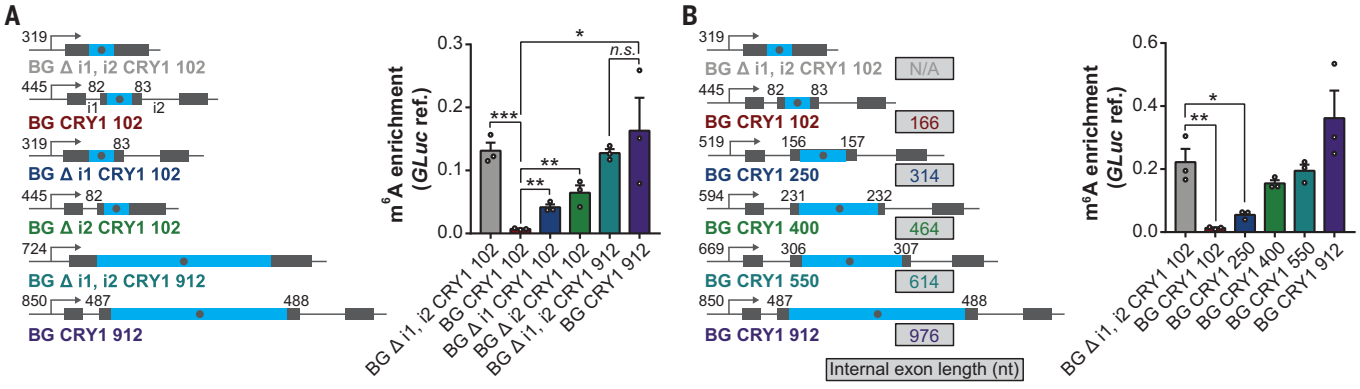


Fig. 2. Pre-mRNA splicing suppresses m^{6}A methylation in average-length exons. (**A** and **B**) (Left) Schematics of specified BG CRY1 constructs. Blue regions indicate sequences derived from the *CRY1* endogenous sequence, and gray regions indicate sequences derived from rabbit BG. The numbers after CRY1 refer to the number of nucleotides of exonic sequence surrounding the *CRY1*-suppressed m^{6}A site in the *CRY1* endogenous mRNA that was cloned into the BG construct. The gray dots in the blue regions denote the suppressed m^{6}A site; the numbers at the left and right of the m^{6}A site show the distance (in nucleotides) between the m^{6}A site and the 3′ and 5′ splice sites, respectively; and the numbers next to the TSS show the distance (in nucleotides) between the m^{6}A site and the promoter. A delta (Δ) denotes deletion of the specified intron(s). Details of each construct are described in the materials and methods. (Right) m^{6}A enrichment at a *CRY1*-suppressed m^{6}A site. Primers amplifying a 62-nt fragment containing the *CRY1*-suppressed m^{6}A site. m^{6}A enrichment was calculated as IP/input normalized to m^{6}A-marked *Gaussia* luciferase (GLuc) RNA spike-in IP/input. Values are means ± SEMs. Statistics by two-tailed t test. *P < 0.05; **P < 0.01; ***P < 0.001. Three biological replicates were used.

be a consequence of suppression of m^{6}A deposition in shorter internal exons, which make up most exons (90% of internal exons are <246 nt) (Fig. 1E). In total, 942 genes containing suppressed m^{6}A sites did not contain any endogenous m^{6}A peaks on their transcripts (fig. S5B). Suppression of these sites appears to involve suppression of m^{6}A deposition rather than active demethylation because binding sites for RBM15—a METTL3-METTL14 methyltransferase complex accessory subunit—were highly enriched near endogenous m^{6}A sites compared with suppressed m^{6}A sites (fig. S6A), whereas FTO and ALKBH5 binding sites were not significantly enriched near suppressed sites (fig. S6B). These results were surprising because previous studies on m^{6}A specificity had mainly focused on activating mechanisms (*7*–*12*). Our MPm^{6}A assay suggests the existence of unknown m^{6}A suppressors that govern global m^{6}A specificity by suppressing m^{6}A deposition.

Pre-mRNA splicing suppresses m^{6}A methylation proximal to splice sites

We next examined the relative enrichment of binding sites for 120 RNA binding proteins (RBPs) near endogenously methylated versus suppressed m^{6}A sites to identify candidate suppressors (*13*). Several spliceosome components

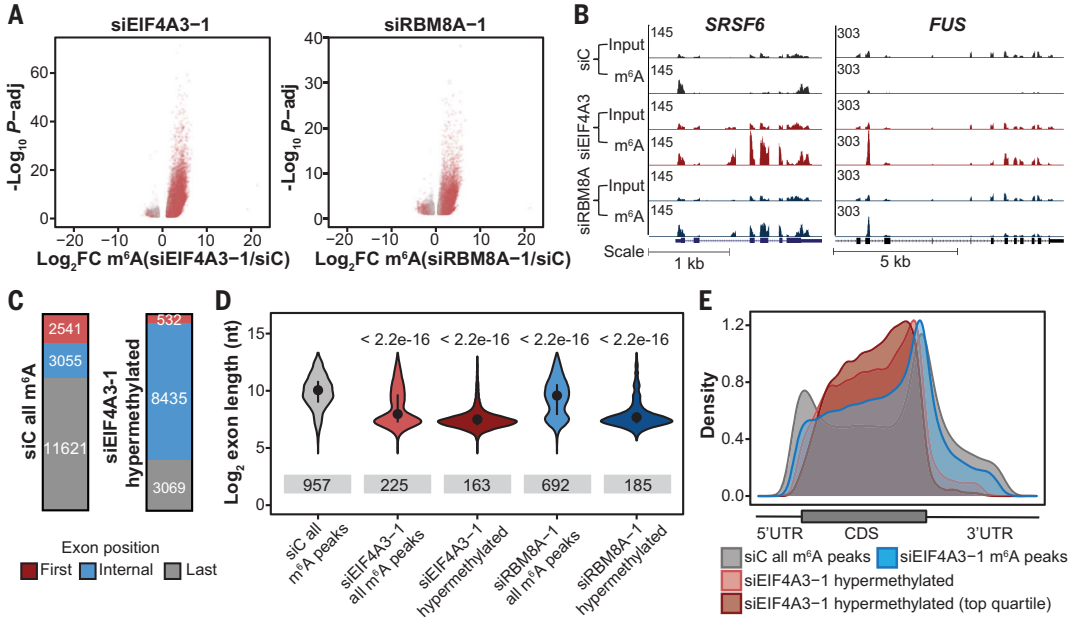


Fig. 3. EJCs protect exon junction–proximal RNA in average-length exons within CDS regions from m^{6}A methylation. (**A**) Differentially methylated regions upon *EIF4A3* KD (left) and *RBM8A* KD (right) in HeLa cells (FDR < 0.1 and $|\log_2 FC| > 1$, where FDR is false discovery rate and FC is fold change). Three biological replicates were used. Gray and red dots indicate differentially methylated regions that overlap and do not overlap m^{6}A peaks in the control cells, respectively. *P*-adj, adjusted *P* value. (**B**) Input and m^{6}A-IP read coverage at *FUS* and *SRSF6* in *EIF4A3* KD, *RBM8A* KD, and control HeLa cells. (**C**) Numbers of *EIF4A3* KD hypermethylated regions (left) and m^{6}A peaks in control cells (right) that reside within first, internal, or last exons. (**D**) Exon lengths for m^{6}A peaks residing within internal exons in control KD, EIF4A3 KD, and RBM8A KD cells and exon lengths of hypermethylated regions residing within internal exons in EIF4A3 and RBM8A KD cells. Dots and bars represent medians and IQRs. Statistics by Wilcoxon rank sum test of indicated group versus the nontargeting siRNA control (siC). Sample sizes for each violin plot from left to right are: *n* = 3166, *n* = 6659, *n* = 8438, *n* = 3817, and *n* = 3827. (**E**) Metagenes of m^{6}A peaks and significantly hypermethylated m^{6}A regions (and top quartile) in *EIF4A3* KD HeLa cells and m^{6}A peaks in control cells.

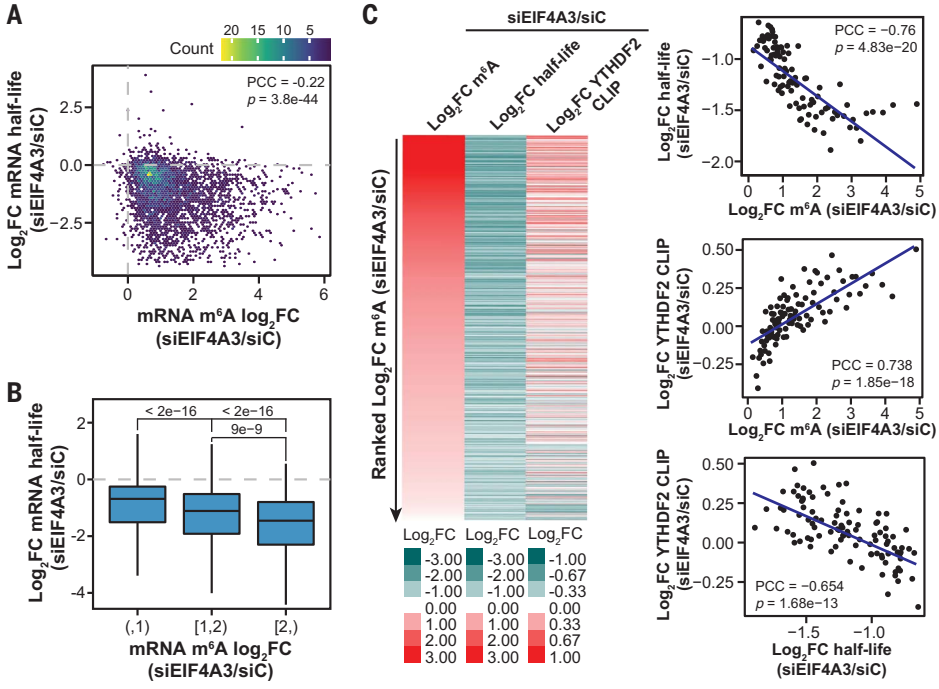


Fig. 4. mRNA m^{6}A hypermethylation upon EJC depletion destabilizes mRNAs. (**A**) Correlation between fold changes in mRNA half-life and m^{6}A level upon *EIF4A3* KD in HeLa cells (*n* = 3840). (**B**) Boxplots showing half-life fold changes of hypermethylated mRNAs upon *EIF4A3* KD in HeLa cells. mRNAs were categorized into three groups according to their methylation changes upon *EIF4A3* KD in HeLa cells. *P* values are from Wilcoxon rank sum test. Sample sizes for each boxplot plot from left to right are: *n* = 1887, *n* = 1201, and *n* = 752. (**C**) (Left) Heatmap showing fold changes in m^{6}A level, mRNA half-life, and YTHDF2 binding upon *EIF4A3* KD in HeLa cells. (Right) Scatter plots showing the correlation among fold changes in m^{6}A level, mRNA half-life, and YTHDF2 binding upon *EIF4A3* KD in HeLa cells. The hypermethylated mRNAs (m^{6}A $\log_2 FC > 0$; *n* = 3424) were categorized into 100 bins on the basis of ranked fold change of m^{6}A level upon *EIF4A3* KD. For (A) and (C), Pearson correlation coefficient (PCC) and *P* values are shown.

(*BUD13*, *SF3B4*, and *EFTUD2*) were significantly enriched near suppressed sites, which suggests that splicing may suppress m^{6}A (fig. S6A). Because suppressed m^{6}A sites in CDS primarily reside within average-length internal exons, we hypothesized that the splicing of average-length internal exons may suppress m^{6}A methylation. To test this, we cloned a suppressed m^{6}A site from an average-length internal exon in the *CRY1* gene (fig. S7A) into a rabbit beta-globin (BG) minigene reporter as well as a version with the introns removed (BG Δi1,i2). First, we cloned the suppressed *CRY1* site and 50/51 nt of the flanking sequence into the internal exon or in the last exon of these constructs. Notably, the spliced construct strongly suppressed methylation of the sequence placed within the internal exon

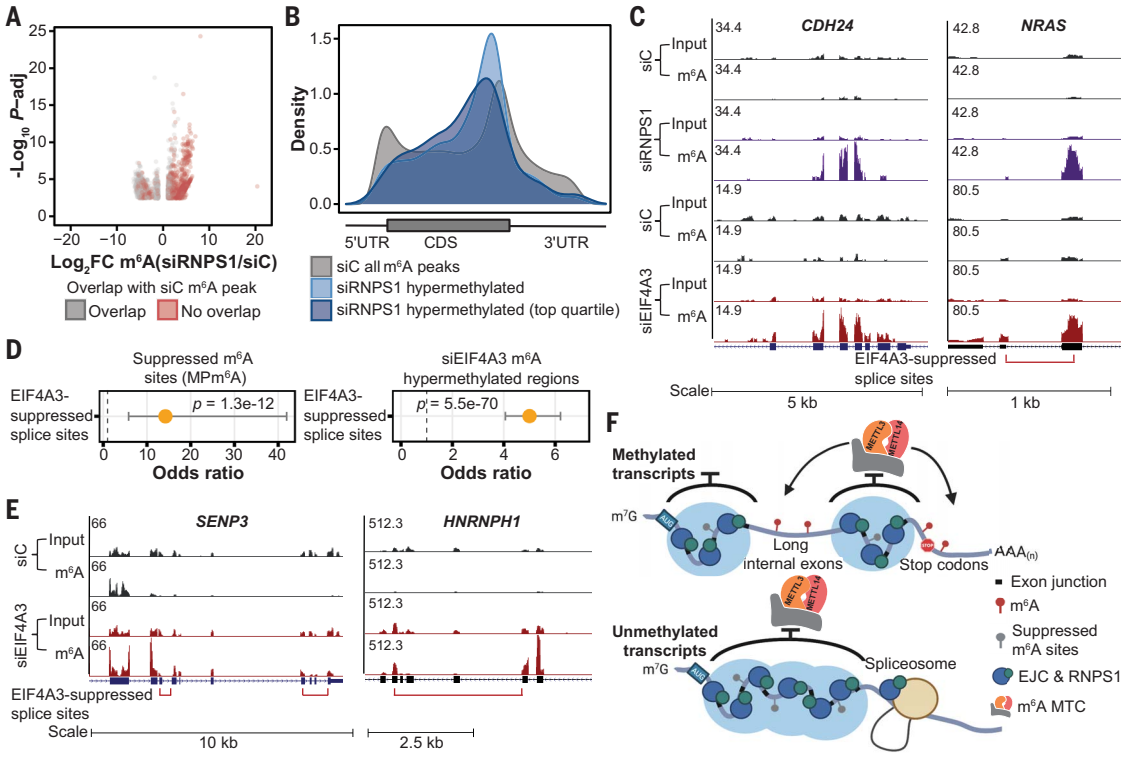


Fig. 5. EJCs and RNPS1 protect proximal RNA regions from aberrant mRNA processing. (**A**) Differentially methylated regions upon *RNPS1* KD in HeLa cells (FDR < 0.1, $|\log_2 FC| > 1$) from three biological replicates. Gray and red dots indicate differentially methylated regions that overlap and do not overlap with m^{6}A peaks in the control KD cells, respectively. (**B**) Metagenes of significantly m^{6}A hypermethylated regions (and top quartile) upon *RNPS1* KD in HeLa cells in comparison with that of all m^{6}A peaks in control cells. (**C**) Input and m^{6}A-IP read coverage at *CDH24* and *NRAS* upon *RNPS1* KD and *EIF4A3* KD, respectively, as well as corresponding controls in HeLa cells. (**D**) Enrichment of suppressed m^{6}A sites (identified from MPm^{6}A) at EIF4A3-suppressed splice sites (left) and enrichment of EIF4A3 KD hypermethylated regions at EIF4A3-suppressed splice sites (right). Statistics were by Fisher's exact test, and dots and bars represent odds ratios and 95% confidence intervals, respectively. (**E**) Input and m^{6}A-IP read coverage at *SENP3* and *HNRNPH1* in *EIF4A3* KD and control HeLa cells. (**F**) Schematic model depicting that EJCs and RNPS1 (and potentially other EJC-associated proteins) protect exon junction–proximal RNA from m^{6}A deposition through local mRNA packaging. For (C) and (E), the red brackets indicate EIF4A3-suppressed splice variant, with the ends of the brackets indicating the suppressed splice junctions.

but not within the last exon (fig. S7B). Removal of either intron (BG Δi1 CRY1 102 or BG Δi2 CRY1 102) resulted in partial loss of suppression, which indicates that splicing of both introns contributed to methylation suppression (Fig. 2A). Consistent with this notion, deletion of all splice sites also resulted in a decrease in m^{6}A suppression (fig. S7C). Cloning in 912 nt of the *CRY1* exonic sequence surrounding the suppressed site into the internal exon (BG CRY1 912), forming a long internal exon, resulted in a loss of suppression (Fig. 2A). We hypothesized that the suppression is dependent on the proximity of the m^{6}A site, located within the center of the exon, to splice sites. Expanding the length of the BG CRY1 102 internal exon by cloning in larger amounts of *CRY1* flanking exonic sequence resulted in a progressive loss of suppression, with a >464-nt internal exon unable to suppress m^{6}A (Fig. 2B and fig. S7D). These results reveal a causal role for pre-mRNA splicing in m^{6}A regulation.

EJCs control m^{6}A epitranscriptome specificity

We next sought to understand the mechanism by which splicing suppresses m^{6}A deposition. EJCs are deposited by spliceosomes onto mRNA ~24 nt upstream of exon-exon junctions and play multifaceted roles in gene expression regulation (*14*, *15*). Notably, two recent studies reported that EJCs efficiently block splicing at proximal aberrant splice sites (*16*, *17*). Additionally, EJCs, together with interacting serine- and arginine-rich (SR) proteins, package and compact mRNA and can protect long stretches of proximal RNA from nuclease accessibility in vitro and block 5′ to 3′ exonuclease degradation in vivo (*18*, *19*). We reasoned that suppressed m^{6}A sites within average-length internal exons are within relatively close proximity to both an upstream and downstream EJCs. Conversely, m^{6}A sites within long internal exons and near stop codons (which generally reside in long last exons) can be hundreds of nucleotides away from the nearest EJC. We therefore hypothesized that EJCs could mediate the splice site–proximal suppression of m^{6}A that we observed.

We knocked down the core EJC factor EIF4A3 in HeLa cells and assessed the effect on m^{6}A deposition transcriptome-wide using m^{6}A methylated RNA immunoprecipitation sequencing (m^{6}A-MeRIP-seq). Notably, 24,350 regions were significantly hypermethylated upon *EIF4A3* knockdown (KD), whereas 3140 regions were hypomethylated (Fig. 3A). Of these hypermethylated regions, 39% exhibited a greater than eightfold increase in m^{6}A enrichment compared with the nontargeting small interfering RNA (siRNA) control. We knocked down RBM8A, another core EJC factor (*20*), and observed similar, though relatively milder, transcriptome-wide m^{6}A changes, with 14,034 significantly hypermethylated regions observed, of which 57% overlapped with hypermethylated regions observed in *EIF4A3* KD (Fig. 3A and fig. S8, A and B). The relatively milder m^{6}A changes upon *RBM8A* KD may result from relatively lower KD efficiency (table S1) or may indicate a stronger requirement of EIF4A3 for suppression. Concordant with these transcriptome-wide m^{6}A changes, using ultrahigh-performance liquid chromatography coupled with triple quadrupole mass spectrometry (UHPLC-QQQ-MS/MS), we found that *EIF4A3* KD increased global levels of m^{6}A in polyadenylated RNA by twofold, whereas *RBM8A* KD resulted in an ~25% increase (fig. S8C).

In total, 94% of hypermethylated regions from *EIF4A3* KD and 82% of hypermethylated regions from *RBM8A* KD did not overlap with m^{6}A peaks identified under the nontargeting siRNA control conditions, which suggests that these regions contain newly methylated, suppressed m^{6}A sites (Fig. 3, A and B, and fig. S8D). Out of 1024 CDS sequences identified by MPm^{6}A to contain suppressed m^{6}A sites, 46%

become methylated upon *EIF4A3* or *RBM8A* KD, including the *CRY1*-suppressed site (fig. S8, E to G) (*21*). Furthermore, *EIF4A3* KD substantially alleviated the previously observed m^{6}A suppression within the internal exon of BG CRY1 102 (fig. S8H).

Consistent with our model, newly methylated and hypermethylated regions were highly enriched in average-length internal exons within CDSs (Fig. 3, C to E, and fig. S8, I and J), with transcriptome-wide increases in m^{6}A enrichment in exon junction–proximal regions observed (fig.S9, A and B) upon *EIF4A3* or *RBM8A* KD. *EIF4A3* KD disrupted m^{6}A epitranscriptome specificity globally, resulting in increased density of m^{6}A in the CDSs relative to the stop codon and attenuated enrichment of m^{6}A peaks in long internal exons (Fig. 3, D and E). It was previously reported that the peak of m^{6}A density near stop codons on metagene plots can be more precisely visualized as an increased enrichment 150 nt past the start of last exons (*6*). *EIF4A3* KD resulted in a global increase in m^{6}A enrichment <150 nt past the start of last exons (fig. S9, A to C), which indicates that EJC suppression of methylation proximal to last exon-exon junctions is responsible for the characteristic m^{6}A peak density near stop codons. Whereas most transcripts exhibited hypermethylation and contained one or more endogenous m^{6}A peaks upon *EIF4A3* KD, more than a thousand transcripts that ordinarily lack endogenous m^{6}A peaks also gained aberrant m^{6}A methylation upon *EIF4A3* KD, revealing a major role for EJCs in suppressing m^{6}A deposition on the subset of transcripts that ordinarily are not subject to m^{6}A regulation (fig. S9, D to F).

The widespread suppression of m^{6}A by the EJCs also implies that many m^{6}A modifications are deposited after splicing, which we confirmed using pulse-chase metabolic labeling experiments and UHPLC-QQQ-MS/MS (supplementary text and fig. S10). Two genes used in gene therapies for mucopolysaccharidosis type II and spinal muscular atrophy—*IDS* and *SMN*, respectively—contain EJC-suppressed m^{6}A sites in their mRNAs. As expected, when these mRNAs were expressed from cDNA constructs, and thus not bound by EJCs, they were significantly hypermethylated relative to the corresponding endogenous mRNAs (fig. S11). Further, long noncoding RNAs (lncRNAs) that contain three or more exons globally exhibit EJC suppression of m^{6}A in internal exons, whereas those with two or less do not (supplementary text and fig. S12). We depleted *EIF4A3* with a different siRNA in HeLa cells and knocked down *EIF4A3* in HEK293T cells as well as in a glioblastoma cancer cell line (U87) that is sensitive to EIF4A3 perturbation (*22*), and we observed similar transcriptome-wide m^{6}A changes in each case (figs. S13 to S15). Altogether, our results indicate that spliceosomes widely suppress m^{6}A methylation through deposition of EJCs that protect proximal RNA from methylation.

EJCs regulate mRNA stability by suppressing m^{6}A methylation

m^{6}A is known to mainly accelerate mRNA degradation through the reader protein YTHDF2 (*23*, *24*). Accordingly, we observed a globally reduced mRNA half-life of hypermethylated transcripts (~90%) upon *EIF4A3* KD in HeLa cells (Fig. 4, A and B). Consistently, we observed generally increased YTHDF2 binding on hypermethylated mRNAs accompanied with the decreased mRNA half-life (Fig. 4C). *YTHDF2* KD could rescue accelerated degradation of YTHDF2 target transcripts upon *EIF4A3* KD (fig. S16). Further, the density of EJC loading on transcripts (estimated by the number of exons within CDS regions per 1 kb) correlated with transcriptome-wide mRNA stability (fig. S17). Higher EJC density on transcripts tended to correlate with reduced m^{6}A methylation and higher mRNA stability, and the strength of this correlation was diminished by *Mettl3* knockout in mouse embryonic stem cells (fig. S17, A and B).

We also found that *METTL3* depletion could generally reduce the expression level changes of hypermethylated genes upon EJC depletion in HeLa cells (supplementary text and fig. S18). This indicates that these EJC-dependent gene expression changes are at least in part mediated by m^{6}A methylation.

Although the vast majority of hypermethylated transcripts were destabilized by *EIF4A3* KD, a small subset of hypermethylated transcripts were stabilized (Fig. 4A). One example is p53, which mediates neurodevelopmental defects in mouse models of EJC haploinsufficiency (*25*). The *TP53* transcript was hypermethylated but also up-regulated upon *EIF4A3* KD. Mechanistically, we observed increased binding to *TP53* mRNA by IGF2BP proteins, which are known to stabilize methylated transcripts (supplementary text and fig. S19). Although the predominant effect of EJC-mediated m^{6}A suppression is to stabilize mRNAs by preventing YTHDF2-mediated decay, in a minority of instances, hypermethylated transcripts can be stabilized by other mechanisms, such as binding by IGF2BPs (*26*).

Consistent with a general role for m^{6}A in promoting translation (*12*, *27*), *EIF4A3* KD led to slightly increased translation efficiency of hypermethylated transcripts, with more highly hypermethylated transcripts exhibiting greater increases in translation efficiency (fig. S20). However, the effect was modest relative to the effects observed on mRNA stability.

Differential m^{6}A methylation across tissues and species through EJC suppression

Our model suggests that the cellular EJC levels may affect global mRNA m^{6}A deposition in different tissues. We observed a negative correlation between *EIF4A3* expression level and global mRNA m^{6}A modification level in 25 different human tissues with available transcriptome-wide m^{6}A profiles (fig. S21A) (*28*). We examined the top 10% of genes with the strongest correlations and found that most of them (>70%) exhibited a negative correlation between m^{6}A and *EIF4A3* levels in different tissues. Further, m^{6}A levels of these genes also negatively correlated with their transcript abundances (fig. S21B). Similar trends were also observed in mouse tissues (fig. S21C). These results further support m^{6}A suppression by EJCs and subsequent mRNA stability regulation in mammalian tissues.

Notably, we observed the lowest *EIF4A3* expressions in brain tissues, which exhibited the highest overall mRNA m^{6}A levels (fig. S21A). We further compared the methylome of the human cerebellum (lowest *EIF4A3* level and highest overall mRNA m^{6}A) with that of the heart (higher *EIF4A3* level and lower overall mRNA m^{6}A). Regions that are hypermethylated in the cerebellum (compared with the heart) reside within short internal exons (fig. S21D), which suggests reduced m^{6}A suppression because of low *EIF4A3* expression in the cerebellum. This association between high m^{6}A level and low *EIF4A3* expression in the cerebellum was attenuated upon depletion of METTL3 (fig. S21C). These observations further indicate that the widespread suppression by EJCs contributes to differential m^{6}A deposition in tissues. We also found that a subset of EJC-suppressed m^{6}A sites physiologically escape suppression in certain tissues through methylation of alternative transcript isoforms. These isoforms contain longer exons and thus altered EJC positioning; methylation of these isoforms generates tissue-specific m^{6}A patterns (supplementary text and fig. S22).

The effect of exon-intron architecture on mRNA stability may have coevolved with YTHDF2 in vertebrates. The strong correlation between EJC loading, represented by the number of exons, and mRNA level across tissues is maintained across humans, mice, and zebrafish but not in flies and worms, which lack YTHDF2 orthologs (supplementary text and fig. S23).

EJCs and peripheral EJC factor RNPS1 protect exon junction–proximal RNA regions from aberrant mRNA processing

We did not observe interactions between the methyltransferase complex and EJC complexes (fig. S24), which suggests that steric hindrance from EJCs, rather than a specific inhibitory interaction, accounts for methylation suppression. Nuclear EJCs bound with the peripheral EJC factor RNPS1 multimerize and associate with a wide variety of SR and SR-like proteins to package and compact mRNA into higher-order, megadalton-scale messenger ribonucleoproteins

(mRNPs) that ensheathe proximal RNA well beyond the canonical EJC deposition sites (*18*, *29*, *30*). Tens to hundreds of nucleotides of proximal RNA could be protected by this megacomplex from nuclease digestion resulting from this packaging (*18*, *31*). To examine whether the mRNA packaging function of the EJCs mediates suppression of proximal methylation, we isolated EJCs and EJC-bound RNA from cellular extracts, digested away physically accessible RNA with in vitro nuclease treatment, and then measured m^6A levels on the EJC-protected RNA footprints (fig. S25, A and B). EJC-protected footprints were strongly depleted of m^6A, indicating that these inaccessible RNA regions are largely protected from m^6A deposition within cells (fig. S25C). EJCs also protected these footprints from in vitro methylation by recombinant METTL3-METTL14 (fig. S25D). This was not because of general inhibition of methyltransferase activity or lack of methylatable sites on the EJC footprints because free, unmethylated RNA spiked into the methylation reaction as well as deproteinized footprints were both robustly methylated (fig. S25, D and E). Therefore, EJCs suppress local m^6A deposition by packaging proximal RNA.

We next investigated whether the peripheral EJC factor RNPS1, which associates with high–molecular weight EJCs in these highly packaged mRNP structures (*29*), plays a role. *RNPS1* KD led to substantial transcript m^6A hypermethylation within average-length internal exons and CDS regions (Fig. 5, A to C, and fig. S26, A to C). We detected fewer hypermethylated regions overall compared with depletion of the core EJC factors; however, we did observe high overlap (45%) between si*RNPS1* hypermethylated regions and si*EIF4A3* and si*RBM8A* hypermethylated regions (Fig. 5C and fig. S26C). By contrast, depletion of UPF1, a central nonsense-mediated decay (NMD) factor that interacts with the EJC in the cytoplasm, did not result in m^6A changes similar to those of the core EJCs (fig. S27).

The ability of EJCs to protect proximal RNA regions from methylation resembles the recently characterized EJC- and RNPS1-mediated suppression of proximal aberrant splice sites and recursive splicing (*16*, *17*). Transcriptome-wide, EJC-suppressed splice sites significantly colocalize with EJC-suppressed m^6A sites (supplementary text; Fig. 4, C to E; fig. S26, D to F; and table S2). Altogether, these results suggest that RNPS1-associated EJCs suppress both local cellular m^6A methylation and splicing through packaging of proximal RNA and point to exon architecture as an important determinant of local RNA accessibility to regulatory machineries. Additionally, beyond components of the m^6A methyltransferase complex, a number of other RBPs also exhibit preferential binding at long internal exons, which suggests that EJCs may regulate mRNA accessibility to a broader range of mRNA regulators in addition to the splicing and m^6A methylation machineries through their mRNA packaging function (supplementary text and fig. S28).

Discussion

Previously identified m^6A effector proteins fall broadly into three categories according to their activities: writers, which catalyze m^6A methylation; readers, which preferentially bind m^6A; and erasers, which reverse m^6A methylation. In this work, we establish the EJCs as a member of a previously unidentified class of m^6A regulators—suppressors, which broadly suppress the deposition of m^6A (fig. S28). EJCs appear to be a major regulator of m^6A deposition that mediate multiple key aspects of global m^6A epitranscriptome specificity, including enrichment of m^6A in long internal exons, depletion of m^6A in short exons, enrichment of m^6A in last exons near stop codons, and methylation selectivity for transcripts that have long internal exons. This mechanism may also explain the high abundance of m^6A on certain noncoding RNAs, such as LINE-1 elements that are generally unspliced and thus not bound by EJCs (*32*, *33*). Further, our systematic analysis of m^6A determinants using MPm^6A may suggest the existence of additional m^6A suppressing pathways, including m^6A suppression within the CDS because *EIF4A3* KD does not appear to completely restore methylation to unspliced levels (supplementary text, fig. S8H, and fig. S29).

Our results point to exon length within transcripts as a functionally relevant element for posttranscriptional gene expression regulation. Mammalian EJCs stably bind most pretranslational mRNAs in the transcriptome at closely spaced intervals. Long internal exons and terminal exons, which usually encode UTRs, are notably free of EJCs. This widespread binding, in conjunction with their mRNA packaging function, appears to uniquely position EJCs to broadly determine mRNA accessibility to regulatory machineries, such as the m^6A methylation and splicing machineries (fig. S30). Our work has relevance for the use of cDNA expression constructs in research studies and gene therapies because the loss of endogenous mRNA exon architecture and EJC protection may result in m^6A hypermethylation (fig. S11), which could modulate gene expression outcome. Finally, our study also suggests that exon length and architecture coevolved with mRNA processing steps as an additional regulatory layer of gene expression.

REFERENCES AND NOTES

1. M. Frye, B. T. Harada, M. Behm, C. He, *Science* **361**, 1346–1349 (2018).
2. W. V. Gilbert, T. A. Bell, C. Schaening, *Science* **352**, 1408–1412 (2016).
3. I. A. Roundtree, M. E. Evans, T. Pan, C. He, *Cell* **169**, 1187–1200 (2017).
4. P. C. He, C. He, *EMBO J.* **40**, e105977 (2021).
5. D. Dominissini *et al.*, *Nature* **485**, 201–206 (2012).
6. S. Ke *et al.*, *Genes Dev.* **29**, 2037–2053 (2015).
7. I. Barbieri *et al.*, *Nature* **552**, 126–131 (2017).
8. A. Bertero *et al.*, *Nature* **555**, 256–259 (2018).
9. L. Fish *et al.*, *Mol. Cell* **75**, 967–981.e9 (2019).
10. B. Slobodin *et al.*, *Cell* **169**, 326–337.e12 (2017).
11. H. Huang *et al.*, *Nature* **567**, 414–419 (2019).
12. Z. Zhang *et al.*, *Nat. Genet.* **52**, 939–949 (2020).
13. E. L. Van Nostrand *et al.*, *Nature* **583**, 711–719 (2020).
14. H. Le Hir, J. Saulière, Z. Wang, *Nat. Rev. Mol. Cell Biol.* **17**, 41–54 (2016).
15. V. Boehm, N. H. Gehring, *Trends Genet.* **32**, 724–735 (2016).
16. V. Boehm *et al.*, *Mol. Cell* **72**, 482–495.e7 (2018).
17. L. Blazquez *et al.*, *Mol. Cell* **72**, 496–509.e9 (2018).
18. G. Singh *et al.*, *Cell* **151**, 750–764 (2012).
19. W.-C. Lee *et al.*, *Plant Cell* **32**, 904–922 (2020).
20. T. Ø. Tange, T. Shibuya, M. S. Jurica, M. J. Moore, *RNA* **11**, 1869–1883 (2005).
21. Y. Xiao *et al.*, *Angew. Chem. Int. Ed.* **57**, 15995–16000 (2018).
22. W. Tang *et al.*, *Mol. Ther. Nucleic Acids* **19**, 905–921 (2020).
23. X. Wang *et al.*, *Nature* **505**, 117–120 (2014).
24. H. Du *et al.*, *Nat. Commun.* **7**, 12626 (2016).
25. H. Mao, J. J. McMahon, Y.-H. Tsai, Z. Wang, D. L. Silver, *PLOS Genet.* **12**, e1006282 (2016).
26. H. Huang *et al.*, *Nat. Cell Biol.* **20**, 285–295 (2018).
27. X. Wang *et al.*, *Cell* **161**, 1388–1399 (2015).
28. J. Liu *et al.*, *Mol. Cell* **77**, 426–440.e6 (2020).
29. J. W. Mabin *et al.*, *Cell Rep.* **25**, 2431–2446.e7 (2018).
30. M. Metkar *et al.*, *Mol. Cell* **72**, 715–726.e3 (2018).
31. H. Le Hir, E. Izaurralde, L. E. Maquat, M. J. Moore, *EMBO J.* **19**, 6860–6869 (2000).
32. J. Liu *et al.*, *Science* **367**, 580–586 (2020).
33. J. Wei *et al.*, *Science* **376**, 968–973 (2022).
34. P. C. He, X. Dou, Custom scripts associated with "Exon architecture controls mRNA m^6A suppression and gene expression", Zenodo (2023); https://doi.org/10.5281/zenodo.7541415.

ACKNOWLEDGMENTS

We thank G. Singh for FLAG-Magoh and FLAG-EIF4A3 cell lines and L. Wu for the BG plasmid. We thank T. Pan, E. Adams, M. Clark, M. Nobrega, J. Staley, and A. Joslin for comments and suggestions. We thank the Genomics Facility of the University of Chicago and the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility for assistance with sequencing. **Funding:** This study was supported by National Institutes of Health grants HG008935 (C.H.), T32 HD007009 (P.C.H.), and F32 CA221007 (B.T.H.). C.H. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** Conceptualization: P.C.H.; Methodology: P.C.H., J.W., X.D., B.T.H., Z.Zh., and C.H.; Formal analysis: P.C.H., X.D., X.Y., and R.L.; Investigation: P.C.H., J.W., X.D., B.T.H., Z.Zh., S.W., R.G., C.L., L.-S.Z., and Z.Zo.; Visualization: P.C.H., X.D., and J.W.; Funding acquisition: P.C.H., B.T.H., and C.H.; Project administration: C.H.; Supervision: M.C. and C.H.; Writing – original draft: P.C.H. and C.H.; Writing – review & editing: P.C.H., J.W., X.D., B.T.H., Z.Zh., S.W., R.L., M.C., and C.H. **Competing interests:** C.H. is a scientific founder and a scientific advisory board member of Accent Therapeutics, Inc.; Aferna Bio, Inc.; and AccuraDX, Inc. The other authors declare no competing interests. **Data and materials availability:** Raw and processed data can be found in the National Center for Biotechnology Information's Gene Expression Omnibus under accession no. GSE162199. Custom scripts are available on Zenodo (*34*). All other data are available in the manuscript or the supplementary materials.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj9090
Materials and Methods
Supplementary Text
Figs. S1 to S30
Tables S1 to S6
References (*35*–*79*)
MDAR Reproducibility Checklist

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

Submitted 9 June 2021; resubmitted 3 October 2022
Accepted 16 January 2023
10.1126/science.abj9090

SOLAR CELLS

Reducing nonradiative recombination in perovskite solar cells with a porous insulator contact

Wei Peng[1]†, Kaitian Mao[1]†, Fengchun Cai[1], Hongguang Meng[1], Zhengjie Zhu[1], Tieqiang Li[1], Shaojie Yuan[1], Zijian Xu[1], Xingyu Feng[1], Jiahang Xu[1], Michael D. McGehee[2], Jixian Xu[1,3]*

Inserting an ultrathin low-conductivity interlayer between the absorber and transport layer has emerged as an important strategy for reducing surface recombination in the best perovskite solar cells. However, a challenge with this approach is a trade-off between the open-circuit voltage (V_{oc}) and the fill factor (FF). Here, we overcame this challenge by introducing a thick (about 100 nanometers) insulator layer with random nanoscale openings. We performed drift-diffusion simulations for cells with this porous insulator contact (PIC) and realized it using a solution process by controlling the growth mode of alumina nanoplates. Leveraging a PIC with an approximately 25% reduced contact area, we achieved an efficiency of up to 25.5% (certified steady-state efficiency 24.7%) in p-i-n devices. The product of V_{oc} × FF was 87.9% of the Shockley-Queisser limit. The surface recombination velocity at the p-type contact was reduced from 64.2 to 9.2 centimeters per second. The bulk recombination lifetime was increased from 1.2 to 6.0 microseconds because of improvements in the perovskite crystallinity. The improved wettability of the perovskite precursor solution allowed us to demonstrate a 23.3% efficient 1-square-centimeter p-i-n cell. We demonstrate here its broad applicability for different p-type contacts and perovskite compositions.

Advanced device architecture and contact design, together with absorber quality improvement, have driven the rise of perovskite solar cells (*1*). Power-conversion efficiencies (PCEs) as high as 25.7% have been achieved for conventional n-i-p devices with the perovskite deposited on the electron transport layer (ETL) (*2*, *3*). The p-i-n cells [an inverted architecture built on the hole transport layer (HTL)] recently also reached certified PCEs of ~24% (*4*, *5*). Although PCEs of p-i-n cells lag behind those of their n-i-p counterparts, they are highly valued, especially for their demonstrated greater stability and higher efficiency in state-of-the-art tandem cells (*6*–*10*). The open-circuit-voltages (V_{oc}s) of these top-performing cells have reached >90% relative to the Shockley-Queisser (S-Q) limit for the corresponding band gap, exceeding the level of record silicon cells (~85%) (*1*, *11*). This high level of V_{oc} reflects the intensive efforts to reduce nonradiative recombination at photocarrier transport interfaces realized by the insertion of ultrathin low-conductivity passivation layers. Low-dimensional perovskites (*4*, *12*–*17*) and low-conductivity organic materials [such as polystyrene, poly(methyl methacrylate) polymer (PMMA), and organometallic molecules] (*3*, *5*, *18*, *19*) are the dominant passivation materials for current top-performing cells. Researchers have also explored dielectric inorganics such as Al_2O_3 and ZrO_2 (*20*–*24*), analogous to the tunnel oxide passivation in silicon cells (*25*–*28*). However, they are far less successful, primarily because of the extreme sensitivity of electric tunneling to the dielectric thickness. It has been reported that a subnanometer increase in the dielectric layer thickness can cause a vast fill factor (FF) reduction because of the series resistance loss (*20*, *26*). This trade-off problem is also generic for a two-dimensional (2D) perovskite and insulating organic passivation layers. This passivation-transport trade-off makes the simultaneous maximization of V_{oc} and FF a challenge (*1*, *18*, *19*). It also makes up-scaling fabrication a further challenge because it is not easy, especially using a solution process, to cover a rough large-area surface uniformly with an ultrathin passivation layer.

Here, we present a contact structure to overcome these two challenges in which the conventional ultrathin (~1 nm) passivation layer is replaced with a thick (~100 nm) dielectric mask that has random nanoscale openings (Fig. 1A). Photocarrier extraction is not compromised because photocarriers travel across the thick dielectric mask through the openings rather than by electrical tunneling. This contact design is called a "porous insulator contact" (PIC). We realized this design using a low-cost and scalable solution process that exhibits high tolerance to variations in dielectric thickness and opening size. Our numerical simulations and device experiment revealed that PIC reduces nonradiative recombination at perovskite's HTL contact through a combined effect of contact area reduction and passivation. Moreover, because PIC clusters are hydrophilic and rough and the preferred HTL materials are hydrophobic, surface wetting and perovskite film quality were improved. The surface recombination velocity (SRV) at the HTL contact was reduced from 64.2 to 9.2 cm/s. The bulk recombination lifetime was increased from 1.2 to 6.0 μs because of improvements in the perovskite crystallinity. Using a PIC that can reduce the contact area by ~25%, we boosted the PCE of a p-i-n device from ~23 to 25.5% (certified steady-state PCE of 24.7%, active area 0.06 cm^2), the highest ever reported for a p-i-n cell (table S1). Its product of V_{oc} × FF is 87.9% of the S-Q limit, rivaling that of the record n-i-p devices. PIC's advanced scalability also allows us to demonstrate 23.3% PCE for a 1-cm^2 p-i-n cell, among the highest for all perovskite cells with active area ≥1 cm^2 (table S2). We demonstrated its broad applicability for different HTLs and perovskite compositions (band gaps of 1.55, 1.57, and 1.65 eV).

PIC design

The PIC design is inspired by but different from the local contact in commercial silicon solar cells (*27*–*31*). For instance, the dielectric passivation layer (Al_2O_3/SiN_x or SiO_2/SiN_x stack with oxide thickness of ~10 nm) in a passivated emitter and rear contact silicon cell is opened by laser firing to localize the metal contact (or heavily doped contact) for carrier extraction (fig. S1A). The typical opening area ratio is ~1% to minimize contact area for recombination and to retain a high FF. The spacing between openings matches the photocarrier diffusion length in silicon on a scale of multiple hundreds of micrometers to one millimeter. Unfortunately, this technique cannot be directly transferred to perovskite solar cells because the photocarrier diffusion length is on a scale of hundreds of nanometers. This nanoscale requirement leads to a fabrication challenge for laser firing or opening etching. For n-i-p perovskite solar cells, nanoscale local contact has been attempted by a periodic TiO_2 nanorod array (pitch of ~300 nm) fabricated using high-resolution electron-beam lithography (*18*). The pinheads of TiO_2 nanorods puncture the ultrathin (<3 nm) PMMA insulating passivation layer to form the local contacts with the perovskite for electron extraction. However, the sophisticated electron-beam lithography microfabrication process makes it difficult for mass production. The nanorod local contact had enhanced FF up to 84.5%, but also showed a slightly reduced V_{oc} because of the increased perovskite/nanorod contact area for recombination (fig. S1B).

In light of these considerations, we sought a PIC design (Fig. 1A and fig. S1C) for p-i-n devices with combined features of (i) a reduced contact area at the perovskite/HTL interface; (ii) opening spacing on a scale of ~100 nm, shorter than the photocarrier diffusion length of perovskite; and (iii) a scalable fabrication

[1]Hefei National Research Center for Physical Sciences at the Microscale, CAS Key Laboratory of Materials for Energy Conversion, Department of Materials Science and Engineering, University of Science and Technology of China, Hefei 230026, China. [2]Chemical and Biological Engineering, University of Colorado, Boulder, CO 80309, USA. [3]Institute of Energy, Hefei Comprehensive National Science Center, Hefei 230051, China.
*Corresponding author. Email: jixianxu@ustc.edu.cn
†These authors contributed equally to this work

process. As illustrated in Fig. 1A, process *a* is the lateral transport of carriers that cannot electrically tunnel through the thick Al_2O_3. Process *b* represents the "through-the-opening" transport of carriers through the channels formed by the infilling perovskite. We selected Al_2O_3 to be the dielectric material for its easy accessibility, but in principle, other insulating materials such as SiO_2, ZrO_2, and insulating organic dielectrics could also be used.

To disentangle the PIC's structural effect and interface passivation effect, we developed a 2D numerical drift-diffusion model (for details, see the materials and methods, fig. S2, and tables S3 and S4) (*18*, *32*) to simulate devices with different dielectric coverages and different surface recombination velocities at the perovskite/HTL interface (tuning the defect density while the defect capture ability remained unchanged). The local insulator was modeled with thickness h = 100 nm to eliminate the electrical injection or tunneling through the dielectric layer. The local insulator's width d = 100 nm is shorter than the photocarriers' diffusion length in the perovskite to ensure their lateral transport between adjacent openings (*18*, *31*). The percentage reduction of the contact area is tuned through the dielectric coverage $[d/(d+s)]^2$, where s is the local opening width. We started by applying PIC in a control device (bare HTL) with a PCE of ~22%, a respectable value for p-i-n cells. As shown in Fig. 1B, we found that the PCE, V_{oc}, and FF all increased steadily with the dielectric coverage, i.e., reduced the perovskite/HTL contact area, which agrees well with previous reports (*26*–*28*, *33*). For PIC devices with different trap densities at the perovskite/HTL interface (fig. S3), we observed the same improvement trend of PCE, V_{oc}, and FF. This independence confirms the PIC's inherent structural effect on the V_{oc} and FF enhancement, providing a pathway to overcoming the passivation-transport trade-off encountered in conventional contacts. A PIC's coverage of 25% can deliver a respectable PCE percentage improvement of ~5% (Fig. 1, B and C, blue line). Increasing the PIC's coverage to 64% (Fig. 1C, orange dashed line) or coating this PIC on a passivated HTL with lower surface trap density (Fig. 1C, red line) can further increase the PCE percentage improvement to exceed 10%. These simulation results provide important guidance for achieving >25% efficient p-i-n solar cells. To verify the applicability of the PIC model, we also conducted simulations for varied insulator widths d (30, 100, and 200 nm) and higher coverage up to 81% (fig. S4), which delivered entirely consistent results. Beyond the perovskite/HTL contact area analyzed in the simulation, the interface passivation effect in real devices can also occur at the perovskite/dielectric contact interface. Therefore, it is rational to regard the above simulation as a conservative estimation. The PIC's reduced contact area, paired with the entire interface passivation, can gain even higher V_{oc} and FF.

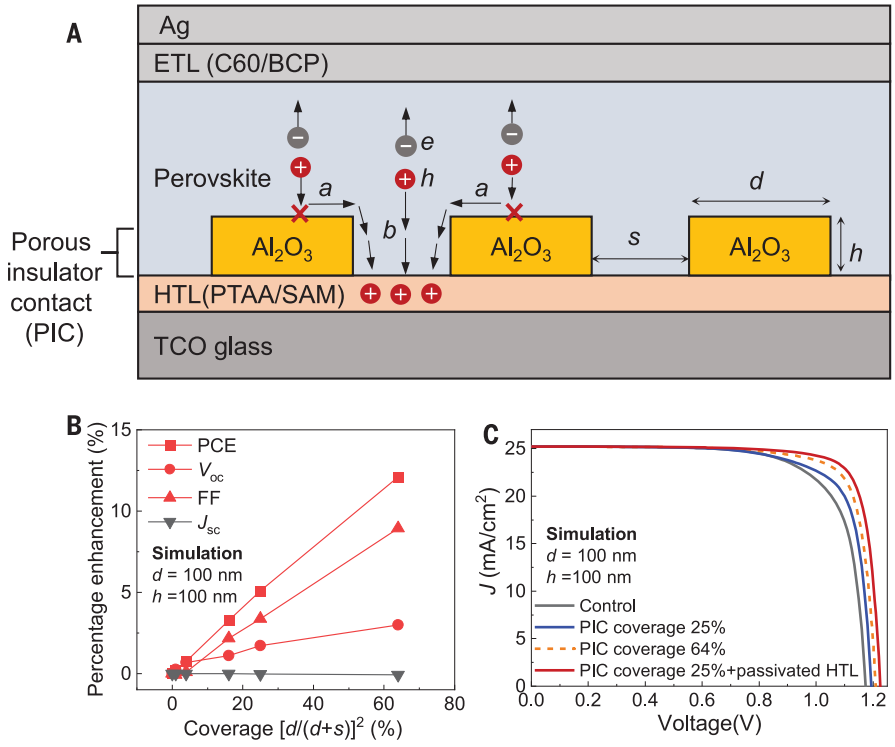


Fig. 1. Concept and simulation of a PIC for a p-i-n device. (**A**) Schematic description of the PIC device structure (not to scale). The parameters *h*, *d*, and *s* represent the local Al_2O_3 dielectric mask's height, width, and local opening width, respectively. Process *a* represents the lateral transport of carriers that cannot electrically tunnel through the thick Al_2O_3. Process *b* represents the "through-the-opening" transport of carriers through the channels formed by the perovskite infill. The insertion of PIC reduces the perovskite/HTL contact area and suppresses interface recombination. The PIC coverage fraction is defined as $[d/(d+s)]^2$. *e* and *h* represent the electron and hole charge carrier, respectively. (**B**) 2D drift-diffusion simulation of a PIC device (d = 100 nm, h = 100 nm) built on a conventional HTL (such as a PTAA polymer) showing consistently increased PCE, V_{oc}, and FF as a function of PIC coverage fraction. The *y* axis shows their percentage enhancement relative to the control (bare HTL without PIC). (**C**) Simulated *J*-*V* curves of the control device (without PIC, solid black line), the PIC devices with coverage of 25% (solid blue line) and 64% (orange dashed line) built on conventional HTL, and the PIC device with coverage of 25% built on a passivated HTL (reduced surface recombination velocity through reduced trap density at interface, solid red line). Simulation results [(B) and (C)] show that the PIC structural effect can be paired with HTL passivation to improve the photovoltaic performance.

PIC fabrication

We implemented the PIC contact design by judicious control over the growth mode of the Al_2O_3 nanoplate clusters during the solution process. The PIC principle requires a high contrast between the local dielectric mask and the adjacent opening area of HTL. Therefore, it is critical to avoid forming a continuous ultrathin dielectric layer that would create a passivation-transport trade-off. The former morphology desirable for PIC is analogous to the "isolated island" that results from Volmer-Weber (V-W) growth (Fig. 2E) (*34*), and the latter is analogous to the "layer-plus-island" morphology that arises from Stranski-Krastanov (S-K) growth (Fig. 2D) (*35*). We first tested the 30-nm Al_2O_3 nanoplate dispersion (d = 30 nm) that has been used for mesoporous scaffolds (*33*). By spin-casting the dispersion with different concentrations (0.7 to 4 mg ml^{-1}), we found that its strong preference was to form continuous layers, as shown in atomic force microscopy (AFM) morphology images (Fig. 2, F to H). The Al_2O_3 coverage fraction rapidly increased to almost 100%, even at a diluted concentration of 2 mg ml^{-1} (Fig. 2M). As per the S-K growth mechanism, we reason that this morphology occurs because the nanoplate-nanoplate interaction is weaker than the nanoplate-substrate interaction (*35*). The weak nanoplate-nanoplate interaction is suggested by the narrow peak of hydrodiameter size in dynamic light scattering (Fig. 2A) and the well-isolated dispersion manner shown in the transmission electron microscope (TEM) images

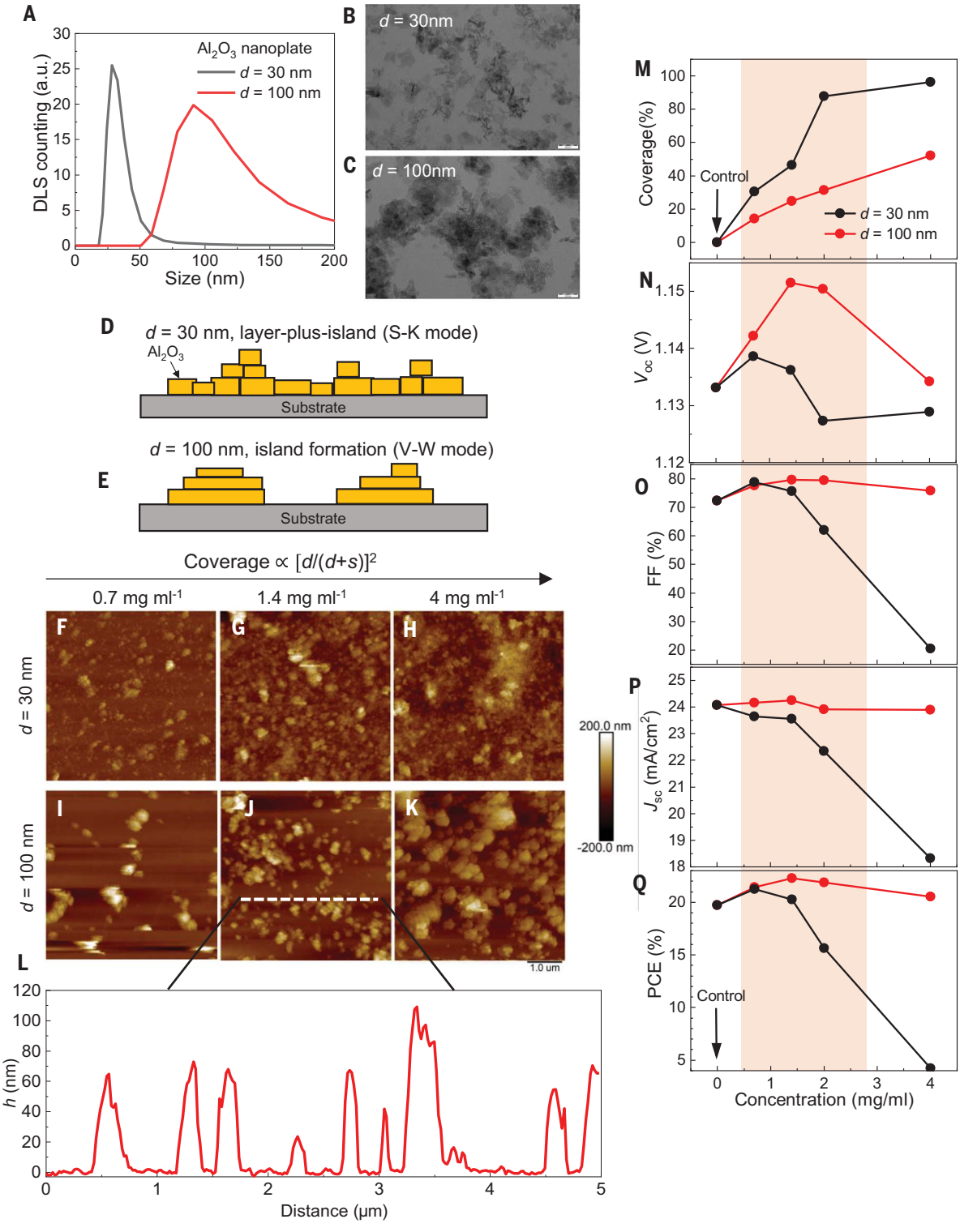


Fig. 2. Realization of PIC contact design through a solution process of Al_2O_3 nanoplate. (**A**) Dynamic light scattering (DLS) measurements of Al_2O_3 nanoplate in isopropanol (1 mg ml^{-1}) with two different hydrodiameter sizes (d = 30 nm and d = 100 nm). (**B** and **C**) TEM images of Al_2O_3 dispersions (~0.1 mg ml^{-1}) with nanoplate size d = 30 nm (B) and d = 100 nm (C). (**D**) Illustration of S-K growth of a "layer-plus-island" using Al_2O_3 nanoplate dispersion with d = 30 nm. (**E**) Schematic description of V-W growth of an "isolated island" using Al_2O_3 nanoplate dispersion with d = 100 nm. (**F** to **K**) AFM images of ITO/HTL/Al_2O_3 stacks using d = 30 nm [(F) to (H)] and d = 100 nm [(I) to (K)] Al_2O_3 nanoplates with a 0.7, 1.4, and 4 mg ml^{-1} dispersion concentration, respectively. (**L**) Line scan from an AFM image of Al_2O_3 islands (d = 100 nm, concentration = 1.4 mg ml^{-1}) after the V-W growth and coarsening procedure. The flat and clean baseline corresponds to the HTL surface at openings. (**M** to **Q**) Evolution of coverage fraction (M), V_{oc} (N), FF (O), J_{sc} (P), and PCE (Q) for two different Al_2O_3 sizes (d = 30 nm versus 100 nm) as a function of the dispersion concentration. The solar cell structure is ITO/PTAA/PIC/perovskite/C60/BCP/Ag. The HTL is the polymer PTAA.

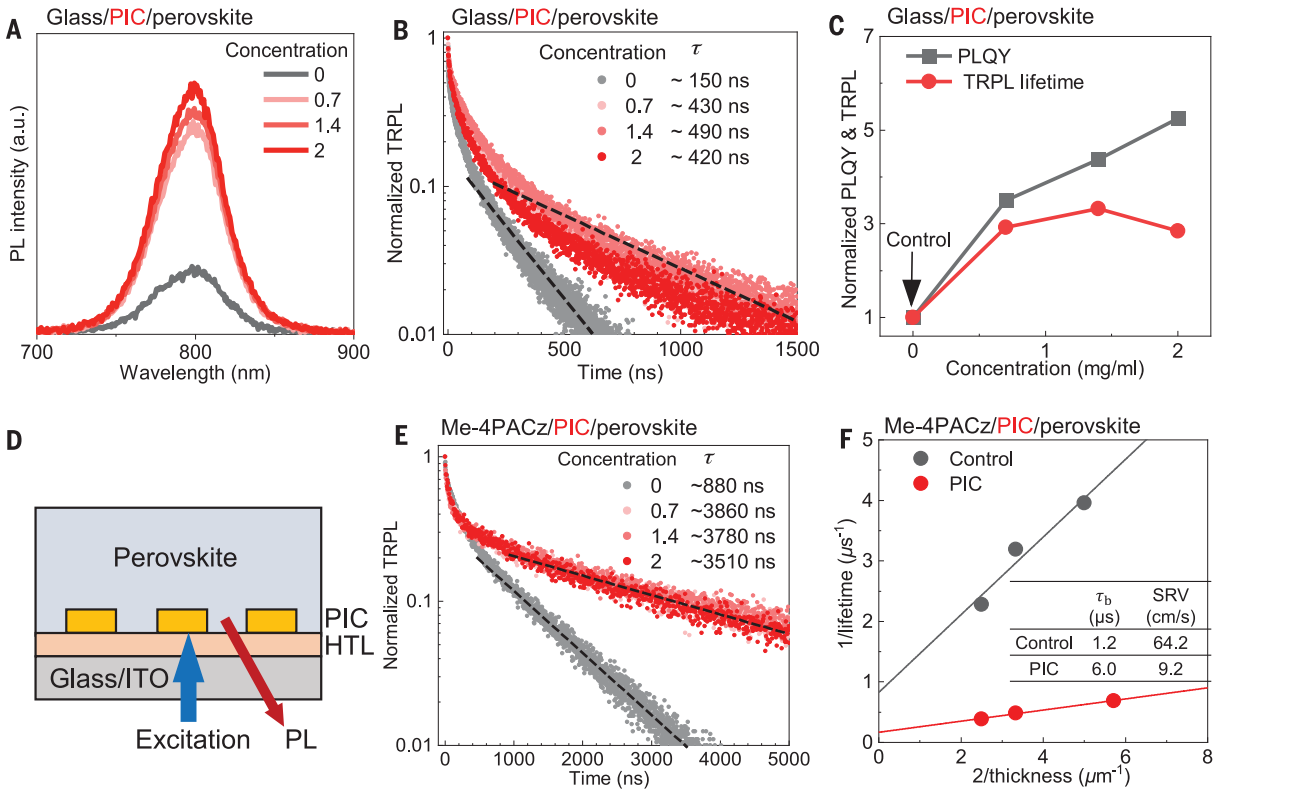


Fig. 3. Impact of PIC on reducing nonradiative recombination. (**A** and **B**) Steady-state photoluminescence (PL) spectra (A) and TRPL decays (B) of the glass/PIC/perovskite stacks. Al_2O_3 nanoplates (d = 100 nm) with concentrations of 0 to 2 mg ml^{-1} were used to form different coverage fractions. The slow monoexponential decay (τ, indicated by the black dashed line) represents the dominance of trap-assisted recombination in the bulk and the surface. The prolonged τ in PIC-containing samples indicates reduced nonradiative recombination. (**C**) Normalized PLQY and τ of glass/PIC/perovskite stacks (indicated as PIC sample) relative to the control sample (PIC concentration = 0). (**D**) Schematic description of the PL/TRPL experiment on the ITO/HTL/PIC/perovskite stacks. (**E**) TRPL decays of ITO/HTL/PIC/perovskite stacks, where the HTL is Me-4PACz. Insertion of PIC at the HTL/perovskite interface can prolong the τ (indicated by the black dashed line) and enhance the PLQY, highlighting the reduced nonradiative recombination and the potential for V_{oc} and FF enhancement. (**F**) Substantially improved perovskite bulk lifetime (τ_b) and SRV in the PIC sample relative to the control revealed by fitting the TRPL transients of varied thickness through a double-sided heterostructure model.

(Fig. 2B). By contrast, the 100-nm Al_2O_3 nanoplates (d = 100 nm) formed islands and clusters in the fashion of V-W growth, as shown in AFM morphology images for the concentration of 0.7 to 4 mg ml^{-1} (Fig. 2, I to K). The broad peak of the hydrodiameter size shown in dynamic light scattering (Fig. 2A) and the clustering manner shown in TEM images (Fig. 2C) of the 100-nm Al_2O_3 suggest a strong nanoplate-nanoplate interaction, providing the driving force for the V-W growth. With increased concentration (0.7 to 4 mg ml^{-1}), we observed growth and coarsening of clusters, resulting in the coverage fraction gradually increasing from ~15 to 50% (Fig. 2M). A representative height profile at a concentration of 1.4 mg ml^{-1} (Fig. 2L) reflects the growth and coarsening of Al_2O_3 islands, with average island width of 160 nm (range 64 to 270 nm), average island height of 64.5 nm (range 25 to 109 nm), and coverage of ~25% (fig. S5). Crucially, the clean and flat background in morphology images, corresponding to the exposed HTL surface, can be attained for concentrations ranging from 0.7 to 4 mg ml^{-1} (Fig. 2, I to K), providing the precondition for the successful realization of PIC.

Solar cell fabrication and testing

We then fabricated p-i-n devices to evaluate the solution-processed Al_2O_3 dielectric mask with different sizes (d = 30 versus 100 nm) and various concentrations (0.7 to 4 mg ml^{-1}) (Fig. 2, M to Q). Following the state-of-the-art p-i-n architecture (*5*, *7*), we used poly[bis(4-phenyl)(2,4,6-trimethylphenyl)amine (PTAA) as the HTL deposited on indium tin oxide–coated glass (ITO), and C60 as the ETL deposited on top of the perovskite surface (for details, see the materials and methods). The perovskite composition was $Cs_{0.05}(FA_{0.95}MA_{0.05})_{0.95}Pb(I_{0.95}Br_{0.05})_3$ (where FA is formamidinium and MA is methyl ammonium) with a band gap of 1.55 eV. For the Al_2O_3 mask with d = 30 nm, we observed slightly increased V_{oc} and FF only for the concentration of 0.7 mg ml^{-1}. With further increased concentration, the FF decreased vastly, in agreement with the trend of short-circuit current density (J_{sc}) loss and the formation of a continuous layer as observed in AFM images (Fig. 2, G and H). By contrast, we observed a significantly broader performance enhancement window for the Al_2O_3 mask with d = 100 nm. With increased concentration from 0.7 to 2mg ml^{-1}, the coverage fraction was consistently increased. V_{oc}, FF, and PCE, as well as the retained J_{sc}, are fully subject to the prediction of the ideal PIC model and consistent with the V-W growth of islands. For the concentration of 4 mg ml^{-1} with an even higher coverage fraction (~50%), the V_{oc} and FF remained higher than the control device but less efficient than the trend extrapolation from the ideal PIC model. To explain this deviation at high concentration, we propose that the increased proportion of oversized islands (i.e., those larger than the photocarrier diffusion length of the perovskite) reduces carrier collection. Given the relatively large size dispersion in d = 100 nm solution, it is also possible that the S-K growth of smaller nanoplates occurs in parallel with V-W growth of larger

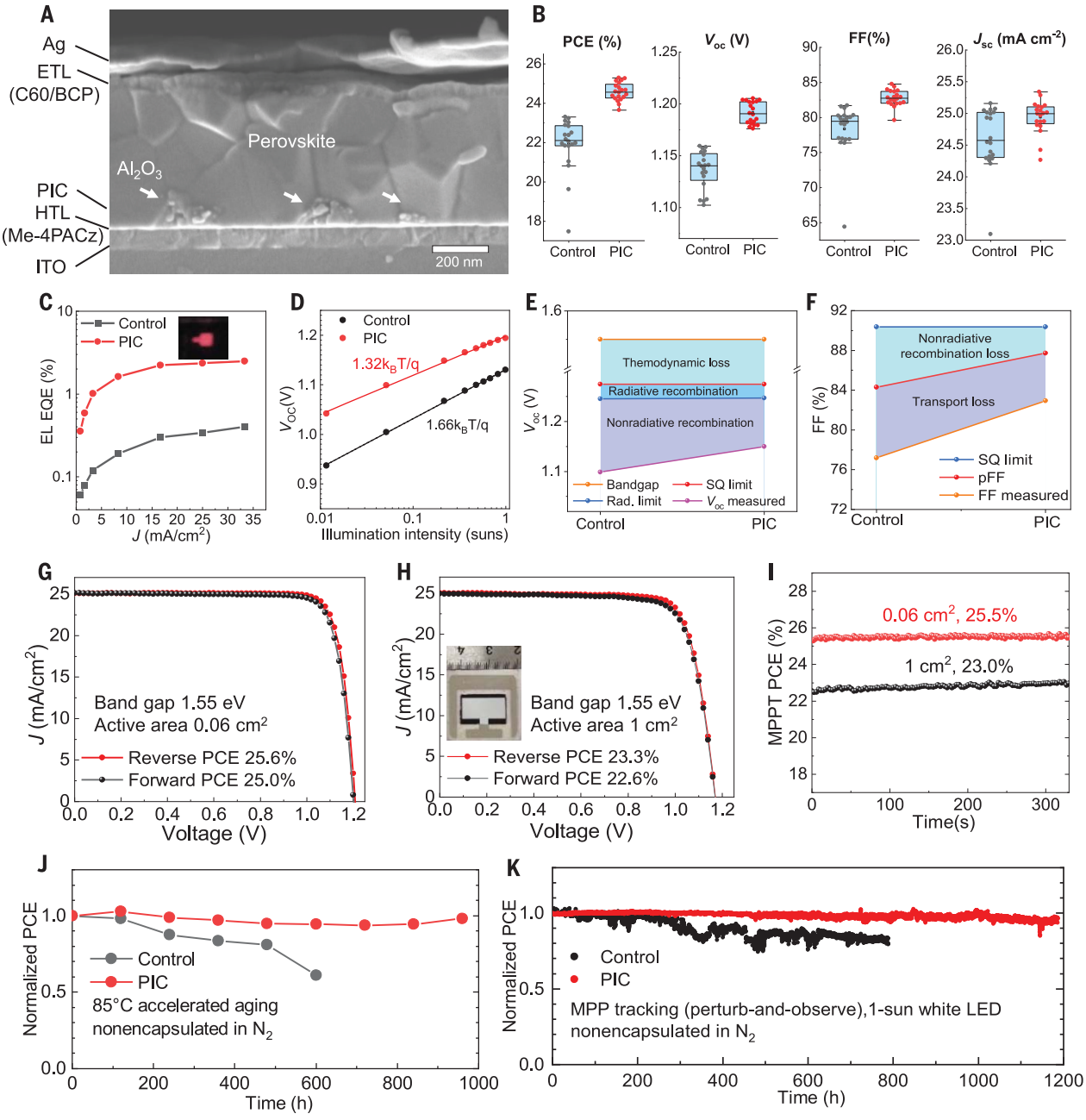


Fig. 4. PV characteristics of PIC-enhanced p-i-n devices. (**A**) Cross-sectional SEM image of the PIC device. The HTL is Me-4PACz coated with PIC (area coverage ~25%). Arrows indicate the Al_2O_3 islands of PIC. (**B**) Distribution of PCE, V_{oc}, FF, and J_{sc} of control and PIC devices. Data are shown for 20 substrates of each type made from the same perovskite precursor and processing batch. The control device uses Me-4PACz without the coating of PIC. (**C**) The external electroluminescence quantum efficiency (EL-EQE) measurements of control and PIC devices under different injection currents. The inset photo shows a PIC solar cell operated as an LED. (**D**) Suns-V_{oc} measurements to deduce the diode ideality factor. (**E** and **F**) Detailed V_{oc} loss analysis (E) and FF loss analysis (F) of the control and PIC devices. (**G** and **H**) *J-V* curves of the champion PIC devices with an active area of 0.06 cm^2 (G) and 1 cm^2 (H). The perovskite absorber's band gap is 1.55 eV. (**I**) Stabilized PCEs of champion PIC devices determined by MPPT for 300 s. (**J** and **K**) Long-term accelerated aging test of thermal stability (continuous heating at 85°C) (J) and continuous MPPT operational stability (1-sun equivalent white LED illumination) (K) of the control and PIC devices. The devices are nonencapsulated and tested in ambient N_2 to reveal their inherent properties.

nanoplates, and its effect becomes non-ignorable at the concentration of 4 mg ml^{-1}. We also note that as the coverage increases, the local opening width can become very narrow (such as <10 nm) and thus may induce non-ideal factors such as the failure of perovskite infill of narrow local openings or the dimensional-limited perovskite in the narrow channels exhibiting charge transport properties differing from the bulk (*36*, *37*). These non-ideal factors, from a positive viewpoint, provide valuable guidelines for further optimization of solution-processed PIC. The above PIC device model can also be more precise if we can find a way to include these non-ideal factors at high coverage.

With these studies, we henceforth used 100-nm Al_2O_3 with a concentration of 1.4 to 2 mg ml^{-1} for the optimized PIC solar cells.

For the concentration of 1.4 mg ml^{-1} with coverage of ~25%, we obtained a maximized PCE percentage improvement of ~12% (19.7 versus 22.2%) (Fig. 2, M to Q), in agreement with the above discussion that the simulator's prediction is conservative. Improvement of V_{oc} and FF could arise from factors other than just the PIC's reduction in contact area: the passivation of perovskite surface at HTL and the improved bulk quality as the perovskite grows on PIC-coated HTL. To evaluate these effects, we compared photoluminescence quantum yield (PLQY) and transient photoluminescence lifetime (TRPL) of glass/perovskite and glass/PIC/perovskite samples in the absence of HTL (Fig. 3, A to C). The sample was illuminated through the glass to be consistent with the solar cell's operational condition (*19*, *38*). PLQY enhancement can be observed for PIC samples of concentrations from 0.7 to 2 mg ml^{-1}. Consistent with PLQY enhancement, we observed a significantly slower TRPL decay in all PIC samples with concentrations of 0.7 to 2 mg ml^{-1}, suggesting suppressed nonradiative recombination in bulk and surfaces.

We then measured the PLQY and TPPL of perovskite coated on PTAA. Consistent with other reports (*6*, *19*), we observed that the biexponential TRPL decay of an ITO/PTAA/perovskite stack was much faster than that of a glass/perovskite stack (fig. S6) because of the additional nonradiative recombination pathways. As pointed out previously (*39*–*42*), the fast initial decay represents the combined process of charge extraction (quenching) and nonradiative recombination. The slow monoexponential decay lifetime [(τ), indicated by the black dashed line in fig. S6] is mainly attributed to trap-assisted recombination in the bulk and surfaces and is among the most widely used and trusted indicators to predict the achievable V_{oc} (for details, see the materials and methods) (*19*, *38*). Following this reasoning, we attributed the similar initial fast decays in the ITO/PTAA/perovskite stack and the ITO/PTAA/PIC/perovskite stack to the retained carrier extraction and the significantly increased τ with the use of PIC to the reduced nonradiative recombination (fig. S6). These TRPL changes agree with the V_{oc} and FF improvement observed in PIC-coated PTAA devices (Fig. 2, N to Q).

We further tested the PIC's applicability on different HTLs. We used a self-assembled monolayer of [4-(3,6-dimethyl-9H-carbazol-9-yl)butyl] phosphonic acid (Me-4PACz) because this has been demonstrated to increase the PL lifetime and enable superior interface passivation to the PTAA/perovskite stack (*6*). The ITO/Me-4PACz/perovskite stack exhibited a biexponential decay consisting of a fast decay and an impressively long τ of ~880 ns (Fig. 3, D and E), corresponding to an achievable V_{oc} of 1.18 V (for details, see the materials and methods) (*19*, *38*). By coating the Me-4PACz with PIC, the τ was further increased by a factor of ~4.4 (Fig. 3E). The τ of ~3800 ns in a Me-4PACz/PIC/perovskite stack corresponds to an achievable V_{oc} of 1.25 V, representing 97.3% of the S-Q limit of V_{oc} (1.285 V) for a 1.55-eV-band gap semiconductor. The increased τ in Me-4PACz/PIC/perovskite stacks was highly consistent for different PIC concentrations (0.7 to 2 mg ml^{-1}). At the same time, the TRPL and PLQY trends in Me-4PACz/PIC/perovskite stacks (fig. S7) agreed well with those observed in glass/PIC/perovskite stacks. This consistency suggests that the passivation effect of PIC and Me-4PACz can be successfully combined in the PIC device because the perovskite is in contact with both the HTL and PIC materials.

We also conducted the differential TRPL analysis to resolve the stages of charge extraction and recombination (*6*, *43*, *44*). The lifetime plateau increased from ~1 to ~4.4 μs upon the inclusion of PIC, reconfirming the substantially reduced nonradiative recombination losses in bulk and surfaces (fig. S8). The steep lifetime rise at early times was maintained in PIC samples, indicating that the local openings in PIC support the fast charge transfer.

To disentangle the bulk and surface recombination, we conducted SRV analysis by measuring TRPL as a function of perovskite thickness (Fig. 3F and fig. S9) following a double-heterostructure model (*43*, *45*–*48*) as follows:

$$\frac{1}{\tau_{eff}} = \frac{1}{\tau_b} + \frac{2SRV}{d} \quad (1)$$

where τ_{eff} is the effective lifetime of TRPL decay, τ_b the bulk lifetime, and d the perovskite thickness (for details, see the materials and methods). The inclusion of PIC at HTL substantially improved the bulk lifetime from 1.2 to 6.0 μs, likely because of the hydrophilic and rough surface of Al_2O_3 clusters that profoundly modify the perovskite's growth and bulk quality, as previously reported (*33*, *49*). This fact is evidenced by the markedly improved wetting and coverage of perovskite on the PIC-coated HTL (fig. S10). The enlargement of the perovskite's apparent grain size is not appreciable, as shown in the scanning electron microscope (SEM) top-view images and AFM morphology images in figs. S11 and S12. However, our grazing incidence wide-angle x-ray scattering (GIWAXS) data (fig. S13, A and B) showed that the PIC sample exhibited higher radially integrated intensities of perovskite (100) and (200) crystal planes relative to the control sample (fig. S13, C to E), consistent with the peak ratios observed in x-ray diffraction measurements (fig. S14). Moreover, the azimuthal integration of (100) and (200) planes showed that the PIC sample's diffraction signals were more oriented and concentrated at azimuth angles around 90° (fig. S13F), suggesting that the presence of PIC promoted its out-of-plane growth. Per previous reports (*50*–*53*), the enhancement and orientation promotion of the (100) plane have been widely linked with reduced defect density and improved carrier lifetime in bulk perovskite, in agreement with the above improvement shown in TRPL, PLQY, and SRV measurements. The SRV at the HTL interface was substantially reduced from 64.2 to 9.2 cm/s, highlighting the opportunity to approach the thermodynamic limit according to previous device simulations (*46*, *54*).

Motivated by this success, we fabricated the p-i-n solar cells of ITO/Me-4PACz/PIC/perovskite/C60/BCP/Ag. The Al_2O_3 isolated islands (coverage of ~25% and concentration of 1.4 mg ml^{-1}) formed on the Me-4PACz layer as shown in cross-sectional SEM images (Fig. 4A). We fabricated 20 solar cells with the PIC and 20 without it in the same batch. The PIC increased the average PCE from 22.1 to 24.5% (Fig. 4B), primarily caused by an increase in V_{oc} and FF. The PIC increased the average V_{oc} from 1.13 to 1.19 V, agreeing with the above prediction from TRPL measurements (1.18 versus 1.25 V). The additional V_{oc} loss in the solar cell relative to the prediction can be linked to the non-idealities of fabrication and additional loss at the ETL contact (*6*, *19*, *55*).

We conducted a quantitative analysis of the V_{oc} loss following the detailed balance theory (for details, see the materials and methods) (*56*, *57*). The external quantum efficiency of electroluminescence of the PIC device (fig. S15)

Table 1. PV parameters of PIC-enhanced p-i-n perovskite solar cells (band gap = 1.55 eV).

Area (cm^2)	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF (%)	PCE (%)	Stabilized PCE* (%)
0.06	1.208	25.08	84.37	25.56	25.50
1	1.172	25.05	79.48	23.30	23.04
0.058	1.203	25.02	82.73	24.90†	24.71†

*Maximum power point tracking for 300 s. †Certified.

increased by a factor of ~7 relative to the control at the injection level close to J_{sc} (Fig. 4C), which is consistent with the PLQY and SRV improvements. The nonradiative recombination loss was reduced by ~50 mV, representing the main contributor to the higher V_{oc} in PIC devices (Fig. 4E). By contrast, the improvement (~2 mV) in radiative recombination loss was secondary, as revealed by the high-sensitivity external quantum efficiency near the band edge (fig. S16).

To quantitatively analyze the enhanced FF in PIC devices (from the average of 78.5 to 82.3%) (Fig. 4B), we used the Suns-V_{oc} method, which involves measuring J–V curves as a function of light intensity (for details, see the materials and methods) (*58*, *59*). The deduced ideality factor was improved from 1.66 to 1.32 (Fig. 4D), in good agreement with nonradiative recombination suppression and FF enhancement. The FF enhancement in PIC devices has considerable contributions from reduced transport loss (4.78 versus 7.11%) and reduced nonradiative recombination loss (2.65 versus 6.05%) (Fig. 4F). The PIC enabled the pseudo-FF, which is the FF in the absence of transport loss, to be 87.7% (fig. S17), which is ~97.1% of the thermodynamic limit (S-Q limit). By contrast, the pseudo-FF for the bare Me-4PACz was only 84.3%. This FF analysis highlights how the PIC overcomes the V_{oc}-FF trade-off by forming a practical, through-the-opening transport through the perovskite infill.

The champion PIC device showed a PCE of 25.6% (V_{oc} = 1.208 V, J_{sc} = 25.08 mA cm^{-2}, and FF = 84.37%) with negligible hysteresis (Fig. 4G). The stabilized PCE was 25.5% after a maximum-power-point tracking for 300 s (Fig. 4I). This efficiency is among the highest ever recorded for perovskite single-junction cells (table S1). The measured V_{oc} and FF are 94 and 93.5% of their S-Q limit for a 1.55 eV solar cell and ~97% of their upper limits predicted from TRPL lifetime (1.25 V) and pseudo-FF (87.7%), respectively. As previously reported (*1*), the product V_{oc} × FF relative to the S-Q limit allows quantifying the V_{oc}-FF trade-off without the influence of different band gaps. Our PIC device can deliver a V_{oc} × FF of 87.9% of the S-Q limit, the highest for a p-i-n device to date and rivaling the level (87.3%) in the record n-i-p device (table S1 and fig. S18).

The certification of one PIC device measured by an independently accredited testing center (the Chinese National PV industry Measurement and Testing Center) gave a PCE of 24.9% (V_{oc} = 1.203 V, J_{sc} = 25.02 mA cm^{-2}, and FF = 82.73%) (fig. S19) and a stabilized PCE of 24.71% after maximum-power-point tracking (MPPT) for 300 s (fig. S20). This confirmed value is the highest certified efficiency among all p-i-n devices (table S1). The J_{sc} measured in our laboratory highly matched the certified J_{sc} within an error of <0.3% and also matched the J_{sc} (24.4 mA cm^{-2}) calculated by integrating the EQE spectrum (fig. S21).

From the device manufacturing perspective, the standard deviation in PCE was reduced from 1.37 to 0.45% by using the PIC (Fig. 4B). The improved homogeneity and yield of perovskite films deposited on the rough and hydrophilic PIC surface allowed us to realize enlarged cells with fewer shunting pathways. The PIC cells with an aperture area of 1 cm^2 reached a PCE of 23.3% (V_{oc} = 1.17 V, J_{sc} = 25.05 mA cm^{-2}, and FF = 79.48%) (Fig. 4H), among the highest values for all perovskite cells with active area ≥1 cm^2 (table S2).

Stability measurements

We examined the long-term stability of PIC devices under accelerated lifetime conditions of continuous heating and illumination following the ISOS consensus protocol (for details, see the materials and methods) (*60*). An SnO_2 layer (15 to 20 nm) made by atomic layer deposition was used to replace the BCP layer for the long-term stability test without a trade-off of PCE. We tested unencapsulated devices in an N_2 atmosphere to reveal their intrinsic properties and to avoid encapsulation modification. The PIC devices retained 98% of the initial PCE after accelerated aging at 85°C for 1000 hours, whereas the control device exhibited ~20% PCE loss (T_{80} lifetime) at ~400 hours (Fig. 4J). Moreover, the PIC devices exhibited minor degradation of <2% after 1000 hours of continuous operation at MPPT (perturb-and-observe tracking) under 1-sun illumination (chamber temperature ~35 to 40°C) (Fig. 4K). The control devices showed no degradation for ~200 hours and then a 20% loss of PCE at ~800 hours. We noted that our p-i-n devices exhibited no burn-in degradation but a slightly increased PCE during the initial stage, which may have been caused by light-induced defect healing of perovskite, as revealed in previous reports (*61*, *62*). This impressive operational stability in PIC devices is consistent with the observed improvement in film wetting, passivation, and carrier transport. It has been revealed that better passivation paired with faster carrier extraction at the perovskite/HTL interface can effectively improve stability because of less carrier accumulation and suppressed phase segregation (*6*, *63*).

Outlook

In light of the successful application of double-sided (both n- and p-type) local contacts in silicon cells, we attempted the PIC device's extension by additionally spin-coating Al_2O_3 on the perovskite surface but did not gain any further improvement in V_{oc} and FF (fig. S22A). We speculate that Al_2O_3 clustering and growth on the perovskite rough surface may differ from that on the HTL flat surface, and the local contact structure at the rear ETL differs from that of the front PIC (see the schematic diagrams in fig. S22, B to E), suggesting that the extension of PIC on the top surface requires appropriate modification. It is also worth noting that the established silicon solar cells' local contacts are typical of a coverage of >90%. Consistently, our simulations (fig. S4) showed that the PCE could be raised by another ~2% if a way could be found to fabricate a nearly ideal PIC structure with coverage increased from 25 to >80%.

The PIC concept can also be generalized to other perovskite compositions. In $FA_{0.95}MA_{0.05}Pb(I_{0.92}Br_{0.08})_3$, a Cs-free composition with a band gap of 1.57 eV, we improved the p-i-n cells' PCE from 22.90 to 25.13% by using PIC, featured with high V_{oc} of 1.2 V and FF of 84.14% (fig. S23). Moreover, we also observed considerable PCE, V_{oc}, and FF improvement in 1.65-eV perovskite cells, which are of high interest for tandem applications (fig. S24). These two demonstrations were not optimized. For instance, the PIC parameters, such as the opening spacing, should be tuned to match the photocarrier diffusion length of a specific perovskite.

REFERENCES AND NOTES

1. A. Polman, M. Knight, E. C. Garnett, B. Ehrler, W. C. Sinke, *Science* **352**, aad4424 (2016).
2. National Renewable Energy Laboratory, "Best research-cell efficiency chart" (NREL, 2023); https://www.nrel.gov/pv/cell-efficiency.html.
3. H. Min *et al.*, *Nature* **598**, 444–450 (2021).
4. H. Chen *et al.*, *Nat. Photonics* **16**, 352–358 (2022).
5. Z. Li *et al.*, *Science* **376**, 416–420 (2022).
6. A. Al-Ashouri *et al.*, *Science* **370**, 1300–1309 (2020).
7. J. Xu *et al.*, *Science* **367**, 1097–1104 (2020).
8. R. Lin *et al.*, *Nature* **603**, 73–78 (2022).
9. K. A. Bush *et al.*, *Nat. Energy* **2**, 17009 (2017).
10. Q. Han *et al.*, *Science* **361**, 904–908 (2018).
11. K. Yoshikawa *et al.*, *Nat. Energy* **2**, 17032 (2017).
12. G. Yang *et al.*, *Nat. Photonics* **15**, 681–689 (2021).
13. J. J. Yoo *et al.*, *Nature* **590**, 587–593 (2021).
14. J. J. Yoo *et al.*, *Energy Environ. Sci.* **12**, 2192–2199 (2019).
15. Q. Jiang *et al.*, *Nat. Photonics* **13**, 2558–2565 (2019).
16. D. Luo *et al.*, *Science* **360**, 1442–1446 (2018).
17. R. Azmi *et al.*, *Science* **376**, 73–77 (2022).
18. J. Peng *et al.*, *Science* **371**, 390–395 (2021).
19. M. Stolterfoht *et al.*, *Nat. Energy* **3**, 847–854 (2018).
20. D. Koushik *et al.*, *Energy Environ. Sci.* **10**, 91–100 (2017).
21. C. Das *et al.*, *Cell Rep. Phys. Sci.* **1**, 100112 (2020).
22. L. Zhang *et al.*, *J. Mater. Res.* **36**, 1855–1865 (2021).
23. S. K. Yadav, *Mater. Chem. Phys.* **268**, 124709 (2021).
24. J. Zhang, Q. Guo, Y. Zhao, J. Duan, Q. Tang, *Chem. Commun.* **58**, 13891–13894 (2022).
25. M. J. Kerr, A. Cuevas, *Semicond. Sci. Technol.* **17**, 35–38 (2001).
26. J. Shewchun, R. Singh, M. A. Green, *J. Appl. Phys.* **48**, 765–770 (1977).
27. T. G. Allen, J. Bullock, X. Yang, A. Javey, S. De Wolf, *Nat. Energy* **4**, 914–928 (2019).
28. F. Feldmann, M. Bivour, C. Reichel, M. Hermle, S. W. Glunz, *Sol. Energy Mater. Sol. Cells* **120**, 270–274 (2014).
29. E. Franklin *et al.*, *Prog. Photovolt. Res. Appl.* **24**, 411–427 (2016).
30. A. W. Blakers, A. Wang, A. M. Milne, J. Zhao, M. A. Green, *Appl. Phys. Lett.* **55**, 1363–1365 (1989).
31. M. A. Green, *Prog. Photovolt. Res. Appl.* **17**, 183–189 (2009).
32. D. Walter *et al.*, *J. Phys. Chem. C* **122**, 11270–11281 (2018).
33. M. M. Lee, J. Teuscher, T. Miyasaka, T. N. Murakami, H. J. Snaith, *Science* **338**, 643–647 (2012).
34. K. Oura, V. G. Lifšic, A. A. Saranin, A. V. Zotov, M. Katayama, *Surface Science: An Introduction* (Springer, 2003).
35. A. Pimpinelli, J. Villain, *Physics of Crystal Growth* (Cambridge Univ. Press, 1998).

36. Z. Meng, D. Guo, J. Yu, K. Fan, *Appl. Surf. Sci.* **430**, 632–638 (2018).
37. Y. Yang *et al*., *J. Mater. Chem. A Mater. Energy Sustain.* **3**, 9103–9107 (2015).
38. W. Tress, *Adv. Energy Mater.* **7**, 1602358 (2017).
39. L. Bertoluzzi *et al*., *Joule* **4**, 109–127 (2020).
40. S. A. L. Weber *et al*., *Energy Environ. Sci.* **11**, 2404–2413 (2018).
41. R. A. Belisle *et al*., *Energy Environ. Sci.* **10**, 192–204 (2017).
42. A. R. Bowring, L. Bertoluzzi, B. C. O'Regan, M. D. McGehee, *Adv. Energy Mater.* **8**, 1702365 (2018).
43. T. Kirchartz, J. A. Marquez, M. Stolterfoht, T. Unold, *Adv. Energy Mater.* **10**, 1904134 (2020).
44. B. Krogmeier, F. Staub, D. Grabowski, U. Rau, T. Kirchartz, *Sustain. Energy Fuels* **2**, 1027–1034 (2018).
45. J. Tong *et al*., *Nat. Energy* **7**, 642–651 (2022).
46. J. Wang *et al*., *ACS Energy Lett.* **4**, 222–227 (2018).
47. A. B. Sproul, *J. Appl. Phys.* **76**, 2851–2854 (1994).
48. D. W. deQuilettes *et al*., *ACS Energy Lett.* **1**, 438–444 (2016).
49. J. You *et al*., *J. Energy Chem.* **38**, 192–198 (2019).
50. B. Park *et al*., *Nat. Energy* **6**, 419–428 (2021).
51. C. Luo *et al*., *Joule* **6**, 240–257 (2022).
52. Y. Zhang *et al*., *Adv. Mater.* **34**, e2107420 (2022).
53. X. Li *et al*., *Energy Environ. Sci.* **15**, 4813–4822 (2022).
54. S. Jariwala *et al*., *Chem. Mater.* **33**, 5035–5044 (2021).
55. P. Caprioglio *et al*., *Adv. Energy Mater.* **9**, 1901631 (2019).
56. J. Liu *et al*., *Nat. Energy* **1**, 16089 (2016).
57. F. Li *et al*., *J. Am. Chem. Soc.* **142**, 20134–20142 (2020).
58. N. Mundhaas *et al*., *Sol. RRL* **3**, 1800378 (2019).
59. Q. Cao *et al*., *Sci. Adv.* **7**, eabg0633 (2021).
60. M. V. Khenkin *et al*., *Nat. Energy* **5**, 35–49 (2020).
61. M. Saliba *et al*., *Science* **354**, 206–209 (2016).
62. E. Mosconi, D. Meggiolaro, H. J. Snaith, S. D. Stranks, F. De Angelis, *Energy Environ. Sci.* **9**, 3180–3187 (2016).
63. Y. Lin *et al*., *Nat. Commun.* **9**, 4981 (2018).

ACKNOWLEDGMENTS

We thank Z. Su, X. Gao, and T. Yang for supporting GIWAXS measurements and the Shanghai Synchrotron Radiation Facility (SSRF) for use of the BL14B1 beamline (project No.2020-SSRF-PT-014606). **Funding:** This work was supported by the National Natural Science Foundation of China (grant 52172246); the National Key R&D Program of China (grant 2021YFF0501900); the Institute of Energy, Hefei Comprehensive National Science Center (grant 21KZS212); the Tencent Foundation through an XPLORER PRIZE to J.X.; and Fundamental Research Funds for the Central Universities (grant WK2060140026). **Author contributions:** Jixian Xu conceived and directed this project. M.D.M., W.P., and Jixian Xu reviewed the experiment, analyzed the data, and wrote the manuscript. All authors discussed the results and commented on the manuscript. W.P. fabricated and characterized the devices. K.M. helped to fabricate and characterize the devices. F.C. developed the numerical model of devices. Z.Z., Jiahang Xu, and Z.X. supported the device fabrication. H.M. and T.L. helped to conduct the stability test. K.M. and S.Y. helped to conduct the V_{oc} loss and FF loss characterization. X.F. contributed to the SEM characterization. T.L. helped to conduct the x-ray diffraction characterization. **Competing interests:** M.D.M. is an adviser to Swift Solar. University of Science and Technology of China has filed patents that cover the insulator contact and device presented in this work. The authors declare no competing interests. **Data and materials availability:** All data are available in the main manuscript or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade3126
Materials and Methods
Figs. S1 to S24
Tables S1 to S4
References (*64*–*85*)
Simulation Model File

Submitted 8 August 2022; accepted 23 January 2023
10.1126/science.ade3126

SOLAR CELLS

Rational design of Lewis base molecules for stable and efficient inverted perovskite solar cells

Chongwen Li[1], Xiaoming Wang[1], Enbing Bi[1], Fangyuan Jiang[2], So Min Park[3], You Li[1], Lei Chen[1], Zaiwei Wang[3], Lewei Zeng[3], Hao Chen[3], Yanjiang Liu[3], Corey R. Grice[1,4], Abasi Abudulimu[1], Jaehoon Chung[1], Yeming Xian[1], Tao Zhu[1], Huagui Lai[5], Bin Chen[3,6], Randy J. Ellingson[1], Fan Fu[5], David S. Ginger[2], Zhaoning Song[1], Edward H. Sargent[3,6,7], Yanfa Yan[1]*

Lewis base molecules that bind undercoordinated lead atoms at interfaces and grain boundaries (GBs) are known to enhance the durability of metal halide perovskite solar cells (PSCs). Using density functional theory calculations, we found that phosphine-containing molecules have the strongest binding energy among members of a library of Lewis base molecules studied herein. Experimentally, we found that the best inverted PSC treated with 1,3-bis(diphenylphosphino)propane (DPPP), a diphosphine Lewis base that passivates, binds, and bridges interfaces and GBs, retained a power conversion efficiency (PCE) slightly higher than its initial PCE of ~23% after continuous operation under simulated AM1.5 illumination at the maximum power point and at ~40°C for >3500 hours. DPPP-treated devices showed a similar increase in PCE after being kept under open-circuit conditions at 85°C for >1500 hours.

Given their high power conversion efficiencies (PCEs), metal halide perovskite solar cells (PSCs) offer a route to lowering the cost of solar electricity (*1*–*4*). However, durability remains a major hurdle along the path to technological relevance (*5*–*7*) and must be assessed through accelerated degradation tests (*8*). Damp heat testing at 85°C in the dark at 85% relative humidity (RH), a test standard for crystalline silicon (Si) and thin-film photovoltaic (PV) modules, has been adopted for accelerating the durability test of PSCs (*9*–*11*). These tests are typically used to evaluate packaging rather than PV material durability. PSCs can also show degradation under photoexcited conditions (*12*), and especially under open-circuit (OC) conditions (*13*), that are more acute than one sees in standardized silicon tests. Mechanistically, such findings are often attributed to ion migration (*14*, *15*) and charge accumulation at interfaces (*9*, *16*, *17*).

We studied the operating stability at 85°C under simulated 1-sun illumination (where 1 sun is defined as the standard illumination at AM1.5, or 1 kW m^{-2}) and OC conditions—important test conditions under which PSCs have been studied to date only to a limited degree (*18*–*20*). Light- and heat-induced degradation in PSCs is related to point defects formed at interfaces and grain boundaries (GBs) (*14*, *21*). Moisture-induced degradation is curtailed using encapsulation (*22*), whereas the passivation of defects at interfaces and GBs within the perovskite film is required to improve the PCE and intrinsic durability of PSCs (*11*, *23*–*26*). The use of phosphorus (P)–, nitrogen (N)–, sulfur (S)– and oxygen (O)–containing Lewis base molecules to form coordinate covalent bonds (dative bonds) that donate electrons to undercoordinated Pb atoms at interfaces and GBs has shown particular promise for increasing PSC durability (*27*–*29*).

Using density functional theory (DFT), we saw evidence that P-containing Lewis base molecules showed the strongest binding with uncoordinated Pb atoms. We thus pursued diphosphine-containing molecules, reasoning that these would provide additional binding and bridging at interfaces and GBs. For our experimental studies, we selected 1,3-bis(diphenylphosphino)propane (DPPP), a diphosphine Lewis base. We found that treating perovskites with a small amount of DPPP improves PCE and durability: Inverted (p-i-n) PSCs after DPPP treatment showed a champion PCE of 24.5%. A DPPP-treated PSC with an initial PCE of ~23% stabilized at ~23.5% after maximum power point tracking (MPPT) under continuous simulated AM1.5 illumination at ~40°C for >3500 hours. DPPP-stabilized PSCs showed no PCE degradation after being kept at OC and 85°C conditions for >1500 hours.

Bonding interactions of Lewis bases

The P, N, S, and O atoms in Lewis base molecules donate electrons to the Lewis acid sites in perovskites, such as the undercoordinated Pb^{2+} at perovskite surfaces, in order to form coordinate covalent bonds. In general, the Lewis basicity, which is inversely proportional to electronegativity, is expected to determine the binding energy and the stabilization of interfaces and GBs. We compared binding energies of the prototypical Lewis bases trimethylphosphine (TMP), trimethylamine (TMA), dimethyl sulfide (DMS), and dimethyl ether (DME), with sp^3 hybridization to the surface of formamidinium lead iodide ($FAPbI_3$) through DFT calculations (Fig. 1A). The calculated binding strength followed the order P > N > S > O, indicating that the electronegativity rule did not strictly apply—a finding we attribute to the remaining lone-pair electrons in the case of S and O after binding with perovskites. We also compared frequently reported Lewis base molecules with the sp^2 oxygen in the carbonyl groups and took acetone and methyl acetate (MeOAc) as examples in Fig. 1A (*26*, *30*). The binding energies were similar to that of DME.

Lewis base molecules have mostly been used to passivate uncoordinated Pb atoms (*24*, *31*–*33*). A Lewis base molecule with two electron-donating atoms can potentially bind and bridge interfaces and GBs, offering the potential to enhance the adhesion and strengthen the mechanical toughness of PSCs and provide additional benefits related to the durability of PSCs. For this reason, we selected DPPP, a diphosphine Lewis base molecule, in seeking to stabilize and passivate PSCs. As shown in Fig. 1B and fig. S1, a DPPP molecule has two P atoms with sp^3 hybridization in tetrahedral coordination. The lone-pair electrons occupy the missing vertex of the tetrahedron, and if donated to Lewis acids (metal cations) to form a covalent bond, the fully tetrahedral coordination would realize and gain more stabilization.

We calculated the binding of DPPP on the surfaces of $FAPbI_3$ both with lead(II) iodide (PbI_2) and formamidinium iodide (FAI) terminations. Although DFT calculations predicted that the FAI-terminated surface would be more stable (*34*), experimental evidence showed that the PbI_2-terminated surface is readily formed during the solvent treatment from depositing subsequent layers (*35*). DFT calculations also showed that DPPP was chemically bonded to the PbI_2-terminated surface through P–Pb bond formation with a binding energy of 2.24 eV but was weakly bonded to the FAI-terminated surface by van der Waals interaction with a binding energy of 1.09 eV (Fig. 1, C and D). Moreover, the calculated binding energy of DPPP with perovskites in two adjacent slabs (3.08 eV) was larger than that in the same slab (2.24 eV) (Fig. 1, D and E). Similarly, the binding energy of DPPP with both the perovskite and NiO_x slabs (4.31 eV) was larger than that in the same NiO_x slab (3.28 eV) (Fig. 1F and fig. S2). Thus, DPPP was predicted to bind, bridge, and stabilize

[1]Department of Physics and Astronomy and Wright Center for Photovoltaics Innovation and Commercialization, The University of Toledo, Toledo, OH 43606, USA. [2]Department of Chemistry, University of Washington, Seattle, WA 98195, USA. [3]The Edward S. Rogers Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. [4]Center for Materials and Sensors Characterization, The University of Toledo, Toledo, OH 43606, USA. [5]Laboratory for Thin Films and Photovoltaics, Empa–Swiss Federal Laboratories for Materials Science and Technology, Dübendorf 8600, Switzerland. [6]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [7]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208, USA.
*Corresponding author. Email: yanfa.yan@utoledo.edu

perovskite GBs and the perovskite/NiO_x interface. DPPP molecules also provided hole transport channels through the P-terminated alkane chain of DPPP (fig. S3).

Synthesis and structure

The interaction between DPPP molecules and Pb^{2+} is observed through the formation of a new adduct when a thin layer of DPPP is deposited on a PbI_2 layer (fig. S4). When fabricating devices, we deposited the FA-based perovskite layer on a DPPP-coated NiO_x hole transport layer. During the growth of perovskite films, some DPPP molecules redissolved and segregated at both the perovskite/NiO_x interface and the perovskite surface regions, as verified by the time-of-flight secondary ion mass spectrometry (TOF-SIMS) depth profiles shown in Fig. 2A. X-ray photoelectron spectroscopy (XPS) revealed that after DPPP treatment, the core levels of the elements in both perovskite and NiO_x shifted (Pb and Ni XPS spectra in Fig. 2, B and C, and the O, C, N, and I spectra in fig. S5). The universal shift of core levels caused by electrostatic interaction indicates the existence of DPPP at both interfaces. The DPPP treatment also slightly improved the crystallinity of perovskite films, as can be seen from the enhancement of grain domain size (fig. S6) and x-ray diffraction (XRD) peak intensity (fig. S7). DPPP treatment did not change the bandgap of the perovskite films (fig. S8). Photoluminescence (PL) and time-resolved PL spectroscopy (TRPL) spectra (Fig. 2, D and E) showed enhanced PL intensity and an ~50% improvement in lifetime, from 0.98 to 1.49 μs, for the DPPP-treated perovskite films, consistent with the expected reduction in nonradiative recombination and defect density upon Lewis base treatment (*29*, *36*). We further verified that DPPP treatment enhanced the mechanical toughness of the perovskite/NiO_x interface. Perovskite films were deposited on a half-cell structure with and without DPPP treatment, protected by a thin polymethyl methacrylate (PMMA) layer, and adhered to a glass plate with epoxy (see details in the supplementary materials and fig. S9). A

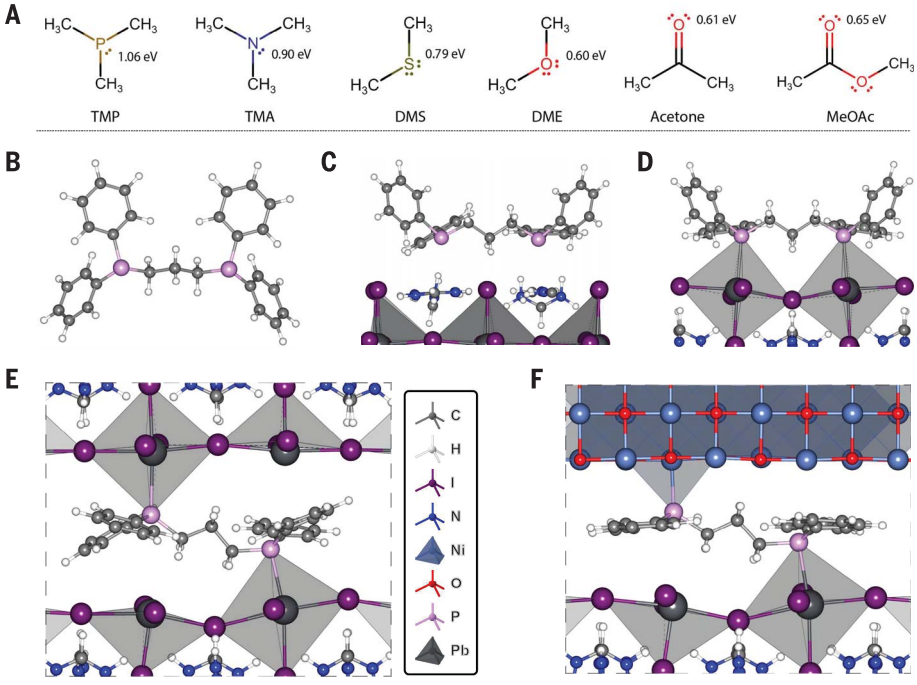


Fig. 1. DFT-calculated DPPP binding with perovskites. (**A**) Chemical structures of prototypical Lewis base molecules. The values shown are the DFT-calculated binding energies (in electron volts) of the Lewis base molecule bonded to the $FAPbI_3$ surface with PbI_2 termination. (**B**) Molecular structure of DPPP. The P atom of DPPP donates a lone-pair electron to the metal cation forming a coordinate covalent bond. Covalent bonding and van der Waals bonding for DPPP bound on $FAPbI_3$ surfaces with (**C**) FAI and (**D**) PbI_2 terminations, respectively. DPPP binds (**E**) two perovskite slabs and (**F**) perovskite and NiO_x substrate through chemical-bond formation between P and Pb or Ni atoms in a Lewis acid-base reaction.

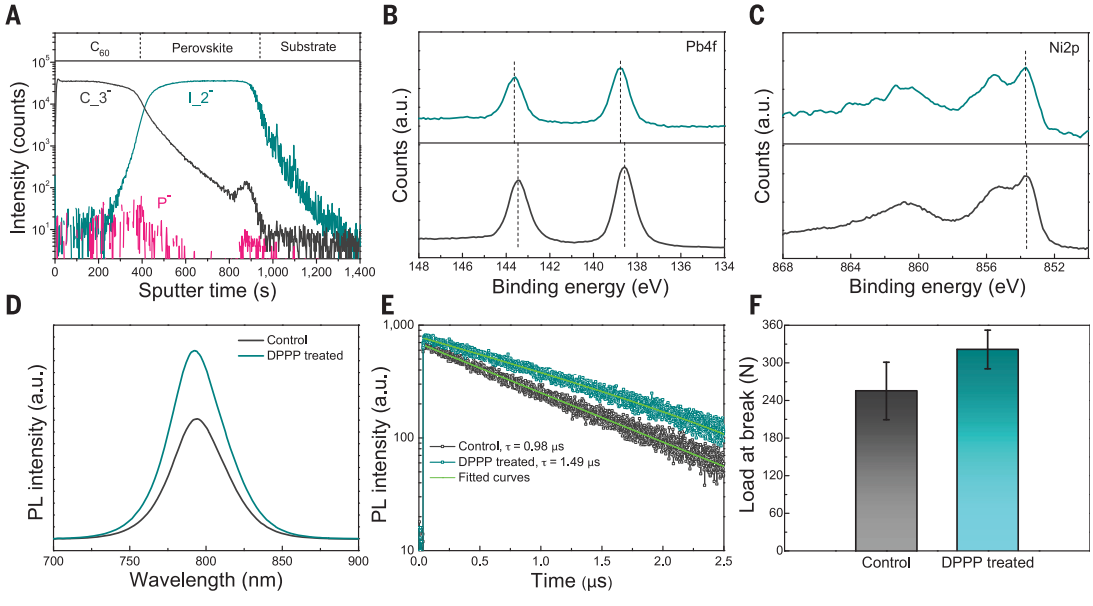


Fig. 2. Effects of DPPP on perovskite film quality and device performance. (**A**) TOF-SIMS depth profile of a DPPP-treated sample showing that DPPP molecules segregate at both interfaces. XPS spectra comparing (**B**) binding energy of Pb-4f core levels of control and DPPP-treated samples and (**C**) binding energy of Ni-2p core levels of NiO_x before and after DPPP treatment. PL (**D**) and TRPL (**E**) spectra of control and DPPP-treated perovskite films measured from the film side. The samples were excited with a continuous-wave 633-nm laser at a fluence of 1.5×10^{17} photons cm^{-2} s^{-1}. (**F**) Histogram and standard deviation values of loads at break of 10 sets of control and DPPP-treated samples. a.u., arbitrary units.

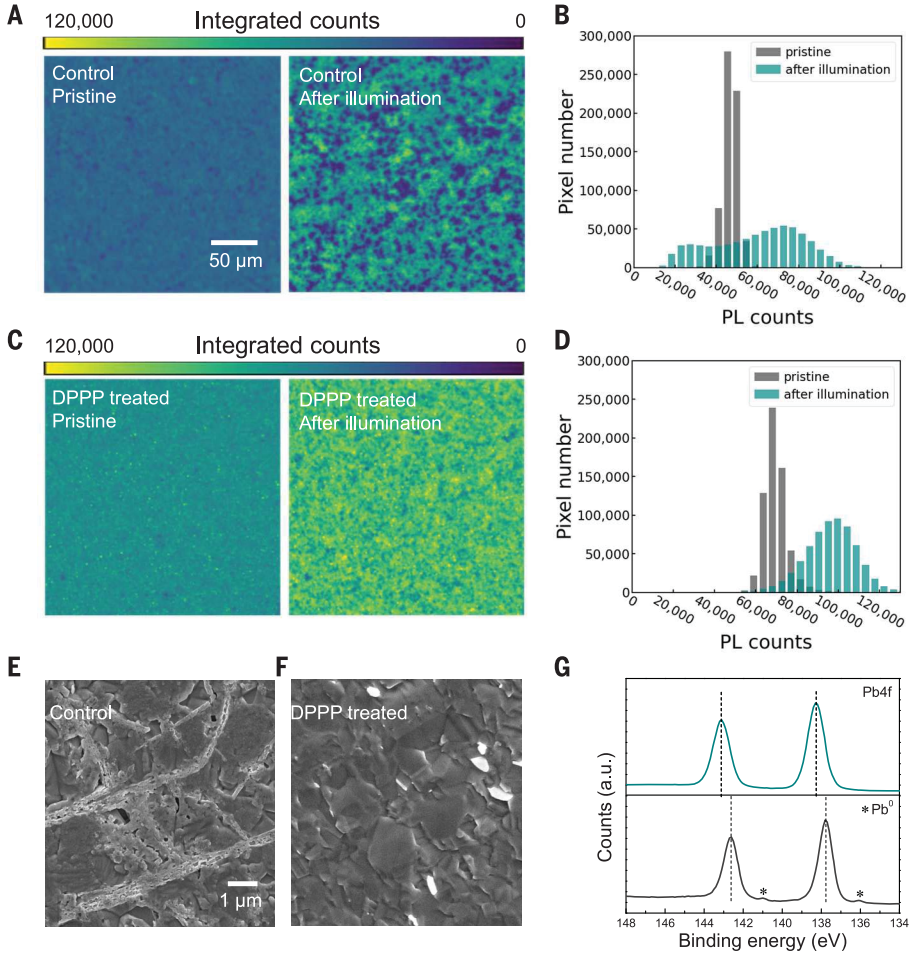


Fig. 3. Characterization of perovskite film stability before and after DPPP treatment. (**A** to **D**) Hyperspectral PL images and histograms of control and DPPP-treated perovskite films before and after light aging. The PL images were taken from the buried interface side, and the PL count at each pixel refers to the integrated PL counts over the whole spectrum. (**E** and **F**) SEM top-view images taken from the buried interface of control and DPPP-treated perovskite films after light soaking. (**G**) XPS measured from the buried interface of control and DPPP-treated perovskite films after light soaking.

tensile load was applied to delaminate the films using the testing machine shown in fig. S10. After delamination, both the NiO_x and perovskite surfaces of the NiO_x/perovskite interface showed P signals, indicating that DPPP remained on both surfaces (fig. S11). The tensile force recorded during the delamination process (Fig. 2F) suggests that DPPP treatment enhanced the mechanical strength of the perovskite/NiO_x interface through the binding of DPPP at the interface.

Hyperspectral PL mapping measured from the buried interfaces and PL intensity histograms (Fig. 3, A to D) revealed an overall higher PL intensity of the DPPP-treated perovskite film that was consistent with the longer PL lifetimes and indicated reduced defect density as compared with the control perovskite films. The PL emission heterogeneity increased, and more dark spots were observed over time in the control film after light aging (fig. S12). These dark spots showed lower PL emission intensity, and we attributed them to the initial sites of photodegradation at localized defects and local heterogeneities, which were more prevalent in the control sample without DPPP. PL decay takes place mostly in regions with low initial PL counts, whereas PL enhancement takes place in regions with high initial PL counts (fig. S13). This observation proves that local PL enhancement or decay in the control sample is associated with local defect heterogeneities. In comparison, the DPPP-treated films exhibited higher uniformity and better light stability, with most of the pixels showing photo-brightening with increased PL counts, rather than photo-decay with decreased PL counts, as was observed for the control perovskite film (Fig. 3, B and D, and fig. S12). Scanning electron microscopy (SEM) images measured from the bottom surface show needle-shaped PbI_2 crystals on the as-prepared control film, which is likely caused by the excess PbI_2 added to the precursor solution (fig. S14A). The as-prepared DPPP-treated film showed distinct layered structures (fig. S14B), which we speculated could possibly be the Lewis acid-base adduct of PbI_2 and DPPP. After aging, the PbI_2 crystals in the control film decomposed and produced pinholes on the grains (Fig. 3E). In contrast, the DPPP-treated films exhibited much-suppressed degradation (Fig. 3F). Time-resolved mass spectroscopy was also conducted on control and DPPP-treated films under illumination to investigate the degradation process. The control sample showed the release of hydrogen iodide (HI) and iodide (I) species (fig. S15), which are by-products of the photoinduced decomposition and trigger irreversible chemical chain reactions that accelerate the decomposition of perovskites (*37*, *38*). In addition to the iodide species, another perovskite decomposition by-product, metallic lead (Pb_0), was also observed on the control film after light soaking. The XPS spectrum of the control film after light soaking showed two Pb_0 peaks at ~136 and ~141 eV (Fig. 3G), whereas no Pb_0 peaks were observed in DPPP-treated film. These results reveal that DPPP treatment could effectively suppress the photodecomposition at the NiO_x/perovskite interface, likely by reducing the density of reactive undercoordinated lead sites (halide vacancies) at the surface and interfaces.

Solar cell fabrication and performance

We fabricated PSCs with the p-i-n configuration of glass/FTO/NiO_x/Me-4PACz/(with or without) DPPP/$FA_{0.95}Cs_{0.05}PbI_3$/PEAI/C_{60}/SnO_2/Ag to show the effect of DPPP treatment on improving device performance and stability {FTO, fluorine-doped tin oxide; Me-4PACz, [4-(3,6-dimethyl-9H-carbazol-9-yl)butyl]phosphonic acid; PEAI, phenethylammonium iodide}. Note that NiO_x was treated by a very thin layer of Me-4PACz to improve the reproducibility (fig. S16). The statistics of our device performance results are shown in fig. S17. A small amount of DPPP (1 or 2 mg/ml) consistently improved the open-circuit voltage (V_{OC}) and fill factor (FF) across the comparison of 40 devices. Both the improvement in V_{OC} and FF are expected to accompany a reduction in surface recombination velocity through a reduction in surface defects (*39*), and they are also in qualitative agreement with the increase in film PL and PL lifetime upon treatment. Higher concentrations of DPPP (4 mg/ml) decreased short-circuit current density (J_{SC}) and FF, which may have resulted from the overreaction between DPPP and perovskites. Figure 4A shows current

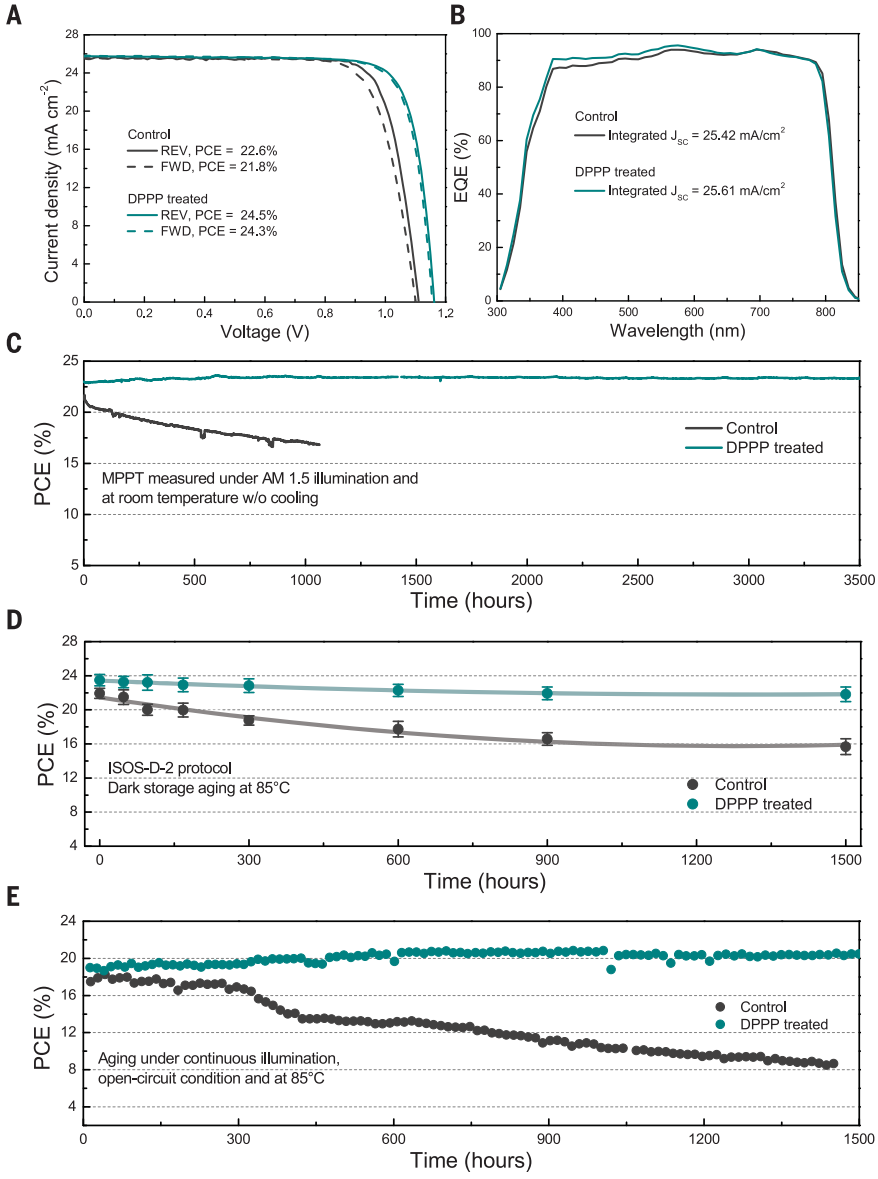


Fig. 4. Performance and stability of control and DPPP-treated devices. (**A**) *J-V* curves of champion control and DPPP-treated devices. REV, reverse scan; FWD, forward scan. (**B**) EQE spectra of the corresponding control and DPPP-treated devices. (**C**) MPPT of control and DPPP-treated devices measured under continuous 1-sun illumination in an N_2 environment at a temperature of ~40°C. (**D**) Thermal stress test of control and DPPP-treated devices aged at 85°C following the ISOS-D-2 protocol. (**E**) Tracking of control and DPPP-treated devices measured at 85°C and under continuous ~0.9-sun illumination and OC condition.

density–voltage (*J-V*) scans of the champion control and DPPP-treated (2 mg/ml) devices with 0.1 cm^2 aperture masks under forward and reverse scans. The PCE of the DPPP-treated device improved from 22.6 to 24.5%, with V_{OC} increasing from ~1.11 to ~1.16 V and FF increasing from ~79 to ~82% (see detailed device parameters in table S1). External quantum efficiency (EQE) spectra (Fig. 4B) verified the J_{SC} values obtained from the *J-V* measurement. The enhanced EQE at short wavelengths also indicates improved carrier extraction at the front interface, which may be associated with the passivated interface achieved through the use of DPPP. We also fabricated PSCs with a larger active area (1.05 cm^2) to validate the benefits of DPPP treatment. The target devices with DPPP treatment also showed overall improvement in device parameters (fig. S18), with the champion target device showing PCEs of 23.9 and 23.6% under reverse and forward scans, respectively (fig. S19 and table S2). The improvement on FF and V_{OC} confirmed the reduction in defect density at the NiO_x/perovskite front interface after DPPP treatment.

For durability tests, we replaced the Ag electrodes with Cr-Cu alloy to avoid instability caused by Ag corrosion and diffusion but slightly lowered the device efficiency (by ~4%) (fig. S20). We used 0.1 cm^2 aperture masks for solar cells when conducting all the stability tests. We first tested the effect of DPPP on device stability by MPPT under continuous 1-sun illumination in an N_2 environment at a temperature of ~40°C. As shown in Fig. 4C, the DPPP-treated device exhibited an initial PCE of ~23%, which increased to ~23.5% after ~450 hours and remained unchanged after 3500 hours. The increased PCE resulted from the increased voltage at maximum power point (fig. S21). The increased voltage is likely due to the light-induced annihilation of halide defects (*10*, *40*–*43*). However, the PCE of the control device decreased to <80% of its initial PCE after 1000 hours. The film area of the control PSC turned dark yellow, indicating the decomposition of perovskite to PbI_2 after 3500 hours (fig. S22). In contrast, the film area of the DPPP-treated device remained dark, indicating a more-stabilized perovskite after DPPP treatment.

We then conducted accelerated durability tests at elevated temperatures. We kept the PSCs in a dark oven at 85°C and in the ambient atmosphere and measured the PCEs periodically for the thermal stress test (ISOS-D-2 protocol). We tested 18 PSCs and summarize their statistics in fig. S23. The average PCEs and their corresponding standard deviations (Fig. 4D) reveal that the DPPP-treated devices retain, on average, ~90% of their initial PCE after 1500 hours, whereas, in contrast, the average PCE of the control devices dropped to <90% after 168 hours.

We conducted a more rigorous accelerated durability test under the OC condition and continuous ~0.9-sun illumination. The devices were kept in an oven at a temperature of 85°C and humidity of ~65% and measured every 208 s. The devices measured at 85°C showed slightly lower PCEs, possibly owing to the negative temperature coefficient (fig. S24). The PCE was stable for the first ~300 hours and then slightly increased to ~108% of its initial value and showed no degradation after 1500 hours (Fig. 4E). However, the control device degraded rapidly after 400 hours, dropping to ~80% of its initial PCE. Two more DPPP-treated PSCs were measured to validate the improved stability by DPPP treatment (fig. S25). The photos of devices after 1500 hours of illumination are shown in fig. S26. The glass-side view of the control film turned gray, which was likely the result of delamination of the perovskite film at the NiO_x/perovskite front interface and the degradation of the perovskite layer. In contrast, the DPPP-treated device showed partial delamination on the edge but remained dark over most of the film area.

Discussion

Taking experimental findings together with DFT studies, we offer that DPPP molecules strengthen the NiO_x/perovskite interface and stabilize the perovskite phase. The robust binding between the NiO_x and perovskite enabled by DPPP modification appears to be an enabler of the stable operation of PSCs under outdoor conditions. The measured stability under accelerated testing conditions indicates a benefit from DPPP in the form of improved device stability and provides ways of realizing commercialization of PSCs.

REFERENCE AND NOTES

1. A. K. Jena, A. Kulkarni, T. Miyasaka, *Chem. Rev.* **119**, 3036–3103 (2019).
2. Z. Li *et al*., *Nat. Rev. Mater.* **3**, 18017 (2018).
3. H. Min *et al*., *Nature* **598**, 444–450 (2021).
4. Y. Zhao *et al*., *Science* **377**, 531–534 (2022).
5. Y. Rong *et al*., *Science* **361**, eaat8235 (2018).
6. L. Meng, J. You, Y. Yang, *Nat. Commun.* **9**, 5265 (2018).
7. P. Holzhey, M. Saliba, *J. Mater. Chem. A* **6**, 21794–21808 (2018).
8. K. Domanski, E. A. Alharbi, A. Hagfeldt, M. Grätzel, W. Tress, *Nat. Energy* **3**, 61–67 (2018).
9. M. V. Khenkin *et al*., *Nat. Energy* **5**, 35–49 (2020).
10. R. Azmi *et al*., *Science* **376**, 73–77 (2022).
11. C. C. Boyd, R. Cheacharoen, T. Leijtens, M. D. McGehee, *Chem. Rev.* **119**, 3418–3451 (2019).
12. M. Kim *et al*., *ACS Energy Lett.* **6**, 3530–3537 (2021).
13. E. Bi, Z. Song, C. Li, Z. Wu, Y. Yan, *Trends Chem.* **3**, 575–588 (2021).
14. K. Domanski *et al*., *Energy Environ. Sci.* **10**, 604–613 (2017).
15. T. Duong *et al*., *ACS Appl. Mater. Interfaces* **9**, 26859–26866 (2017).
16. B. Chen *et al*., *Adv. Mater.* **31**, e1902413 (2019).
17. Y. Yuan, J. Huang, *Acc. Chem. Res.* **49**, 286–293 (2016).
18. L. Shi *et al*., *ACS Appl. Mater. Interfaces* **9**, 25073–25081 (2017).
19. S. Yang *et al*., *Science* **365**, 473–478 (2019).
20. Y.-H. Lin *et al*., *Science* **369**, 96–102 (2020).
21. S. Bai *et al*., *Nature* **571**, 245–250 (2019).
22. Z. Li *et al*., *Science* **376**, 416–420 (2022).
23. T.-H. Han *et al*., *Nat. Commun.* **10**, 520 (2019).
24. M. Zhu *et al*., *Mater. Horiz.* **7**, 2208–2236 (2020).
25. D. W. deQuilettes *et al*., *ACS Energy Lett.* **1**, 438–444 (2016).
26. C. Shi *et al*., *Adv. Funct. Mater.* **32**, 2201193 (2022).
27. Z. Yang *et al*., *Adv. Funct. Mater.* **30**, 1910710 (2020).
28. N. K. Noel *et al*., *ACS Nano* **8**, 9815–9821 (2014).
29. N. Ahn *et al*., *J. Am. Chem. Soc.* **137**, 8696–8699 (2015).
30. A. Seidu, M. Dvorak, J. Järvi, P. Rinke, J. Li, *APL Mater.* **9**, 111102 (2021).
31. S. Tan *et al*., *J. Am. Chem. Soc.* **143**, 6781–6786 (2021).
32. D. W. deQuilettes *et al*., *Science* **348**, 683–686 (2015).
33. Z. Song *et al*., *Sustain. Energy Fuels* **2**, 2460–2467 (2018).
34. F. Fu *et al*., *Energy Environ. Sci.* **12**, 3074–3088 (2019).
35. J. Wang *et al*., *ACS Energy Lett.* **4**, 222–227 (2019).
36. J. A. Christians *et al*., *Nat. Energy* **3**, 68–74 (2018).
37. W. Nie *et al*., *Nat. Commun.* **7**, 11574 (2016).
38. G. Kresse, J. Furthmüller, *Phys. Rev. B* **54**, 11169–11186 (1996).
39. G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **6**, 15–50 (1996).
40. P. E. Blöchl, *Phys. Rev. B* **50**, 17953–17979 (1994).
41. J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
42. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).
43. A. R. M. Alghamdi, M. Yanagida, Y. Shirai, G. G. Andersson, K. Miyano, *ACS Omega* **7**, 12147–12157 (2022).

ACKNOWLEDGMENTS

C.L. acknowledges D. Luo for important discussions about the XPS analysis. **Funding:** This material is based on work supported by the US Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office awards DE-EE0008970 and DE-EE0008753 and by the US Air Force Research Laboratory under agreement FA9453-21-C-0056. The contributions of F.J. and D.S.G., focusing on hyperspectral imaging for cell metrology, are based primarily on work supported by EERE under the Solar Energy Technologies Office (award DE-EE0009528) as well as institutional support from the B. Seymour Rabinovitch Endowment and the state of Washington. DFT calculations were supported by the Center for Hybrid Organic-Inorganic Semiconductors for Energy (CHOISE), an Energy Frontier Research Center funded by the Office of Basic Energy Sciences, Office of Science, within the US Department of Energy and the National Science Foundation under contract DMR-1807818. The DFT calculations were performed using computational resources sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory and the DOS calculations used resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract DE-AC02-05CH11231 using NERSC award BES-ERCAP0017591. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views expressed are those of the authors and do not reflect the official guidance or position of the United States government, the Department of Defense, or of the United States Air Force. The appearance of external hyperlinks does not constitute endorsement by the United States Department of Defense (DoD) of the linked websites, or the information, products, or services contained therein. The DoD does not exercise any editorial, security, or other control over the information you may find at these locations. Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2022-3776. E.H.S. acknowledges support from the US Department of the Navy, Office of Naval Research (grant NO0014-20-1-2572). **Author contributions:** C.L. and Y.Y. conceived of the idea. Y.Y. supervised the projects and process. C.L., S.M.P., and E.B. fabricated perovskite films and devices for characterization and performance measurement. X.W. and Y.X. carried out DFT calculations. C.L., L.C., T.Z., and L.Z. carried out SEM, ultraviolet-visible, and XRD measurements and data analysis. Z.S. carried out time-resolved mass spectroscopy measurements and data analysis. F.J. and D.S.G. carried out hyperspectral microscope measurements and associated data analysis. C.L. and J.C. prepared NiO_x substrates. Y.Li carried out stability tests and data analysis. Z.W., Y.Liu, and H.C. carried out XPS measurements and data analysis. H.L. and F.F. carried out the TOF-SIMS measurements and data analysis. C.R.G. carried out tensile force measurements and data analysis. A.A. and R.J.E. carried out the PL and TRPL measurements and data analysis. C.L., X.W., Z.S., and Y.Y. wrote the first draft of the manuscript. E.H.S., Y.Y., C.L., Z.S., and B.C. reviewed and edited the manuscript. All authors discussed the results and contributed to the manuscript revisions. **Competing interests:** C.L. and Y.Y are inventors on a patent application (no. 17038731) related to this work filed by the University of Toledo. The other authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade3970
Materials and Methods
Figs. S1 to S26
Tables S1 and S2
References

Submitted 16 August 2022; resubmitted 18 September 2022
Accepted 13 January 2023
10.1126/science.ade3970

BIOPHOTONICS

A tunable reflector enabling crustaceans to see but not be seen

Keshet Shavit[1], Avital Wagner[1], Lukas Schertel[2,3], Viviana Farstey[4], Derya Akkaynak[4,5], Gan Zhang[1]†, Alexander Upcher[6], Amir Sagi[7,8], Venkata Jayasurya Yallapragada[9], Johannes Haataja[2]*, Benjamin A. Palmer[1]*

Many oceanic prey animals use transparent bodies to avoid detection. However, conspicuous eye pigments, required for vision, compromise the organisms' ability to remain unseen. We report the discovery of a reflector overlying the eye pigments in larval decapod crustaceans and show how it is tuned to render the organisms inconspicuous against the background. The ultracompact reflector is constructed from a photonic glass of crystalline isoxanthopterin nanospheres. The nanospheres' size and ordering are modulated to tune the reflectance from deep blue to yellow, enabling concealment in different habitats. The reflector may also function to enhance the acuity or sensitivity of the minute eyes by acting as an optical screen between photoreceptors. This multifunctional reflector offers inspiration for constructing tunable artificial photonic materials from biocompatible organic molecules.

Many pelagic animals (such as jellyfish, mollusks, and fish) use transparent bodies to appear invisible in the ocean as a defense against predation (*1*). However, to enable vision, these organisms require highly conspicuous eye pigments that enhance their risk of being detected (*2*, *3*). Transparent organisms have developed ingenious strategies to circumvent this trade-off between seeing and not being seen (*4*), including reducing the size and visibility of eyes (*5*) and retinas (*6*) or using mirrored irises (*7*) to conceal eye pigments. In addition to possessing condensed retinas (*6*), larval crustaceans use an alternative strategy: a reflector overlying the opaque eye pigments whose color is matched to the background, providing crypsis (*2*, *3*). This reflector produces a characteristic eyeshine distinct from the reflectance of tapeta lucida displayed by many other organisms. The eyeshine phenomenon is present in many larval *Malacostracans* (*8*–*11*), but the nature of the photonic structure is unknown.

We show that the eyeshine of larval decapods is produced by a photonic glass [a disordered assembly of wavelength-sized, dielectric spheres that express resonant scattering behavior (*12*–*14*)], constructed from high refractive index nanospheres (*15*, *16*). Deep blue to yellow-green eyeshine in different species is produced by tuning the size of the crystalline nanospheres, enabling crypsis in aquatic habitats with different optical properties. We observed that some decapod species also change their eyeshine color when subjected to different light intensities (dark and light adaptation). This is achieved through a modulation in the nanosphere ordering and suggests an ability to dynamically match their eyeshine to a changing background. The reflector may also function to screen cross-talk between adjacent photoreceptors, potentially mitigating against the visual costs of having extremely small eyes.

To investigate the eyeshine phenomenon, we studied a model species of freshwater prawn, *Macrobrachium rosenbergii* (Fig. 1A), as well as larval crustaceans collected from the Gulf of Aqaba (Fig. 1, B to I, and supplementary materials). The collected prawn, shrimp, lobster, stomatopod, and crab larvae have completely transparent bodies but exhibit striking blue, green, silver, and yellow-green eyeshine (Fig. 1).

Ultrastructural properties of the reflector

To determine the structural origin of the eyeshine, the brilliant yellow-green eyes of *M. rosenbergii* were imaged by means of optical (Fig. 2, A to C) and cryogenic scanning electron microscopy

[1]Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel. [2]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K. [3]Department of Physics, University of Fribourg, 1700 Fribourg, Switzerland. [4]The Interuniversity Institute for Marine Sciences, Eilat 8810302, Israel. [5]Hatter Department of Marine Technologies, University of Haifa, Haifa 3498838, Israel. [6]Ilse Katz Institute for Nanoscale Science & Technology, Ben-Gurion University of the Negev, Be'er Sheba 8410501, Israel [7]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel. [8]The National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel. [9]Department of Physics, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India.
***Corresponding author. Email: jh2225@cam.ac.uk (J.H.); bpalmer@bgu.ac.il (B.A.P.)**
†Present address: College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 730000, China.

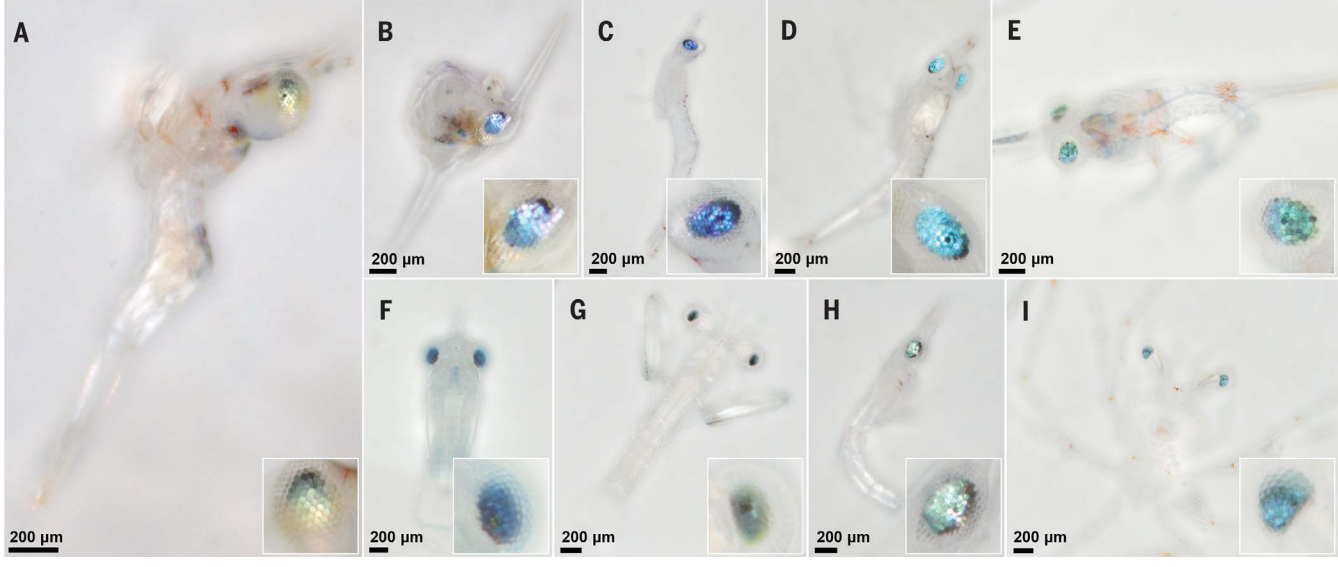


Fig. 1. Eyeshine reflectance in larval crustaceans. Polarized light micrographs of decapod and stomatopod larvae. (Insets) Higher-magnification images of the eyes. (**A**) A giant freshwater prawn, *M. rosenbergii* (development stage; zoea 2) (supplementary materials, materials and methods). (**B**) Zoea of a brachyuran crab. (**C** and **D**) Zoeae of caridean shrimp. (**E**) Zoea of a porcelain crab. (**F** and **G**) Stomatopod larvae. (**H**) Zoea of an axiid shrimp. (**I**) Phyllosoma of an achelate lobster.

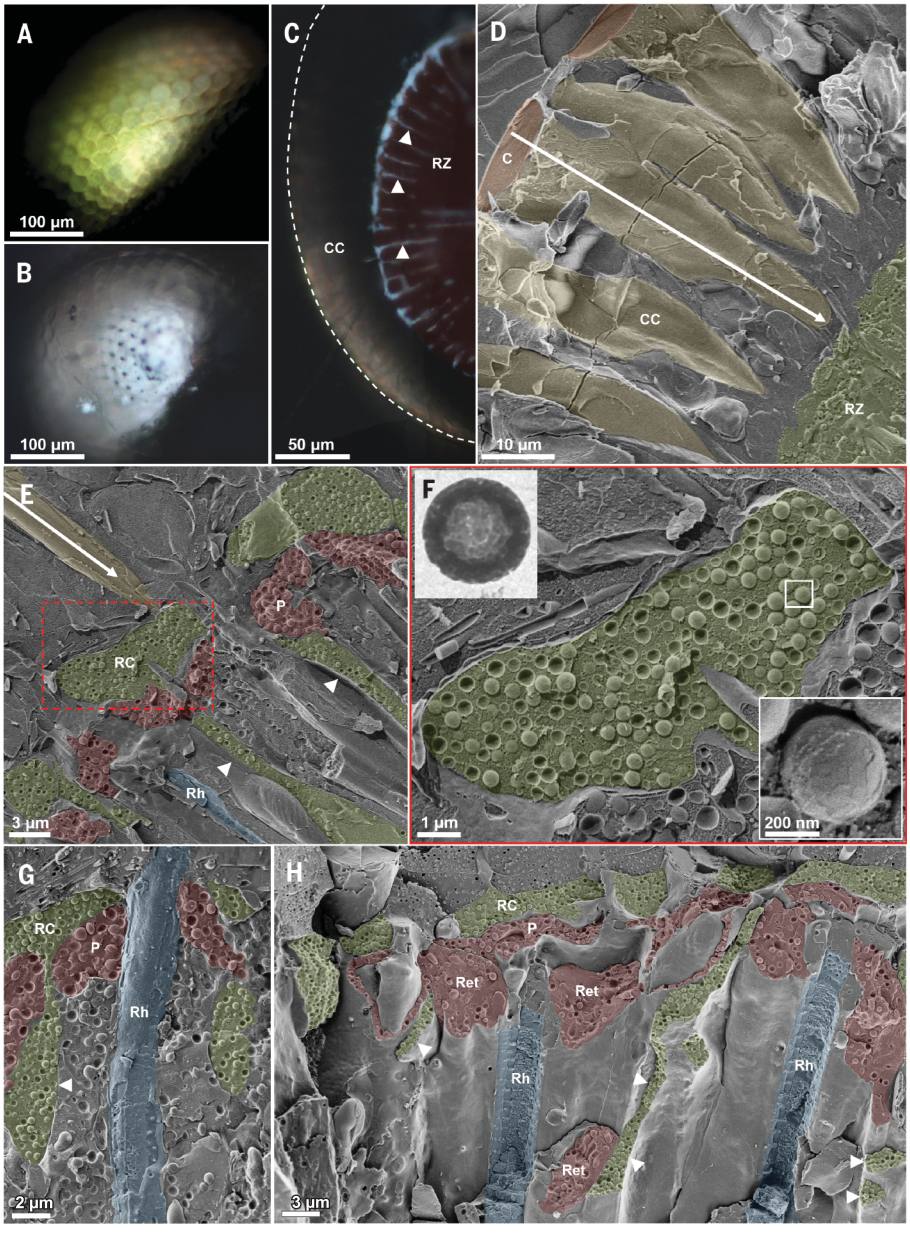


Fig. 2. Ultrastructure of the eyeshine reflector in *M. rosenbergii* larvae. (**A** to **C**) Polarizing optical micrographs of (A) a live eye (zoea 5), exhibiting yellow-green reflectance; (B) a cross section through a fixed eye (zoea 7); and (C) a longitudinal eye section viewed perpendicular to the eye's optic axis (zoea 7). Arrowheads indicate projections of the reflective cells extending into the retinal zone. Dashed trace denotes the boundary of the cornea. (**D** to **H**) Pseudo-colored cryo-SEM micrographs of longitudinal eye sections (zoea 4). (D) Low-magnification image of the eye showing the cornea (C; orange), crystalline cone (CC; yellow), and retinal zone (RZ; green). The arrow indicates the direction of the incident light. (E) Higher-magnification image showing the reflector cell (RC; green), screening pigment cells (P; red), and rhabdoms (Rh; blue). Projections of the reflective cells extend down the sides of the rhabdoms and retinal cells. (F) Region of interest in (E) showing a reflective cell comprising arrays of nanospheres. (Insets) (Top left) TEM and (bottom right) cryo-SEM micrographs of individual core-shell nanospheres that comprise concentric lamellae of isoxanthopterin crystal plates (*18*, *19*). The scale is the same in both insets. [(G) and (H)] As shown from two different angles, the reflector cells closely envelope the screening pigment, rhabdoms (Rh), and retinal cells (Ret). (G) is viewed perpendicular to the eyes optic axis, and (H) is at a more obtuse angle that shows the view along the ommatidia. In (H), the smooth, nonvesicular material surrounding the rhabdoms, pigment cells, and reflective cells are membranes. The labeling for the various eye structures in (C) to (H) was based on the anatomical descriptions provided in (*9*).

(cryo-SEM) (Fig. 2, D to H). Larval decapods have apposition eyes, constructed from hexagonal eye units called ommatidia (Fig. 2A) (*9*, *11*, *17*). The upper part of each ommatidium contains a proteinaceous body (the crystalline cone) that channels light onto the photoreceptors (Fig. 2, C to E). Polarizing optical micrographs viewed along the eye's optic axis reveal a highly reflective material at the base of the crystalline cones. The reflective material is absent from the optical pathway, allowing light from the cones to pass to the underlying photoreceptors (Fig. 2B). The reflective material also contains projections that extend into the retinal zone (Fig. 2C, arrowheads) that comprises the retinal cells, rhabdoms [the photoreceptive unit of crustacean eyes (*9*)], screening pigment cells, and nerve connections.

Cells (fig. S1) at the base of the cones contain assemblies of ~400-nm particles (Fig. 2, E and F). The particles are composed of crystalline isoxanthopterin (figs. S2 and S3) and are identical to those in the reflective tapetum of adult decapods (*18*, *19*). The extreme refractive index (n = 1.96) (*18*, *19*) nanospheres are constructed from a spherulitic assembly of single-crystal isoxanthopterin plates, arranged in lamellae around a hollow core (Fig. 2F, insets, and fig. S3). The anisotropic orientation of refractive indices arising from this spherulitic birefringence enhances scattering efficiency (*18*, *19*). Isoxanthopterin particles were present in all shrimp (fig. S4) and prawns we investigated, but a different reflective material is used in stomatopods or crabs (figs. S5 to S7). We attribute the eyeshine of *M. rosenbergii* and the marine shrimp to the isoxanthopterin-containing cells. The presence of isoxanthopterin in both the larval and adult shrimp reflectors suggests that the eyeshine reflector may be a developmental "precursor" to the adult tapetum (fig. S8) (*9*). Thus, a single material fulfills two distinct optical functions during development according to the changing visual ecology of the organism: the requirement for crypsis in larvae and dim light vision in adults (*17*).

On the dorsal side of the eye, the reflective cells form a reflective sheath around the absorbing pigment cells, effectively shielding them from view (Fig. 2, E, G, and H, and fig. S4). Conversely, the ventral side of the eye appears dark because the absorbing pigment lies on top of the reflective cells (fig. S9). Because larvae swim upside-down (*9*), the dorsal-lying reflective cells point downward, countershading the conspicuous eye pigments when viewed from below (*9*). The elongated projections of the reflective cells seen in optical micrographs (Fig. 2C) form a separation between adjacent rhabdoms (Fig. 2, E, G, and H) (*9*). The cellular extensions that interdigitate the rhabdoms are found in both the dorsal and ventral sides. This suggests that the reflective cells may serve a secondary function that directly enhances

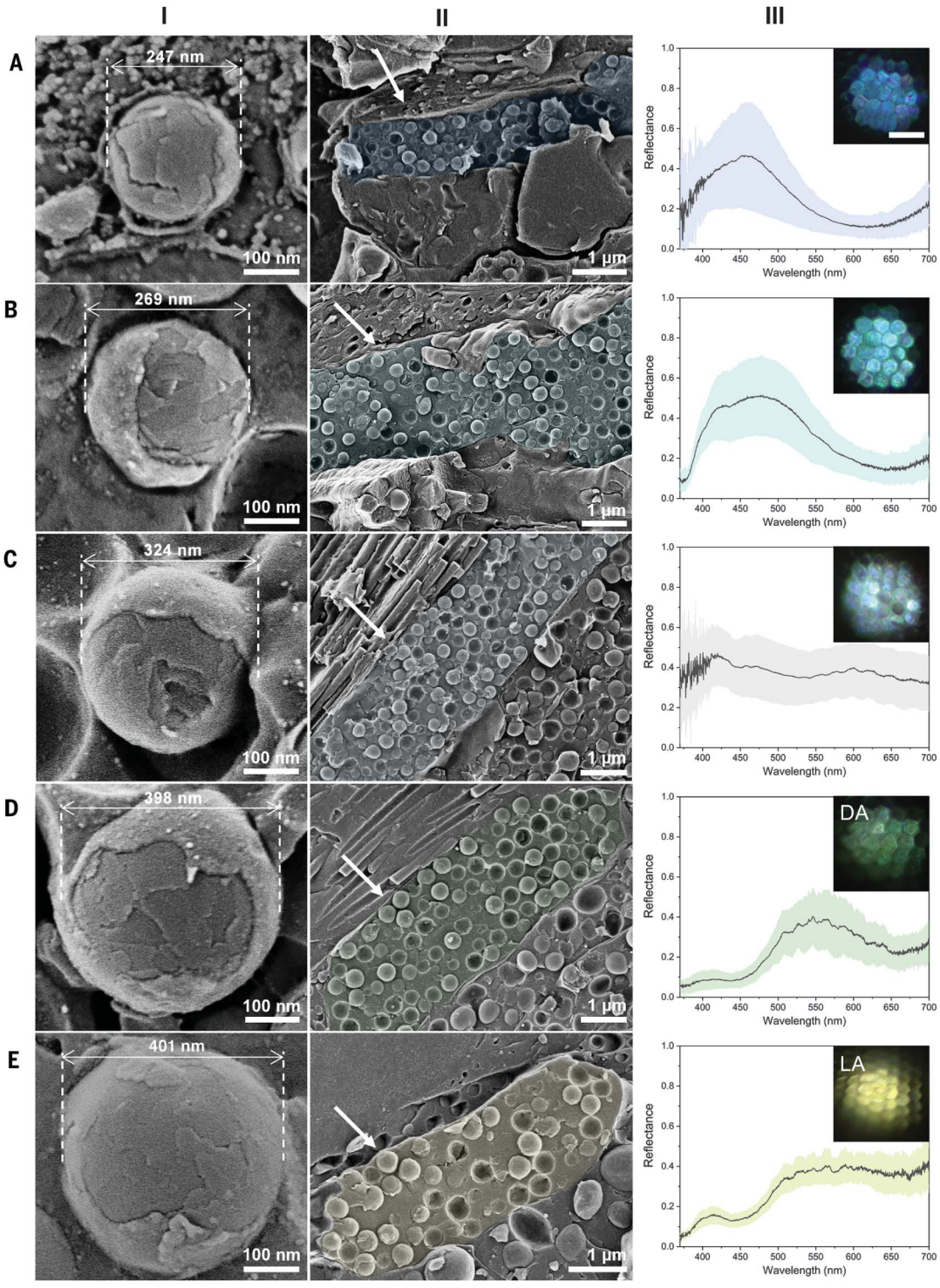


Fig. 3. The reflectance and structural properties of reflective cells in different shrimp and prawn larvae. Column I shows cryo-SEM images of single nanoparticles from a reflector cell. Column II shows representative cryo-SEM images of the reflector cells. White arrows indicate incident light direction. Column III shows the reflectance spectra from the eyes. Pastel shading indicates ±SD of the average reflectance across numerous specimens in the same color group (table S1). (Insets) Polarizing optical micrographs of the eyes. Scale bar, 50 µm and applies to all insets in column III. (**A** to **C**) Zoeae of different species of caridean shrimp from the Gulf of Aqaba. (**D**) Dark-adapted (DA) *M. rosenbergii* (zoea 6). (**E**) Light-adapted (LA) *M. rosenbergii*. Average particle sizes for the color groups are (A) blue, 247 nm (±7 nm, number of particles measured N = 398); (B) turquoise, 269 nm (± 15 nm, N = 400); (C) silvery blue, 324 nm (± 14 nm, N = 338); and (D) and (E) yellow-green (*M. rosenbergii*), DA, 398 nm (±30 nm, N = 179), and LA, 401 nm (±27 nm, N = 48). Average 2D filling fractions of the color groups are (A) blue, 57% (± 6%); (B) turquoise, 64% (± 5%); (C) silvery blue, 63% (± 7%); and (D) and (E) yellow-green (*M. rosenbergii*), DA, 63 ± 9%, and LA, 64 ± 5%.

vision—either as an additional screen between photoreceptors (*20*) to improve acuity or as a lateral tapetum-like structure that improves sensitivity during activity in dim light environments (*21*).

Correlation between the eyeshine color and nanosphere size

To rationalize the variation in eyeshine colors in the various shrimp and prawn species, we performed correlative cryo-SEM–reflectance measurements (Fig. 3). The eyeshine reflectance was measured and the shrimp grouped according to the position and shape of their reflectance peak(s). Four groups were assigned: (i) blue, exhibiting a narrow reflectance maximum around 450 nm (Fig. 3A); (ii) turquoise, exhibiting a broad reflectance band with maximum at 400 to 500 nm (Fig. 3B); (iii) silvery blue, exhibiting broadband reflectance with minor peaks at ~410 and ~600 nm (Fig. 3C); and (iv) yellow-green, exhibiting a broad reflectance maximum from 500 to 700 nm (*M. rosenbergii*) (Fig. 3, D and E). Cryo-SEM images of the reflective cells in the different groups revealed a clear correlation between the eyeshine color and the size of the component nanospheres. The reflectance shifted to longer wavelengths as the nanosphere diameter increased from ~250 nm in blue-eyed shrimp to ~400 nm in

the yellow-green *M. rosenbergii* (Fig. 3, A to E, and table S1).

Adaptation of the eyeshine reflector to camouflage in different habitats

These eyeshine colors appear to be well adapted to the background habitats of the different organisms. Larval marine shrimp exhibit blue to silvery blue eyeshine reflectance. The calculated eyeshine radiance of these organisms spectrally matches with the water radiance of the clear blue waters of the Gulf of Aqaba at different depths (Fig. 3, A to C, and figs. S10 to S15). Moreover, *M. rosenbergii*, *Macrobrachium amazonicum* (*10*), and *Palaemonetes pugio* (*9*), which inhabit estuaries [characterized by high organic content and turbid yellow water (*22*)], possess yellow-green eyeshine reflectance (Fig. 3, D and E, and fig. S16).

Feller and Cronin found (*3*) that the blue-green eyeshine of stomatopods was well matched to deeper depths in their marine environment and that depth was the key determinant in the visibility of the eyes. In our experiments, the marine shrimp shown in Fig. 3 were collected at night from shallow water (10 to 25 m), where they migrate from depth to feed (*23*). During the day, decapod larvae migrate to deeper depths to avoid predation from visual predators (*23*, *24*). In accordance with Feller and Cronin, our data suggest that the eyeshine of the marine shrimp can provide crypsis against the background of their daytime habitats at various depths (fig. S10 to S15).

Simulating the optical properties of the eyeshine reflector

To elucidate the optical mechanism underlying the eyeshine, we note that the upper limit (supplemenatry materials, materials and methods) of the two-dimensional (2D) particle filling fraction in the reflective cells was typically (Fig. 3 and fig. S4) between 50 and 70% (Fig. 3, column II), which is much lower than that of a close-packed photonic crystal (*25*). The nanospheres thus lack long-range periodicity but display the short-range positional ordering of a photonic glass structure. The scattering of photonic glasses is determined by an interplay between (i) the scattering properties of single particles (the form factor, F) and (ii) correlative scattering by particle ensembles (the structure factor, S) (Fig. 4A) (*26*). The absence of strong resonant peaks in form factor calculations indicates that single-particle scattering does not contribute appreciably to the optical response (Fig. 4, A and B). Conversely, strong structural peaks in the visible and infrared (IR) spectra in structure factor calculations may explain the experimental reflectance (Fig. 4C). We used molecular dynamics (MD) simulations to generate photonic glasses with particle sizes akin to those in the eyes and with a variety of filling fractions and structures (Fig. 4 and figs. S17 and S18). We simulated reflectance spectra using the finite-difference time domain (FDTD) technique (*18*, *27*) for each observed particle size (Fig. 4D and figs. S17 and S18). The simulated spectra reveal major peaks in the visible and IR (Fig. 4D), corresponding to the second- and first-order structural peaks predicted with numerical calculations, respectively

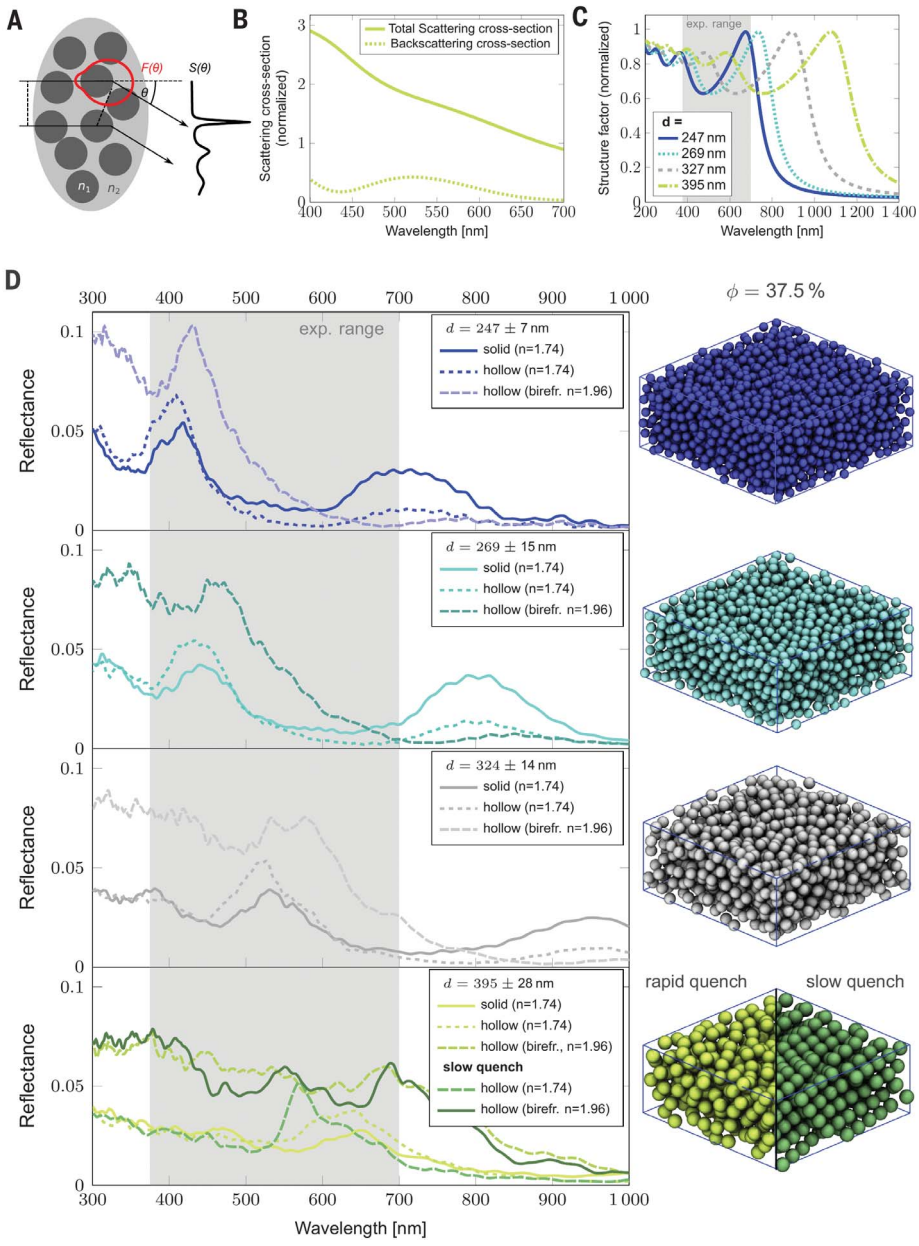


Fig. 4. Optical modeling of the eyeshine phenomena. (**A**) A schematic of the form $F(\theta)$ and structure factor $S(\theta)$ contributions to the scattering of a photonic glass. (**B**) Total (solid trace) and backscattering (dotted trace) cross sections for a spherical particle found in *M. rosenbergii* (diameter, 395 nm; refractive indices of particle and medium, respectively, $n_{particle} = 1.96$, $n_{medium} = 1.33$) calculated with Mie theory (*34*). No resonance peak in the visible spectrum is observed in the total scattering in contrast to the experimental reflectance (Fig. 3). The backscattering spectra exhibits a low-intensity peak in the visible spectrum, which is much broader than in the experimental spectra. (**C**) Percus-Yevick structure factor (*35*) for the experimentally observed particle sizes with a 3D filling fraction of 50%, by using the Garnett effective refractive index for the system (*36*). The first-order structural peak is in the red and IR, whereas a second-order peak lies in the visible spectrum. (**D**) MD and FDTD simulations of 2- by 5- by 5-μm photonic glass slabs by using a lower-bound estimate (37.5%) of the 3D filling fraction. (Left) Simulated spectra of solid (n = 1.74), hollow (n = 1.74), and hollow-birefringent particles (n = 1.96) for the four experimentally observed particle sizes. (Right) Snapshot of the MD-simulated particle configurations. For the 395-nm particle, a composite image is shown for rapid (disordered; yellow) and slow quench (more crystalline; green) states, simulating the dark-light adaption observed in *M. rosenbergii*.

(Fig. 4C). The position of the visible peak correlates with the experimental reflectance and red-shifts with increasing particle size (Fig. 4D and fig. S19). Simulations using hollow and hollow-birefringent spheres led to an enhancement in the intensity of the visible peak and a suppression of the IR peak compared with those of solid isotropic particles. Although the IR reflectance is unlikely to be biologically relevant, particle hollowness and birefringence are seemingly used as a means of enhancing scattering efficiency in the visible spectrum. The increased intensity of the birefringent nanospheres is likely due to their higher refractive index (n = 1.96), resulting from the anisotropic alignment of the component refractive indices in the spheres (*18*). The reflectance intensities obtained from simulations are lower than the experimental spectra (Fig. 3) because of the small size of the computed photonic glasses (fig. S20) and experimental factors (such as specular reflection).

Further evidence for the role of structural correlations in the optical response came from the observation that the eyeshine of *M. rosenbergii* changes reversibly in different light conditions. Dark-adapted (DA) larvae have green eyeshine, with a large reflectance peak at ~560 nm (Fig. 3D). Upon light adaptation, the peak broadens, resulting in silvery yellow eyeshine (Fig. 3E). This dynamically changing eyeshine indicates that *M. rosenbergii* may spectrally match its eyeshine appearance to a temporally changing background, which may be relevant during diel vertical migrations (fig. S16) (*23*). Cryo-SEM imaging shows that the particle size, filling fraction, and absorbing pigment distribution in DA and light-adapted (LA) eyes are constant (fig. S21 and table S1). However, upon dark adaptation, spatial correlations in the particle arrangement emerge, which is manifest in the presence of disordered multilayers consisting of three to five layers of particles (Fig. 3D and fig. S21).

To rationalize the light-dark adaptation behavior, MD simulations were performed in which the 395-nm particles were allowed to equilibrate during a slow quench to achieve a more well-ordered particle packing (Fig. 4D and fig. S22). In comparison with the normal glassy state, simulations that used hollow and hollow-birefringent spheres in this configuration led to an emergence of a reflectance peak around 550 to 600 nm, resembling the changes in DA and LA *M. rosenbergii* (Fig. 3, D and E). The position of this peak aligns with that of an ordered face-centered cubic lattice of such particles, suggesting the structural origin of this reflectance peak. In general, the simulated spectra exhibit narrower peaks than the experimental ones because the latter represent the convoluted optical response from multiple cells encompassed in the illumination volume (an average over variations in filling fraction, sample thickness, degree of disorder, and particle polydispersity).

Discussion

Our results show that decapod eyeshine reflectance is generated by a photonic glass constructed from crystalline isoxanthopterin nanospheres, whose location overlying the screening pigments is consistent with a camouflage function (*3*). Reflectors are widely used in animal eyes, and their function depends on their location, nanostructure, and composition (*28*). Concave and radial mirrors form images in scallop (*29*) and decapod crustacean eyes (*17*, *19*), respectively, and in many organisms, tapetal reflectors enhance sensitivity by back-scattering light to the photoreceptors (*17*). Mirrored irises in fish prevent unfocused light entering the eye to preserve visual acuity (*7*). Reflectors in the cornea of insects (*30*) or embedded within the photoreceptors of crustaceans (*21*) spectrally filter light impinging on the retinas. Ocular reflectors can also perform nonvisual functions. For example, basally located *stratum argenteum* reflectors in fish eyes enhance camouflage by reducing the visibility of the pupil (*31*).

A distinguishing feature of the eyeshine reflector described here is its tunability and compactness. Across different larval shrimp and prawn species, the eyeshine color is modulated by tuning the size of the isoxanthopterin nanospheres in the photonic glass. Within a single species, eyeshine color can be changed by tuning the particle ordering. Feller and Cronin predicted (*3*) that because larval crustacean eyes are minute, the underlying photonic structure must be ultracompact and efficient, unlike the disordered multilayer reflectors of fish scales (*32*), which occupy space unavailable in the larval eyes. However, the most efficient photonic structures—ordered quarter-wave stacks and photonic crystals (*25*)—produce iridescence (*25*), meaning that the color would vary with viewing angle, which is not usually (*33*) advantageous for crypsis. The use of a photonic glass, comprising high-index isoxanthopterin particles, enables efficient, angle-independent colors to be produced, simultaneously fulfilling the demands of crypsis (from all angles) and compactness. There is a strong selective pressure for transparent prey animals to reduce the size and visibility of their eyes (*5*, *6*). However, small eyes have lower resolution, and thus there is an intrinsic trade-off between the visibility of the eyes and the ability to see (*3*). The presence of reflective cell projections between the rhabdoms suggests that the eyeshine reflector may also function to screen adjacent photoreceptors (*20*). The enhanced screening efficiency provided by a reflective material (compared with an absorbing screening pigment) may enable the rhabdoms to be more tightly packed, mitigating against the cost to visual acuity of having minute eyes. Our studies elucidate design strategies for tuning the properties of biocompatible optical materials: by controlling the size, ordering, hollowness, and birefringence of the component particles.

REFERENCES AND NOTES

1. S. Johnsen, *Biol. Bull.* **201**, 301–318 (2001).
2. T. W. Cronin, M. J. Bok, C. Lin, *Integr. Comp. Biol.* **57**, 1139–1150 (2017).
3. K. D. Feller, T. W. Cronin, *J. Exp. Biol.* **217**, 3263–3273 (2014).
4. S. Johnsen, *Ann. Rev. Mar. Sci.* **6**, 369–392 (2014).
5. R. Svanbäck, F. Johansson, *Proc. Biol. Sci.* **286**, 20182625 (2019).
6. D. E. Nilsson, *Nature* **302**, 818–821 (1983).
7. D. Gur *et al*., *Adv. Sci.* **5**, 1800338 (2018).
8. S. Harzsch, J. Benton, R. R. Dawirs, B. Beltz, *J. Neurobiol.* **39**, 294–306 (1999).
9. J. K. Douglass, R. B. Forward Jr., *Cell Tissue Res.* **258**, 289–300 (1989).
10. K. Anger, L. Hayd, *Aquat. Biol.* **7**, 19–30 (2009).
11. T. W. Cronin, R. N. Jinks, *Am. Zool.* **41**, 1098–1107 (2001).
12. L. Schertel *et al*., *Adv. Opt. Mater.* **7**, 1900442 (2019).
13. S. Magkiriadou, J. G. Park, Y. S. Kim, V. N. Manoharan, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **90**, 062302 (2014).
14. G. Jacucci, S. Vignolini, L. Schertel, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23345–23349 (2020).
15. M. J. Henze, O. Lind, B. D. Wilts, A. Kelber, *J. R. Soc. Interface* **16**, 20180785 (2019).
16. H. Noh *et al*., *Adv. Mater.* **22**, 2871–2880 (2010).
17. M. F. Land, D. E. Nilsson, *Animal Eyes* (Oxford Univ. Press, ed. 2, 2012).
18. B. A. Palmer *et al*., *Nat. Nanotechnol.* **15**, 138–144 (2020).
19. B. A. Palmer *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2299–2304 (2018).
20. E. Hallberg, R. Elofsson, *J. Crustac. Biol.* **9**, 359–372 (1989).
21. K. D. Feller *et al*., *Curr. Biol.* **29**, 3101–3108.e4 (2019).
22. D. G. Bowers, D. Evans, D. N. Thomas, K. Ellis, P. J. B. Williams, *Estuar. Coast. Shelf Sci.* **59**, 13–20 (2004).
23. R. B. Jr., *Oceanogr. Mar. Biol. Annu. Rev.* **26**, 1–393 (1988).
24. J. Marshall, J. Kent, T. Cronin, in *Adaptive Mechanisms in the Ecology of Vision*, S. N. Archer, M. B. A. Djamgoz, E. R. Loew, J. C. Partridge, S. Vallerga, Eds. (Springer, 1999), pp. 285–327.
25. S. Kinoshita, S. Yoshioka, J. Miyazaki, *Rep. Prog. Phys.* **71**, 076401 (2008).
26. S. Fraden, G. Maret, *Phys. Rev. Lett.* **65**, 512–515 (1990).
27. B. D. Wilts, K. Michielsen, H. De Raedt, D. G. Stavenga, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4363–4368 (2014).
28. B. A. Palmer, D. Gur, S. Weiner, L. Addadi, D. Oron, *Adv. Mater.* **30**, e1800006 (2018).
29. B. A. Palmer *et al*., *Science* **358**, 1172–1175 (2017).
30. D. G. Stavenga *et al*., *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **203**, 23–33 (2017).
31. M. Santon, P.-P. Bitton, U. K. Harant, N. K. Michiels, *Sci. Rep.* **8**, 7368 (2018).
32. T. M. Jordan, J. C. Partridge, N. W. Roberts, *Nat. Photonics* **6**, 759–763 (2012).
33. K. Kjernsmo *et al*., *Curr. Biol.* **30**, 551–555.e3 (2020).
34. C. F. Bohren, D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, 2008).
35. J. K. Percus, G. J. Yevick, *Phys. Rev.* **110**, 1–13 (1958).
36. J. C. M. Garnett, *Philos. Trans. Royal Soc. A* **203**, 385–420 (1904).
37. K. Shavit *et al*., Data and code for: "A tunable reflector enabling crustaceans to see but not be seen". Zenodo (2023); https://doi.org/10.5281/zenodo.7566611.

ACKNOWLEDGMENTS

We acknowledge M. Pavan for assistance with Raman spectroscopy; T. Levy, Y. Molcho, and R. Manor for valuable discussions on the manuscript; and E. Sestieri for help in shrimp collection. **Funding:** This work was supported by ERC Starting Grant (grant 852948, "CRYSTALEYES") awarded to B.A.P.; B.A.P. is the Nahum Guzik Presidential Recruit and a recipient of the 2019 Azrieli Faculty Fellowship. This work was also supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement (893136) awarded to J.H. and a Swiss National Science Foundation Grant (grant 40B1-0_198708) awarded to L.S. Electron microscopy studies were supported by the Ilse Katz Institute for Nanoscale Science & Technology at Ben-Gurion University of the

Negev. **Author contributions:** Conceptualization: K.S., V.J.Y., and B.A.P. Methodology: K.S., L.S., V.J.Y., J.H., and B.A.P. Investigation: K.S., A.W., L.S., V.F., D.A., G.Z., A.U., J.H., and B.A.P. Funding acquisition: B.A.P. Supervision: A.S. and B.A.P. Writing – original draft: K.S., L.S., J.H., and B.A.P. Writing – review and editing: K.S., A.W., L.S., V.F., D.A., J.H., and B.A.P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Raw data and relevant code can be found at (*37*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add4099
Materials and Methods
Figs. S1 to S22
Table S1
References (*38*–*50*)
MDAR Reproducibility Checklist

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

Submitted 9 June 2022; accepted 24 November 2022
10.1126/science.add4099

NEUROSCIENCE

Psychedelics promote neuroplasticity through the activation of intracellular 5-HT2A receptors

Maxemiliano V. Vargas[1,2], Lee E. Dunlap[2,3]†, Chunyang Dong[4]†, Samuel J. Carter[2,3], Robert J. Tombari[2,3], Shekib A. Jami[5], Lindsay P. Cameron[1], Seona D. Patel[3], Joseph J. Hennessey[6], Hannah N. Saeger[2,7], John D. McCorvy[6], John A. Gray[2,5,8], Lin Tian[2,5,9], David E. Olson[2,3,5,9]*

Decreased dendritic spine density in the cortex is a hallmark of several neuropsychiatric diseases, and the ability to promote cortical neuron growth has been hypothesized to underlie the rapid and sustained therapeutic effects of psychedelics. Activation of 5-hydroxytryptamine (serotonin) 2A receptors (5-HT2ARs) is essential for psychedelic-induced cortical plasticity, but it is currently unclear why some 5-HT2AR agonists promote neuroplasticity, whereas others do not. We used molecular and genetic tools to demonstrate that intracellular 5-HT2ARs mediate the plasticity-promoting properties of psychedelics; these results explain why serotonin does not engage similar plasticity mechanisms. This work emphasizes the role of location bias in 5-HT2AR signaling, identifies intracellular 5-HT2ARs as a therapeutic target, and raises the intriguing possibility that serotonin might not be the endogenous ligand for intracellular 5-HT2ARs in the cortex.

Dysregulation of the cortex has been hypothesized to play an important role in the pathophysiology of mental illnesses such as depression and often manifests as structural changes, including decreased dendritic arbor complexity and reduced dendritic spine density (*1*–*3*). Traditional antidepressants, such as selective serotonin reuptake inhibitors (SSRIs), can rescue these deficits after chronic treatment, although it seems that their effects may be independent of serotonin and perhaps involve the activation of tropomyosin receptor kinase B (TrkB) signaling (*4*, *5*). A class of therapeutic compounds known as psychoplastogens (*6*) are differentiated from SSRIs by their ability to produce both rapid and sustained effects on structural plasticity and behavior after a single administration (*7*). Psychoplastogens include both ketamine and serotonergic psychedelics, although their primary targets are distinct (*7*).

Psychedelics are 5-hydroxytryptamine (serotonin) 2A receptor (5-HT2AR) agonists that can lead to profound changes in perception, cognition, and mood (*8*). Recent evidence suggests that they promote cortical structural and functional neuroplasticity through activation of 5-HT2ARs (*9*, *10*). The mechanism by which 5-HT2AR activation leads to changes in neuronal growth is still poorly defined, although it appears to involve TrkB, mechanistic target of rapamycin (mTOR), and AMPA receptor signaling (*11*). It is currently unclear why some 5-HT2AR ligands can promote neuroplasticity and produce sustained therapeutic behavioral responses in the absence of hallucinogenic effects, whereas other 5-HT2AR agonists do not promote plasticity at all (*12*–*15*). Indeed, serotonin itself does not produce psychedelic-like effects on neuronal growth when administered to cortical cultures (*9*). This enigmatic finding cannot be easily explained by traditional biased agonism because serotonin is a balanced agonist of the 5-HT2AR that exhibits high potency and efficacy for activating both heteromeric guanine nucleotide–binding protein (G protein) and β-arrestin pathways (*16*, *17*).

Unlike psychedelics, the physicochemical properties of serotonin prevent it from entering cells by passively diffusing across nonpolar membranes (*8*). Thus, we reasoned that another form of functional selectivity, known as location bias, might explain the difference in cellular signaling elicited by serotonin and psychedelics (*18*, *19*). Here, we leveraged both chemical design and genetic manipulation to test the hypothesis that activation of an intracellular population of 5-HT2ARs is necessary for 5-HT2AR ligands to induce cortical structural plasticity and produce antidepressant-like behavioral responses.

Lipophilicity correlates with psychoplastogenicity

To firmly establish the role of 5-HT2AR activation in psychedelic-induced spinogenesis, we administered 5-methoxy-*N*,*N*-dimethyltryptamine (5-MeO) to wild-type (WT) and 5-HT2AR knockout (KO) mice (*20*) and assessed structural and functional changes in layer 5 pyramidal neurons of the prefrontal cortex (PFC) 24 hours later (fig. S1, A to C). Golgi-Cox staining revealed that 5-MeO increased spine density in both male and female animals, and this effect was absent in 5-HT2AR KO mice (fig. S1, A and B). Furthermore, ex vivo electrophysiology confirmed that 5-HT2AR activation is necessary for 5-MeO to produce sustained increases in both the frequency and amplitude of spontaneous excitatory postsynaptic currents (sEPSCs) (fig. S1C).

Next, we determined how the structures of 5-HT2AR ligands affect their abilities to promote neuronal growth by treating embryonic rat cortical neurons with serotonin, tryptamine (TRY), 5-methoxytryptamine (5-MeO–TRY), and their corresponding *N*-methyl and *N*,*N*-dimethyl congeners (Fig. 1A) before assessing neuronal morphology by means of Sholl analysis (*21*). Ketamine was used as a positive control because of its known ability to induce structural plasticity in this assay (*9*). These structure-activity relationship (SAR) studies revealed that increasing N-methylation led to an enhanced ability to promote neuronal growth, with the *N*,*N*-dimethyl compounds increasing dendritic arbor complexity to the greatest extent (Fig. 1, B to D).

Increasing N-methylation is known to affect the efficacy of 5-HT2AR signaling, so we attempted to correlate psychoplastogenic efficacy across a range of 5-HT2AR ligands with efficacy in a traditional [^{3}H]–inositol phosphates (IP) accumulation assay (Fig. 1E) (*22*). Notably, we did not observe a positive correlation between psychoplastogenic effects and ligand efficacy. Indeed, there seemed to be a non-significant inverse correlation between [^{3}H]-IP accumulation and dendritogenesis efficacy (Fig. 1E). To avoid potential issues associated with the amplification of secondary messengers, we used psychLight2, a fluorescent biosensor that is capable of directly detecting changes in 5-HT2AR conformation (*14*). PsychLight2 efficacy closely mirrored that observed by using [^{3}H]-IP accumulation assays, although psychLight2 efficacy exhibited an even stronger anticorrelation with dendritogenesis efficacy (P = 0.06) (Fig. 1F). Lastly, we used bioluminescence resonance energy transfer (BRET) assays to directly measure G_q activation or β-arrestin-2 recruitment (fig. S2) (*23*). Both measures of 5-HT2AR efficacy exhibited a strong negative correlation with psychoplastogenicity (Fig. 1, G and H). Moreover, both G_q activation and β-arrestin recruitment correlated well with psychLight efficacy [coefficient of determination (R^2) = 0.9; P < 0.0001 and P = 0.0006, respectively] (Fig. 1I). Negative correlation between 5-HT2AR efficacy and psychoplastogenicity should be interpreted with caution because this relationship may only apply to tryptamine-based ligands or compounds that exhibit a threshold level of 5-HT2AR activation.

[1]Neuroscience Graduate Program, University of California, Davis, Davis, CA 95618, USA. [2]Institute for Psychedelics and Neurotherapeutics, University of California, Davis, Davis, CA 95618, USA. [3]Department of Chemistry, University of California, Davis, Davis, CA 95616, USA. [4]Biochemistry, Molecular, Cellular, and Developmental Biology Graduate Program, University of California, Davis, Davis, CA 95616, USA. [5]Center for Neuroscience, University of California, Davis, Davis, CA 95618, USA. [6]Department of Cell Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [7]Pharmacology and Toxicology Graduate Program, University of California, Davis, Davis, CA 95616, USA. [8]Department of Neurology, School of Medicine, University of California, Davis, Sacramento, CA 95817, USA. [9]Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis, Sacramento, CA 95817, USA.
*Corresponding author. Email: deolson@ucdavis.edu
†These authors contributed equally to this work.

Nitrogen methylation of 5-HT2AR ligands structurally related to serotonin is known to result in partial agonism (*22*), but it also has a profound effect on their physicochemical properties. These compounds display a wide range of lipophilicities that ranged from highly polar molecules such as serotonin to relatively nonpolar compounds such as *N*,*N*-dimethyltryptamine (DMT). Using calculated LogP (cLogP) values, we observed a significant positive correlation with psychoplastogenic effects—more lipophilic agonists exhibited greater abilities to promote structural plasticity than polar compounds (Fig. 1J). This relationship was evident within the TRY, serotonin, and 5-MeO–TRY scaffolds. The finding that lipophilicity was a better predictor of psychoplastogenicity than 5-HT2AR activation led us to hypothesize that an intracellular pool of 5-HT2ARs in cortical neurons might be responsible for psychedelic-induced neuronal growth.

Primary localization of 5-HT2ARs in cortical neurons is intracellular

Although most G protein–coupled receptors (GPCRs) are believed to be localized primarily to the plasma membrane, several exhibit substantial intracellular localization (*24*–*27*). In vitro and ex vivo experiments have also established the existence of large intracellular pools of 5-HT2ARs in various cell types in the absence of a ligand (*28*–*31*). To compare 5-HT2AR localization patterns between cell types, we expressed a Myc–5-HT2AR–enhanced cyan fluorescent protein (ECFP) construct in both human embryonic kidney 293T (HEK293T) cells and cortical neurons, performed live-cell imaging, and assessed colocalization with a membrane dye (Cellbrite Steady) that labels the extracellular side of the plasma membrane. Because overexpression of tagged receptor constructs might alter trafficking and localization, we included several controls. We used a β2 adrenergic receptor tagged with ECFP (β2AR-ECFP) as a GPCR that is canonically described as being plasma membrane bound, and we used an ECFP construct to establish the localization of a fluorescent protein that is not tagged to a GPCR (*32*, *33*).

When expressed in HEK293T cells that were cultured in the absence of serum, 5-HT2ARs and β2ARs exhibited similar cellular expression patterns (Fig. 2A) and possessed correlation coefficients with the plasma membrane marker that were not statistically different (Fig. 2B). However, these two GPCRs displayed distinct localization patterns in neurons. In cortical neurons, β2AR expression was more highly correlated with the plasma membrane marker than was 5-HT2AR expression (Fig. 2B), which demonstrates that overexpression of a GPCR in cortical neurons does not necessarily lead to intracellular localization. The extent of overexpression was similar for 5-HT2ARs and β2ARs in both HEK293T cells and cortical neurons (fig. S3A). We performed these localization experiments without serum in the culturing media because serum contains serotonin, which can lead to agonist-induced changes in trafficking and localization (fig. S3, B and C). In both neurons and HEK293T cells,

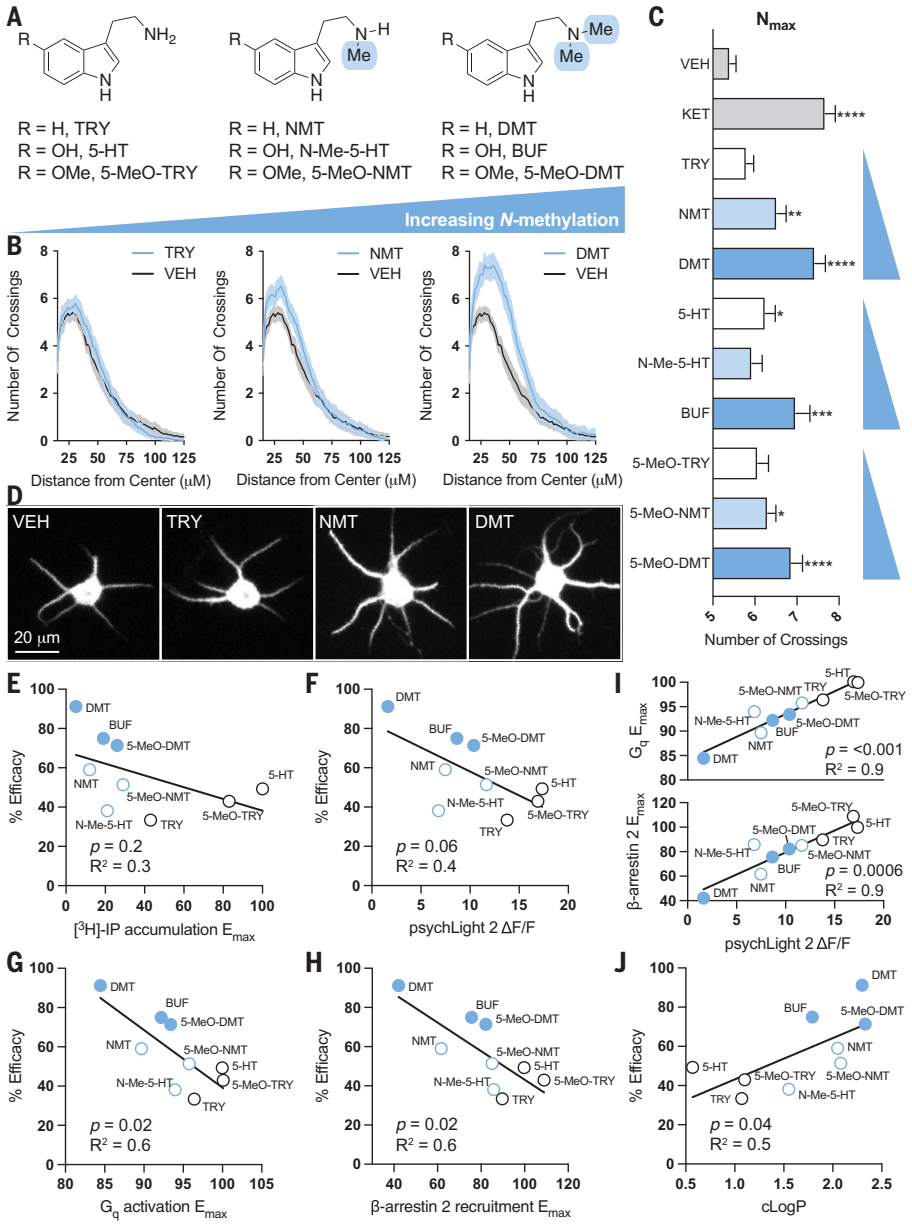


Fig. 1. Compound-induced neuronal growth correlates with ligand lipophilicity. (**A**) Chemical structures of serotonin, TRY, and 5-MeO–TRY as well as their corresponding *N*-methyl and *N*,*N*-dimethyl analogs. 5-HT, serotonin; BUF, bufotenin; Me, methyl; N-Me-5-HT, *N*-methylserotonin; NMT, *N*-methyltryptamine. (**B**) Sholl analysis demonstrates that increasing N-methylation leads to a concomitant increase in dendritic arbor complexity. The shaded area represents 95% confidence intervals. Sholl plots were generated from rat embryonic cortical neurons (DIV6) treated with compounds (10 μM). VEH, vehicle. (**C**) Maximum numbers of crossings (N_{max}) of the Sholl plots in (B) (N = 45 to 64 neurons per treatment). Error bars represent standard error of the mean. KET, ketamine. (**D**) Widefield images of rat embryonic cortical neurons (DIV6) treated with compounds (10 μM). (**E** to **J**) Correlation plots of Sholl analysis percent efficacy (N_{max} values relative to 10 μM ketamine as the positive control) versus [^{3}H]-IP accumulation (E), activation of psychLight2 (F), G_q activation (G), β-arrestin recruitment (H), or calculated LogP (J). PsychLight2 activation correlates well with both G_q activation and β-arrestin recruitment (I). Compounds were treated at 10 μM. Data for [^{3}H]-IP accumulation were obtained from literature values (*22*). $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$, as compared with VEH controls [one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparisons test]. R^2 values were calculated through simple liner regression. ΔF/F, change in fluorescence intensity relative to the baseline fluorescence intensity; E_{max}, maximum effect.

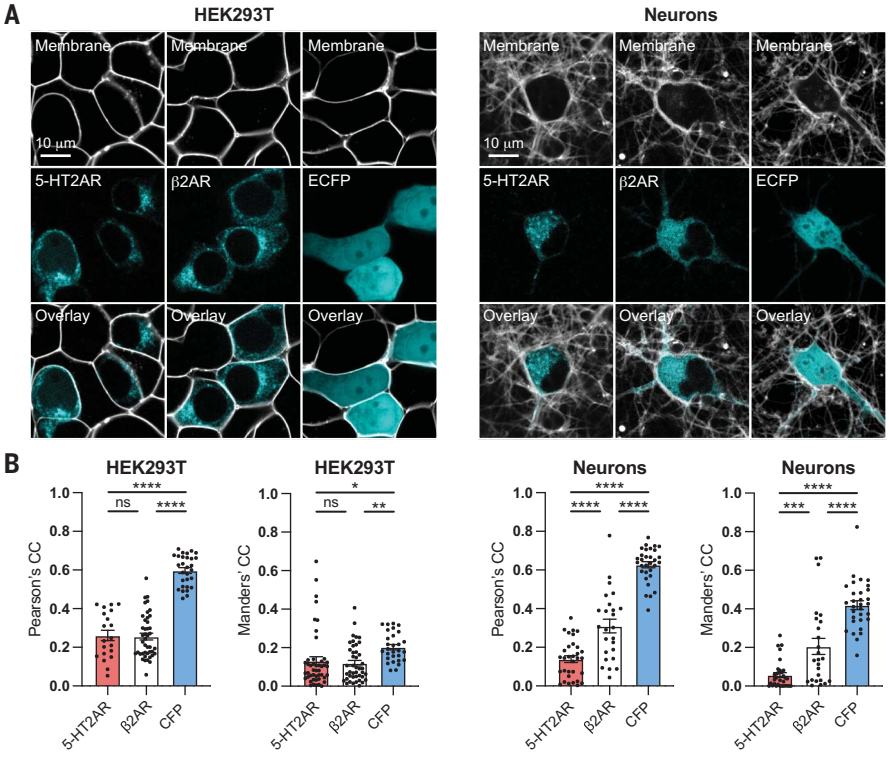


Fig. 2. Cortical neurons express intracellular 5-HT2A receptors. (**A**) Live-cell images of HEK293T cells and rat embryonic cortical neurons (DIV6) expressing Myc–5-HT2AR–CFP, β2AR-ECFP, or ECFP. Signals from the fluorescent protein and fluorescent plasma membrane marker (Cellbrite Steady) are shown in cyan and white, respectively. The expression patterns of 5-HT2ARs and β2ARs are comparable in HEK293T cells. However, in neurons, β2ARs exhibit higher expression on the plasma membrane than 5-HT2ARs. The expression of ECFP is diffuse in both HEK293T cells and cortical neurons. (**B**) Pearson's correlation coefficients (Pearson's CCs) and Manders' colocalization coefficients (Manders' CCs) quantify the extent of colocalization between MYC–5-HT2AR–CFP, β2AR-ECFP, or ECFP and the fluorescent plasma membrane marker (N = 20 to 43 cells per group). Error bars represent standard error of the mean. ns is not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$; bars indicate comparisons between data (one-way ANOVA followed by Tukey's multiple comparisons test).

the expression patterns of the tagged GPCRs were markedly distinct from ECFP, which confirms that the GPCR component of the constructs dictates cellular localization. Expression of a FLAG–5-HT2AR construct in rat cortical neurons produced a similar intracellular localization pattern, which suggests that these tags do not substantially alter the localization patterns of the 5-HT2AR (*34*).

Bright punctate staining within neurons suggested that 5-HT2ARs were localized within intracellular compartments (Fig. 2A), so we performed additional immunocytochemistry experiments to assess the overlap with markers of various subcellular organelles (fig. S4). We observed a high level of 5-HT2AR colocalization with Rab5 and Rab7 in both HEK293T cells and neurons. Ras-related guanosine triphosphatases (GTPases) of the Rab family are known to regulate intracellular transport of GPCRs (*35*). We observed a very large difference in 5-HT2AR colocalization with the Golgi apparatus between neurons and HEK293T cells, with the former exhibiting substantially higher correlation coefficients (fig. S4). The Golgi apparatus is a key regulator of GPCR signaling, and ligands can either passively diffuse into Golgi-localized receptor pools or gain access by facilitated transmembrane transport (*18*, *36*, *37*). Signaling from within the Golgi can be distinct, as is the case for opioid receptors (*38*).

To ensure that high 5-HT2AR colocalization with the Golgi was not an artifact of overexpression, we assessed native 5-HT2AR expression in neurons using one of the few validated 5-HT2AR antibodies (*30*). To confirm the antibody's specificity, we performed in-house validation by overexpressing 5-HT2ARs in HEK293T cells (fig. S5A) and we used mouse 5-HT2AR KO cortical neurons (fig. S5B). Next, we imaged rat embryonic cortical neurons that expressed only native 5-HT2ARs or overexpressed a Myc–5-HT2AR–ECFP construct. Longitudinal and transverse line scans indicated that Myc–5-HT2AR–ECFP expression closely mirrored the native localization pattern of 5-HT2ARs (fig. S5, C and D).

Membrane permeability is required for psychedelic-induced neuroplasticity

Given that cortical neurons have a large pool of intracellular 5-HT2ARs, we next used chemical tools to determine whether activation of this intracellular population was essential for psychedelics to promote structural neuroplasticity. Chemical modification of the membrane-permeable ligands DMT, psilocin (PSI), and ketanserin (KTSN) converted them into highly charged membrane-impermeable congeners *N*,*N*,*N*-trimethyltryptamine (TMT), psilocybin (PSY), and methylated ketanserin (MKTSN), respectively (Fig. 3A). All of the charged species exhibited negative cLogP scores (fig. S6A) but retained affinity for 5-HT2ARs, as determined by radioligand competition binding experiments (fig. S6B). In psychLight2 assays, the membrane-impermeable analogs displayed comparable efficacies to their uncharged parent molecules with reduced potencies (fig. S6, C and D).

Next, we treated freshly dissected rat embryonic cortical neurons with DMT (1 μM) and PSI (1 μM) as well as their membrane-impermeable congeners in the presence and absence of electroporation. Electroporation creates temporary openings in the plasma membrane, which enables highly charged molecules to pass through and access the intracellular space. Although the membrane-permeable 5-HT2AR agonists were able to promote dendritogenesis regardless of whether electroporation was applied, the membrane-impermeable compounds could only promote neuronal growth when applied with electroporation (Fig. 3, B and C). Similarly, the membrane-permeable 5-HT2AR antagonist KTSN (treated in 10-fold excess at 10 μM) blocked DMT-induced plasticity both with and without electroporation; however, the membrane-impermeable antagonist MKTSN (10 μM) could only block DMT-induced neuronal growth when it was applied with electroporation (Fig. 3, B and C). By using more mature neurons (DIV15), we demonstrated that the membrane-permeable 5-HT2AR agonists increased dendritic spine density—another measure of structural neural plasticity—whereas the membrane-impermeable agonists did not (Fig. 3, D and E). KTSN was able to inhibit the effects of DMT and PSI on dendritic spine density, whereas MKTSN was not (fig. S7).

To further establish that membrane-permeable and -impermeable ligands target different populations of 5-HT2ARs, we performed an experiment in HEK293T cells that express psychLight1. The psychLight1 construct was chosen over psychLight2 because the former lacks an endoplasmic reticulum (ER) export sequence, which results in greater intracellular localization (*14*). PsychLight1-expressing HEK293T cells were pretreated with either

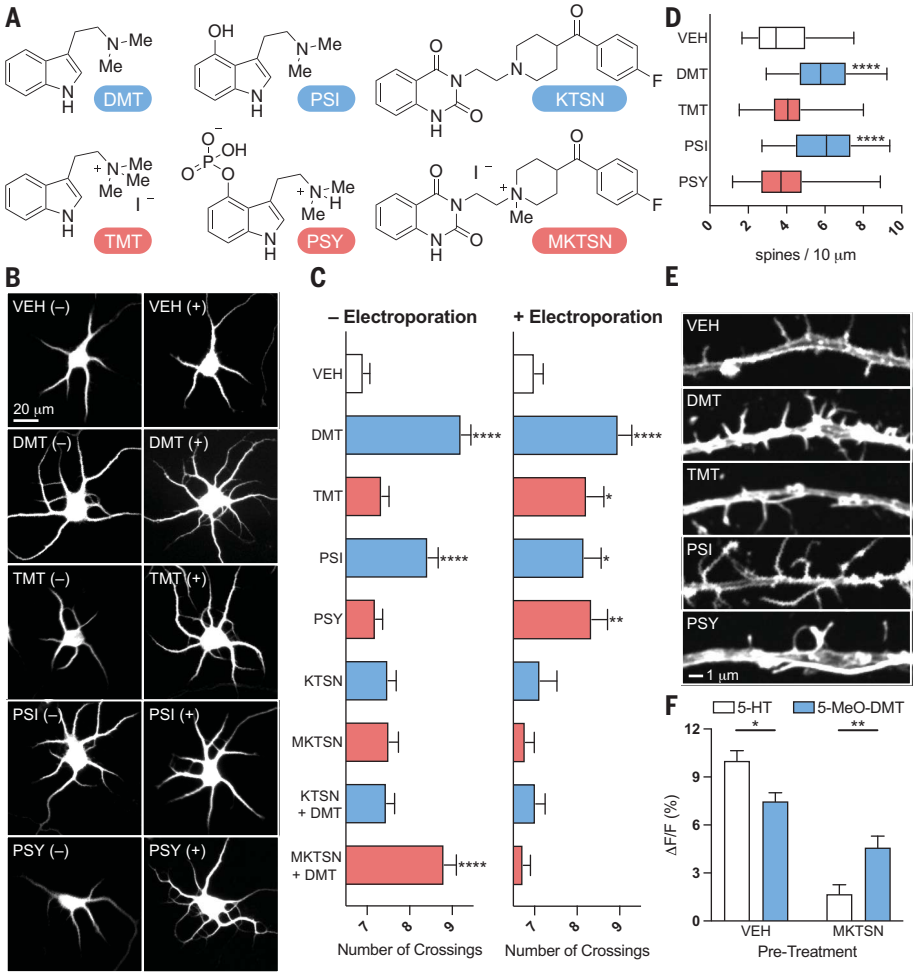


Fig. 3. Intracellular 5-HT2ARs mediate structural plasticity induced by serotonergic psychoplastogens. (**A**) Structures of membrane-permeable (blue) and impermeable (red) compounds. (**B**) Widefield images of rat embryonic cortical neurons (DIV6) that were administered compounds with (+) and without (–) electroporation. (**C**) N_{max} values obtained from Sholl plots. Membrane-impermeable analogs of psychedelics (1 μM) only promote growth when administered with electroporation (N = 35 to 110 neurons per treatment). Unlike KTSN (10 μM), MKTSN (10 μM) only blocks psychedelic-induced plasticity when administered with electroporation (N = 45 to 109 neurons per treatment). (**D**) Membrane-impermeable agonists of 5-HT2ARs (1 μM) cannot promote spinogenesis in cultured embryonic rat cortical neurons (DIV15) (N = 30 to 34 neurons per treatment). (**E**) Confocal images of treated cortical neurons (DIV15). (**F**) HEK293T cells that expressed psychLight1 were pretreated with VEH or MKTSN (10 μM) before the administration of serotonin (10 μM) or 5-MeO–DMT (10 μM) (N = 41 to 54 cells per treatment). Error bars in (C) and (F) represent standard error of the mean. *p < 0.05, **p < 0.01, and ****p < 0.0001, as compared with VEH controls in (D) and (C) (one-way ANOVA followed by Dunnett's multiple comparisons test) or between indicated pairs of data in (F) (two-way ANOVA followed by Šídák's multiple comparisons test). For box-and-whisker plots in (D), the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values.

vehicle or MKTSN to selectively antagonize the population of 5-HT2ARs that were expressed on the plasma membrane. Next, changes in psychLight1 fluorescence were measured after administration of the membrane-impermeable ligand serotonin or its membrane-permeable congener 5-MeO–DMT. Because serotonin is a full agonist, it induced a stronger response in the absence of antagonist compared with the partial agonist 5-MeO–DMT. Pretreatment with MKTSN resulted in a much larger reduction in the serotonin-induced psychLight1 signal compared with that induced by 5-MeO–DMT (Fig. 3F). Pretreatment with MKTSN could nearly fully antagonize the effect of serotonin. In sharp contrast, MKTSN only partially blocked the ability of 5-MeO–DMT to turn on psychLight1 fluorescence.

Because serotonin and 5-MeO–DMT exhibit different lipophilicities (cLogP values of 0.57 and 2.33, respectively), we reasoned that their abilities to displace [^{3}H]–D-lysergic acid diethylamide (LSD) bound to the 5-HT2AR would depend on whether those receptors were exposed to the extracellular environment. Thus, we performed radioligand competition binding experiments using intact HEK293T cells that expressed Myc–5-HT2AR or psychLight2 as well as membrane preparations obtained from these systems. The inhibition constant (K_i) values for serotonin and 5-MeO–DMT were nearly identical when using membrane preparations or intact PSYLI2 cells, with HEK293T cells that stably expressed a 5-HT2AR construct with an ER export sequence resulting in a large proportion of the 5-HT2ARs being exposed to the extracellular environment (fig. S8). However, 5-MeO–DMT was an order of magnitude more potent than serotonin when using intact HEK293T cells that expressed large populations of both plasma membrane–bound and intracellular 5-HT2ARs (fig. S8).

To confirm that serotonin and 5-MeO–DMT can target distinct populations of 5-HT2ARs, we performed inositol monophosphate (IP1) assays in cortical neurons and HEK293T cells that expressed the Myc–5-HT2AR–ECFP construct. When the assay was performed in HEK293T cells that expressed Myc–5-HT2AR–ECFP, which display a large proportion of plasma membrane–bound 5-HT2ARs, serotonin and 5-MeO–DMT had comparable potencies and efficacies (fig. S9A). When the same experiment was performed in rat cortical neurons, serotonin failed to elicit an agonist response, although 5-MeO–DMT remained a potent agonist (fig. S9B).

Cellular import of serotonin leads to structural plasticity and antidepressant-like effects

If activation of intracellular 5-HT2ARs in cortical neurons is sufficient to promote structural plasticity, we hypothesized that serotonin should be able to promote cortical neuron growth if given access to the intracellular space. To test this hypothesis, we first treated cortical neurons with serotonin (1 μM) in the presence and absence of electroporation. Unlike ketamine (1 μM), serotonin was only able to promote cortical neuron growth when applied with electroporation (Fig. 4A). Next, we took advantage of the serotonin transporter (SERT), which can import serotonin from the extracellular environment (*39*). Endogenous expression of SERT is typically restricted to presynaptic terminals of neurons that emanate from the raphe, and thus, rat cortical neurons do not express appreciable levels of the transporter (*40*, *41*). Embryonic rat cortical neurons were electroporated with an enhanced yellow fluorescent protein (EYFP)–tagged SERT construct under the control of the calcium/calmodulin-dependent protein kinase II (CaMKII) promoter to restrict expression to excitatory pyramidal neurons. Sparse transfection resulted in cultures

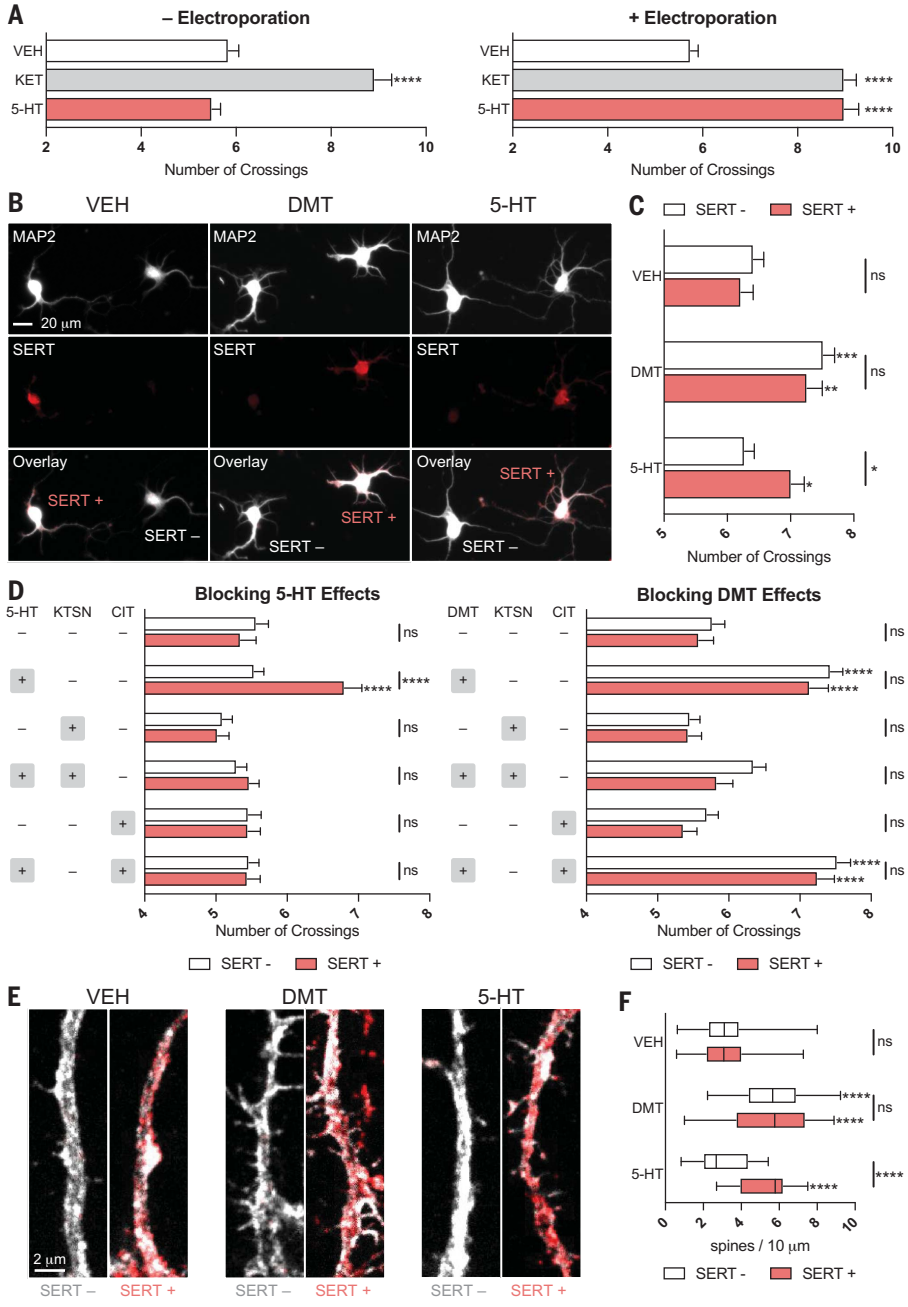


Fig. 4. Cellular uptake of serotonin induces structural plasticity in vitro. (**A**) Rat embryonic cortical neurons were administered compounds (1 μM) with (+) and without (–) electroporation. N_{max} values obtained from Sholl plots (DIV6) demonstrates that unlike ketamine, serotonin only promotes growth when administered with electroporation (N = 49 to 59 neurons per treatment). (**B**) Widefield images of embryonic rat cortical cultures (DIV6) sparsely transfected with CaMKII-SERT-EYFP and treated with compounds. (**C**) N_{max} values obtained from Sholl plots demonstrate that serotonin (10 μM) can only increase the dendritic arbor complexity of SERT-positive neurons, whereas DMT (10 μM) can promote the growth of both SERT-positive and SERT-negative neurons (N = 69 to 93 neurons per treatment). (**D**) KTSN pretreatment (10 μM) blocks the plasticity-promoting effects of serotonin (1 μM) in SERT-positive neurons and DMT (1 μM) in both SERT-positive and SERT-negative neurons. CIT (10 μM) only blocks the plasticity-promoting effects of serotonin (1 μM) in SERT-positive neurons (N = 59 to 100 neurons per treatment). (**E**) Confocal images of CaMKII-SERT-EYFP positive and negative dendrites (DIV15) treated with compounds (10 μM). (**F**) Serotonin only promotes spine growth in SERT-positive neurons (N = 11 to 35 neurons per treatment). Error bars in (A), (C), and (D) represent standard error of the mean. ns is not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$, as compared with VEH controls in (A) (one-way ANOVA followed by Dunnett's multiple comparisons test) or compared with VEH controls from the same genotype in (C), (D), and (F) (two-way ANOVA followed by Tukey's multiple comparisons test). For box-and-whisker plots in (F), the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values.

that expressed both SERT-positive and SERT-negative neurons (Fig. 4B), which enabled us to compare the effects of serotonin on neurons capable of importing the monoamine with those that could not within the same cultures.

Treatment with the membrane-permeable psychedelic DMT (10 μM) resulted in greater dendritic arbor complexity and increased spine density in both SERT-positive and SERT-negative neurons (Fig. 4, B to F). Only SERT-positive neurons treated with serotonin (10 μM) displayed increased dendritogenesis and spinogenesis (Fig. 4, B to F). To ensure that serotonin-induced changes in structural neural plasticity were due to the engagement of intracellular 5-HT2Rs through serotonin importation from SERT, we performed similar experiments in the presence of the selective SERT inhibitor citalopram (CIT) and KTSN. When these experiments were performed in the presence of CIT (10 μM), the plasticity-promoting effects of serotonin (1 μM) on SERT-positive neurons were blocked (Fig. 4D). By contrast, CIT had no effect on the ability of DMT (1 μM) to promote the growth of SERT-positive or SERT-negative cortical neurons (Fig. 4D). In the presence of KTSN (10 μM), neither serotonin nor DMT could promote neuronal growth, which confirms that DMT and intracellular serotonin promote plasticity by means of 5-HT2AR receptors in vitro (Fig. 4D).

To determine whether intracellular serotonin could promote the growth of cortical neurons in vivo, we injected the medial PFC (mPFC) of Thy1–enhanced green fluorescent protein (EGFP) mice with either CaMKII-SERT-mCherry or CaMKII-mCherry. The mPFC was chosen as the injection site because it exhibits high levels of 5-HT2AR expression (*28*) and has been implicated in the antidepressant-like effects of psychoplastogens (*42*). After 3 weeks to enable construct expression (Fig. 5, A and B), both groups were administered (±)-para-chloroamphetamine [PCA, 5 mg/kg intraperitoneally (ip)], a selective serotonin-releasing agent (*43*). Dendritic spine density on mPFC pyramidal neurons was assessed 24 hours later. Animals that expressed SERT in the mPFC displayed significantly higher densities of dendritic spines after PCA administration as compared with mCherry controls (Fig. 5, C and D). Notably, PCA is not a 5-HT2AR agonist (fig. S10A), does not directly promote the growth of SERT-positive or SERT-negative cortical neurons in culture (fig. S10B), and does not induce a head-twitch response (HTR) in WT mice (fig. S10C).

Evidence suggests that psychoplastogen-induced structural plasticity in the mPFC might be related to sustained antidepressant-like effects in rodents (*44*). To probe for an antidepressant-like response that might be linked to neuroplasticity, we injected an adeno-associated virus (AAV) that contained a CaMKII-SERT-EYFP or CaMKII–green fluorescent protein (GFP)

construct into the mPFC of WT C57BL/6J mice (Fig. 5E and fig. S10D). After 3 weeks to allow for construct expression, both groups of mice underwent a novelty induced locomotion (NIL) test, and no differences in locomotion were observed (Fig. 5F). After 24 hours, mice were subjected to a forced swim test (FST) (*7*, *45*). Again, we observed no differences between the mice that expressed SERT and those that expressed the GFP control (Fig. 5G). Two days later, we administered PCA (5 mg/kg ip), waited 24 hours, and then performed a second FST (Fig. 5G). The SERT-expressing mice exhibited a statistically significant HTR immediately after the administration of PCA as compared with the GFP control mice (fig. S10E), and they also displayed a significant reduction in immobility in the FST 24 hours after administration (Fig. 5G).

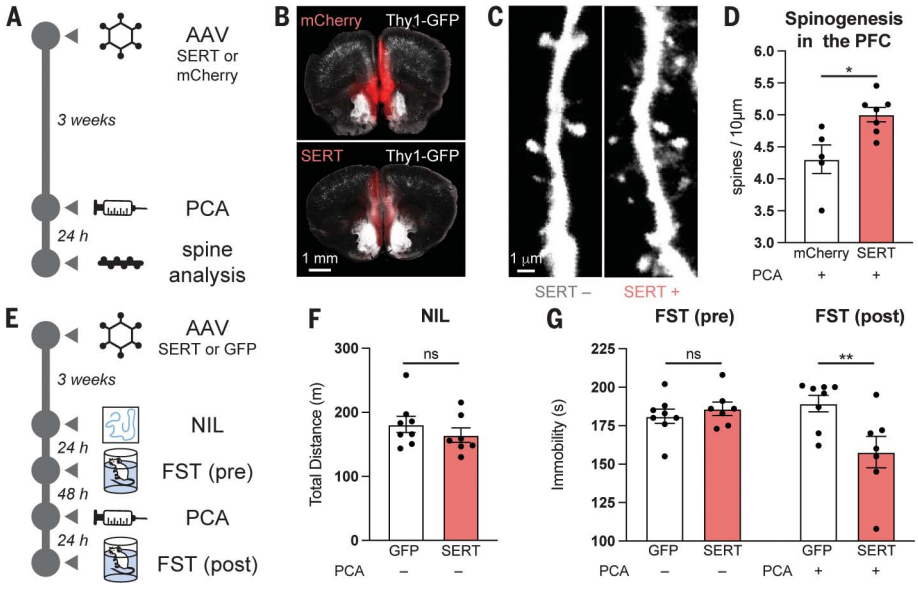


Fig. 5. Cellular uptake of serotonin produces antidepressant-like effects in vivo. (**A**) Schematic that displays experimental design for measuring spine density in Thy1-EGFP mice after administration of a serotonin-releasing agent. (**B**) Histology images of the mPFC of Thy1-EGFP mice that express CaMKII-mCherry or CaMKII-SERT-mCherry. (**C**) Confocal images of dendritic spines in the mPFC of mice treated with PCA (5 mg/kg ip). (**D**) Mice that express CaMKII-SERT-mCherry display increased dendritic spine density after PCA treatment. (**E**) Schematic that displays experimental design for determining the sustained antidepressant-like effects of serotonin in mice that express SERT in the mPFC. (**F**) NIL demonstrates no difference between CaMKII-SERT-EYFP– and CaMKII-GFP–expressing mice. (**G**) CaMKII-SERT-EYFP– and CaMKII-GFP–expressing mice exhibit no differences in the FST. After PCA (5 mg/kg ip) administration, CaMKII-SERT-EYFP–expressing mice display a sustained antidepressant-like effect. ns is not significant, $*p < 0.05$, and $**p < 0.01$, as compared between indicated pairs of data in (D) and (F) (two-tailed unpaired Student's *t* test) or (G) (two-way repeated measures ANOVA followed by Šídák's multiple comparisons test). Error bars in (D), (F), and (G) represent standard error of the mean.

Discussion

Although GPCRs are traditionally viewed as initiators of signal transduction that originates at the plasma membrane, increasing evidence suggests that GPCR signaling from intracellular compartments can play important roles in cellular responses to drugs. Recently, location bias has been proposed to explain signaling differences between endogenous membrane-impermeable peptide ligands and membrane-permeable ligands of opioid receptors (*38*). Moreover, distinct ligand-induced signaling has been observed for plasma membrane–localized and intracellular populations of δ-opioid receptors (*46*). Here, we extend the concept of location bias to ligands of the 5-HT2AR.

A substantial proportion of 5-HT2ARs in cortical neurons are localized to the Golgi, and intracellular compartments such as the Golgi are slightly acidic compared with the cytosol and extracellular space. Thus, it is possible that protonation of psychedelics within the Golgi leads to retention and sustained signaling, which results in neuronal growth, even after transient stimulation (*47*). Persistent growth after the drugs have been removed from the extracellular space is a hallmark of serotonergic psychoplastogens (*47*, *48*). Although the mechanistic details that link intracellular 5-HT2AR activation to cortical neuron growth have not been fully elucidated, they are likely to involve AMPA receptor, TrkB, and mTOR signaling, as previously established (*9*). Future studies should examine the detailed signaling interplay between these proteins.

In addition to promoting psychedelic-induced structural neuroplasticity, the intracellular population of 5-HT2ARs might also contribute to the hallucinogenic effects of psychedelics. When we administered a serotonin-releasing agent to WT mice, we did not observe a HTR. However, the same drug was able to induce a HTR in mice that expressed SERT on cortical neurons of the mPFC, a brain region that is known to be essential for the HTR (*49*). Thus, activation of intracellular cortical 5-HT2ARs may play a role in the subjective effects of psychedelics. This hypothesis is further supported by previous work that demonstrates that a high dose of the serotonin precursor 5-hydroxytryptophan (5-HTP) induces a HTR in WT mice, which can be blocked by an N-methyltransferase inhibitor that prevents the metabolism of 5-HTP to *N*-methyltryptamines (*50*). Inhibition of N-methyltransferase failed to block the HTR induced by 5-MeO–DMT (*50*). Taken together, this work emphasizes that accessing intracellular 5-HT2ARs is important for 5-HT2AR agonists to produce a HTR.

Our results demonstrate that membrane permeability is essential for a ligand to activate 5-HT2ARs in cortical neurons; however, our experiments did not distinguish between intracellular signaling or the possibility of psychedelics acting as pharmacological chaperones. Others have hypothesized that GPCR ligands may act as pharmacological chaperones, which facilitates their export to the plasma membrane where they could presumably engage in canonical signaling (*51*). Thus, future studies should determine whether intracellular 5-HT2AR signaling is distinct from 5-HT2AR signaling at the plasma membrane.

Although intracellular expression of 5-HT2ARs within cortical pyramidal neurons has been known for some time, it is unclear at present why the subcellular localization of these receptors differs greatly between neurons and HEK293T cells (*28*, *29*, *31*). One possibility is that 5-HT2ARs form heterodimeric complexes that affect cellular trafficking (*31*). Thus, by dictating 5-HT2AR localization, cellular context could influence responses to endogenous neuromodulators and/or exogenous drugs, which potentially results in circuit-specific effects of 5-HT2AR ligands.

Intracellular signaling has been hypothesized to contribute to the pharmacological properties of a diverse range of compounds that includes nicotine, ketamine, and SSRIs (*51*–*53*). Like psychedelics, these compounds are weak bases with pK_a (where K_a is the acid dissociation constant) values ranging from 7 to 10. Given that the antidepressant mechanisms of ketamine and SSRIs have not been definitively established, it is intriguing to speculate that they also might promote cortical neuron growth

by binding to intracellular targets. Perhaps other antidepressants affect the function of scaffolding proteins within the cell interior to modulate neuronal growth phenotypes.

Without facilitated transport across the plasma membrane, serotonin cannot induce psychedelic-like effects on neuronal morphology. Although it is possible that serotonin could alter cortical neuron physiology by activating cell-surface 5-HT2ARs, this receptor pool does not seem to be involved in 5-HT2AR–induced structural plasticity. Our results raise the intriguing possibility that serotonin may not be the endogenous ligand for the population of 5-HT2ARs expressed inside cortical neurons. Alternative ligands could include methylated congeners of serotonin or TRY because these compounds have greater abilities to cross nonpolar membranes. Endogenous psychedelics such as DMT, 5-MeO–DMT, and bufotenin have been identified in a variety of species, including humans, and have long been hypothesized to play roles in diseases such as schizophrenia (*54*). However, they are rapidly degraded in vivo, which makes their detection by classic analytical methods quite challenging. The use of more modern analytical techniques has improved detection of these analytes (*55*), with a recent study demonstrating that the concentration of DMT in the cortex was comparable to that of serotonin (*56*). The possibility that endogenous psychedelics play a role in health or disease should therefore be thoroughly investigated.

REFERENCES AND NOTES

1. R. S. Duman, G. K. Aghajanian, G. Sanacora, J. H. Krystal, *Nat. Med.* **22**, 238–249 (2016).
2. H. Qiao *et al*., *Neural Plast.* **2016**, 8056370 (2016).
3. R. M. Shansky, C. Hamo, P. R. Hof, B. S. McEwen, J. H. Morrison, *Cereb. Cortex* **19**, 2479–2484 (2009).
4. T. Rantamäki *et al*., *Neuropsychopharmacology* **32**, 2152–2162 (2007).
5. P. C. Casarotto *et al*., *Cell* **184**, 1299–1313.e19 (2021).
6. D. E. Olson, *J. Exp. Neurosci.* **12**, 1179069518800508 (2018).
7. M. V. Vargas, R. Meyer, A. A. Avanes, M. Rus, D. E. Olson, *Front. Psychiatry* **12**, 727117 (2021).
8. A. C. Kwan, D. E. Olson, K. H. Preller, B. L. Roth, *Nat. Neurosci.* **25**, 1407–1419 (2022).
9. C. Ly *et al*., *Cell Rep.* **23**, 3170–3182 (2018).
10. M. de la Fuente Revenga *et al*., *Cell Rep.* **37**, 109836 (2021).
11. D. E. Olson, *Biochemistry* **61**, 127–136 (2021).
12. L. E. Dunlap *et al*., *J. Med. Chem.* **63**, 1142–1155 (2020).
13. L. P. Cameron *et al*., *Nature* **589**, 474–479 (2021).
14. C. Dong *et al*., *Cell* **184**, 2779–2792.e18 (2021).
15. D. Cao *et al*., *Science* **375**, 403–411 (2022).
16. L. M. Luttrell, S. Maudsley, L. M. Bohn, *Mol. Pharmacol.* **88**, 579–588 (2015).
17. K. Kim *et al*., *Cell* **182**, 1574–1588.e19 (2020).
18. R. Irannejad *et al*., *Nat. Chem. Biol.* **13**, 799–806 (2017).
19. M. A. Mohammad Nexhady, J. C. Nezhady, S. Chemtob, *iScience* **23**, 101643 (2020).
20. N. V. Weisstaub *et al*., *Science* **313**, 536–540 (2006).
21. D. Ristanović, N. T. Milosević, V. Stulić, *J. Neurosci. Methods* **158**, 212–218 (2006).
22. B. J. Ebersole, I. Visiers, H. Weinstein, S. C. Sealfon, *Mol. Pharmacol.* **63**, 36–43 (2003).
23. R. H. J. Olsen *et al*., *Nat. Chem. Biol.* **16**, 841–849 (2020).
24. J. C. Bermak, M. Li, C. Bullock, Q. Y. Zhou, *Nat. Cell Biol.* **3**, 492–498 (2001).
25. U. E. Petäjä-Repo, M. Hogue, A. Laperriere, P. Walker, M. Bouvier, *J. Biol. Chem.* **275**, 13727–13736 (2000).
26. S. E. Jacobsen *et al*., *J. Biol. Chem.* **292**, 6910–6926 (2017).
27. C. A. Purgert *et al*., *J. Neurosci.* **34**, 4589–4598 (2014).
28. V. Cornea-Hébert, M. Riad, C. Wu, S. K. Singh, L. Descarries, *J. Comp. Neurol.* **409**, 187–209 (1999).
29. V. Cornea-Hébert *et al*., *Neuroscience* **113**, 23–35 (2002).
30. C. L. Schmid, K. M. Raehal, L. M. Bohn, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1079–1084 (2008).
31. R. Toneatti *et al*., *Sci. Signal.* **13**, eaaw3122 (2020).
32. J. E. Saffitz, S. B. Liggett, *Circ. Res.* **70**, 1320–1325 (1992).
33. Q. Fu, Y. K. Xiang, *Prog. Mol. Biol. Transl. Sci.* **132**, 151–188 (2015).
34. A. C. Magalhaes *et al*., *Nat. Neurosci.* **13**, 622–629 (2010).
35. J. L. Seachrist, S. S. Ferguson, *Life Sci.* **74**, 225–235 (2003).
36. C. A. Nash, W. Wei, R. Irannejad, A. V. Smrcka, *eLife* **8**, e48167 (2019).
37. R. Irannejad *et al*., *Nature* **495**, 534–538 (2013).
38. M. Stoeber *et al*., *Neuron* **98**, 963–976.e5 (2018).
39. S. Aggarwal, O. V. Mortensen, *Curr. Protocols Pharmacol.* **79**, 12.16.1–12.16.17 (2017).
40. F. C. Zhou, Y. Sari, J. K. Zhang, *Brain Res. Dev. Brain Res.* **119**, 33–45 (2000).
41. C. Lebrand *et al*., *J. Comp. Neurol.* **401**, 506–524 (1998).
42. M. Fuchikami *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8106–8111 (2015).
43. T. Sharp, T. Zetterström, L. Christmanson, U. Ungerstedt, *Neurosci. Lett.* **72**, 320–324 (1986).
44. R. N. Moda-Sava *et al*., *Science* **364**, eaat8078 (2019).
45. D. A. Slattery, J. F. Cryan, *Nat. Protoc.* **7**, 1009–1014 (2012).
46. S. E. Crilly, W. Ko, Z. Y. Weinberg, M. A. Puthenveedu, *eLife* **10**, e67478 (2021).
47. C. Ly *et al*., *ACS Pharmacol. Transl. Sci.* **4**, 452–460 (2020).
48. L. X. Shao *et al*., *Neuron* **109**, 2535–2544.e4 (2021).
49. J. González-Maeso *et al*., *Neuron* **53**, 439–452 (2007).
50. C. L. Schmid, L. M. Bohn, *J. Neurosci.* **30**, 13513–13524 (2010).
51. H. A. Lester, J. M. Miwa, R. Srinivasan, *Biol. Psychiatry* **72**, 907–915 (2012).
52. B. J. Henderson, H. A. Lester, *Neuropharmacology* **96**, 178–193 (2015).
53. H. A. Lester, L. D. Lavis, D. A. Dougherty, *Am. J. Psychiatry* **172**, 1064–1066 (2015).
54. S. A. Barker, E. H. McIlhenny, R. Strassman, *Drug Test. Anal.* **4**, 617–635 (2012).
55. S. A. Barker, J. Borjigin, I. Lomnicka, R. Strassman, *Biomed. Chromatogr.* **27**, 1690–1700 (2013).
56. J. G. Dean *et al*., *Sci. Rep.* **9**, 9333 (2019).
57. M. V. Vargas *et al*., Data set for Intracellular 5-HT2A Paper, Figshare (2022); https://doi.org/10.6084/m9.figshare.21669635.

ACKNOWLEDGMENTS

We thank C. Ly, A. C. Greb, L. P. Cameron, and W. C. Duim for performing early pilot studies; A. Avanes for assistance with radioligand binding studies; K. Zito for providing Thy1-GFP breeders; J. González-Maeso for providing the Myc–5-HT2AR–ECFP plasmid; R. Iyer for performing high-resolution mass spectrometry analysis; and P. Beal for use of his CLARIOStar plate reader. We also thank C. Nichols for advice about the radioligand binding experiments. **Funding:** This work was supported by funds from the National Institutes of Health (NIH) (R01GM128997 to D.E.O., R35GM133421 to J.D.M., and U01NS120820, U01NS115579, and 2R01MH101214-06 to L.T.), three NIH training grants (T32GM099608 to M.V.V., T32GM113770 to R.J.T., and T32MH112507 to H.N.S.), Human Frontier (L.T.), the Camille and Henry Dreyfus Foundation (D.E.O.), a sponsored research agreement with Delix Therapeutics (D.E.O.), and a University of California (UC) Davis Provost's Undergraduate Fellowship (S.J.C.). This project used the Biological Analysis Core of the UC Davis MIND Institute Intellectual and Development Disabilities Research Center (U54 HD079125). The Nikon High Content Analysis Spinning Disk Confocal microscope used in this study was purchased using NIH Shared Instrumentation Grant 1S10OD019980-01A1. We thank the MCB Light Microscopy Imaging Facility, which is a UC Davis Campus Core Research Facility, for the use of this microscope. Funding for the nuclear magnetic resonance spectrometers was provided by the National Science Foundation (NSF CHE-04-43516) and NIH (08POES 05707C). Analysis for this project was performed in the UC Davis Campus Mass Spectrometry Facilities, with instrument funding provided by the NIH (1S10OD025271-01A1). Several of the drugs used in this study were provided by the National Institute on Drug Abuse (NIDA) Drug Supply Program. **Author contributions:** M.V.V. performed most of the in vitro experiments, including the dendritogenesis, spinogenesis, subcellular colocalization, IP1, neuromics antibody validation, and psychLight assays. C.D. cloned and validated key reagents for the in vivo experiments, performed the surgeries, and the perfusions. M.V.V. performed the small-molecule electroporation experiments, key pilot experiments imaging HEK293T cells and neurons, and brain-slice imaging with assistance from C.D., and M.V.V. performed the behavioral experiments. L.E.D. performed the N-methylation SAR dendritogenesis experiments, calculated cLogP values, and synthesized TMT and MKTSN. R.J.T. and S.J.C. performed the radioligand binding studies. H.N.S. performed culturing of 5-HT2AR KO cultures. L.P.C. and S.D.P. performed the Golgi staining experiments. S.A.J. performed the electrophysiology experiments. J.J.H. performed BRET assays of 5-HT2AR activation. J.D.M., J.A.G., L.T., and D.E.O. supervised various aspects of this project and assisted with data analysis. D.E.O. conceived the project and wrote the manuscript with input from all authors. **Competing interests:** D.E.O. is a co-founder of Delix Therapeutics, Inc., serves as the chief innovation officer and head of the scientific advisory board, and has sponsored research agreements with Delix Therapeutics. Delix Therapeutics has licensed technology from UC Davis. The sponsors of this research were not involved in the conceptualization, design, decision to publish, or preparation of the manuscript. **Data and materials availability:** Data are available at Figshare (*57*). Custom written data analysis codes are available through GitHub (see methods for details). All materials are available upon request or are commercially available. TMT and MKTSN are available from D.E.O. and psychLight is available from L.T. under material agreements with UC Davis. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf0435
Materials and Methods
Figs. S1 to S10
References (*58*, *59*)
MDAR Reproducibility Checklist

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

Submitted 27 September 2022; accepted 9 January 2023
10.1126/science.adf0435

ELECTROCHEMISTRY

Continuous-flow electrosynthesis of ammonia by nitrogen reduction and hydrogen oxidation

Xianbiao Fu†, Jakob B. Pedersen†, Yuanyuan Zhou†, Mattia Saccoccio, Shaofeng Li, Rokas Sažinas, Katja Li, Suzanne Z. Andersen, Aoni Xu, Niklas H. Deissler, Jon Bjarke Valbæk Mygind, Chao Wei, Jakob Kibsgaard, Peter C. K. Vesborg, Jens K. Nørskov*, Ib Chorkendorff*

Ammonia is a critical component in fertilizers, pharmaceuticals, and fine chemicals and is an ideal, carbon-free fuel. Recently, lithium-mediated nitrogen reduction has proven to be a promising route for electrochemical ammonia synthesis at ambient conditions. In this work, we report a continuous-flow electrolyzer equipped with 25–square centimeter–effective area gas diffusion electrodes wherein nitrogen reduction is coupled with hydrogen oxidation. We show that the classical catalyst platinum is not stable for hydrogen oxidation in the organic electrolyte, but a platinum-gold alloy lowers the anode potential and avoids the decremental decomposition of the organic electrolyte. At optimal operating conditions, we achieve, at 1 bar, a faradaic efficiency for ammonia production of up to 61 ± 1% and an energy efficiency of 13 ± 1% at a current density of −6 milliamperes per square centimeter.

Ammonia (NH_3) is one of the essential feedstocks to produce fertilizers, pharmaceuticals, polymers, and other chemicals. In addition, it is an ideal, zero-carbon energy carrier. In 2021, a global total of 182 million metric tons of ammonia was produced in centralized industrial plants through the Haber-Bosch process, which is considered one of the most impactful technological achievements of the 20th century (*1*–*3*). However, the Haber-Bosch process consumes ~1% of the global annual energy output and generates ~1.3% of global carbon dioxide (CO_2) emissions, which are mainly associated with hydrogen (H_2) production from steam methane reforming or coal gasification (*4*). The high pressure (150 to 200 bar) needed to achieve reasonable ammonia concentrations at the high temperatures (350° to 450°C) needed for catalytic activity requires substantial capital investments, which means that ammonia production is very centralized. Unfortunately, the desirable renewable energy sources for the production process and the use of fertilizers are both decentralized, which means that there is great incentive to make smaller and delocalized production units. This would also make ammonia production less vulnerable to political instabilities and make it desirable even if energy consumption is increased. (*2*, *5*). The electrochemical ammonia synthesis directly from nitrogen (N_2) and water, powered by renewable energy, would enable distributed ammonia production in smaller-scale devices, which would bring tremendous economic and social advantages, such as decreasing the price of fertilizer in developing countries and remote areas lacking transportation networks or infrastructure and realizing a neutral carbon footprint (*6*, *7*).

Recently, despite the substantial efforts that have been devoted toward electrochemical ammonia synthesis in aqueous conditions, this field has been heavily hampered by the ambiguity of the nitrogen source for NH_3 (*8*–*10*). Several protocols and rigorous quantitative investigations have revealed ubiquitous contamination (e.g., nitrogen oxides, nitrate, nitrite, and organic nitrogen compounds) in feed gases ($^{14}N_2$ and $^{15}N_2$), catalysts, and electrolytes, which suggests that ammonia production from N_2 reduction remains unproven in aqueous systems (*9*, *10*). The lithium-mediated nitrogen reduction reaction (Li-NRR) in nonaqueous electrolytes has been validated to produce ammonia from N_2 reduction through a rigorous procedure (i.e., gas purification and quantitative isotope measurements) by our group (*10*). The Li-NRR was first investigated in an alcoholic solution of lithium halide in 1930 by Fichter *et al*. and was further optimized by Tsuneto *et al*. by changing the electrolyte to tetrahydrofuran (THF) mixed with ethanol (EtOH) and increasing nitrogen pressure to improve the current efficiency (*11*–*13*). However, it seems that none of these works were followed up thoroughly, and they vanished in a sea of false positives that left open the question of whether molecular nitrogen was the definitive source of the nitrogen in measured ammonia. Since our work (*10*), several strategies—such as potential cycling (*14*), the addition of oxygen as a promoter (*15*), the use of ionic liquid as a proton shuttle (*16*), increasing the surface area of electrodes (*17*, *18*), regulating electrode-electrolyte interfaces (*19*), and increasing nitrogen pressure (*13*–*19*)—have been developed by several groups to improve the ammonia faradaic efficiency (FE), current density, and ammonia production rates.

Although substantial progress has been made to achieve high FE and current density, most Li-NRR investigations have been performed in a batch-type electrochemical cell (i.e., one-compartment cells and autoclaves), where nitrogen must be dissolved in the electrolyte to participate in the reaction even at high N_2 pressure (e.g., 5 to 50 bar). This type of configuration is mass transport limited with respect to nitrogen because of the low solubility of N_2 in nonaqueous electrolytes (*20*, *21*). Moreover, most Li-NRR studies have the limitation of using a sacrificial solvent as a proton donor and face difficulties in scaling up production in batch reactors. Therefore, introducing a hydrogen oxidation reaction (HOR) on the anode (with the hydrogen sourced from water splitting) is a promising approach to providing a sustainable hydrogen source for ammonia synthesis. Suryanto *et al*. tried to introduce a phosphonium proton shuttle to resolve the sacrificial solvent problem at 0.5 bar H_2 and 19.5 bar N_2 in an autoclave (*16*). However, in the batch cell, the mass transfer limitation of H_2 is an issue, and H_2 also has the potential to react with metallic lithium to form lithium hydride (LiH), which impedes the capacity to activate nitrogen at room temperature. To circumvent transport limitations, Lazouski *et al*. first proposed using a stainless-steel cloth (SSC) as the gas diffusion electrode (GDE) operated in a three-compartment cell and achieved an ammonia FE of 35%, but this system was only stable for 5 to 8 min at 20 mA/cm^2 (*20*). Despite flowing H_2 on the anode side, there is no evidence to demonstrate that hydrogen in NH_3 is derived from the HOR because of the high cell voltage of 20 to 30 V, which also leads to energy efficiency (EE) of only 1.4 to 2.8% (*20*, *22*). For the initial investigations (*14*, *15*, *17*, *18*), we used the pseudo-EE first defined by Lazouski *et al*. in the Li-NRR system (*20*). We call it pseudo-EE because it is not demonstrated whether the protons are coming from the hydrogen or from sacrificial agents, and calculated EE did not include the energy required for the synthesis of H_2. From this point forward, we will instead use the EE as measured using hydrogen as the proton source (more details and discussion are available in the supplementary materials). To date, there are three main challenges to Li-NRR at 1 bar N_2 pressure for practical applications: (i) low FE and EE (both only a few percent in the one-compartment cell at 1 bar N_2) owing to mass transfer limitation and high cell voltage, (ii) unsustainable proton source from sacrificial solvent, and (iii) poor stability in a GDE-equipped electrolyzer.

Li-NRR paired with HOR can avoid a sacrificial solvent as a proton source and can also lower the cell voltage to improve the ammonia selectivity and EE (tables S1 to S4). A crucial part of implementing a successful HOR was

Department of Physics, Technical University of Denmark, Kongens Lyngby, Denmark.
†These authors contributed equally to this work.
*Corresponding author. Email: jkno@dtu.dk (J.K.N.); ibchork@fysik.dtu.dk (I.C.)

to replace the archetype Pt anode with a PtAu alloy, which showed high activity and long-term stability toward hydrogen oxidation in the organic electrolytes. To support this pivotal change of anode material, operando mass spectrometry of deuterium (D_2) oxidation demonstrated that hydrogen oxidation can continuously provide protons for ammonia synthesis and confirmed the recyclability of EtOH as a proton shuttle. Directly feeding the N_2 and H_2 to the electrolyte-GDE interface to participate in the reactions circumvents mass transfer limitations. Based on our PtAu alloy catalyst and electrolyzer design, the continuous-flow reactor with a 25-cm^2 effective area exhibited ammonia FE of up to 61 ± 1% and EE of 13 ± 1% at ambient pressure and temperature.

Design of the flow cell

First, we constructed a three-compartment, continuous-flow reactor in which an SSC-based GDE was positioned between a gas flow field (5 cm by 5 cm) and an electrolyte chamber where gaseous reactants could be directly fed to one side of the SSC electrode while electrolyte was fed on the other side (Fig. 1A and fig. S1). The configuration of the continuous-flow electrolyzer is depicted in Fig. 1B. Generally, the process of Li-NRR in a continuous-flow electrolyzer is as follows: A lithium ion (Li^+) diffuses from the bulk electrolyte through the solid-electrolyte interphase (SEI), and Li^+ is electrochemically reduced into metallic Li on the electrode and subsequently reacts with N_2 to form lithium nitride (Li_3N), which is protonated by a proton shuttle (EtOH) to continuously release ammonia. The sustainable source of protons is provided by hydrogen oxidation on the anode. The working electrode (WE) (cathode) and counter electrode (CE) (anode) were SSC with different pore sizes and PtAu catalysts on the SSC substrate (PtAu/SSC), respectively (figs. S2 to S4). The pseudo–reference electrode (RE) was a Pt wire. The N_2 and H_2 used in the experiments were cleaned by purifiers (NuPure) such that all labile N-containing compounds were down to the parts per trillion by volume (ppt-v) level. The N_2 and H_2 gas flow rates were controlled using a mass flow controller (Brooks Instrument) and were set to 75 standard cm^3/min. The solution of 1 M lithium tetrafluoroborate ($LiBF_4$) in THF contained varying concentrations of 0 to 1 vol % EtOH as an electrolyte. A syringe pump was used to flow the electrolyte with 1 ml/min flow rate. A pressure gradient of 15 mbar between the gas inlet and outlet of the continuous-flow reactor was set by a water column above the gas outlet, which prevented the electrolyte from flooding into the gas side of the GDE (fig. S5). The produced NH_3 was distributed in the electrolyte, electrode deposits, and gas phase–trapped acid solution. The ammonia concentrations were quantified by ion chromatography

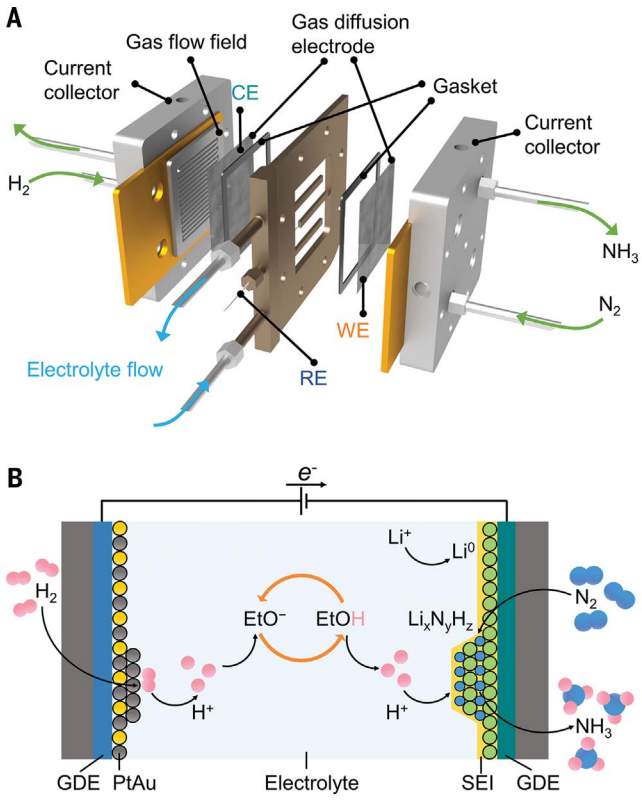


Fig. 1. Schematic and configuration of the continuous-flow reactor for electrochemical ammonia synthesis. (**A**) Expanded view of the continuous-flow electrolyzer configurations. Nitrogen and hydrogen gases are directly fed into the GDE-electrolyte interface to react. (**B**) Schematic process of the Li-NRR in a continuous-flow electrolyzer. The metallic Li from the reduction of Li^+ on the cathode reacts with N_2 to generate the lithium nitride ($Li_xN_yH_z$), which subsequently is protonated with the proton source to produce ammonia. The sustainable protons are produced from HOR on the PtAu anode catalysts and shuttled by EtO^-.

(fig. S6). When the gas-phase NH_3 trapped in hydrochloric acid solution was evaporated, high-purity NH_4Cl was obtained (fig. S7).

So far, Pt is the most active monometallic HOR catalyst in an aqueous electrolyte system (*23*). Only a few publications have investigated the HOR in organic electrolytes, and these have suggested a rapid deactivation of the catalyst unlike the classical observations in the aqueous system (*24*–*27*). Specifically, in organic electrolytes, the HOR catalytic activity of Pt was deactivated in a few minutes because of blockage of the active sites by the organic molecules or intermediates (*24*). Recently, Hodgetts *et al.* examined the activity deactivation for a series of potential HOR catalysts and found even the least-deactivated PtRu catalyst was also deactivated in the initial 20 min (*27*). Therefore, developing a long-term stable HOR catalyst is highly desired but poses a challenge for the Li-NRR system. In the field of electro-oxidation of small molecules, Pt was easily poisoned by carbon monoxide (CO) or other intermediates. Among the Pt-based, bimetallic alloy catalysts, PtAu catalysts improve the tolerance toward CO poisoning and exhibit good catalytic activity and stability because of the ensemble, ligand, and strain effect (*28*–*31*). The Pt modified with Au clusters can enhance the stability of the oxygen reduction reaction by increasing the Pt oxidation potential (*32*). Inspired by these studies, we believed that the PtAu catalyst had the potential to be one of the most stable catalysts for the HOR in organic electrolytes because of the beneficial suppression of the adsorption of organic poison species by Au.

Investigations of the HOR

Density functional theory (DFT) calculations were performed elucidating the effect of organic electrolytes, and the mean-field microkinetic modeling approach was applied (calculation details are given in the materials and methods) (*33*–*48*). The DFT results show that the adsorption free energy of possible intermediates—e.g., H, CO, THF, and furan—is weakened on the PtAu surface (figs. S8 to S10) compared with the pure Pt surface. The weaker binding of hydrogen on the PtAu surface results in a 1– to 2–order of magnitude–higher HOR production rate than that on the pure Pt surface (Fig. 2A). Moreover, the calculations indicate that the selectivity toward the HOR over THF oxidation on the PtAu surface is 5 to 6 orders of magnitude higher than that on a pure Pt surface, which markedly lowers the rate of the unwanted THF oxidation side reactions (fig. S11). CO could be produced during the EtOH oxidation by breaking the C-C bond between CH_x species and CO (CH_x-CO). Pure Pt shows facile predicted kinetics for catalyzing cleavage of the CH_2-CO bond, whereas on PtAu, there is a nontrivial CH_2-CO dissociation barrier under ambient conditions (fig. S12). We prepared Pt and PtAu electrodes on the SSC substrate (Pt/SSC and PtAu/SSC) by the

hydrogen bubble templated method (figs. S13 to S15) and the PtAu thin film on titanium (Ti) foil by magnetron sputtering (fig. S16). The HOR performance was evaluated in 1 M $LiBF_4$ in THF electrolytes using a one-compartment cell in an Ar-filled glovebox. The cyclic voltammetric (CV) curves of Pt/SSC and PtAu/SSC show a hydrogen oxidation peak at 0.2 V versus Pt pseudo-RE and a diffusion-limited current density of 3.8 mA/cm^2 because of the mass transfer limitation of H_2 (fig. S17). Not unexpectedly, the Pt/SSC deactivated in a few minutes during chronopotentiometry (CP), and the anode potential increased to 1 V versus Pt, where THF oxidation takes place (Fig. 2B). By contrast, the PtAu/SSC stays stable at low potentials of ~0.3 V versus Pt over 5 hours, proving the high activity and longer-term stability of the HOR in the given electrolyte (Fig. 2B). In the actual system, the electrolyte will contain produced ammonia during the Li-NRR process. Thus, the HOR stability of Pt/SSC and PtAu/SSC was examined in 1 M $LiBF_4$ with the presence of 40 mM NH_3. The HOR potential of PtAu/SSC stays around 0.5 V versus Pt for 5 hours, whereas Pt/SSC stays at 1.5 V versus Pt (fig. S18).

After long-term stability tests of PtAu, angle-resolved x-ray photoelectron spectrometer (XPS) results show that Pt was enriched on the surface of the PtAu catalyst (Fig. 2C and figs. S19 to S24). DFT calculations suggest that even a trace amount of CO, methyl, and furan leads to Pt atoms segregating onto the surface (Fig. 2D and fig. S10), and the underlayer of Au further weakens the adsorption strength of organic species (figs. S25 to S27). Moreover, the Au underlayer promotes the formation of a Pt(111)-like overlayer (*30*, *49*). The H-binding energy, a typical HOR reactivity descriptor, is −0.13 eV on the PtAu surface and −0.09 eV on the Pt√19×√19/Au surface (fig. S26), whereas the dissociation barrier of CH_2-CO is 0.85 eV on PtAu surface and 1.29 eV on the Pt√19×√19/Au surface (fig. S12). Thus, the Au underlayer has a minimal effect on the intrinsic HOR activity, and the Pt overlayer still exhibits the anti–CO poison effect of the PtAu surface. Additionally, PtAu exhibits a much lower activity of ammonia electro-oxidation compared with Pt, which indicates that PtAu could avoid the consumption of ammonia on the anode in the Li-NRR system (figs. S28 and S29).

Operando mass spectrometry of D_2 oxidation

To prove the combination of NRR with HOR, operando mass spectrometry with D_2 isotope studies were performed in the continuous-flow electrolyzer (Fig. 3 and figs. S30 to S39). Contrary to our previous studies (*14*), where the reaction relied on sacrificial solvent oxidation as the proton source, the electrolyzer is provided with hydrogen on the anode for the HOR. When flowing D_2 over the anode while measuring products at the cathode, H-containing products stemming from electrolyte reactions can be distinguished from D-containing products from D_2 oxidation (Fig. 3B and figs. S31 and S32). Initially, the cathode is fully covered in fresh electrolyte yielding mainly H-containing products, such as NH_3 and NH_2D. But as the experiment progresses, more D-containing products are formed (Fig. 3C), eventually leading to fully deuterated ammonia detected as

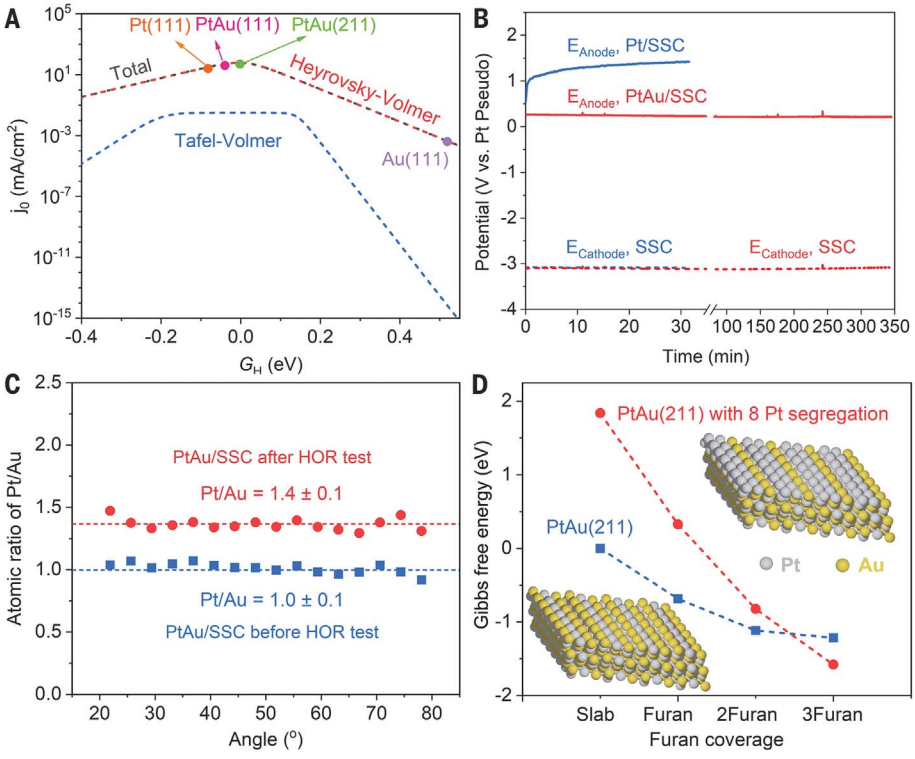


Fig. 2. Investigations of hydrogen oxidation catalysts. (**A**) Calculated exchange current densities (j_0) as a function of hydrogen adsorption free energy (G_H). (**B**) Experimental CP of PtAu/SSC and Pt/SSC at 2 mA/cm^2 current density in 1 M $LiBF_4$ in THF with H_2 bubbling. (**C**) The measured atomic ratio of Pt/Au before and after the long-term HOR stability test of PtAu/SSC using angle-resolved XPS. Error bars represent the standard deviation from 16 different angle channels. (**D**) The calculated Gibbs formation free energy of PtAu(211) and PtAu(211) with 8-Pt segregation to the surface with different furan coverage.

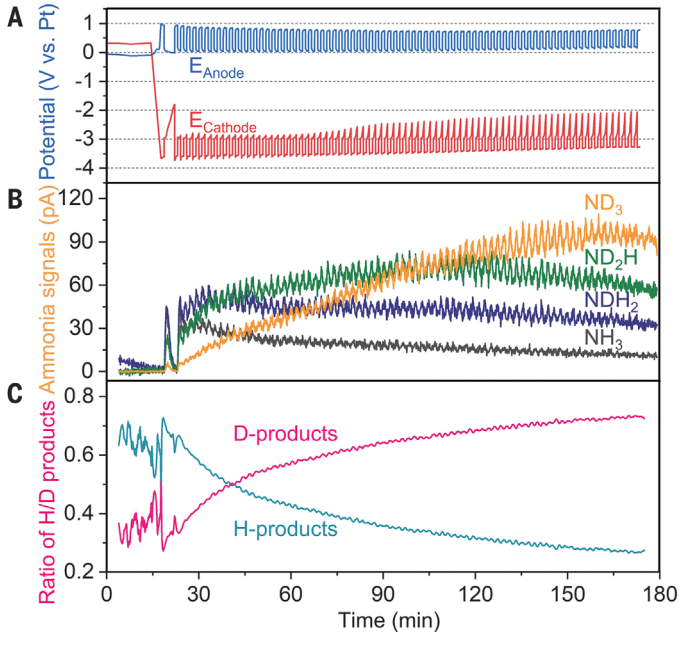


Fig. 3. Operando mass spectrometry of cathodic ammonia gas products with isotope-labeled deuterium. (**A**) Applied potentials during potential cycling at −6 mA/cm^2. (**B**) Measured ammonia products with varying amounts of hydrogen and deuterium. (**C**) Relative amounts of H and D in measured ammonia. As the experiment progresses, H is replaced with D in the proton shuttle (EtOH), leading to increased deuterated products.

ND_3. This unambiguously proves that coupled NRRs and HORs are occurring and demonstrates the capacity of EtOH as a proton shuttle. Additionally, the H_2 outlet stream was measured by operando mass spectrometry, and no other gaseous anode by-products were observed except H_2 and evaporated THF.

Trade-off strategy of proton- and nitrogen-limited region

Our previous studies have demonstrated a potential cycling strategy, where switching potential between applied potential and open circuit voltage (OCV) can improve the performance of Li-NRR (*14*). The optimal potential cycling condition is 1 min for Li deposition and 1 min for resting at OCV determined by changing the cycling time or the resting (OCV) potential at −2 V versus Pt determined by the modulo battery technique (Fig. 4A and figs. S40 to S42). Li-NRR coupled with HOR at the anode leads to the average anode potential staying around 0.6 V versus Pt, which contributes to much lower cell voltages of 4.3 V (Fig. 4A and fig. S43) compared with those from previous studies (*20*, *21*). The proton shuttle (EtOH) concentration effect was evaluated in the continuous-flow electrolyzer (figs. S44 to S48 and table S1), and the optimal EtOH concentration was determined to be 0.25 vol % (equal to 43 mM), which was different from the 100 mM optimal found by Lazouski *et al.* in a one-compartment cell (*50*). As shown in Fig. 4B, a FE of 61 ± 1% (from n = 3 repeated experiments) was achieved on the 30-μm SSC at optimal experiment conditions (0.25 vol % EtOH in 1 M $LiBF_4$ THF with typical potential cycling). The promotion effect of potential cycling shows opposite effects in the proton and nitrogen diffusion limited region. This is in line with our theoretical atomistic kinetic model analysis. During the resting period with potential cycling, in the proton-limited region, the deposited lithium metal reacts with N_2 gas to form lithium nitride, which couples with Li dissolution to produce NH_3 to improve the performance. In the N_2-limited region, the deposited lithium metal reacts with protons to form lithium hydride, which couples with Li dissolution to produce H_2, leading to decreasing NH_3 selectivity. Therefore, the selectivity is improved with the potential cycling in the proton-limited region but is decreased with potential cycling in the N_2-limited region. (fig. S49). The long-term stability of the Li-NRR system in a continuous-flow electrolyzer was also examined for 10 hours (fig. S50). The pore size effect of the SSC-based GDE was also investigated. Compared with the 30-μm SSC, the N_2 gas can access the electrolyte side through the bigger pores in the SSC mesh (43 and 61 μm) because of the imbalance of the electrolyte and gas. Nonetheless, a FE of up to 67 ± 2% and EE of 14 ± 1% (from n = 3 repeated experiments) was achieved on the 43-μm SSC (fig. S51). Moreover, about half of the ammonia was distributed in the gas phase, which is desirable because of the cheaper and easier separations compared with ammonia accumulating in the electrolyte (Fig. 4C).

To confirm the importance of the HOR on the anode and the role of EtOH, control experiments were carried out by varying the conditions, such as Ar, H_2, and N_2 gas feeding into the anode or cathode and with or without

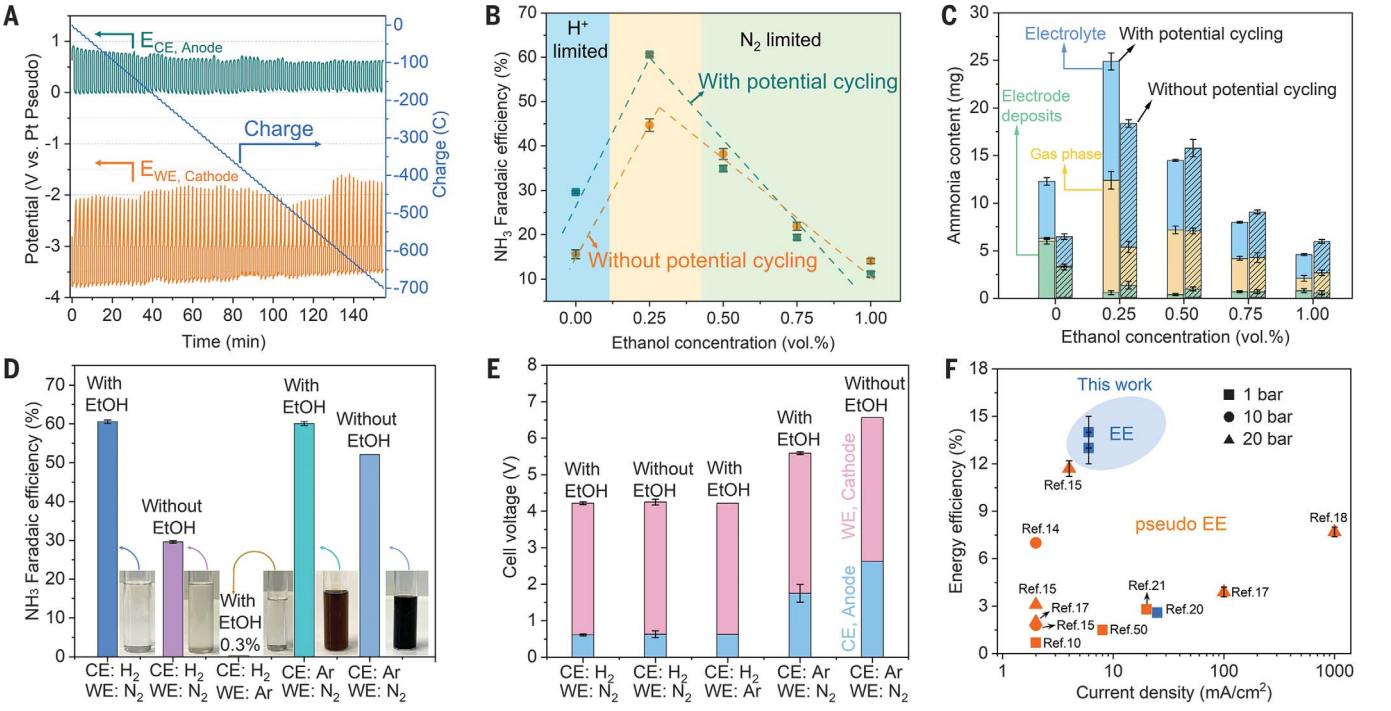


Fig. 4. Potential cycling method and trade-off strategy of proton- and nitrogen-limited region to enhance the performance of Li-mediated ammonia synthesis. (**A**) The CP of 30-μm SSC at the current density of −6 mA/cm^2. Potential cycling condition is −6 mA/cm^2 for 1 min and 0 mA/cm^2 for 1 min. All potentials reported are shown without ohmic drop (iR) correction. (**B**) The effect of EtOH concentration on the ammonia FE with or without potential cycling. Each FE was obtained by passing a total charge of 700 C at a current density of −6 mA/cm^2 on 30-μm SSC. (**C**) The effect of different EtOH concentrations on the distribution of produced ammonia in the electrolyte, gas phase, and electrode deposits. The produced ammonia was obtained by passing a total charge of 700 C and not normalized by reaction time. (**D**) Ammonia FE under control experimental conditions (different gas feeding into anode or cathode and with or without EtOH) with a typical potential cycling method. The corresponding insets are photos of electrolytes after passing 700 C or 500 C charge. (**E**) The average anode and cathode potential under control experimental conditions. (**F**) Comparison of current density and EE with previous accounts in the literature. It should be noted that (*20*) used H_2 as the anode reactant but did not demonstrate that the hydrogen in ammonia is from HOR. Notice this point (*20*) had to be corrected to consider the energy used for the synthesis of hydrogen because that was not done originally (*20*). Error bars represent the standard deviation from at least three independent measurements.

EtOH (Fig. 4D). When Ar flows on the anode side, the dark electrolytes indicate that THF is heavily decomposed or polymerized because of the high anode potentials of 1.7 V versus Pt (with EtOH) and 2.6 V versus Pt (without EtOH), whereas using an H_2 flow shows clear electrolytes with the lower anode potential of 0.6 V versus Pt (Fig. 4, D and E, and fig. S52). This suggests that the HOR occurs on the anode and protects the electrolytes, consistent with nuclear magnetic resonance (NMR) analysis (figs. S53 and S54). Moreover, an FE of 0.3% is observed using Ar as the cathode gas, which indicates that the produced ammonia is from N_2 reduction when feeding N_2 (Fig. 4D). Although the Li-NRR is a well-proven method, $^{15}N_2$ isotope–labeled experiments were also performed to confirm that the produced NH_3 originates from $^{15}N_2$ (figs. S39 and S55). Under similar conditions, the isotope experiment showed about the same performance as a $^{14}N_2$ experiment (fig. S56). The current densities, FEs, and EEs of previous publications are summarized in Fig. 4F and fig. S57 for comparison with this work. Previous works (*16*, *20*) did use H_2 but did not prove that the hydrogen in the ammonia is coming from the HOR. The current density of this work is lower than that reported in recent studies (*17*–*19*), but we demonstrate that the anode potential can be reduced substantially and that hydrogen in the produced ammonia is originating from the oxidation of feeding H_2 and not from the sacrificial solvent. We also highlight the whole system's efficiency and the importance of introducing the HOR on the anode to provide a sustainable hydrogen source. Future work should focus on obtaining high EE and FE at a higher current density to increase ammonia production rates. After more than 10 cycles of reuse for the CP test, the x-ray diffraction (XRD) patterns of PtAu anode catalysts show no visible changes, which confirms the stability of the PtAu catalysts (fig. S58). Previous studies have demonstrated that the SEI layer could modulate the ionic conductivity to affect the performance of Li-NRR (*15*, *18*). To reveal the key role of EtOH in the composition and formation of SEI, the flow cell was run in an Ar-filled glovebox under three conditions—i.e., after 700 C of CP test with EtOH and after a linear sweep voltammetry (LSV) test with or without EtOH. The XRD, XPS, and scanning electron microscope (SEM) transfer systems were used to avoid air and moisture exposure (figs. S59 to S61). The SEM images show the uniform SEI layer formed after the LSV test with EtOH, and fewer deposition species without EtOH indicate the key role of EtOH in SEI formation and Li deposition behavior (figs. S62 and S63). After 700 C of the CP test, the cathode is fully covered by the porous layer in which the main component of lithium fluoride (LiF) is confirmed by XRD and depth-profiling XPS (figs. S64 to S66). After the LSV test, the F 1s peak at 685.5 eV in depth-profiling XPS spectra is attributed to LiF, whereas the peak at 687.5 eV is attributed to $LiBF_4$ (figs. S67 to S70). All the results suggest that the main composition of SEI is LiF, which also matches previous accounts (*18*).

Very recently, 95% FE and a current density of 1 A/cm^2 were reported using autoclaves (15 or 20 bar) by investigation of the electrolyte (*19*) and construction of high–surface area electrodes (*18*), respectively. Although a FE of almost 100% (*19*) and a high current density of 1 A/cm^2 (*18*) have been achieved by sacrificing solvent as protons, it is quite far from the industrial application of Li-mediated ammonia synthesis. Achieving high FE and current density has more importance only when the anode truly uses hydrogen or water as the sustainable hydrogen source. Undeniably, introducing those strategies found in the one-compartment cell into the continuous-flow reactor is worth exploring to put the Li-NRR closer to practical applications. Although the results reported here represent an important advance in the stable PtAu catalyst for HOR to improve the operation stability of the flow-cell system and continuous-flow electrochemical synthesis of ammonia by Li-NRR, this work does not solve all the problems of Li-mediated ammonia synthesis at the level of industrial application. Future research on Li-NRR and HOR should aim at improving the current density, optimizing the mass transport of H_2 and N_2, and precisely modulating the pressure gradient between gas and liquid over the GDE. The main goal should be achieving high FE and EE at industrial current density in the pilot-scale flow cell. Sustainable H_2 can be provided by water electrolysis powered by renewable electricity. Recycling the anode outlet H_2 could be a solution to improve the utilization efficiency of H_2 at the anode. Our findings provide a solid foundation and a guide for these explorations for further scale up in the future.

REFERENCES AND NOTES

1. United States Geological Survey (USGS), "Nitrogen (Fixed)—Ammonia" (Mineral Commodity Summaries, January 2022); https://pubs.usgs.gov/periodicals/mcs2022/mcs2022-nitrogen.pdf.
2. J. G. Chen *et al*., *Science* **360**, eaar6611 (2018).
3. B. H. R. Suryanto *et al*., *Nat. Catal.* **2**, 290–296 (2019).
4. J. W. Erisman, M. A. Sutton, J. Galloway, Z. Klimont, W. Winiwarter, *Nat. Geosci.* **1**, 636–639 (2008).
5. A. J. Martín, T. Shinagawa, J. Pérez-Ramírez, *Chem* **5**, 263–283 (2019).
6. J. M. McEnaney *et al*., *Energy Environ. Sci.* **10**, 1621–1630 (2017).
7. D. R. MacFarlane *et al*., *Joule* **4**, 1186–1205 (2020).
8. G. Qing *et al*., *Chem. Rev.* **120**, 5437–5516 (2020).
9. J. Choi *et al*., *Nat. Commun.* **11**, 5546 (2020).
10. S. Z. Andersen *et al*., *Nature* **570**, 504–508 (2019).
11. F. Fichter, P. Girard, H. Erlenmeyer, *Helv. Chim. Acta* **13**, 1228–1236 (1930).
12. A. Tsuneto, A. Kudo, T. Sakata, *Chem. Lett.* **22**, 851–854 (1993).
13. A. Tsuneto, A. Kudo, T. Sakata, *J. Electroanal. Chem.* **367**, 183–188 (1994).
14. S. Z. Andersen *et al*., *Energy Environ. Sci.* **13**, 4291–4300 (2020).
15. K. Li *et al*., *Science* **374**, 1593–1597 (2021).
16. B. H. R. Suryanto *et al*., *Science* **372**, 1187–1191 (2021).
17. K. Li *et al*., *ACS Energy Lett.* **7**, 36–41 (2022).
18. S. Li *et al*., *Joule* **6**, 2083–2101 (2022).
19. H. L. Du *et al*., *Nature* **609**, 722–727 (2022).
20. N. Lazouski, M. J. Chung, K. Williams, M. L. Gala, K. Manthiram, *Nat. Catal.* **3**, 463–469 (2020).
21. X. Cai *et al*., *iScience* **24**, 103105 (2021).
22. C. Ampelli, *Nat. Catal.* **3**, 420–421 (2020).
23. K. C. Neyerlin, W. B. Gu, J. Jorne, H. A. Gasteiger, *J. Electrochem. Soc.* **154**, B631–B635 (2007).
24. W. C. Barrette, D. T. Sawyer, *Anal. Chem.* **56**, 653–657 (1984).
25. D. S. Silvester, L. Aldous, C. Hardacre, R. G. Compton, *J. Phys. Chem. B* **111**, 5000–5007 (2007).
26. M. F. Suárez-Herrera, M. Costa-Figueiredo, J. M. Feliu, *Langmuir* **28**, 5286–5294 (2012).
27. R. Y. Hodgetts, H. L. Du, T. D. Nguyen, D. MacFarlane, A. N. Simonov, *ACS Catal.* **12**, 5231–5246 (2022).
28. J. Suntivich *et al*., *J. Am. Chem. Soc.* **135**, 7985–7991 (2013).
29. T. Duan, R. G. Zhang, L. X. Ling, B. J. Wang, *J. Phys. Chem. C* **120**, 2234–2246 (2016).
30. P. P. Lopes *et al*., *Nat. Mater.* **19**, 1207–1214 (2020).
31. Y. J. Deng, V. Tripkovic, J. Rossmeisl, M. Arenz, *ACS Catal.* **6**, 671–676 (2016).
32. J. Zhang, K. Sasaki, E. Sutter, R. R. Adzic, *Science* **315**, 220–222 (2007).
33. W. Kohn, L. J. Sham, *Phys. Rev.* **140**, A1133–A1138 (1965).
34. G. Kresse, J. Furthmüller, *Phys. Rev. B Condens. Matter* **54**, 11169–11186 (1996).
35. S. R. Bahn, K. W. Jacobsen, *Comput. Sci. Eng.* **4**, 56–66 (2002).
36. B. Hammer, L. B. Hansen, J. K. Nørskov, *Phys. Rev. B Condens. Matter* **59**, 7413–7421 (1999).
37. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).
38. P. E. Blöchl, *Phys. Rev. B Condens. Matter* **50**, 17953–17979 (1994).
39. G. Kresse, D. Joubert, *Phys. Rev. B Condens. Matter* **59**, 1758–1775 (1999).
40. H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* **13**, 5188–5192 (1976).
41. G. Henkelman, B. P. Uberuaga, H. Jonsson, *J. Chem. Phys.* **113**, 9901–9904 (2000).
42. G. Henkelman, H. Jonsson, *J. Chem. Phys.* **111**, 7010–7022 (1999).
43. J. K. Nørskov *et al*., *J. Phys. Chem. B* **108**, 17886–17892 (2004).
44. M. T. Tang, X. Y. Liu, Y. F. Ji, J. K. Norskov, K. R. Chan, *J. Phys. Chem. C* **124**, 28083–28092 (2020).
45. A. M. Patel, S. Vijay, G. Kastlunger, J. K. Nørskov, K. Chan, *J. Phys. Chem. Lett.* **12**, 5193–5200 (2021).
46. H. A. Hansen, J. Rossmeisl, J. K. Nørskov, *Phys. Chem. Chem. Phys.* **10**, 3722–3730 (2008).
47. O. Vinogradova, D. Krishnamurthy, V. Pande, V. Viswanathan, *Langmuir* **34**, 12259–12269 (2018).
48. Á. Valdés, Z. W. Qu, G. J. Kroes, J. Rossmeisl, J. K. Nørskov, *J. Phys. Chem. C* **112**, 9872–9879 (2008).
49. P. P. Lopes *et al*., *J. Electroanal. Chem.* **819**, 123–129 (2018).
50. N. Lazouski, Z. J. Schiffer, K. Williams, K. Manthiram, *Joule* **3**, 1127–1139 (2019).
51. X. Fu, J. B. Pedersen, Y. Zhou, J. K. Nørskov, I. Chorkendorff, Data for "Continuous flow electrosynthesis of ammonia by nitrogen reduction and hydrogen oxidation," dataset, Technical University of Denmark (2023); https://doi.org/10.11583/DTU.21728168.v1.

ACKNOWLEDGMENTS

We thank floor managers B. P. Knudsen, J. E. Sørensen, and M. R. Vinther for helping with the connection of laboratory gas lines and the build of the mass spectrometer for isotope studies. We also thank the NMR center of the Technical University of Denmark (DTU). We acknowledge computational resources at Computerome from the Danish E-Infrastructure Cooperation (DeiC). **Funding:** We gratefully acknowledge the funding by Villum Fonden, part of the Villum Center for the Science of Sustainable Fuels and Chemicals (V-SUSTAIN grant 9455); Innovationsfonden (E-ammonia grant 9067-00010B); the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 741860); and MSCA European Postdoctoral Fellowships (Eelctro-Ammonia Project 101059643). **Author contributions:** Conceptualization: J.B.P., X.F., M.S., Y.Z., I.C., and J.K.N. Data curation: X.F., J.B.P., Y.Z., S.L., M.S., K.L., and C.W.

Formal analysis: X.F., J.B.P., Y.Z., and R.S. Investigation: X.F., J.B.P., Y.Z., S.L., and M.S. Methodology – equipment design: M.S., J.B.P., X.F., and S.Z.A. Visualization: X.F., J.B.P., and Y.Z. Supervision: I.C., J.K., P.C.K.V., and J.K.N. Writing – original draft: X.F., J.B.P., and Y.Z. Writing – review & editing: X.F., J.B.P., Y.Z., S.Z.A., S.L., A.X., N.H.D., J.B.V.M., J.K.N., and I.C. **Competing interests:** A patent application titled "Flow cell for electrochemical ammonia synthesis" was submitted on 9 September 2022 (application no. EP22194879) regarding the robust PtAu anode catalysts and record-high FE of up to 67 ± 2% and EE of 14 ± 1% at 1 bar of nitrogen described in this paper. M.S., J.B.P., X.F., S.Z.A., R.S., S.L., Y.Z., K.L., J.K., P.C.K.V., J.K.N., I.C., J.B.V.M., and N.H.D. are inventors on the patent application, submitted by DTU. The authors declare no other competing interests. **Data and materials availability:** Data are provided in the supplementary materials. The raw data are archived at DTU (*51*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf4403
Materials and Methods
Supplementary Text
Figs. S1 to S70
Tables S1 to S9
References (*52*–*59*)

Submitted 21 October 2022; accepted 24 January 2023
10.1126/science.adf4403

GEOPHYSICS

Ultralow frictional healing explains recurring slow slip events

Srisharan Shreedharan[1,2]*, Demian Saffer[1,3]*, Laura M. Wallace[1,4], Charles Williams[4]

Plate motion on shallow subduction megathrusts is accommodated by a spectrum of tectonic slip modes. However, the frictional properties and conditions that sustain these diverse slip behaviors remain enigmatic. Frictional healing is one such property, which describes the degree of fault restrengthening between earthquakes. We show that the frictional healing rate of materials entrained along the megathrust at the northern Hikurangi margin, which hosts well-characterized recurring shallow slow slip events (SSEs), is nearly zero (<0.0001 per decade). These low healing rates provide a mechanism for the low stress drops (<50 kilopascals) and short recurrence times (1 to 2 years) characteristic of shallow SSEs at Hikurangi and other subduction margins. We suggest that near-zero frictional healing rates, associated with weak phyllosilicates that are common in subduction zones, may promote frequent, small-stress-drop, slow ruptures near the trench.

Over the past 2.5 decades, a rapidly growing body of observations has revealed a continuum of slip behavior on tectonic plate boundary faults globally, spanning time scales from seconds to years (*1*). These modes of slip include fast elastodynamic earthquakes, tectonic tremors, low-frequency earthquakes (LFEs), very-low-frequency earthquakes (VLFEs), slow slip events (SSEs), and aseismic creep. The recognition of these diverse slip modes has sparked a rethinking of the underlying mechanics of earthquakes and fault slip and raised fundamental questions about in situ conditions, rheology, and processes along the subduction interface that govern this spectrum of behavior (*1*, *2*).

SSEs, in particular, have emerged as a potentially important mode of slip on the shallow subduction zone megathrust. During SSEs, the fault slips slowly over a period of days, weeks, or even months (*1*), rather than in seconds as in typical elastodynamic earthquakes. In some cases, SSEs may be accompanied by tremors, LFEs, and VLFEs, which radiate seismic energy (*1*, *3*) and are detected seismologically.

Numerous hypotheses have been proposed to explain this spectrum of fault slip modes, including highly elevated pore-fluid pressures (*4*, *5*), dilatant strengthening (*6*), transitional frictional properties (*7*–*10*), low-rigidity fault rocks (*11*), and lithological and geometric heterogeneities within the fault zone (*12*). These hypotheses center around material properties that control the nucleation of an instability on a fault and its ability (or inability) to grow into a dynamic rupture. A relatively unexplored yet equally fundamental control on earthquake occurrence is frictional healing, which describes the ability of a fault to restrengthen and store elastic strain energy between events (*2*, *10*). In the absence of a mechanism for restrengthening, a fault will fail only by aseismic creep. Frictional healing of an earthquake source region is particularly important because it plays a primary role in controlling rupture properties, including stress drop, recurrence interval, and potentially slip mode and amount (*2*, *10*).

Measuring the frictional properties of rocks in the fault zones hosting these complex slip behaviors has proved challenging because the SSE source depths are often inaccessible to direct sampling or drilling (*13*). One notable exception is the Hikurangi subduction zone offshore New Zealand, where regularly recurring shallow (<15 km) SSEs are well documented. We report on laboratory measurements of frictional healing conducted on samples of rock types involved in slow slip, obtained by drilling at the offshore Hikurangi margin. We then integrate these measurements with models of interevent tectonic loading and observations of SSE stress drop. We show that the very low healing rates of fault zone materials along the megathrust, which are typical of clay-rich faults in both subduction zones and other geological settings, are a key ingredient for the occurrence of recurring, frequent, small-stress-drop, slow slip transients.

Tectonic setting of the northern Hikurangi margin

Westward subduction of the Pacific plate beneath the Australian plate occurs along the northern Hikurangi Trough at 50 to 60 mm/year (*14*) (Fig. 1, A and B). The subducting plate is primarily composed of the Hikurangi Plateau, a Cretaceous large igneous province, and is overlain by ~1 km of volcaniclastic, pelagic, and siliciclastic sediments (*12*). The northern Hikurangi margin represents an ideal natural laboratory to investigate the origin of slow earthquake phenomena, owing to the shallow source depth of regularly recurring SSEs, which has enabled a host of high-resolution near-source geophysical investigations, as well as acquisition of host rock samples by drilling (*15*). The shallow SSEs here originate at depths of 2 to 15 km (*12*, *13*), in a structurally complex region of seamount subduction (Fig. 1C) and intervening areas of high-amplitude seismic reflectivity and low seismic velocity (*5*, *16*), with the latter usually inferred to represent elevated pore-fluid pressure (*16*). These SSEs propagate to within a few kilometers of, and possibly to, the trench (*13*); recur every 1 to 2 years; and last for a few weeks (Fig. 1, D and E), during which the plate interface experiences slip rates of ~1 cm/day (*15*, *17*). The SSEs are characterized by small stress drops (<10 kPa) and a range of energy release equivalent to that for moment magnitude (M_w)~6 to 7 earthquakes (*18*). Microseismicity and tremor have been documented along and above the plate interface during and after the shallow SSEs (*15*). The megathrust in this region has also hosted two tsunami earthquakes (long-duration elastodynamic earthquakes with slow rupture velocities of ~1 km/s) in 1947 (*19*). Thus, the shallow megathrust here is home to a rich variety of slip modes that appear to interact in complex ways in space and time.

International Ocean Discovery Program (IODP) Expedition 375 sampled Cretaceous volcaniclastic conglomerates overlying the subducting Hikurangi Plateau at Site U1520, located ~10 km east of the trench (Fig. 1, A to C). Seismic imaging indicates that these materials are entrained along and eventually form the shallow megathrust here (*12*, *20*). The margin is characterized by low heat flow, with temperatures on the megathrust expected to remain <150°C at a distance of ≳70 to 100 km from the trench (*21*). Thus, the frictional and rheological properties of the volcaniclastic material recovered at Site U1520 are likely representative of those along the shallow megathrust (to ~10- to 12-km depth), well into the source region of regularly recurring SSEs.

Frictional healing, fault failure, recurrence, and stress drop

Frictional healing describes the ability of a material to restrengthen frictionally over time and is a prerequisite for repeating stick-slip (seismic) failure of faults because it allows interseismic coupling and elastic strain accumulation (*22*). Healing is thought to reflect a time-dependent increase in asperity contact area at the micro- to nanoscopic scales (*22*). In the laboratory, frictional healing is commonly measured via slide-hold-slide (SHS) tests (*23*) and typically increases log-linearly with time (*22*). Healing (β) is usually reported as the

[1]Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA. [2]Department of Geosciences, Utah State University, Logan, UT, USA. [3]Department of Geological Sciences, Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA. [4]GNS Science, Lower Hutt, New Zealand.
*Corresponding author. Email: srisharan.shreedharan@usu.edu (S.S.); demian@ig.utexas.edu (D.S.)

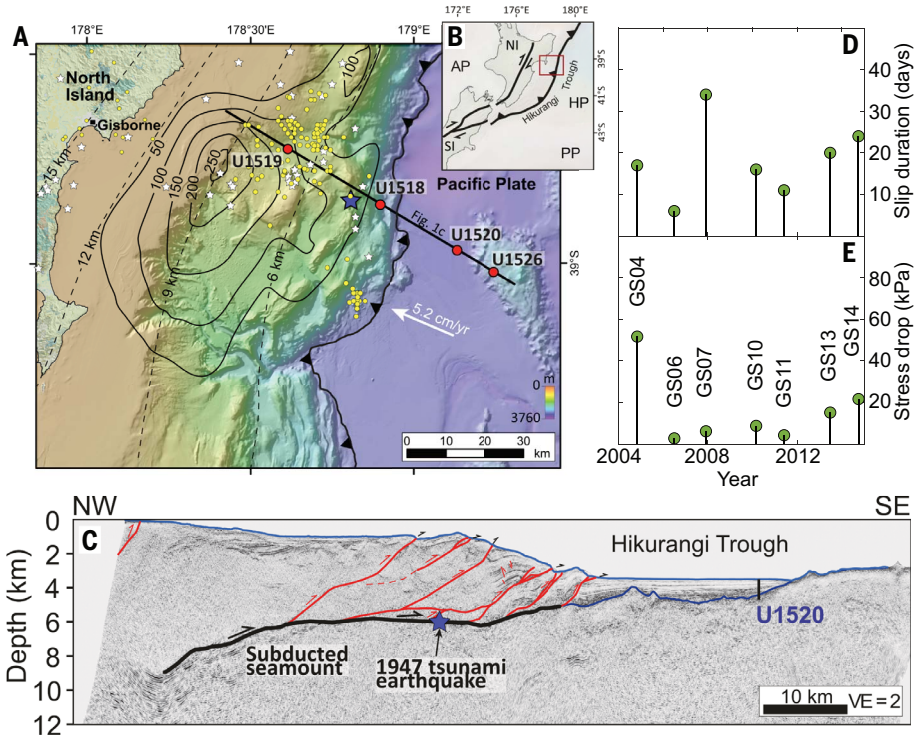


Fig. 1. Tectonic setting of the Gisborne area shallow slow slip events. (**A**) Bathymetric map of the northern Hikurangi margin showing the trench (bold black curve), depth to the plate interface (dashed contours), slip contours of the 2014 SSE in millimeters (black, as labeled), location of the 05CM-04 seismic line (black line), and locations of IODP expedition 375 drill sites (red circles) [modified from (*12*)]. Hypocenter of the 1947 tsunami earthquake (blue star), tremor activity (small yellow circles), and microseismicity (white stars) are also shown. (**B**) Regional tectonic map showing the Pacific plate subducting beneath the Australian plate, forming the northern Hikurangi margin. (**C**) Interpretation of the 05CM-04 seismic line [modified from (*12*)] showing the plate interface (bold black line), active upper-plate splay faults (red lines), and location and depth extent of IODP drill site U1520. The blue star at a depth of ~6 km represents the hypocenter of the 1947 tsunami earthquake. (**D**) Slip duration and (**E**) Spatiotemporally averaged stress drop of the quasiperiodic SSEs rupturing the Gisborne patch from 2002 to 2016 (table S2).

rate of increase in fault strength (or friction) per logarithmic (10-fold) increase in time.

We analyze the SHS experiments conducted by (*10*) on Cretaceous volcaniclastic sediments from Site U1520 (Fig. 2A). Modal compositions of these samples (table S1) (*24*) determined from x-ray diffraction indicate that they contain >55 wt % clay minerals and that the clay is dominantly smectite. These experiments were performed on water-saturated gouges under an effective normal stress of 25 MPa, pore pressures of 5 to 10 MPa, and for driving velocities of 1 to 30 μm/s. These experimental boundary conditions were selected to best represent appropriate conditions of stress and sliding velocity at shallow SSE source depths along the subduction megathrust. The healing rates inferred for these materials ($\beta < 10^{-4}$, Fig. 2B) represent some of the lowest values of fault healing reported for fault rocks and are more than 10 to 100 times below those commonly reported for carbonates and tectosilicates (*25*). Our observations are consistent with near-zero healing rates documented in phyllosilicate-rich synthetic and natural fault gouges, including smectite and serpentinite (*25*, *26*) (Fig. 2B).

We compare these frictional healing rates to average stress drops reported for repeating shallow SSEs that ruptured the Gisborne patch (*18*) (table S2) (Figs. 1 and 2). Seismic evidence indicates that the materials participating in shallow SSEs here may be compositionally similar to the input materials that were frictionally tested (*12*, *20*). Although we have incomplete knowledge of the exact in situ effective stress (σ_{eff}) along the shallow SSE source fault (*2*, *12*, *15*), indirect observations strongly suggest significant overpressure along the plate interface (*2*, *5*). To honor these observations and also allow for reasonable uncertainty in σ_{eff}, we convert the friction drop from SHS experiments to a stress drop (Fig. 2C) by considering a range of pore-fluid overpressure values along the plate interface of λ = 0.8 to 0.95 (where $\lambda = P_f/P_l$, and P_f and P_l are the fluid and lithostatic pressures, respectively) (*23*) (fig. S1).

The laboratory data for the frictionally weak volcaniclastic conglomerates are in excellent agreement with the geodetically defined stress drops for the Gisborne shallow SSEs (Fig. 2C). Together, these analyses suggest near-zero frictional restrengthening rates. In comparison, repeating earthquakes along the Calaveras fault exhibit substantially higher stress drops (1 to 10 MPa) that are strongly correlated with repeating earthquake recurrence intervals (Fig. 2C), increasing by nearly 2 MPa per 10-fold increase in time (*27*). SHS experiments on the predominantly quartzofelspathic material recovered from the Calaveras fault indicate correspondingly high frictional healing rates of ~0.007 (*28*), at least two orders of magnitude higher than the healing rates we report for the Hikurangi megathrust materials (Fig. 2B). Along the San Andreas Fault, experimentally estimated frictional healing rates on materials recovered from drilling are also in agreement with both the recurrence and stress drop of small repeating earthquakes along bounding faults and with the creeping behavior of the central deforming zone (*25*, *26*, *28*).

Competition between loading and healing on the shallow subduction megathrust

To further explore the role of the experimentally observed near-zero healing rates in governing event recurrence and strain accumulation and release, we consider the balance between tectonic loading and healing rates. We compute loading rates along the plate interface using a two-dimensional (2D) numerical model of the margin that incorporates heterogeneous elastic properties, which are well constrained by geophysical surveys (*29*) (Fig. 3, A and B) (*23*). We then compare these loading rates to the event-averaged stress drop and recurrence intervals of the ~10-year sequence of observed Gisborne slip events (Fig. 3C). The latter assumes that all of the strain energy accumulated during an interevent period is released during slip.

The numerical model predicts tectonic stressing rates ranging from 30 kPa/year at 95 km from the trench (near the downdip edge of the SSE region) to 0.3 kPa/year at 30 km (the updip end of the likely SSE source region), with an average loading rate of ~4.5 kPa/year over the outer ~65 km of the megathrust (Fig. 3B). Event-based estimates of stress accumulation rates using observed stress drops and recurrence intervals for each SSE range between 1 to 30 kPa/year, with an average of ~9 kPa/year across the seven events (Fig. 3, B to C). The two estimates of tectonic loading rates are in excellent agreement and suggest that on average, the loading rate along the shallow megathrust is ~4 to 10 kPa/year. The slip distributions for most offshore Gisborne SSEs are poorly constrained, with the exception of the most recent

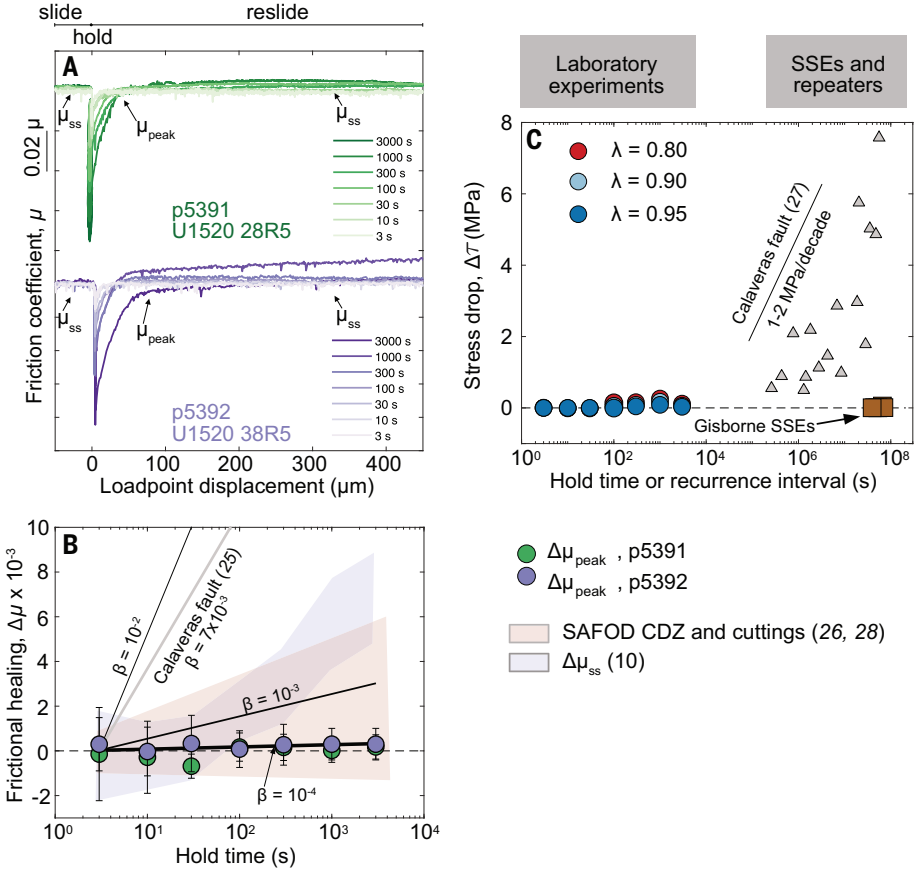


Fig. 2. Frictional healing constrained from laboratory experiments. (**A**) Overlay of multiple SHS protocols with different durations (3 to 3000 s) for two phyllosilicate-rich sediment samples from Site U1520. μ_{ss} represents the pre- and posthold steady-state values of the friction coefficient, and μ_{peak} represents the location of the (near-zero) peak friction upon reshear after a hold. (**B**) The change in friction (healing) between μ_{peak} and μ_{ss} shows little or no correlation with hold time. Error bars quantify electrical noise in the load cell and represent the limit of resolution of the friction drops. Fiducial lines show three different healing rates representing a range of values reported in prior studies (*25*), as well as values encompassed by the experiments reported here. Gray line represents the friction drops from SHS experiments conducted on samples from the Calaveras fault (*25*), where seismic repeaters have been observed. Pink area shows the range of values reported from SHS experiments conducted on smectite-rich San Andreas Fault Observatory at Depth (SAFOD) Central Deforming Zone (CDZ) cuttings (*26*, *28*). Light blue region represents changes in steady-state friction coefficient from pre- to posthold reported by (*10*). (**C**) Changes in stress drop versus recurrence interval for the Gisborne SSEs and versus hold time for the experimental data. Stress drops for experimental data are calculated as the product of friction drop and effective stress for three different degrees of overpressure. Gray triangles represent stress drops estimated from the seismic source for Calaveras repeaters (*27*).

events, for which seafloor geodetic data were available (*13*). This makes a precise comparison of the stress drops in the events with the predicted loading rates for an identical fault region difficult. Nonetheless, our results are consistent with similarly small loading rates in the outer forearc at the Nankai margin (*30*).

Frictional restrengthening increases logarithmically with time (as $\beta \times \sigma_{eff}$), whereas the shear stress borne by the fault increases, to first order, linearly with time as a function of tectonic loading, which in turn is a function of rock and fault elastic properties, geometry, and plate convergence rate (Figs. 2 and *3*). At a given location, and in the absence of additional loading—for example, associated with slip of adjacent fault patches—the fault reaches a threshold for failure when the accumulated stress exceeds the fault strength (or when the interevent loading exceeds the amount of frictional restrengthening; Fig. 4, A to C). In this simplified framework, the time to failure represents the recurrence interval and the stress at failure is the average stress drop of a quasiperiodic event. Here, we combine our numerically determined loading rates (Fig. *3*) with a restrengthening model to understand the size and recurrence of time- and slip-predictable events both through time at specific locations (Fig. 4, A to C) and as a function of position along the megathrust (Fig. 4, D to F).

Both tectonic loading and fault healing increase with depth along the plate interface (Figs. 2 and 3). As noted above, without strong constraints on σ_{eff}, we consider a range of likely pore-fluid pressure ratios based on geological and geophysical inferences (*2*, *5*) (Fig. 4 and fig. S2). We evaluate the competition between the loading rate and healing—and thus the expected recurrence and stress drops of repeating fault failure events—as a function of distance from the trench by considering the spatial distribution of excess strength at different times. We do so for a range of likely pore-fluid pressures (and by proxy σ_{eff}), as well as for a range of β that are consistent with our data. Because our experimental data provide only an upper bound on β, we explore values in the range of $\beta = 3 \times 10^{-5}$ to 3×10^{-6}, representing up to 1.5 orders of magnitude lower than this upper bound.

We illustrate the competition between loading and restrengthening at three representative locations along the plate interface, i.e., 25, 50, and 75 km from the trench (Fig. 4, A to C) for an effective stress profile corresponding to $\lambda = 0.9$. At 75 km from the trench, well within the SSE source region, the plate interface experiences tectonic loading of ~3.5 kPa/year (Fig. 3B). For our upper bound of β (3×10^{-5}), we predict a recurrence interval of 2.3 years, with events having average stress drops of ~8 kPa; the lowest value of β (3×10^{-6}) that we consider yields events with recurrence intervals of 3 to 4 months and <1 kPa stress drops (Fig. 4A). Closer to the trench, at a distance of 25 km, the loading rate on the plate interface is much lower (0.3 kPa/year). As a result, for our upper bound on β, we predict considerably longer times to failure (15 years) and stress drops of 4 kPa (Fig. 4C). Our lower bound on β yields recurrence intervals of ~1.3 years and stress drops of ~0.4 kPa.

Low β ($<10^{-4}$), coupled with the slow loading rates on the shallow plate interface, suggest that the fault is likely to fail in repeating, frequent (1- to 3-year) events, with small maximum total slip (on the order of a few to ~10 cm), on the basis of the slip deficit accrued during the interevent time and on low stress drops (on the order of a few to ~10 kPa). Although β and σ_{eff} are not perfectly constrained, other combinations of these parameters (fig. S2) yield similar outcomes. The observed recurrence and stress drops in the Gisborne SSEs are most consistent with scenarios presenting modest-to-high fluid overpressure ($\lambda > 0.8$), although near-lithostatic pressures are not required (fig. S2). Also, Because the rate of fault restrengthening ($\beta \times \sigma_{eff}$) scales with effective normal stress, this framework provides a mechanism, in addition to conditional frictional stability (*2*, *7*), that links

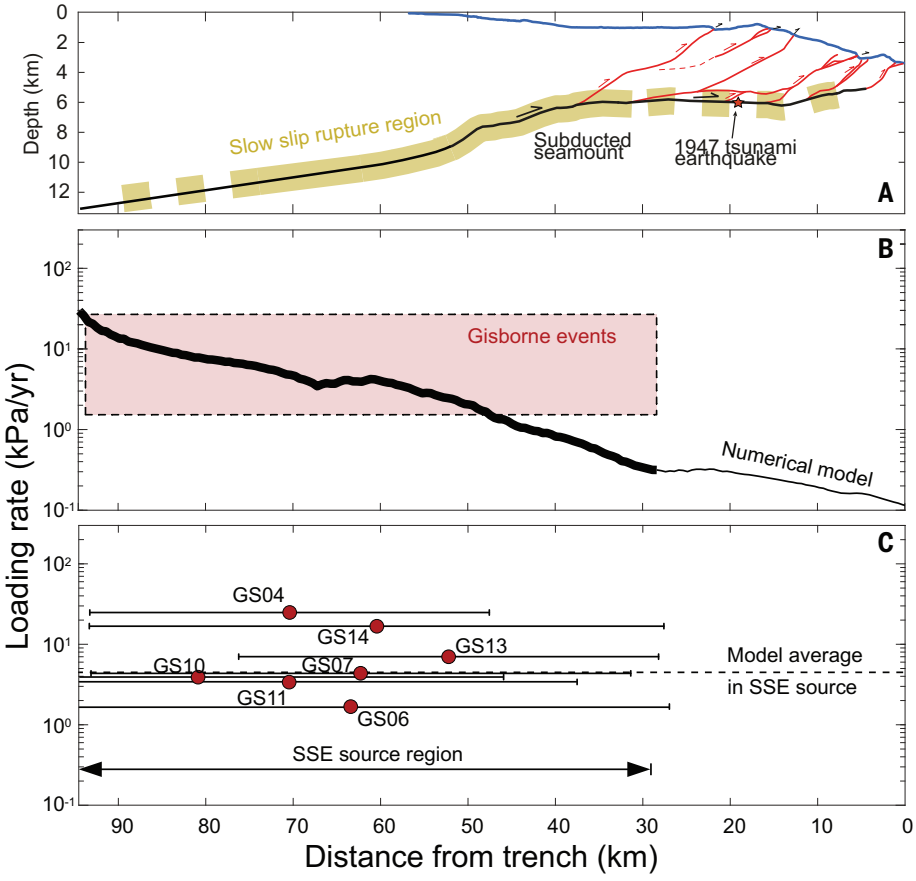


Fig. 3. Inter-SSE loading rates in the northern Hikurangi margin. (**A**) An illustration of the 05CM-04 line showing upper-plate splay faults (red curves), the hypocenter of the 1947 tsunami earthquake (red star), a subducting seamount, and the likely shallow SSE rupture region (thick yellow bars). (**B**) Annual tectonic loading rates (for convergence at 50 mm/year) along the 05CM-04 seismic transect, computed from 2D finite-element numerical model in Pylith (*29*). Loading rates increase from ~0.15 kPa/year near the trench to ~30 kPa/year at a depth of 15 km. (**C**) Tectonic loading rates constrained from the Gisborne SSEs as the ratio of average stress drops and interevent times. The black (~4.5 kPa/year) dashed line shows average loading rates from models.

elevated pore pressure with frequent small-stress-drop events, as is commonly postulated or inferred for SSEs (*5*).

We evaluate the spatial variation of this competition in more detail by considering the strength excess along the plate interface, defined as the difference between strength (or amount of restrengthening) and accumulated tectonic stress at different times (Fig. 4, D to F), i.e., after 1, 2, and 3 years, for an effective stress profile corresponding to $\lambda = 0.9$ (fig. S1). A positive strength excess represents a fault that has healed more than it has been loaded tectonically, indicating that it is still far from failure. Conversely, a negative strength excess indicates a fault that has already failed because it has experienced more tectonic stressing than restrengthening. For the case where $\beta < 3 \times 10^{-5}$, the strength excess is nearly zero to a depth of ~8.5 km (corresponding to distances ~0 to 50 km from the trench), indicating that the ultralow healing rates of the SSE source at such shallow depths may allow it to fail frequently (i.e., on a time scale of months or shorter) and with very small (sub-kilopascal) stress drops (Fig. 4, D and E). Further landward from the trench and deeper, the strength excess reaches zero after 1 to 3 years, which is in good agreement with the recurrence intervals of the shallow SSEs here.

An additional complexity in extrapolating healing rates to natural faults arises because some studies indicate that healing (β) may vary with effective stress, temperature, and lithology (*10*, *31*, *32*). This variation is captured across the suite of cases that we explore for values of β ranging from 3×10^{-6} to 3×10^{-5} (Fig. 4, A to F). Some studies indicate that experimental healing rates could increase with larger hold times (>3000 s) via pressure solution or other mechanisms, particularly in strong quartzofeldspathic minerals (*32*, *33*), whereas others indicate that healing rates may reduce over long time scales, particularly for clay-rich materials as the quality and quantity of contact area becomes saturated (*28*). Here, in accordance with (*25*, *28*), we consider a constant healing rate in mapping from the experimental to natural SSE time scales. In the constant β model (Fig. 4G, left), stress drops are relatively constant at ~6 to 10 kPa regardless of the downdip location of the SSE source, but there is considerable variability in predicted recurrence intervals. If β increases with depth (Fig. 4G, right), we predict frequent (<9- to 14-months recurrence) SSEs with stress drops of <1 kPa near the trench, and larger, less-frequent SSEs further downdip (~10 kPa, 1- to 3-year recurrence) at 10- to 15-km depth. The latter are detectable by traditional onshore geodetic instruments (*15*), whereas the former may be detectable only by offshore borehole observatories near the trench (*34*). These small, frequent, slow transients, if real, raise the possibility of a spectrum of slip spanning from quasiperiodic creep or undulating slip on the shallowest reaches of the megathrust. These transients originate near the trench because, although loading is small, healing is negligible. This would result in a continuum from recurring larger SSEs farther from the trench transitioning to more frequent, smaller events near the trench. Moreover, depending on the timescale and nature of observations, these small, slow, and frequently repeating transients could present themselves as aseismic creep, particularly in the absence of continuous, near-field, and high-precision offshore monitoring; in the limit as healing approaches zero, slip will occur via stable creep. In other words, when healing and loading are equivalently small, such as near the trench, aseismic creep and SSEs may be indistinguishable from one another.

The small stress drops and small total slip predicted by low healing rates are also physically consistent with fault failure via slow earthquakes or SSEs as opposed to rapid slip (*35*). Small stress drops and slip limit the stored elastic energy available for the acceleration of slip, therefore favoring slip in slow motion and over a long duration (*35*, *36*). Additionally, SSEs are thought to involve substantial ductile deformation that extends beyond a narrow slip zone (*37*), which would dissipate a substantial amount of energy off-fault, further reducing the energy available for propagating a dynamic rupture on-fault (*38*), and in turn, keeping any instability slow. Taken together, the recurring small stress drop events of <10 kPa (*18*) and velocity-neutral frictional stability of the volcaniclastic materials (*10*) indicate an inability of the SSE source region to heal between SSEs and an inability to nucleate large, dynamic earthquakes characterized by large fault slip and rapid slip velocities.

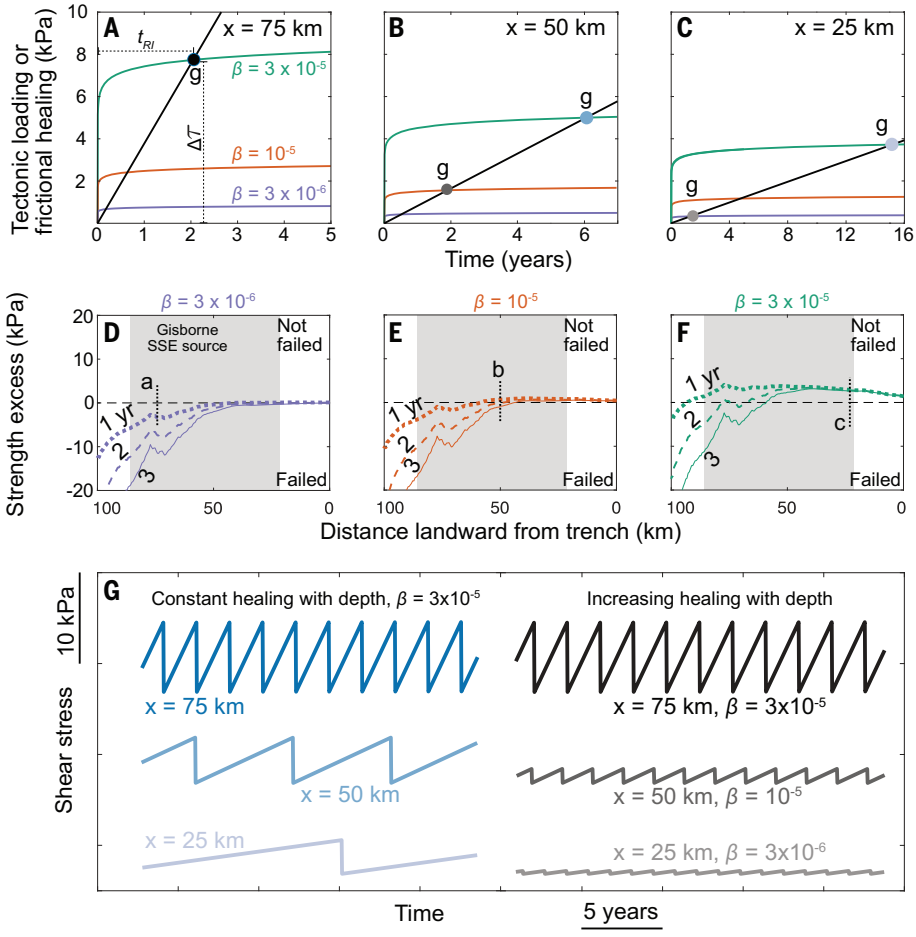


Fig. 4. A framework to quantify competition between tectonic loading and frictional healing. Tectonic loading as shown in Fig. 3 (black lines) and fault restrengthening (colored curves) as a function of time, for three different assumed healing rates (3×10^{-6}, purple; 1×10^{-5}, orange; and 3×10^{-5}, green) at (**A**) 75 km, (**B**) 50 km, and (**C**) 25 km from the trench. Strength excess as a function of distance from the trench, after 1 year (dotted curve), 2 years (dashed curve), and 3 years (solid curve) for healing rates of (**D**) 3×10^{-6}, (**E**) 1×10^{-5}, and (**F**) 3×10^{-5} [same color scheme as in panels (A) to (C)]. (**G**) Simplified models of recurrence and stress drop for two scenarios of depth-varying healing, showing shear-stress variations over time. Constant healing with depth (left, light to dark blue) results in small-stress-drop events with large recurrence intervals near the trench; increasing healing rate with depth (right, gray to black) results in very-small-stress-drop events (~1 kPa) every few weeks or months. Effective stress corresponds to $\lambda = 0.90$ for all panels.

The prediction that a near-zero strength excess is attained simultaneously over a broad region of the plate interface (e.g., Fig. 4, D and E) also implies that the shallow megathrust reaches failure over a broad region. This is consistent with observations of shallow SSEs having a large slip patch with small total slip (*35*), because the low loading rates combined with low healing rates never allow for a large accumulation of slip deficit. In contrast, ordinary earthquakes with similar amounts of slip (a few centimeters) commonly nucleate over much smaller areas (*35*).

In addition to hosting SSEs, the northern Hikurangi margin (*12*, *13*) and other shallow plate boundaries (*2*, *34*, *39*, *40*) also host tsunami earthquakes, microseismicity, LFEs, VLFEs, and tremor. Carbonates, which are characterized by substantially larger β (*10*) (compare to Fig. 2) and capable of nucleating instabilities even at modestly elevated temperatures (*41*, *42*), are also present in the North Hikurangi SSE source region, though with less abundance than the volcaniclastics (*10*, *12*, *41*). This heterogeneity of lithology may provide one explanation for the diverse slip behaviors observed along the margin. For example, it raises the possibility that infrequent tsunami earthquakes could nucleate locally where strong and unstable materials such as carbonates are embedded and/or abundant within the broader SSE region (*12*). In this scenario, such ruptures could then propagate through the weak, phyllosilicate-rich volcanoclastic-rich portions of the megathrust (*43*), producing tsunami earthquakes (*19*). Alternatively, small carbonate- or quartz-filled fractures along the décollement may be loaded during an SSE. This increased loading, coupled with the elevated healing rates of these materials, could result in tremorgenic failure that is spatiotemporally colocated with SSE slip (*34*, *44*).

We integrate experimental measurements of healing and healing rates on relevant megathrust materials, a numerical model of fault loading, and observed shallow SSE source properties. Taken together, they indicate that low healing rates on weak faults are a key ingredient in the generation of frequent, low-stress-drop, recurring slip events with small amounts of total slip, such as those observed at the northern Hikurangi margin and many plate boundaries globally. Because the low rates of healing limit the stored strain and stress (and strain energy) available in the system, materials with low healing rates are likely to host SSEs, creep, and other slow earthquake phenomena. Furthermore, because restrengthening is controlled by the product of frictional healing and effective normal stress, our results highlight an additional pathway by which high pore pressure (or low effective stress) may directly influence the size, frequency, and nature of failure, and ultimately modulate the mode of failure in a host of tectonic and geologic settings, particularly at shallow depths (*1*). Weak phyllosilicate minerals, which typically exhibit low healing rates, are common along shallow subduction megathrusts (*2*), raising the possibility that shallow SSEs and near-trench creep may be the norm at subduction zones and that the comparatively limited observations of shallow SSEs at subduction zones globally may simply be a consequence of our inability to detect them with land-based geodetic networks.

REFERENCES AND NOTES

1. Z. Peng, J. Gomberg, *Nat. Geosci.* **3**, 599–607 (2010).
2. D. M. Saffer, L. M. Wallace, *Nat. Geosci.* **8**, 594–600 (2015).
3. G. C. Beroza, S. Ide, *Annu. Rev. Earth Planet. Sci.* **39**, 271–296 (2011).
4. D. Wei, Y. Sandwell, Y. Fialko, *J. Geophys. Res.* **114**, B07402 (2009).
5. D. Bassett, R. Sutherland, S. Henrys, *Earth Planet. Sci. Lett.* **389**, 1–13 (2014).
6. P. Segall, M. Rubin, M. Bradley, R. Rice, *J. Geophys. Res.* **115**, B12305 (2010).
7. J. R. Leeman, D. M. Saffer, M. M. Scuderi, C. Marone, *Nat. Commun.* **7**, 11104 (2016).
8. H. S. Rabinowitz *et al.*, *Geochem. Geophys. Geosyst.* **19**, 2973–2990 (2018).
9. K. Im, D. Saffer, C. Marone, J.-P. Avouac, *Nat. Geosci.* **13**, 705–710 (2020).
10. S. Shreedharan *et al.*, *Geochem. Geophys. Geosyst.* **23**, e2021GC010107 (2022).
11. Y. Ito, K. Obara, *Geophys. Res. Lett.* **33**, 9 (2006).
12. P. M. Barnes *et al.*, *Sci. Adv.* **6**, eaay3314 (2020).
13. L. M. Wallace *et al.*, *Science* **352**, 701–704 (2016).
14. L. M. Wallace, J. Beavan, R. McCaffrey, D. Darby, *J. Geophys. Res.* **109**, B12406 (2004).
15. L. M. Wallace, *Annu. Rev. Earth Planet. Sci.* **48**, 175–203 (2020).
16. R. Bell *et al.*, *Geophys. J. Int.* **180**, 34–48 (2010).
17. L. M. Wallace, J. Beavan, *J. Geophys. Res.* **115**, B12402 (2010).
18. M. J. Ikari *et al.*, *Geochem. Geophys. Geosyst.* **21**, e2019GC008717 (2020).

19. R. Bell, C. Holden, W. Power, X. Wang, G. Downes, . *Earth Planet. Sci. Lett.* **397**, 1–9 (2014).
20. A. C. Gase, N. L. Bangs, H. J. A. Van Avendonk, D. Bassett, S. A. Henrys, *Geology* **50**, 1145–1149 (2022).
21. A. Antriasian *et al*., *Geophys. J. Int.* **216**, 1177–1190 (2019).
22. J. H. Dieterich, B. D. Kilgore, *Pure Appl. Geophys.* **143**, 283–302 (1994).
23. Materials and methods are available as supplementary materials.
24. M. B. Underwood, "Data report: Clay mineral assemblages within biocalcareous and volcaniclastic inputs to the Hikurangi subduction zone, IODP Expedition 372B/375 Sites U1520 and U1526, offshore New Zealand." *Proceedings of the International Ocean Discovery Program* 372B/375 (2021).
25. B. M. Carpenter, M. J. Ikari, C. Marone, *J. Geophys. Res. Solid Earth* **121**, 1183–1201 (2016).
26. M. M. Scuderi, B. M. Carpenter, *Geophys. J. Int.* **231**, 290–305 (2022).
27. C. Marone, J. E. Vidale, W. L. Ellsworth, *Geophys. Res. Lett.* **22**, 3095–3098 (1995).
28. B. M. Carpenter, D. M. Saffer, C. Marone, *Geology* **40**, 759–762 (2012).
29. C. A. Williams, L. M. Wallace, *Geophys. Res. Lett.* **45**, 7462–7470 (2018).
30. K. Tamaribuchi, M. Ogiso, A. Noda, *J. Geophys. Res. Solid Earth* **127**, e2022JB024403 (2022).
31. G. C. McLaskey, A. M. Thomas, S. D. Glaser, R. M. Nadeau, *Nature* **491**, 101–104 (2012).
32. H. Yasuhara, C. Marone, D. Elsworth, *J. Geophys. Res.* **110**, B06310 (2005).
33. M. J. Ikari, B. M. Carpenter, C. Vogt, A. J. Kopf, *Earth Planet. Sci. Lett.* **450**, 164–172 (2016).
34. E. Araki *et al*., *Science* **356**, 1157–1160 (2017).
35. E. E. Brodsky, J. Mori, *Geophys. Res. Lett.* **34**, 16 (2007).
36. Y. Ida, *Bull. Seismol. Soc. Am.* **63**, 959–968 (1973).
37. Å. Fagereng *et al*., *Geology* **47**, 872–876 (2019).
38. A. Chounet, M. Vallée, M. Causse, F. Courboulex, *Tectonophysics* **733**, 148–158 (2018).
39. M. Radiguet *et al*., *J. Geophys. Res.* **117**, B040305 (2012).
40. Y. J. Tan, D. Marsan, *Sci. Adv.* **6**, eabb2489 (2020).
41. A. M. Eijsink, M. J. Ikari, *Geochem. Geophys. Geosyst.* **23**, e2022GC010369 (2022).
42. R. M. Kurzawski, M. Stipp, A. R. Niemeijer, C. J. Spiers, J. H. Behrmann, *Nat. Geosci.* **9**, 717–722 (2016).
43. H. S. Rabinowitz, H. M. Savage, P. J. Polissar, C. D. Rowe, J. D. Kirkpatrick, *Nat. Commun.* **11**, 533 (2020).
44. Å. Fagereng, F. Remitti, R. H. Sibson, *Tectonophysics* **510**, 381–386 (2011).
45. S. Shreedharan, Data for Hikurangi SSE and Healing Paper, Zenodo (2022); https://doi.org/10.5281/zenodo.7246267.

ACKNOWLEDGMENTS

We thank L. Lavier and D. C. Bolton for insightful discussions. **Funding:** S.S. was funded by the UTIG Palisades Postdoctoral Fellowship. D.M.S. acknowledges funding from a US Science Support Program post-expedition award and the Scott Petty Jr. Director's Chair endowment. L.W. acknowledges funding from the New Zealand MBIE Endeavor Research Fund (contract C05X1605) and the Marsden Fund (grant 20-GNS-001). **Author contributions:** Conceptualization: S.S., D.M.S. Methodology: S.S., D.M.S., L.W., and C.W. Investigation: S.S., D.M.S., L.W., and C.W. Visualization: S.S., D.M.S. Funding acquisition: S.S., D.M.S., and L.W. Writing – original draft: S.S. Writing – review and editing: S.S., D.M.S., L.W., and C.W. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text, supplementary materials, or from Zenodo (*45*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf4930
Materials and Methods
Supplementary Text
Figs. S1 to S3
Tables S1 and S2
References (*46*–*53*)

Submitted 25 October 2022; accepted 25 January 2023
10.1126/science.adf4930

EPIGENETICS

Menin "reads" H3K79me2 mark in a nucleosomal context

Jianwei Lin[1]†, Yiping Wu[1]†, Gaofei Tian[1]†, Daqi Yu[2]†, Eunjeong Yang[4,5], Wai Hei Lam[6], Zheng Liu[1], Yihang Jing[3], Shangyu Dang[2], Xiucong Bao[4]*, Jason Wing Hon Wong[4,5]*, Yuanliang Zhai[6]*, Xiang David Li[1,3]*

Methylation of histone H3 lysine-79 (H3K79) is an epigenetic mark for gene regulation in development, cellular differentiation, and disease progression. However, how this histone mark is translated into downstream effects remains poorly understood owing to a lack of knowledge about its readers. We developed a nucleosome-based photoaffinity probe to capture proteins that recognize H3K79 dimethylation (H3K79me2) in a nucleosomal context. In combination with a quantitative proteomics approach, this probe identified menin as a H3K79me2 reader. A cryo–electron microscopy structure of menin bound to an H3K79me2 nucleosome revealed that menin engages with the nucleosome using its fingers and palm domains and recognizes the methylation mark through a π-cation interaction. In cells, menin is selectively associated with H3K79me2 on chromatin, particularly in gene bodies.

Methylation of lysine at various sites on histones regulates gene activation or repression, depending on which residues are modified and the degree of methylation (i.e., mono-, di-, or trimethylation) (*1*, *2*). Effector (or "reader") proteins are key players that recognize histone methylation and translate this information into downstream gene regulation events to control gene expression (*3*). So far, readers of almost all histone methylation marks have been identified except for methylation at histone H3 lysine 79 (H3K79). Although several "royal family" modules containing proteins have been implicated to interact with this mark, robust evidence for the recognition is lacking (*4*–*7*). In mammals, DOT1L is the methyltransferase that methylates H3K79 (mono-, di-, and trimethylation) (*8*, *9*). The DOT1L-mediated H3K79 methylation is enriched at actively transcribed genes and is involved in the regulation of transcriptional elongation, DNA damage repair, and cell cycle progression (*10*–*14*). In higher eukaryotes, the H3K79 methylation level is exquisitely controlled during embryonic development and hematopoiesis (*15*). Perturbation of H3K79 methylation in mouse embryo leads to cardiovascular defects and anemia (*16*). Aberrant hypermethylation of H3K79 was found to drive leukemogenesis in mixed lineage leukemia (MLL) (*17*). Despite the important roles of H3K79 methylation, understanding of how this histone mark is translated downstream to regulate these processes is hindered by a lack of knowledge about the readers of H3K79 methylation.

[1]Department of Chemistry, University of Hong Kong, Hong Kong SAR, China. [2]Division of Life Science, Hong Kong University of Science and Technology, Hong Kong SAR, China. [3]Greater Bay Biomedical InnoCenter, Shenzhen Bay Laboratory, Shenzhen, China. [4]School of Biomedical Sciences, University of Hong Kong, Hong Kong SAR, China. [5]Centre for Oncology and Immunology, Hong Kong Science Park, Hong Kong SAR, China. [6]School of Biological Sciences, University of Hong Kong, Hong Kong SAR, China.
*Corresponding author. Email: xiangli@hku.hk (X.D.L.); zhai@hku.hk (Y.Z.); jwhwong@hku.hk (J.W.H.W.); baoxc@hku.hk (X.B.)
†These authors contributed equally to this work.

Design and synthesis of photocrosslinkable nucleosomes with H3K79me2 mark

Identification of H3K79 methylation readers is challenging because the posttranslational modification (PTM)–mediated protein-protein interactions (PPIs) can be weak and transient. Recognition of H3K79 and its methylation is dependent on nucleosome context, because this residue is in the core region of the nucleosome within the well-defined quaternary structure. Indeed, DOT1L methylates H3K79 only in an intact nucleosome but not in H3 peptides or proteins (*8*). Therefore, a peptide-based biochemical "pull-down" may not be suitable. We thus designed a multifunctional synthetic nucleosome (probe N1) that carries a site-specific dimethylation on H3K79 at a stoichiometric level. The probe also consists of a cleavable photoaffinity handle containing a photoreactive group (i.e., diazirine) that can convert noncovalent weak interactions into irreversible covalent linkages upon ultraviolet (UV) irradiation, a biorthogonal handle (i.e., alkyne) that facilitates the isolation of cross-linked proteins, and a disulfide linkage that allows the release of cross-linked peptides for mass spectrometry (MS) analysis (Fig. 1, A and B). To ensure photocrosslinking efficiency and specificity, this trifunctional handle was placed at H3D81 position. This location is close to H3K79 and less likely to interfere with the histone-histone or histone-DNA interactions within the nucleosome, according to the crystal structure of the nucleosome (*18*) (fig. S1A).

To prepare probe N1, we first performed a multistep semi-synthesis to assemble *Xaenopus laevis* histone H3K79me2 (D81ADdis-Cys, C110A) (probe H1) through expressed protein ligation (*19*) (Fig. 1C). The N-terminal fragment (F1, residues 1 to 74) was produced recombinantly as an intein fusion to yield an α-thioester at its C-terminal residue, whereas the central segment containing K79me2 (F2, residues 75 to 80, A75C) was synthesized through Fmoc solid-phase peptide synthesis as an acyl hydrazide. The C-terminal fragment (F3, residues 81 to 135, D81C, C110A) was prepared recombinantly as a SUMO fusion, which was cleaved by Ulp1 to expose the N-terminal cysteine. The three fragments were joined by native chemical ligation (NCL) (*20*) in an N-to-C manner. After the first ligation, desulfurization yielded a new N-terminal fragment containing K79me2 (F1-2, residues 1 to 80), leaving the inert acyl hydrazide unaffected. Subsequent activation of the C-terminal acyl hydrazide by $NaNO_2$ oxidation and 4-mercaptophenylacetic acid (MPAA) thiolysis generated a new α-thioester for the second NCL to produce H3K79me2 (D81C, C110A) (fig. S1B). Taking advantage of the ligation "scar," in which the cysteine at residue 81 is the only cysteine in the ligation product, we installed the trifunctional handle at this position through an asymmetric disulfide exchange reaction with a 5-nitropyridyl-2-sulfenyl (Npys)–activated diazirino alkynyl hetero disulfide (Npys-Dzyne) to produce the desired product, H3K79me2 (D81ADdis-Cys, C110A) (probe H1). The identity of the product and the site-specific incorporation of the dimethyl lysine together with the trifunctional handle were confirmed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Fig. 1, D and E). As a control protein probe, unmodified histone H3 (D81C, C110A) was expressed and purified from bacteria, followed by functionalization with Npys-Dzyne (fig. S1, C and D). Using a similar strategy, we also prepared H3K79me2 (C110A) without photocrosslinker (fig. S1, E and F). The synthesized dimethylated and unmodified H3 protein probes were then incorporated into mononucleosomes to generate probes N1 and N1c, respectively (Fig. 1, A and F).

Identification of menin as an H3K79me2-dependent nucleosome binder

We performed a standard cross-linking–assisted and SILAC-based protein identification (CLASPI) (*21*) experiment using probes N1 and N1c in both a "forward" and "reverse" manner. Proteins that preferentially bound to the K79-methylated nucleosome (probe N1) over the unmodified nucleosome (probe N1c) should have a higher SILAC H/L (heavy/light) ratio in the forward setting and a higher L/H ratio in the reverse setting (Fig. 2A). Most captured proteins in the CLASPI experiment showed little difference in binding to probes N1 and N1c. However, menin was preferentially enriched by probe N1 (Fig. 2, B and C),

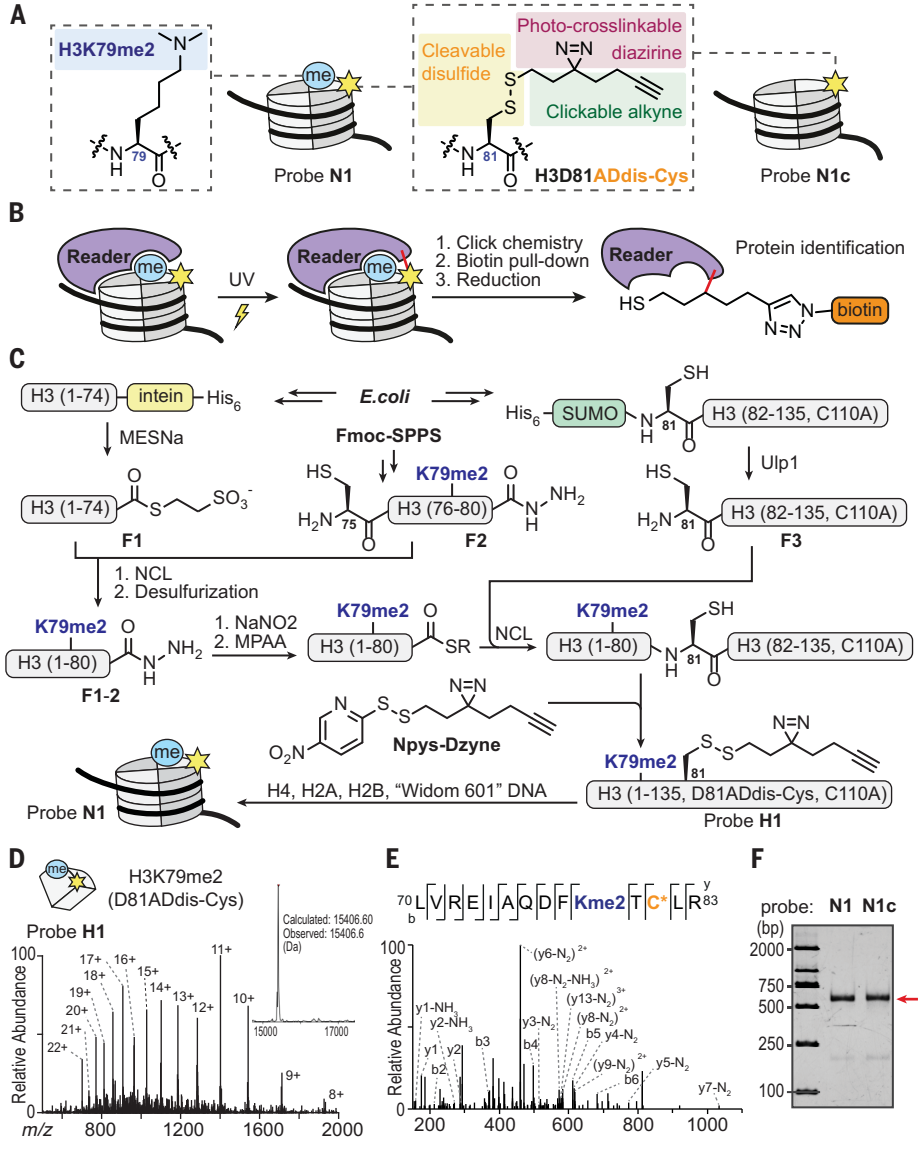


Fig. 1. Development of nucleosome-based photoaffinity probes. (**A**) Illustration and partial structure of probes N1 and N1c. (**B**) Workflow for H3K79me2 "reader" identification using probe N1. (**C**) Synthetic route to generating histone H3K79me2 (D81ADdis-Cys, C110A) (probe H1). (**D**) ESI-MS analysis of probe H1. (**E**) A representative MS/MS spectrum of probe H1, ^{70}LVREIAQDF Kme2 T ADdis-Cys LR^{83}. C*, ADdis-Cys. (**F**) The quality of purified nucleosome probes N1 and N1c examined by native polyacrylamide gel electrophoresis. The gel was stained by ethidium bromide. Red arrow, reconstituted nucleosome probes. 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.

suggesting that it may recognize the H3K79me2 mark on the nucleosome. The selective capture and enrichment of endogenous menin by the methylated nucleosome probe N1 was confirmed by the immunoblotting analysis (Fig. 2D).

Menin is associated with H3K79me2 nucleosomes in cells

To further validate the interaction between menin and the H3K79me2 nucleosome in cells, we overexpressed FLAG-tagged menin in human embryonic kidney 293T (HEK293T) cells and examined whether it could act as a "bait" to pull down nucleosomes in an H3K79 methylation-dependent manner. We isolated nuclei from the cells and treated them with micrococcal nuclease (MNase) to generate the mononucleosomes, which were then subjected to immunoprecipitation using anti-FLAG M2 magnetic beads (fig. S2A). The immunoblotting analysis using the anti-H3 antibody revealed that the mononucleosomes were indeed enriched by FLAG-menin, confirming the association of menin with the nucleosome. The amount of mononucleosomes enriched by FLAG-menin decreased after cells were treated with EPZ5676, a DOT1L inhibitor (*22*) that reduces cellular H3K79me2 level. By contrast, more nucleosomes were enriched by FLAG-menin when the H3K79me2 level was increased by overexpressing DOT1L in the cells (fig. S2B). These results indicate that the interaction between menin and the nucleosome is likely mediated by H3K79 dimethylation.

Menin selectively binds to H3K79me2 nucleosomes in vitro

We next examined whether menin could directly and selectively bind to H3K79me2 nucleosome. To this end, a photocrosslinking experiment was performed with the recombinant menin protein in vitro. As expected, probe N1 captured menin more efficiently than probe N1c (Fig. 2E), validating the methylation-dependent association of menin to the nucleosome. Menin was robustly captured by the nucleosome-based probe N1 but was not captured by the H3 protein-only probe H1 (fig. S2C), indicating that the binding of menin to H3K79me2 requires an intact nucleosome. The selective binding of menin to H3K79me2 nucleosome was also confirmed by nucleosome-based probes N2 and N2c without photoaffinity handles (Fig. 2F and fig. S2, D and E). To determine whether menin formed a stable complex with the H3K79me2 nucleosome, we performed an electrophoretic mobility shift assay (EMSA) using H3K79me2 nucleosome probe N3, which contains Cy5-labeled DNA (Fig. 2G and fig. S2, F and G). We observed a new band with lower mobility, indicating the formation of a stable menin-nucleosome complex with probe N3. The dissociation constant was estimated to be ~6.4 μM (Fig. 2H).

Structural overview of menin bound to H3K79me2 nucleosome

To understand the molecular basis for how menin recognizes the H3K79me2 mark on nucleosome, we reconstituted the menin-H3K79me2 nucleosome complex (hereafter "Menin-H3K79me2") (fig. S3, A to C), cross-linked by GraFix (*23*), for structure determination using cryo–electron microscopy (cryo-EM). We obtained an H3K79me2 nucleosome structure at 2.9-Å resolution and a Menin-H3K79me2 nucleosome structure at ~3.2-Å resolution (fig. S3, D to O). In the Menin-H3K79me2 structure, menin binds to only one face of the nucleosome disk (fig. S3D and movie S1), covering almost half of the nucleosome face (Fig. 3, A to D) and adopting a conformation similar to that of a previous crystal structure of menin (*24*).

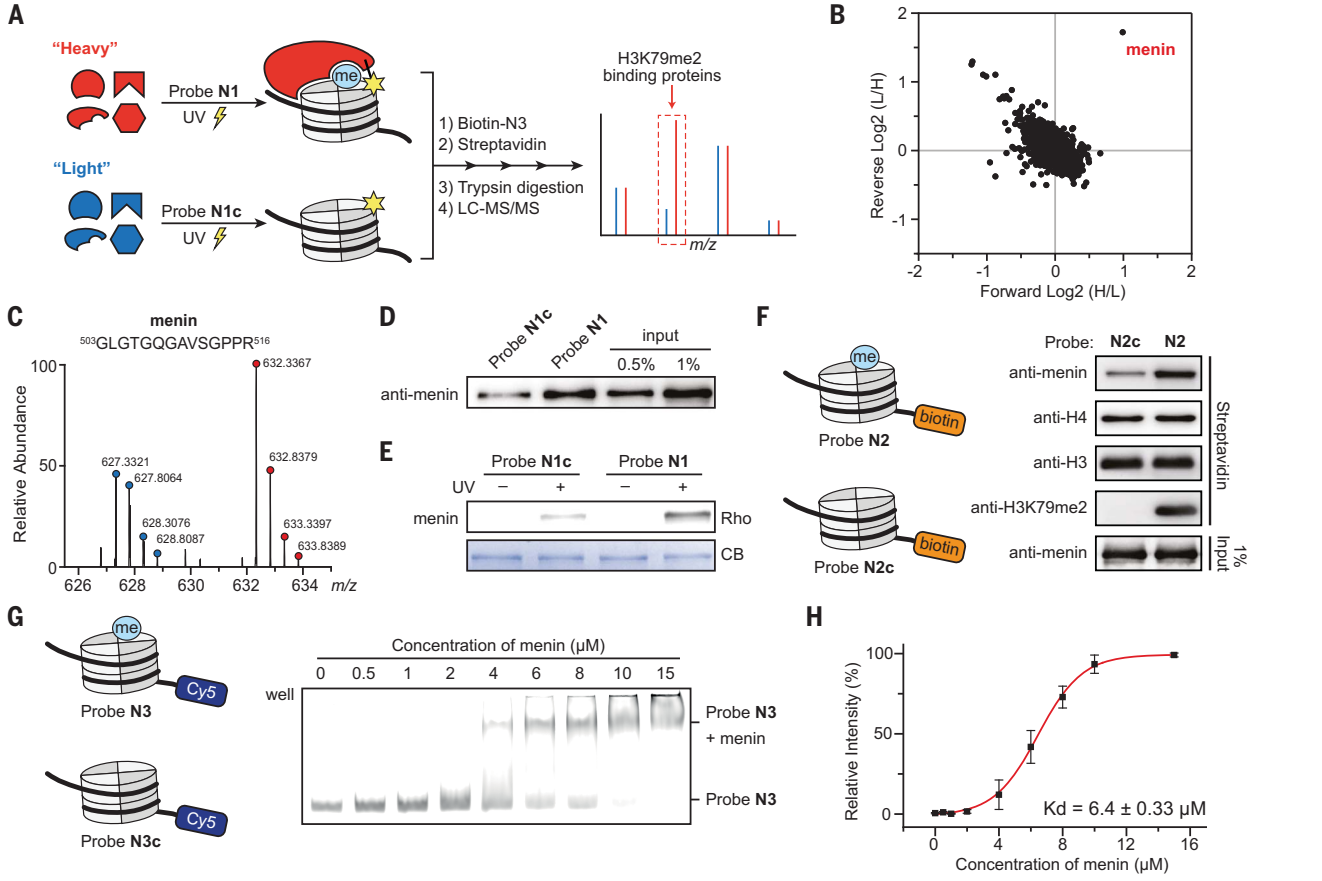


Fig. 2. Identification of menin as an H3K79me2-binding protein in a nucleosomal context. (**A**) Workflow of the CLASPI approach to identify H3K79me2 binders. (**B**) Two-dimensional plot showing that menin stands out from the forward and reverse CLASPI experiments. Data from table S1 are used. (**C**) A representative MS1 spectrum from menin in CLASPI experiments. (**D**) Pull-down of menin by probes N1 and N1c in nuclear extract. (**E**) Photocrosslinking assay of recombinant menin by probes N1 and N1c. Rho, rhodamine fluorescence signal; CB, Coomassie blue staining. (**F**) Non-crosslinking pull-down revealed interaction between menin and nucleosomes. (**G**) EMSA result of Cy5-labeled H3K79me2 nucleosome probe N3 titrated with menin. (**H**) Measurement for the binding affinity of menin toward the H3K79me2 nucleosome in the EMSA assay. Relative intensities of the complex band were plotted against the concentration of menin. Error bars represent mean ± SE (n = 3 biological replicates).

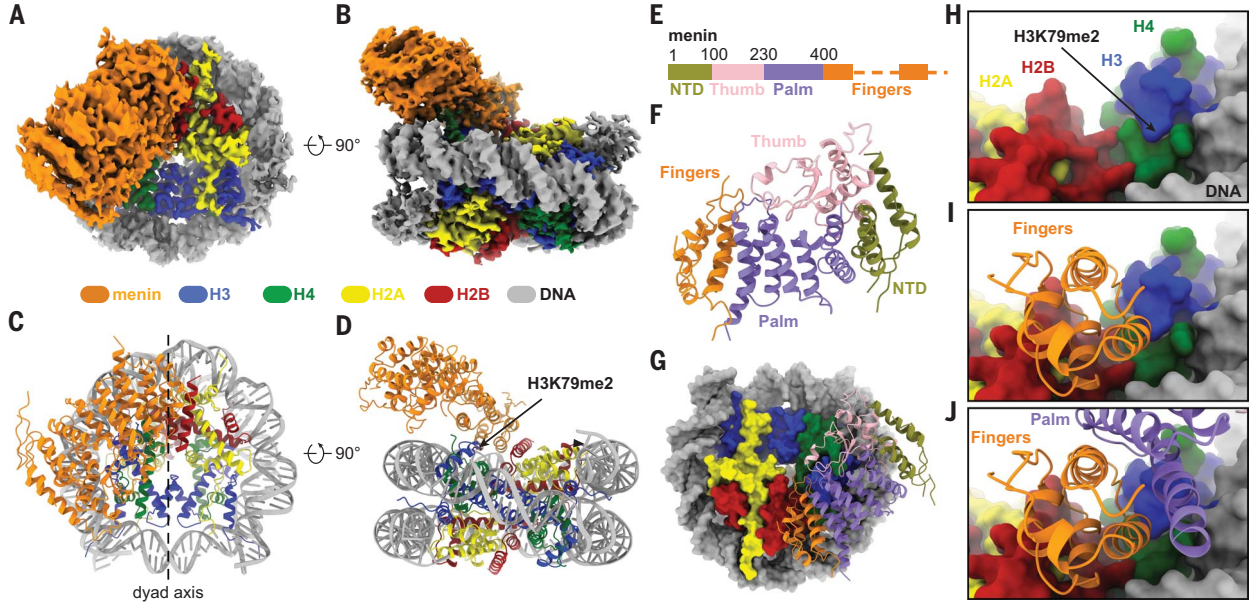


Fig. 3. Overall structure of menin-H3K79me2 nucleosome complex. (**A** and **B**) Top (A) and side (B) views of the segmented cryo-EM map of menin-H3K79me2 nucleosome complex. (**C** and **D**) Same as (A) and (B), respectively, but shown with the atomic model. (**E**) Schematic domain organization of menin. (**F**) Overview of the nucleosome-bound menin structure shown in a cartoon view. The different domains are colored as indicated. (**G** to **J**) H3K79me2 recognition mediated by the fingers and palm domains of menin.

H433 of menin is a key residue for the recognition of H3K79me2 via a π-cation interaction

Menin binds to H3K79me2 nucleosome through its fingers and palm domains (Fig. 3, E to J). The fingers domain engages with the C-terminal α helix of H2B (H2B αC) mainly through three hydrogen bonds on top of the four-helix bundle between H2B and H4 (Fig. 4, A and B, and fig. S3P). This docking is further reinforced by the palm domain through its C-terminal long helix forming four hydrogen bonds with the H3 L1 loop and α1 helix (Fig. 4, C and D, and fig. S3Q). These interactions together place menin in a position to capture the side chain of

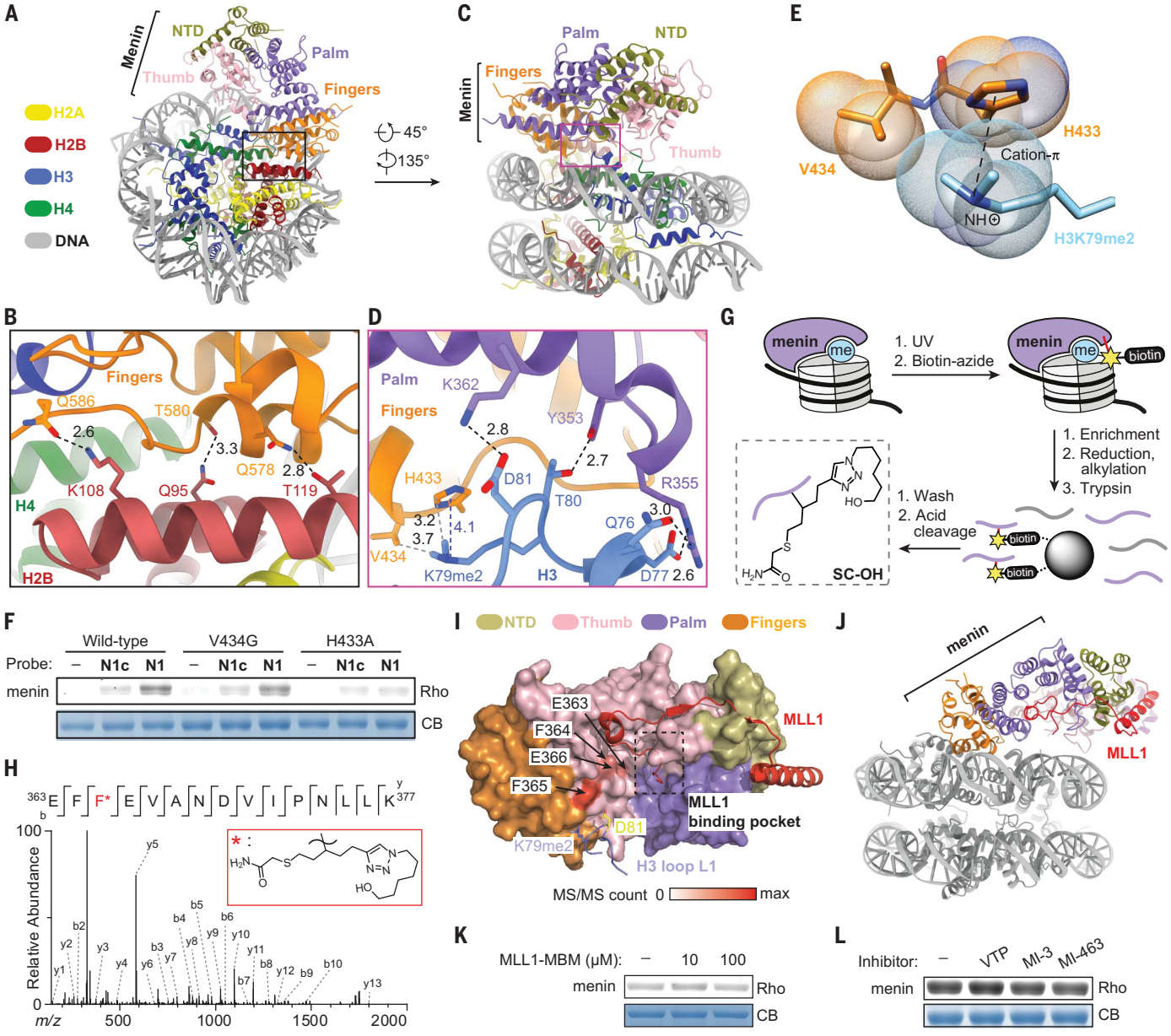


Fig. 4. Menin "reads" H3K79me2 through a π-cation interaction with H433 from its fingers domain. (**A** and **B**) The fingers domain of menin engages with the C-terminal α helix of H2B. Magnified view of the boxed region highlights the interaction details. The hydrogen bonds are indicated by dashed lines. Interatomic distances are shown in angstroms. (**C** and **D**) Menin's palm domain interacts with H3 L1 loop through forming four hydrogen bonds, while H433 from its fingers domain recognizes H3K79me2 through π-cation interaction. (**E**) Magnified view of the π-cation interaction between menin H433 and H3K79me2. The side chains of H433 and V434 and the methylammonium moieties of H3K79me2 are highlighted by the dotted van der Waals radius representation. (**F**) Photocrosslinking assay of recombinant wild-type, V434G, and H433A menin by probe N1 and N1c. Rho, rhodamine fluorescence signal; CB, Coomassie blue staining. (**G**) The workflow used to map the H3K79me2-binding sites of menin. (**H**) A representative MS/MS spectrum for identified SC-OH–labeled menin peptide, ^{363}EFF(SC-OH)EVANDVIPNLLK^{377}. (**I**) Mapping of the identified H3K79me2-binding sites of menin to our structure in surface representation. H3 Loop L1 is shown in cartoon view except that H3K79me2 and H3D81 (where the photocrosslinker is incorporated) are shown in stick format. The heatmap is generated based on the MS/MS counts of "SC-OH" labeled residues (data from table S2). MLL1 from the crystal structure of menin-MLL1-LEDGF (PDB: 3U88), superimposed with our structure, shown in cartoon view. (**J**) Superimposition of the structures of menin-H3K79me2 nucleosome and menin-MLL1-LEDGF (PDB: 3U88) using menin as a reference. (**K** and **L**) Photocrosslinking assay of menin by probe N1 in the presence of MLL1-MBM (K) or menin-MLL1 inhibitors (L).

the H3K79me2 (Figs. 3 and 4). It thus stabilizes the methylammonium moieties of H3K79me2 by both π-cation and hydrophobic interactions with H433 and V434 from its fingers domain (Fig. 4, D and E). Of note, menin binding does not induce obvious conformational changes in histones (fig. S4). Consistent with this observation, the H3K79me2 nucleosome probe N1 showed a largely reduced cross-linking efficiency toward a menin H433A mutant, when compared with the wild-type (Fig. 4F). The unmodified nucleosome probe N1c, which captured menin at much lower efficiency (Fig. 2E), showed little preference for capturing the wild-type or the mutated menin (Fig. 4F). Mutation of V434 to glycine, however, did not affect the binding, suggesting that the hydrophobic interaction mediated by V434 is not essential for H3K79me2 recognition (Fig. 4F). Together, these data demonstrate the key role of H433 of menin in the recognition of the H3K79me2 mark on nucleosome. Our attempt to assemble menin onto the unmodified nucleosome for structural determination was unsuccessful

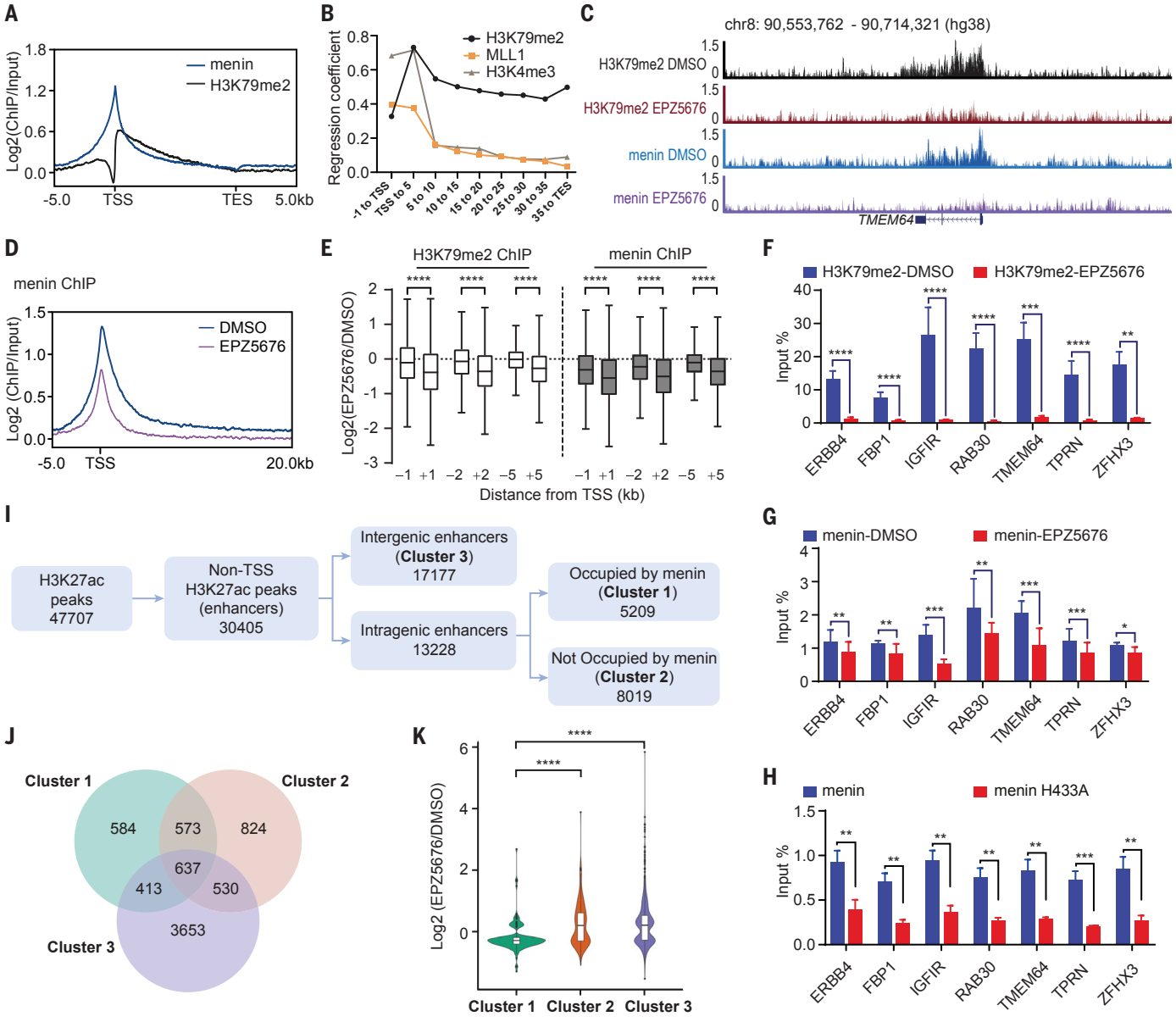


Fig. 5. Menin colocalizes with H3K79me2 at intragenic enhancers and regulates gene transcription. (**A**) Menin and H3K79me2 $\log_2$(ChIP-seq/Input) signal across the gene body. (**B**) Linear regression coefficient of H3K79me2, H3K4me3, and MLL1 ChIP-seq signal/input against MEN1 ChIP-seq/input from the promoter (−1 kb to TSS) through the gene body. (**C**) ChIP-seq signal of H3K79me2 and menin with dimethyl sulfoxide (DMSO) and EPZ5676 treatment at the *TMEM64* locus. (**D**) Menin $\log_2$(ChIP-seq/Input) signal with (EPZ5676) and without (DMSO) Dot1L inhibitor treatment across the gene body. (**E**) Box plot illustrating average $\log_2$(EPZ5676/DMSO) signal over 1, 2, and 5 kb upstream and downstream regions from the TSS. (**F** and **G**) ChIP-qPCR analyses showing the changes in chromatin loci of H3K79me2 (F) and menin (G) on the indicated intragenic regions upon EPZ5676 treatment (n = 6 biological replicates). (**H**) ChIP-qPCR analyses of cells stably expressing wild-type menin or menin H433A on the indicated intragenic regions (n = 6 biological replicates). (**I**) Schematic diagram illustrating how different categories of enhancers are defined. (**J**) Venn diagram showing overlap of genes associated with enhancers from different clusters. (**K**) Violin plot of RNA-seq data representing differential gene expression [$\log_2$(EPZ5676/DMSO)] with specific genes from each cluster (****P < 0.0001). Only significantly differentially expressed genes (adjusted P value < 0.05) are included. Error bars indicate ±SE.*P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

(fig. S3, R and S), likely owing to a largely reduced stability of the menin-nucleosome complex in the absence of H3K79 methylation. Unlike readers of lysine methylation on histone tails, which often capture the methylammonium moiety using an "aromatic cage" constituted with two to four aromatic residues (*25*), menin recognizes H3K79me2 by π-cation interaction with only one aromatic residue, H433. This binding is likely further enhanced by the extensive interactions of menin with histones H3 and H2B of the nucleosome to confer a specific H3K79me2 recognition in a nucleosomal context.

Menin binds H3K79me2 nucleosome and MLL using two different pockets

To provide another line of structural information, we mapped the binding site of menin using our photocrosslinking strategy (*26*). The trifunctional handle of probe N1 enabled crosslinking of the H3K79me2-binding region of menin for enrichment and allowed the release of the cross-linked fragment for MS analysis (Fig. 4G). Notably, only one peptide containing four putative cross-linked residues (E363, F364, F365, E366) was identified in the MS analysis (Fig. 4H), indicating that the binding was highly specific. We mapped these cross-linked residues onto the cryo-EM structure of menin-H3K79me2 nucleosome. The cross-linked residue with the highest MS/MS count, F365, was indeed close to the H3D81 position where the photoreactive warhead was installed (fig. S5, A to C). Considering that the photocrosslinking reaction is highly sensitive to distance, the in-solution structural features of menin at its H3K79me2-binding pocket revealed by the photocrosslinking assay agree well with the solved cryo-EM structure. Of note, the binding sites were located at the boundary between the palm domain and the fingers domain but distal to a conserved binding pocket of several known menin-interacting proteins such as mixed lineage leukemia protein–1 (MLL1) (*24*) (Fig. 4I).

In addition, we also performed a structural comparison between the crystal structure of menin-MLL1-LEDGF (PDB: 3U88) (*24*) and our menin-H3K79me2 nucleosome structure. From Fig. 4J and fig. S5, D to I, it is apparent that the overall architectures of menin from these two structures are similar, suggesting that nucleosome binding does not cause a major conformational change in menin. Whereas menin is recruited onto the H3K79me2 nucleosome via its fingers and palm domains, MLL1-LEDGF mainly associates with the palm and NTD domains. MLL1 embeds its N-terminal menin-binding motif (MBM) deeply into the peptide binding pocket of the palm domain of menin, further extending its remaining part onto the surface of the NTD domain to reach LEDGF on the side far away from the site for nucleosome binding. This structural analysis suggests that menin is capable of binding to both nucleosome and MLL1-LEDGF at the same time without steric clash. Indeed, MBM of MLL1 (^{6}RWRFPARPGTTGGGGGGGRR^{25}) and known menin-MLL1 inhibitors (VTP50469, MI-3, and MI-463) (*27*–*29*) did not abolish the menin-H3K79me2 nucleosome interaction in our photocrosslinking experiment (Fig. 4, K and L). Additionally, the interaction between menin and MLL1 was also not affected by mutating menin's H3K79me2-recognition residue (H433A) (fig S5, J and K), corroborating that menin has two independent binding pockets for MLL1 and H3K79me2 nucleosome. Together, these results suggest that the NTD and palm domains of menin might be important for recruiting menin's associating factors such as JunD or MLL1 to chromatin (*24*).

Menin is associated with H3K79me2 across gene bodies

To examine whether menin recognizes the H3K79me2 mark on native chromatin, we performed chromatin immunoprecipitation sequencing (ChIP-seq) in MCF-7 cells. Consistent with previous studies (*30*, *31*), the H3K79me2 mark was found to peak immediately downstream of the transcription start site (TSS), and the signal diminished throughout the gene body (Fig. 5A and fig. S6A). Menin was previously reported to associate with MLL1 and thereby localize at promoter regions (*27*, *32*). Indeed, menin occupied the promoters and peaked at the TSSs of many active genes in our analysis (Fig. 5A and fig. S6A). However, regression analysis showed that correlation between menin and MLL1–H3K4me3 was only observed around the TSSs. By contrast, the menin signal was strongly predictable by H3K79me2 across the gene body (Fig. 5B).

To further investigate the effects of H3K79me2 on the binding of menin to chromatin, we treated cells with EPZ5676, a DOT1L inhibitor, which resulted in a near-complete global loss of H3K79me2 as detected by immunoblotting (fig. S6, B and C) and by reference-normalized ChIP-seq (Fig. 5C and fig. S6D). The loss of H3K79me2 had no effect on the expression of menin (fig. S6, B and C), but attenuated the interaction between menin and chromatin (Fig. 5, C and D). Consistently, menin binding downstream of TSSs was more sensitive to EPZ5676 treatment compared with menin peaks in the promoter region upstream of the TSS (Fig. 5E). ChIP-qPCR (quantitative polymerase chain reaction) confirmed that the loss of H3K79me2 strongly interfered with the recruitment of menin to the gene bodies of target genes (Fig. 5, F and G), and modestly affected the genomic localization of menin at the promoters (fig. S6, E and F). To further validate that the identified menin binding sites were H3K79me2 dependent, we performed ChIP-qPCR with cells stably expressing either the hemagglutinin (HA)–tagged wild-type menin or the equivalent H433A mutant (fig. S6G). In line with our in vitro biochemical data, the H433A mutation largely disrupted menin's association with chromatin at the selected intragenic regions (Fig. 5H).

Menin is involved in transcriptional regulation through binding to H3K79me2 at potential intragenic enhancers

It was previously proposed that H3K79me2 could activate the expression of target genes through the maintenance of enhancer–promoter interactions (*30*, *33*). Given the correlation between menin and H3K79me2 at genic regions, we explored whether the H3K79me2-menin interaction could mark functional enhancers. To this end, we identified putative enhancers from the presence of H3K27ac peaks (*34*, *35*), then classified them as intragenic or intergenic. We further subcategorized the intragenic enhancers into two groups on the basis of menin occupancy (Fig. 5I and fig. S7A). We correlated these enhancers to their overlapping gene or closest gene based on genomic distance (Fig. 5J). As many genes were regulated by multiple enhancers in each group, we focused on genes that were specifically associated with each type of enhancer: intragenic MEN1-bound (cluster 1), intragenic MEN1-free (cluster 2), and intergenic (cluster 3) (table S3). Most cluster 1 regions overlapped with cis-regulatory regions annotated by ENCODE (Encyclopedia of DNA Elements), and some also show evidence of expression of enhancer RNA (fig. S7B). On the basis of differential expression, we found that EPZ5676 treatment led to significant down-regulation of genes in cluster 1 containing menin-marked enhancers compared with the other clusters, as well as randomly selected gene sets with matched expression levels in untreated cells (Fig. 5K and fig. S7, C to E). Consistent with the transcriptome analysis, EPZ5676 treatment significantly blocked the expression of genes in cluster 1 (fig. S7F). Overall, these data render the possibility that menin might recognize the H3K79me2 mark at intragenic enhancers to mediate gene transcription.

In this study, we identified menin as a bona fide reader of H3K79me2. Menin is known to be involved in transcriptional regulation by interacting with various transcription factors and chromatin regulatory proteins (*36*–*39*). Our study therefore adds a mechanistic possibility that menin recruits its interacting proteins to H3K79me2-marked chromatin regions for gene regulation. Considering that MLL1 is a histone H3K4 methyltransferase, we envision that future studies will reveal whether and how menin may mediate cross-talk between H3K4me3 and H3K79me2 on gene regulation.

REFERENCES AND NOTES

1. E. L. Greer, Y. Shi, *Nat. Rev. Genet.* **13**, 343–357 (2012).
2. A. Jambhekar, A. Dhall, Y. Shi, *Nat. Rev. Mol. Cell Biol.* **20**, 625–641 (2019).
3. K. Hyun, J. Jeon, K. Park, J. Kim, *Exp. Mol. Med.* **49**, e324 (2017).
4. Y. Huyen *et al*., *Nature* **432**, 406–411 (2004).
5. R. Alpatov *et al*., *Cell* **157**, 869–881 (2014).
6. H. Wu *et al*., *PLOS ONE* **6**, e18919 (2011).
7. M. Sabra, P. Texier, J. El Maalouf, P. Lomonte, *J. Cell Sci.* **126**, 3664–3677 (2013).
8. Q. Feng *et al*., *Curr. Biol.* **12**, 1052–1058 (2002).
9. F. van Leeuwen, P. R. Gafken, D. E. Gottschling, *Cell* **109**, 745–756 (2002).
10. K. Wood, M. Tellier, S. Murphy, *Biomolecules* **8**, 11 (2018).
11. B. Zhu *et al*., *Nat. Commun.* **9**, 259 (2018).
12. T. P. Wakeman, Q. Wang, J. Feng, X. F. Wang, *EMBO J.* **31**, 2169–2181 (2012).
13. W. Kim, M. Choi, J. E. Kim, *Cell Cycle* **13**, 726–738 (2014).
14. E. R. Barry *et al*., *Stem Cells* **27**, 1538–1547 (2009).
15. Y. Feng *et al*., *Blood* **116**, 4483–4491 (2010).
16. B. Jones *et al*., *PLOS Genet.* **4**, e1000190 (2008).
17. K. M. Bernt *et al*., *Cancer Cell* **20**, 66–78 (2011).
18. X. Lu *et al*., *Nat. Struct. Mol. Biol.* **15**, 1122–1124 (2008).
19. T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6705–6710 (1998).
20. J. S. Zheng, S. Tang, Y. K. Qi, Z. P. Wang, L. Liu, *Nat. Protoc.* **8**, 2483–2495 (2013).
21. X. Li *et al*., *J. Am. Chem. Soc.* **134**, 1982–1985 (2012).
22. S. R. Daigle *et al*., *Blood* **122**, 1017–1025 (2013).
23. B. Kastner *et al*., *Nat. Methods* **5**, 53–55 (2008).
24. J. Huang *et al*., *Nature* **482**, 542–546 (2012).
25. S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
26. J. Lin, X. Bao, X. D. Li, *Mol. Cell* **81**, 2669–2681.e9 (2021).
27. A. V. Krivtsov *et al*., *Cancer Cell* **36**, 660–673.e11 (2019).
28. J. Grembecka *et al*., *Nat. Chem. Biol.* **8**, 277–284 (2012).
29. D. Borkin *et al*., *Cancer Cell* **27**, 589–602 (2015).
30. L. Godfrey *et al*., *Nat. Commun.* **10**, 2803 (2019).
31. G. Nassa *et al*., *Sci. Adv.* **5**, eaav5590 (2019).
32. K. M. A. Dreijerink *et al*., *Cell Rep.* **18**, 2359–2372 (2017).
33. L. Godfrey *et al*., *Leukemia* **35**, 90–106 (2021).
34. N. D. Heintzman *et al*., *Nat. Genet.* **39**, 311–318 (2007).
35. N. D. Heintzman *et al*., *Nature* **459**, 108–112 (2009).
36. A. Yokoyama *et al*., *Cell* **123**, 207–218 (2005).
37. S. K. Agarwal *et al*., *Cell* **96**, 143–152 (1999).
38. S. Jin *et al*., *J. Clin. Invest.* **120**, 593–606 (2010).
39. H. Sowa *et al*., *J. Biol. Chem.* **279**, 40267–40275 (2004).

ACKNOWLEDGMENTS

Funding: This work was supported by funds from the Hong Kong Research Grants Council Areas of Excellence Scheme (AoE/P-705/16 to X.D.L.); Collaborative Research Fund (CRF C7028-19G to X.D.L., CRF C7009-20G to Y.Z.) and the General Research Fund (GRF 17121120, 17310122 to X.D.L.); Excellent Young Scientists Fund of China (Hong Kong and Macau) (21922708 to X.D.L.); Science and Technology Innovation Committee of Shenzhen (SGDX20201103095201O1 to X.D.L.); Centre for Oncology and Immunology under the Health@InnoHK Initiative funded by the Innovation and Technology Commission, the Government of Hong Kong SAR, China (J.W.H.W); and the K. C. Wong Educational Foundation (GJTD-2020-06). The EM dataset was collected at the Biological Cryo-EM Center, generously supported by a donation from the Lo Kwee Seong Foundation, at HKUST. Cryo-EM data processing was performed in workstations and the cluster HPC3 supported by the Information Technology Services Center (ITSC) at HKUST. We thank Y. Zhang (Biological Cryo-EM Center, HKUST) and P. W. Luo (ITSC, HKUST) for their support. D.Y. is supported by an LKS fellowship. **Author contributions:** Design of experiments: X.D.L., J.L., Y.W., X.B., and G.T. Probe synthesis: Y.W., J.L., Y.J., and Z.L. Mass spectrometry sample preparation and data analysis: Y.W., X.B., and J.L. Validation of H3K79me2-menin interaction in vitro and in cells: J.L., Y.W., and G.T. Cryo-EM sample preparation: J.L., W.L., D.Y., Y.Z., and S.D. Cryo-EM data analysis: D.Y., W.L., Y.Z., and S.D. ChIP-seq and RNA-seq sample preparation: X.B. and G.T. ChIP-seq and RNA-seq data analysis: J.W.H.W. and E.Y. Writing – original draft: Y.W., Y.Z., X.B., and G.T. Writing – review and editing: X.D.L., Y.Z., J.W.H.W., X.B., and J.L. Visualization: J.L., Y.Z., Y.W., D.Y., J.W.H.W., E.Y., X.B., and G.T. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The menin-H3K79me2 nucleosome cryo-EM model and map reported in this paper have been deposited in the Protein Data Bank (PDB) and Electron Microscopy Data Bank (EMDB), respectively, with accession numbers 8GPN (PDB) and EMD-34195 (EMDB). The raw sequencing data for ChIP-seq and RNA-seq analyses presented in this paper have been deposited in Gene Expression Omnibus with the accession number GSE202593. All other data are available in the main text or the supplementary materials. Newly created materials from this study may be requested from the corresponding authors. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adc9318
Materials and Methods
Figs. S1 to S7
Tables S1 to S5
Data S1 and S2
References (*40*–*58*)
MDAR Reproducibility Checklist
Movie S1

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

Submitted 10 May 2022; resubmitted 29 August 2022
Accepted 5 January 2023
10.1126/science.adc9318

REVIEW SUMMARY

IMMUNOLOGY

The extracellular matrix and the immune system: A mutually dependent relationship

Tara E. Sutherland, Douglas P. Dyer, Judith E. Allen*

BACKGROUND: The extracellular matrix (ECM) forms a dynamic structure around cells that is essential for the supply of environmental factors, mechanical support, and protection of tissues. It includes components such as fibrillar proteins, glycosaminoglycans (GAGs), proteoglycans, and mucus. The molecular, physical, and mechanical properties of the ECM regulate immune cell mobility, survival, and function. In turn, the immune system maintains and regulates healthy matrix and restores matrix integrity after injury. A dysregulated ECM–immune system partnership contributes to most diseases. Exploring the complex interconnectivity between ECM biology and immune cells has the potential to help treat disease and maintain healthy aging.

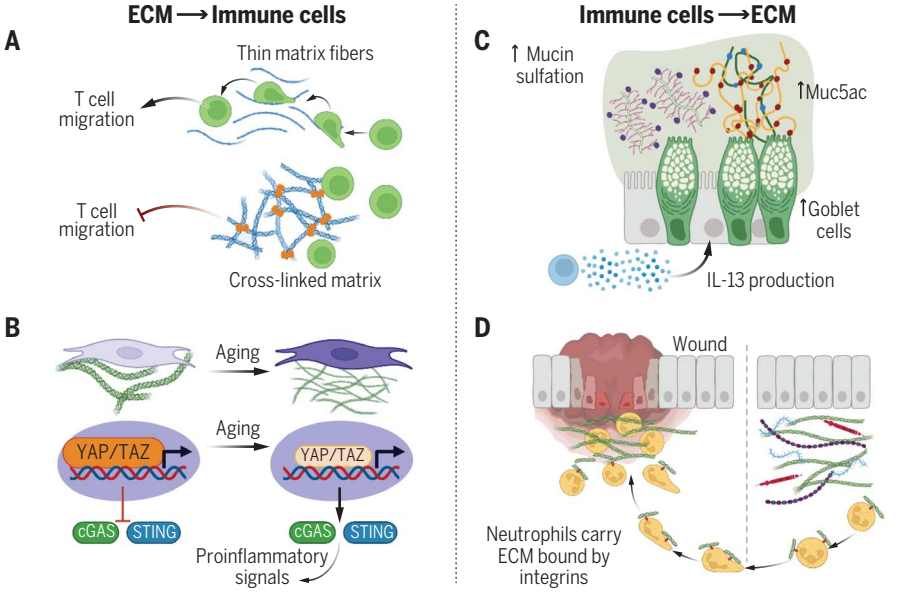
ADVANCES: Immune cells are perpetually in contact with the ECM, yet the potential consequences of these interactions often remain unexplored. One function of the ECM is to guide immune cell movement and positioning. T cells, for example, move through sites containing thin ECM fibers in preference to more densely cross-linked collagen matrices, whereas heparan sulfate proteoglycans within the vasculature and tissue parenchyma bind and present chemokines to form gradients that direct cell movement. During inflammation, injury, infection, or even aging, ECM components can be released to act as "danger signals." Conversely, the breakdown of the ECM by matrix-degrading enzymes can generate immunoregulatory fragments. Critically, because cytokines are often bound to GAGs, ECM changes can regulate cytokine availability or activity.

During aging and fibrotic diseases, changes to the fibrillar collagen network result in pathological tissue stiffness and loss of mechanical compliance. Additionally, increases in the GAG hyaluronan contribute to altered ECM mechanical properties that accompany aging and disease. There is increasing evidence that immune cell function is regulated by mechanosensing receptors such as Piezo1, and these ECM changes have major impacts on immune function. Furthermore, a decline in the activity of the mechanosensing transcriptional activators YAP and TAZ during physiological aging results in failure to down-regulate inflammation. Thus, ECM composition actively regulates immune processes, but immune signals will themselves regulate ECM composition, reflecting an essential bidirectional dialog.

One prominent way that the immune system regulates the ECM is through transforming growth factor–β (TGF-β), which promotes myofibroblast differentiation and collagen production and inhibits matrix-degrading metalloproteinases. Furthermore, type 2 cytokines, particularly interleukin-13 (IL-13), have emerged as modulators of ECM quantity and quality, which includes regulating the mucosal barrier. Moreover, direct biophysical interaction with chemokines or cytokines can alter ECM structure and/or function. For instance, CXCL4 (PF4) functions by binding to GAGs rather than by directly binding to chemokine receptors. This can lead to remodeling of the cell surface ECM and signaling through proteoglycans.

Immune cells control ECM not only through the production of cytokines and chemokines but also by direct synthesis of ECM components and the enzymes that break them down. Enzymatic remodeling of the rigid basement membrane by tissue-infiltrating myeloid cells, for example, can provide routes for lymphocytes to follow. Macrophages, which are pivotal in ECM turnover through receptor-mediated uptake of and degradation of collagen, also produce collagens that may provide templates for tissue remodeling. Neutrophils can pull and carry preexisting matrix from nearby sites to wound beds early in the tissue repair process to reestablish new ECM scaffolds.

OUTLOOK: The ECM, long considered an inert scaffold, can now be seen as a highly dynamic partner to the immune system. In this review, we aim to highlight the absolute interdependence of these systems with consequences for therapy. For example, immune cell–based therapies may fail if they are placed in diseased matrix that itself drives pathology. Ultimately, to answer the most pressing questions in tissue health, it will be critical that immunologists and matrix biologists work together. ■



Interconnectivity between the ECM and the immune system. (**A**) Physical and molecular properties of the ECM control immune cell positioning and migration within a tissue during pathology. (**B**) Aging not only alters the mechanical properties of ECM, which are sensed by the immune system, but also reduces mechanosensors that down-regulate proinflammatory pathways. (**C** and **D**) The immunomodulatory cytokine IL-13 can remodel the mucus barrier (C), whereas immune cells themselves can carry matrix across tissues to help build an ECM scaffold for repair (D). (Figure created with BioRender.)

The list of author affiliations is available in the full article online.
*Corresponding author. Email: judi.allen@manchester.ac.uk
Cite this article as T. E. Sutherland *et al.*, *Science* **379**, eabp8964 (2023). DOI: 10.1126/science.abp8964

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abp8964

REVIEW

IMMUNOLOGY

The extracellular matrix and the immune system: A mutually dependent relationship

Tara E. Sutherland[1,2], Douglas P. Dyer[1,3], Judith E. Allen[1]*

For decades, immunologists have studied the role of circulating immune cells in host protection, with a more recent appreciation of immune cells resident within the tissue microenvironment and the intercommunication between nonhematopoietic cells and immune cells. However, the extracellular matrix (ECM), which comprises at least a third of tissue structures, remains relatively underexplored in immunology. Similarly, matrix biologists often overlook regulation of complex structural matrices by the immune system. We are only beginning to understand the scale at which ECM structures determine immune cell localization and function. Additionally, we need to better understand how immune cells dictate ECM complexity. This review aims to highlight the potential for biological discovery at the interface of immunology and matrix biology.

The molecular, physical, and mechanical properties of the extracellular matrix (ECM) regulate immune cell mobility, survival, and function. In turn, the immune system is required to maintain and regulate healthy matrix and restore matrix integrity after injury. Thus, the ECM and the immune system work together to preserve tissue health, and if this partnership is compromised, disease can develop. However, matrix biology and immunology are typically studied independently. Progress in understanding and treating disease requires that we embrace the complexity not only of different immune cell-cell interactions but also of their interactions with the matrix that surrounds them, recognizing that the matrix is as complex as the cellular systems it regulates.

The ECM accounts for more than a third of our body mass, and its dysregulation is the direct or indirect cause of most major chronic diseases. A prime example of this is fibrosis, which accounts for 45% of all deaths (*1*) and can be defined as dysregulated matrix that leads to the formation of dysfunctional scar tissue and end-stage organ failure. To understand the immunological basis for these diseases, we therefore need an understanding of the interconnectivity between the ECM and immune function. Indeed, immune cell therapies may fail because the cells have been transferred into a diseased matrix that recapitulates a pathogenic signaling process (*2*). Consequently, a knowledge of matrix biology is essential to aid immunological discovery and unravel mechanisms driving complex diseases. In this review, we highlight the intimate and essential connection between the ECM and immune cells and identify emerging areas of research that are ripe for discovery.

[1]Wellcome Centre for Cell-Matrix Research, Lydia Becker Institute for Immunology & Infection, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Center, University of Manchester, Manchester M13 9PT, UK. [2]School of Medicine, Medical Sciences and Dentistry, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. [3]Geoffrey Jefferson Brain Research Centre, Manchester Academic Health Science Centre, Northern Care Alliance NHS Group, University of Manchester, Salford M6 8HD, UK.
*Corresponding author. Email: judi.allen@manchester.ac.uk

What is the ECM?

The ECM forms a structure around cells that is essential for the supply of nutrients and local and systemic factors, as well as the removal of waste. It also provides different levels of mechanical support specific to different tissues. The ECM is produced and assembled by the cells living within it and contains components that can be either long-lived or rapidly turned over, making the ECM dynamic and sensitive to local and systemic changes. Diverse ECM functions are reflected by a varied composition of a hierarchy of protein and glycan structures. These include fibrillar molecules such as collagen (which give tissue strength and resilience), glycosaminoglycans (GAGs) and proteoglycans (which form hydrated gels that act as a cushion within and around tissues), and mucus (which protects our barrier surfaces). Additionally, many matrix-associated molecules are involved in holding ECM structures together and/or connecting them to the cell surface. Together, ECM components not only provide dynamic tissue organization and integrity but also are signaling molecules in their own right, participating in and driving many biological functions. Although many ECM macromolecules are common across tissues, the complex structures that they form, along with unique modifications, are highly tissue specific (Fig. 1). The biochemistry of ECM molecules, including collagen (*3*), sulfated GAGs (*4*–*6*), nonsulfated GAGs (*6*–*8*), mucins (*9*), and the ECM more broadly (*10*, *11*), have been extensively reviewed previously. From an immunological standpoint, matrices also contain a varied spectrum of other secreted proteins, including cytokines, chemokines, and growth factors, that can be active, latent, or concealed and are an implicit part of immune cell regulation. It is also notable that cells are often intimately connected to the unique matrix surrounding them, and every cell to varying degrees is coated in a glycocalyx, a complex network of sugar-rich molecules either free or bound to protein or lipid (*12*).

How does the ECM regulate immune cell function?

Immune cells are in contact with the ECM from their early embryonic or bone marrow origins through to effector functions in the tissues where they transit or reside, yet the potential consequences of these interactions often remain unexplored. For example, given the complex structure and function of the ECM, is it easier for a cell to move through a loose, hydrated matrix or to use molecular handles (e.g., integrins) to latch onto and pull through a thicker matrix? The best understood of these processes for immunologists is the process of leukocyte extravasation during an inflammatory event (*13*). However, even here, immunologists rarely consider the challenges required for cells to cross through the different ECM environments involved. These include the endothelial glycocalyx and the basement membrane, as well as interstitial matrix lying beneath endothelial cells that line the blood vessel wall. Furthermore, once immune cells have penetrated the endothelial lining and entered the tissues, the ECM continues to direct immune cell migration. For both migrating and tissue-resident cells, the ECM not only helps to determine their ultimate location but also regulates their survival and function.

The endothelial glycocalyx shield

The most studied ECM barrier in the context of leukocyte trafficking and positioning is the basement membrane. This complex network of collagen (types IV and VIII) and laminin, which is bridged by the glycoprotein nidogen and proteoglycans (e.g., perlecan) (Fig. 1), lies underneath the vasculature and surrounds most tissues (*14*). However, the first ECM structure encountered by trafficking immune cells is the endothelial glycocalyx lining the vascular endothelium (Fig. 2) (*15*, *16*). The glycocalyx can extend from 200 nm up to 2 μm into the blood vessel lumen and is largely composed of proteoglycans and GAGs. Within the glycocalyx, ECM components close to the endothelial surface (within 50 nm) facilitate interactions with endothelial adhesion molecules on circulating leukocytes. This structure of the

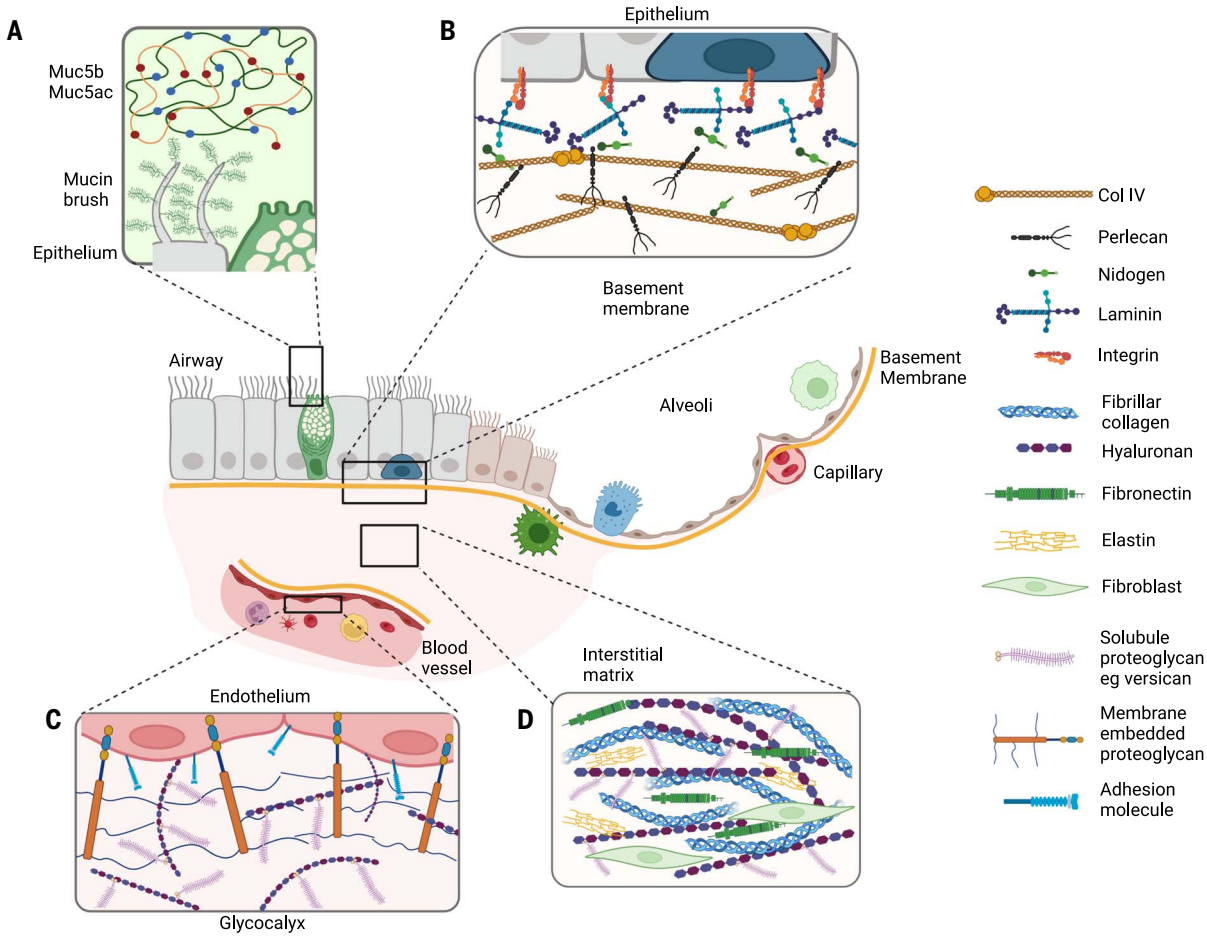


Fig. 1. ECMs in the healthy lung. Shown is an overview of a healthy lung airway and the alveolar structure, including capillaries, blood vessels, and the various ECMs that can be found within the tissue architecture. (**A**) Airway epithelial cells are protected by a periciliary layer containing mucins that are tethered to the cilia of epithelial cells and a gel-like mucus mesh of secreted mucins Muc5ac and Muc5b. (**B**) Basement membrane ECM forms a sheet-like layer that coats the basal surface of epithelial and endothelial cells. Collagen IV fibers are important for maintaining the normal architecture of the tissue. Laminin chains form into a cruciform structure linking cells through integrin binding to the basement membrane, and they connect to collagen fibers directly or through interactions with HS proteoglycans (e.g., perlecan) and proteins (e.g., nidogen), overall contributing to an ECM network. (**C**) The glycocalyx is a brushlike layer that lines the luminal surface of endothelial cells and is rich in membrane-bound proteoglycans such as syndecan, glycosaminoglycans such as HA, and soluble chondroitin sulfate proteoglycans such as versican. (**D**) The interstitial lung matrix is primarily maintained by fibroblasts and made up of the core proteins type I and III fibrillar collagens and elastin arranged in a three-dimensional network, providing structural support and flexibility of the lung during homeostasis. Fibronectin helps to orientate new collagen fibers and can aid cell anchoring to the ECM through integrin binding. HA, the largest and most abundant nonsulfated GAG, helps to assemble, hydrate, and stabilize connective tissue ECM and can be decorated with proteoglycans such as versican. Molecules are not shown to scale.

glycocalyx raises a key challenge to our current understanding of leukocyte migration. Specifically, how can leukocytes interact with endothelial adhesion molecules when they are covered by a glycocalyx "blanket" that is too thick to be simply bridged by endothelial and leukocyte adhesion molecules (Fig. 2) (*15*, *17*, *18*)? Enzymatic digestion of glycocalyx components has been shown to reveal endothelial adhesion molecules and to facilitate increased leukocyte-endothelial interactions (*15*, *16*, *19*). Degradation of the glycocalyx is a critical feature of sepsis, where tumor necrosis factor (TNF) is a key inducer of endothelial heparanase, leading to exposure of adhesion molecules and facilitating neutrophil trafficking (*20*, *21*). However, the role of the endothelial glycocalyx in the trafficking of leukocytes other than neutrophils is poorly understood. The identification of glycocalyx fragments in the serum of patients with sepsis, cardiovascular disease, and, more recently, COVID-19 (*19*, *22*–*24*) suggests that breakdown of the glycocalyx likely contributes to disease. However, a caveat is that these fragments may be released from sources other than the endothelium, because the specific glycocalyx content and structure across different vessel types and tissues are not well characterized. A deeper analysis of glycocalyx structures throughout the body will enable us to better understand the consequences of its degradation during inflammation.

The matrix as a guide to immune cell movement and position

Multiple physical and chemical properties of the ECM guide cell migration (*25*). In numerous settings, leukocytes have been shown to navigate through the path of least resistance, often through "holes" in the matrix that have been studied extensively (*26*, *27*). For example, T cells will preferentially move through tissue sites that are characterized by thin fibers of collagen and fibronectin, deliberately avoiding denser matrices that are made more substantial by lysyl oxidases that cross-link collagen proteins (*28*). The inhibition of lysyl oxidases, which weakens the fibrillar collagen network, improves T cell access to tumors (*29*)

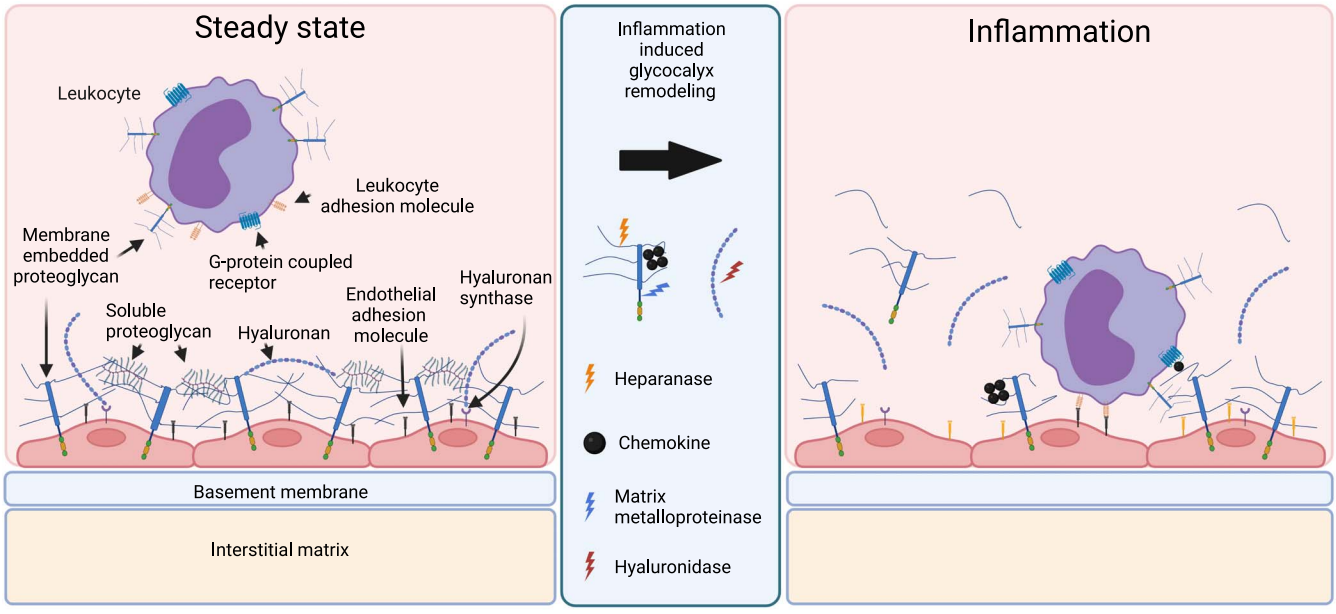


Fig. 2. The endothelial glycocalyx shield is remodeled during inflammation to facilitate leukocyte rolling on the endothelium. The endothelial glycocalyx is a thick (200 to 2000 nm) barrier on the blood-exposed endothelial surface largely composed of ECM proteoglycans. This barrier forms a "blanket" over the endothelial adhesion molecules (e.g., P-selectin), thus blocking interaction with adhesion molecules on circulating leukocytes (e.g., PSGL1) and preventing aberrant leukocyte recruitment. After an inflammatory stimulus (e.g., TNF), endothelial cells and leukocytes themselves can produce factors that remodel the glycocalyx barrier to facilitate endothelial-leukocyte interactions. Specifically, heparanase cleaves the GAG sugar side chains of proteoglycans, matrix metalloproteinases cleave their protein cores, and hyaluronidase cleaves HA, increasing their levels in the circulation. Additionally, chemokines may cross-link GAG chains so that they occupy less space within the glycocalyx. Together, these mechanisms enable leukocytes to access the endothelial surface, initiate rolling, and subsequently enter underlying tissues.

and increases the speed and travel distance of innate lymphoid cells during lung inflammation (*30*). Moreover, collagen fibers can directly influence cell motility, as shown by the ability of type 1 collagen to alter the cytoskeletal organization and shape of innate lymphoid cells toward a promigratory exploratory phenotype (*30*). In inflamed skin, confirmational changes to fibronectin fibers result in enhanced integrin binding. These altered ECM fibers can arrest the interstitial migration of T cells, causing cellular accumulation at perivascular sites (*31*). The mechanical stretching properties of native ECM fibers can further regulate integrin-mediated cell binding and biochemical signaling (*32*). Once positioned within a tissue environment, the necessity for communication with the ECM has been highlighted by the identification of key receptors for ECM on a variety of myeloid cell subpopulations. LAIR1, a receptor that binds glycine-proline-hydroxyproline repeats in collagen, is needed for the survival, proliferation, and differentiation of monocytes and interstitial macrophages in the lung (*33*).

Hyaluronan (HA), a major structural component of the ECM, is an enormous hydrophilic GAG (>10^6 D) made up of unbranched repeating units of D-*N*-acetyl glucosamine and D-glucuronic acid. Through interactions with its receptors, HA facilitates immune cell function in part by directing localization. HA is found in the glycocalyx of many cells. In some cases, such as in the alveolar macrophage, it is constitutively bound to the cell surface through CD44, an HA receptor found on most leukocytes (*34*). With a rapid turnover compared with other ECM components (*35*, *36*), HA production is tightly balanced and controlled locally through CD44-mediated cellular uptake and degradation by hyaluronidases [for review, see (*37*)]. HA-CD44 interactions can be essential for leukocyte migration and location, as shown in the HA-rich liver sinusoids, where CD44 binding to HA allows neutrophil adhesion in the absence of other adhesion molecules (*38*). After antigen-induced activation, T cells will bind HA through enhanced expression of CD44 (*39*), possibly honing their recruitment to specific inflamed tissue sites (*40*). Lyve-1, another receptor for HA, is found on macrophages that have specific functions in the maintenance of the vasculature. Here, HA dictates Lyve-1$^+$ macrophage location along the blood vessel wall (*41*). Additionally, Lyve-1 expression on the lymphatic endothelium acts as a dock allowing dendritic cells, and potentially other immune cells that are endogenously coated in HA, to adhere and traffic through the lymphatics (*42*). It is not simply the presence of HA and its receptors that determines cellular location, but also the specific composition of the HA matrix. For example, Lyve-1 only binds HA that is correctly organized and cross-linked by molecules such as TSG-6, which is up-regulated during inflammation (*43*). Additionally, versican, a large chondroitin sulfate–containing proteoglycan, can bind to HA with high affinity, forming distinct HA matrices that often accumulate during inflammation (*44*). TSG-6 cross-linking of HA (Fig. 3) and versican-HA aggregates can also enhance cellular interactions with CD44 (*45*, *46*), adding another level of ECM control to leukocyte adhesion and/or migration. Moreover, HA receptors can act in tandem. CD44 alters the spatial organization of the HA glycocalyx on dendritic cells to allow binding to Lyve-1 on the lymphatic endothelium, thus regulating the efficiency of dendritic cell entry to the lymphatics (*47*).

Similarly, heparan sulfate (HS) proteoglycans help to direct the recruitment and positioning of neutrophils and monocytes through direct interactions between their GAG side chains and P and L selectins on immune cells (*48*, *49*). The ECM can also act indirectly as a cell guide by "presenting" interacting proteins. For example, HS proteoglycans bind and present chemokines within the vasculature, as well as the tissue parenchyma, to form gradients that direct cell movement (*50*, *51*). This interaction is also important in decoding complex chemokine signals. For example, the CXCR3 ligands CXCL9, CXCL10, and CXCL11 are often expressed together but have very different affinities for HS, which allows discrete localization

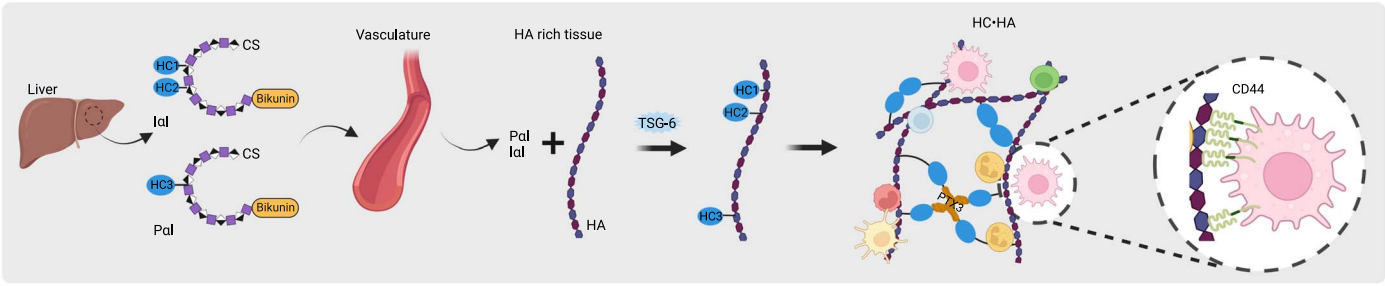


Fig. 3. HC cross-linked HA matrix. IαI and pre-IαI [composed of bikunin and HC proteins attached to a chondroitin sulfate (CS) backbone] are assembled in the liver and secreted into the circulation with transfer of HCs during inflammation catalyzed by tumor necrosis factor–stimulated gene 6 (TSG-6). Formation of HC•HA leads to cross-linking of HA chains through interactions with HCs and through HCs with accessory proteins such as pentaxin 3 (PTX3). Cross-linked HC•HA can stabilize a pathological HA matrix that has enhanced affinity for HA receptor CD44 and thus increased adhesion for leukocytes.

(*52*). HS proteoglycans also bind cytokines, but the consequences for immune cell location and function of these interactions remain relatively unexplored.

The inflammation-altered matrix

A fundamental feature of the ECM is that it is altered after inflammation, injury, or infection and with age (*53*–*55*). For example, HA increases in the lung after influenza infection (*55*) and during aging (*54*). Many different molecules bind to HA chains, forming large, hydrated matrix scaffolds (Figs. 1C and 3). During injury and inflammation, HA can become covalently modified with heavy chains (HCs) from the inter-alpha-inhibitor (IαI) family of proteoglycans (*56*), forming cross-linked HA matrices mediated by HC-HC and pentraxin-HA interactions (*57*, *58*) (Fig. 3). Although such HC•HA matrices are critical for some normal biological processes, they are generally only made in tissues during inflammatory conditions. Cross-linked HC•HA can have increased affinity for CD44 and binding to leukocytes (*59*, *60*), which can lead to the retention of immune cells within the pathogenic tissue ECM, thus prolonging inflammation (*61*). However, the contribution of an HA matrix to any immune process will depend on its composition and the specific tissue or disease context (*56*). These large, hydrated HA structures, including HA bound to versican, can also directly restrict cell movement. Two matrix proteases, ADAMTS4 and ADAMTS5, are both needed during influenza infection to break down versican, remodeling the ECM such that a conduit is created through which $CD8^+$ T cells can travel (*62*, *63*). The failure to provide passage for $CD8^+$ T cells in *Adamts4*-deficient mice (*62*) prevents immune pathology but compromises host protection in *Adamts5*-deficient mice (*63*).

Breakdown of the ECM by proteases such as the ADAMTS family members generates fragments that are themselves immunoregulatory [for review, see *44*)]. For example, HA becomes fragmented at sites of inflammation and injury, leading to the buildup of low-molecular-weight HA, which may antagonize normal HA functions through competition for receptor binding (*64*). Although the specific immunoregulatory roles of low-molecular-weight HA remain to be elucidated, *Cd44*-deficient mice, which fail to clear it, die from lung failure after bleomycin-induced fibrosis (*65*). Consistent with the role of HA in regulating immune cell function, HA binding through the CD44 receptor on activated regulatory T cells (T_{regs}) promotes interleukin-2 (IL-2) production, persistence of FoxP3 expression, and enhanced IL-10 production (*66*, *67*). Other ECM components released by tissue damage have well-established roles as "danger signals" (*68*–*70*). Toll-like receptor 2 (TLR2) and TLR4, in particular, recognize proteoglycans such as biglycan, aggrecan, and versican, as well as various GAGs. Macrophages can be activated through TLR2 and TLR4 by heparanase-mediated GAG cleavage, and this is associated with atherosclerotic plaque progression (*71*). Given that the ECM is one of the first tissue components damaged by infection, it makes sense that its degradation acts as a danger signal to the immune system.

The ECM is a rich source of chemokines, cytokines, and growth factors, the availability of which is controlled by the ability of immune cells to access them. Indeed, many immune cells, including T cells, express heparanase, which digests HS to release bound cytokine (*72*). However, damage or remodeling of the ECM caused by tissue injury or matrix-degrading enzymes can also cause the release or activation of cytokines normally bound to GAGs (*73*). The bioavailability of transforming growth factor–β (TGF-β), one of the most important cytokines in the regulation of the ECM, is itself tightly controlled by its regulated release from the ECM (*73*). The mechanical state of ECM fibers, such as their capacity to stretch or unfold, can also dictate cytokine availability or activity (*32*). For example, the large latent TGF-β complex stored in the ECM can be activated by mechanical forces, and a stiffer matrix can reduce the threshold for TGF-β1 activation (*74*). Thus, the capacity of inflammation to alter the chemical composition and mechanical properties of the ECM will modify an enormous range of immune processes and needs to be considered when investigating the initiation or resolution of an inflammatory response.

Mechanical changes with age and disease

Many tissues lose mechanical compliance with age and during the course of fibrotic disease. Although alterations to the fibrillar collagen network are a major determinant of pathological tissue stiffness, other ECM changes alter the mechanical properties of the ECM with age and disease (*75*), including increased HA (*54*, *55*, *61*). The transcriptomes of tissue-resident alveolar macrophages change with age, not because of the age of the cell but rather because of signals from the aging lung ECM (*54*). The ability of immune cells to respond to mechanical cues is now well established (*76*), and many (if not most) fundamental immune processes will likely be influenced by the biophysical properties of the matrix environment. For example, an emerging paradigm suggests that stiffer, less-elastic tissues promote an inflammatory response in macrophages (*77*). However, numerous signals integrate to determine macrophage activation state, and more compliant substrates can increase sensitivity to proinflammatory stimuli (*78*). Because most work on this topic has involved either in vitro analysis or the use of biomaterials in vivo, we still lack an understanding of the contribution of natural tissue–specific variation in ECM elastic and tensile properties (e.g, soft brain and hard bone). The remodeling of the ECM after injury, such as the deposition of thicker and denser collagen fibrils in scar tissue, is strongly implicated as a driver of further disease progression (*2*, *79*). However, the unique biophysical properties of native ECM fibers are very difficult to replicate synthetically (*32*), so the specific functional consequences for local immune cells remain to be discovered.

An intimate link between mechanosensing of the ECM and regulation of inflammatory signals can drive aging (*80*). The transcriptional activators YAP and TAZ transduce mechanical signals, and their activity declines with physiological aging in stromal and contractile cells. YAP-TAZ signaling inhibits the proinflammatory immune cGAS-STING pathway, and when this restraint is lifted, cell senescence proceeds with the emergence of aging-related tissue degeneration (*80*). The identification of specific receptors involved in mechanotransduction is also providing us with a new understanding of the impact of tissue biophysical properties on immune cell function. For example, Piezo1, a mechanosensitive ion channel protein, has been demonstrated in vivo to regulate macrophage responses to physical forces (*81*). More broadly, there is good evidence that Piezo1 regulates immune cell function (*82*, *83*). Critically, this is a two-way communication: Whereas tissue stiffness regulates immune cell function, immune signals can alter tissue stiffness in a wide variety of mechanisms, from regulating collagen production and cross-linking (see below) to inflammation-induced HA matrices (*56*).

How do immune cells regulate ECM composition?

ECM regulation by chemokines and cytokines

The cytokine with the most evident influence on shaping the ECM is TGF-β, with its ability to promote myofibroblast differentiation, induce collagen production, and inhibit the matrix metalloproteinases that degrade matrix (*75*). However, type 2 cytokines, in particular IL-13, have emerged as another layer of cytokine control over matrix quantity and quality of the ECM. Consistent with type 2 cytokines being mediators of tissue repair and remodeling (*84*), type 2 cytokines regulate the production of collagen-degrading enzymes (*85*), the production of collagen by macrophages (*86*), and the uptake and degradation of collagen (*87*). IL-4 and IL-13, which both signal through the IL-4 receptor alpha (IL4Rα), appear to be central players in the maintenance of the ECM—but also in ECM dysregulation. IL4Rα signaling induces dermal, synovial, and lung fibroblasts to produce more collagen, as well as more hepatic stellate cells, which are the primary source of aberrant ECM during liver fibrosis (*88*–*93*). Indeed, IL-13 signaling to fibroblasts is sufficient to drive liver fibrosis (*94*). IL4Rα signaling can also regulate the amount of collagen cross-linking by signaling to macrophages. IL4Rα-dependent RELMα production by macrophages is critical for vascular integrity and repair of the skin by regulation of the collagen cross-linking enzyme lysyl hydroxylase in fibroblasts (*95*). In addition to effects on collagen during ECM remodeling, IL-13 signaling can directly induce HA accumulation in the lungs, a key mechanism recently identified as a driver of severe COVID-19 pathology (*96*).

Although the finding that IL-13 regulates HA is new (*96*), it is well documented that IL-13 regulates mucus, another sugar-rich "gel-like" secreted ECM structure. IL-13–driven goblet cell hyperplasia and enhanced mucus production are known to be essential steps in protection against gastrointestinal nematodes (*97*). However, less well known is that IL-13 alters not only the quantity but also the composition of the mucus, with consequences for disease outcome. For example, IL-13 specifically induces the mucin Muc5AC, the overproduction of which is an essential feature of allergic airway inflammation (*98*) and which is required for nematode expulsion (*99*). IL-13 also increases mucin sulfation, which makes it more difficult for the helminth *Trichuris muris* to establish a successful infection (*100*). Perhaps the most profound impact of immune alterations on the mucus barrier is the key role that sugars play in regulating the microbiome. Host GAGs and mucins are food sources for commensal bacteria, and the ability of bacterial species to use these sugars will regulate the gut microbial composition (*101*). For example, *Bacteroides theatiotaomicron* produces sulfatases that enable it to use specific mucins (*102*) and metabolize GAGs (*103*). In an elegant example of this interaction, the opportunistic bacterium *Pseudomonas aeruginosa* induces a type 2 immune response in the lung, stimulating the production of Muc5AC, which the bacterium then uses as an energy source to enhance colonization. This type 2 enhancement loop additionally results in allergic sensitization (*104*).

Cytokines such as TGF-β and IL-13 regulate the ECM by driving ECM production or enzymatically regulating its composition, whereas direct biophysical interaction with chemokines or cytokines can alter ECM structure and/or function. Some proteins are thought to bind to GAGs through less-specific electrostatic-driven interactions, whereas others bind more specifically on the basis of shape complementarity (*5*). The nature of these protein-GAG interactions may dictate functional biological outcomes. For example, chemokines display highly variable affinities for GAGs (*105*, *106*), suggesting distinct biological functions of these interactions that go beyond using GAGs as a platform for a chemotactic gradient (*48*). Although it makes sense that weaker (i.e., less specific) interactions will facilitate localization, particularly in the context of the bloodstream, certain chemokines (e.g., CXCL4, also known as PF4) bind with incredibly high affinity. Recent work has suggested that some chemokines, such as CXCL4, primarily function by binding to GAGs rather than by binding directly to chemokine receptors (*107*). This can be mediated by signaling through cell surface proteoglycans and/or remodeling the ECM glycocalyx to facilitate leukocyte-endothelial interactions (*108*). For instance, high-affinity binding by CXCL4 leads to the cross-linking of HS chains, resulting in shrinking of the thickness of the glycocalyx (Fig. 2). This reduced depth may allow better cellular access to the endothelial surface. In addition to chemokines, several cytokines and growth factors have been shown to bind to and remodel GAGs, such as interferon-γ, the function of which is regulated by this interaction (*109*). We still have much to learn about the mechanistic role of cytokine-GAG interactions, particularly to better understand how GAGs control cytokine availability and/or present cytokines to their receptors.

ECM remodeling by immune cells

Immune cells typically migrate using non-proteolytic mechanisms, which ensures that matrix barriers are not catastrophically compromised. However, to transverse the rigid basement membranes, activated macrophages use membrane type I matrix metalloproteinase (MT1-MMP) as well as cytoskeletally generated physical forces (*110*). Neutrophils can be found coated in laminin after passing through holes in the basement membrane, suggesting a level of localized remodeling (*111*). Additionally, without the expression of elastase, neutrophils cannot breach the vascular basement membrane and migrate into the interstitial matrix (*112*), which may also affect the subsequent migration of other cells such as eosinophils (*113*). Indeed, although matrix-degrading enzymes involved in cell trafficking can be produced by resident stromal fibroblasts (*62*), they are often made by the immune cells themselves. During influenza infection, neutrophil-derived MMP9 is required for neutrophils to access the infection site and to control viral replication (*114*), whereas ECM degradation by MT1-MMP, which is highly expressed on infiltrating myeloid cells, compromises tissue integrity and promotes secondary bacterial infection (*115*). These studies highlight the possibility that "pioneer" cells recruited early in recruitment can modify the ECM to facilitate and control subsequent waves of immune cells, which is consistent with the known role of neutrophils in directing T cell migration (*116*). However, myeloid cells are not alone in their capacity to remodel the ECM, because T cells with intrinsic loss of granzyme B (which degrades several ECM components) have reduced extravasation in vivo (*117*). These studies highlight inflammatory events involving local remodeling of the basement membrane that allows other cells to gain access into the tissue. However, the capacity of neutrophils to remodel the ECM goes beyond enzyme release. In the early stages of tissue repair, neutrophils use cell surface integrins to pull and carry preexisting matrix from nearby tissue, where it is incorporated into the wound-bed matrix to reestablish a new ECM scaffold (*118*).

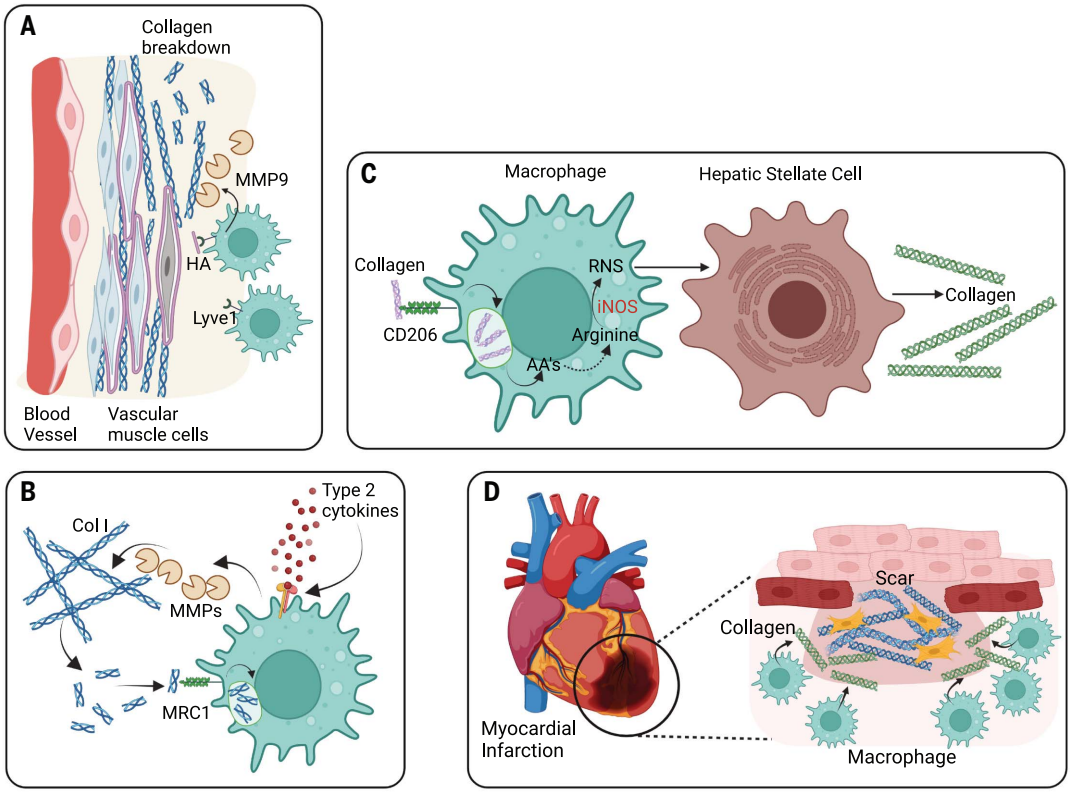


Fig. 4. Macrophages regulate collagen production and breakdown. (**A**) Macrophages located around the outer region of the blood vessel wall express Lyve-1, a receptor that recognizes and binds HA, a component of the ECM that can be found on the pericellular coat of cells such as vascular smooth muscle cells. This Lyve-1–HA interaction initiates MMP9-mediated proteolysis of collagen fibers, thereby regulating the composition of the arterial ECM and thus maintaining the tone and diameter of the artery (*41*). (**B**) Collagen degradation by MMP9 also occurs after macrophage activation by type 2 cytokines. Fragments of collagen bind through mannose receptor 1 (MRC1) and are taken up intracellularly into lysosomes within the macrophage, where they are further processed for degradation (*87*). (**C**) Collagen degradation mediated through MRC1 signaling can lead to an increase in free amino acids (AAs) intracellularly, which drives arginine biosynthesis and the subsequent generation of reactive nitrogen species (RNS) through inducible nitric oxide synthase (iNOS) activity. Release of RNS has been shown to activate profibrotic pathways in pancreatic stellate cells, resulting in collagen synthesis (*119*). (**D**) Macrophages themselves have also been shown to directly produce different collagen molecules. During myocardial infarction, macrophages at the site of injury contribute to the deposition of collagen, specifically at the periphery of the scar (*120*).

Macrophage remodeling of the ECM

Beyond the secretion of collagen-degrading enzymes, macrophages play a pivotal role in mature ECM turnover through receptor-mediated uptake of collagen and routing to lysosomes for degradation (Fig. 4). Enhanced collagen turnover occurs through the up-regulation of members of the mannose receptor family, which target specific collagen types for degradation in a process enhanced by type 2 cytokines (*87*). However, in some contexts, collagen scavenging by macrophages can enhance rather than limit collagen deposition. The superabundance of intracellular arginine generated by collagen breakdown promotes a metabolic switch toward reactive nitrogen species, which triggers further collagen deposition (Fig. 4) (*119*).

Critically, macrophages are not just mediators of matrix removal, they also directly produce collagen themselves. Myofibroblasts are typically the major producers of fibrillar collagen during scar formation and ECM remodeling, and the functional purpose of macrophage collagen production remains unclear. In models of heart injury, macrophages produce collagen, which contributes to scar formation (*120*). This has been observed in both mice and zebrafish, which suggests that collagen deposition is an evolutionarily conserved property of macrophages. This is further supported by evidence of collagen production by macrophages in *Drosophila* (*121*). Macrophages activated by numerous signals, including IL-10, IL-13, and TGF-β, secrete collagen VI (*86*), a molecule that binds different ECM components and plays a role in organizing the three-dimensional tissue architecture (*122*). Although the specific role for macrophages versus fibroblasts is not yet apparent, macrophages produce several ECM components (e.g., fibronectin, laminin, osteopontin, and versican) and may provide an initial scaffold after injury, before the fibroblasts take over. In a pancreatic cancer model, macrophages produced a subset of collagen isoforms distinct from fibroblasts, providing evidence that macrophages fine-tune the fibrotic response (*123*). However, in a colorectal cancer model, collagen-producing macrophages actually outnumbered the fibroblasts and were directly responsible for the remodeling of the tumor microenvironment (*124*).

The past decade has seen a deepening expansion of knowledge demonstrating critical tissue-specific functions of resident macrophages that act to maintain tissue integrity and identity (*125*). It seems very likely that this will turn out to include the ability to define the unique composition of the different ECM niches. Indeed, the intimate mechanistic relationship between particular macrophage subsets and fine-tuning of the collagen matrix is now being revealed. For example, Lyve-1$^+$ macrophages, which localize to the collagen-rich adventitial layer of large blood vessels, are able to degrade collagen 1 produced by vascular smooth muscle cells primarily through MMP-9 release (*41*) (Fig. 4A). This control of collagen production requires Lyve-1 binding to HA on the vessel wall and is needed to prevent the ECM remodeling and arterial stiffness associated with cardiovascular disease. The regulation of ECM turnover by Lyve-1$^+$ macrophages is likely

to apply more broadly, as suggested by the ability of Lyve-1$^+$ macrophages to regulate both collagen and HA levels in the mammary gland (*126*).

Summary

Although this review is divided into matrix impact on immune function and vice versa, the reality is naturally more complex, and many of the examples that we describe reflect an active cross-talk between matrix components and the immune system. For example, implanted ECM scaffolds made from decellurized porcine tissue were shown to induce a type 2 immune response (*127*, *128*), illustrating the capacity of the ECM alone to regulate an immune response. In turn, the type 2 response that it elicits promotes an eosinophil and macrophage program that helps to drive ECM deposition (*84*). The development of targeted therapies for any immune-mediated condition must therefore consider both the immune cell and the physical environment that will surround it. Advances in techniques to spatially map where immune cells preferentially reside within the tissue are transforming our understanding of cell-cell interactions (*129*). However, to truly understand how immune cells behave and function in a healthy versus diseased or aging tissue, these methodologies need to incorporate the complex and changing ECM structures that define the tissue niche, which will include tackling the glycome (*130*). Ultimately, to answer the most pressing questions in tissue health, it will be critical that immunologists and matrix biologists work together.

REFERENCES AND NOTES

1. N. C. Henderson, F. Rieder, T. A. Wynn, Fibrosis: From mechanisms to medicines. *Nature* **587**, 555–566 (2020). doi: 10.1038/s41586-020-2938-9; pmid: 33239795
2. M. W. Parker *et al*., Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J. Clin. Invest.* **124**, 1622–1635 (2014). doi: 10.1172/JCI71386; pmid: 24590289
3. K. E. Kadler, C. Baldock, J. Bella, R. P. Boot-Handford, Collagens at a glance. *J. Cell Sci.* **120**, 1955–1958 (2007). doi: 10.1242/jcs.03453; pmid: 17550969
4. D. Soares da Costa, R. L. Reis, I. Pashkuleva, Sulfation of glycosaminoglycans and its implications in human health and disorders. *Annu. Rev. Biomed. Eng.* **19**, 1–26 (2017). doi: 10.1146/annurev-bioeng-071516-044610; pmid: 28226217
5. D. Xu, J. D. Esko, Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* **83**, 129–157 (2014). doi: 10.1146/annurev-biochem-060713-035314; pmid: 24606135
6. R. P. Richter, N. S. Baranova, A. J. Day, J. C. Kwok, Glycosaminoglycans in extracellular matrix organisation: Are concepts from soft matter physics key to understanding the formation of perineuronal nets? *Curr. Opin. Struct. Biol.* **50**, 65–74 (2018). doi: 10.1016/j.sbi.2017.12.002; pmid: 29275227
7. M. I. Tammi, A. J. Day, E. A. Turley, Hyaluronan and homeostasis: A balancing act. *J. Biol. Chem.* **277**, 4581–4584 (2002). doi: 10.1074/jbc.R100037200; pmid: 11717316
8. A. J. Day, C. A. de la Motte, Hyaluronan cross-linking: A protective mechanism in inflammation? *Trends Immunol.* **26**, 637–643 (2005). doi: 10.1016/j.it.2005.09.009; pmid: 16214414
9. C. E. Wagner, K. M. Wheeler, K. Ribbeck, Mucins and their role in shaping the functions of mucus barriers. *Annu. Rev. Cell Dev. Biol.* **34**, 189–215 (2018). doi: 10.1146/annurev-cellbio-100617-062818; pmid: 30296390
10. N. K. Karamanos *et al*., A guide to the composition and functions of the extracellular matrix. *FEBS J.* **288**, 6850–6912 (2021). doi: 10.1111/febs.15776; pmid: 33605520
11. G. Liu *et al*., Therapeutic targets in lung tissue remodelling and fibrosis. *Pharmacol. Ther.* **225**, 107839 (2021). doi: 10.1016/j.pharmthera.2021.107839; pmid: 33774068
12. L. Möckl, The emerging role of the mammalian glycocalyx in functional membrane organization and immune system regulation. *Front. Cell Dev. Biol.* **8**, 253 (2020). doi: 10.3389/fcell.2020.00253; pmid: 32351961
13. L. Sorokin, The impact of the extracellular matrix on inflammation. *Nat. Rev. Immunol.* **10**, 712–723 (2010). doi: 10.1038/nri2852; pmid: 20865019
14. R. Jayadev *et al*., A basement membrane discovery pipeline uncovers network complexity, regulators, and human disease associations. *Sci. Adv.* **8**, eabn2265 (2022). doi: 10.1126/sciadv.abn2265; pmid: 35584218
15. H. H. Lipowsky, Role of the glycocalyx as a barrier to leukocyte-endothelium adhesion. *Adv. Exp. Med. Biol.* **1097**, 51–68 (2018). doi: 10.1007/978-3-319-96445-4_3; pmid: 30315539
16. A. Marki, J. D. Esko, A. R. Pries, K. Ley, Role of the endothelial surface layer in neutrophil recruitment. *J. Leukoc. Biol.* **98**, 503–515 (2015). doi: 10.1189/jlb.3MR0115-011R; pmid: 25979432
17. A. A. Constantinescu, H. Vink, J. A. E. Spaan, Endothelial cell glycocalyx modulates immobilization of leukocytes at the endothelial surface. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1541–1547 (2003). doi: 10.1161/01.ATV.0000085630.24353.3D; pmid: 12855481
18. P. Sundd, M. K. Pospieszalska, L. S.-L. Cheung, K. Konstantopoulos, K. Ley, Biomechanics of leukocyte rolling. *Biorheology* **48**, 1–35 (2011). doi: 10.3233/BIR-2011-0579; pmid: 21515934
19. R. Uchimido, E. P. Schmidt, N. I. Shapiro, The glycocalyx: A novel diagnostic and therapeutic target in sepsis. *Crit. Care* **23**, 16 (2019). doi: 10.1186/s13054-018-2292-6; pmid: 30654825
20. E. P. Schmidt *et al*., The pulmonary endothelial glycocalyx regulates neutrophil adhesion and lung injury during experimental sepsis. *Nat. Med.* **18**, 1217–1223 (2012). doi: 10.1038/nm.2843; pmid: 22820644
21. H. Kataoka *et al*., Structural behavior of the endothelial glycocalyx is associated with pathophysiologic status in septic mice: An integrated approach to analyzing the behavior and function of the glycocalyx using both electron and fluorescence intravital microscopy. *Anesth. Analg.* **125**, 874–883 (2017). doi: 10.1213/ANE.0000000000002057; pmid: 28504989
22. M. Yamaoka-Tojo, Vascular endothelial glycocalyx damage in COVID-19. *Int. J. Mol. Sci.* **21**, 9712 (2020). doi: 10.3390/ijms21249712; pmid: 33352699
23. S. Banerjee, J. G. Mwangi, T. K. Stanley, R. Mitra, E. E. Ebong, Regeneration and assessment of the endothelial glycocalyx to address cardiovascular disease. *Ind. Eng. Chem. Res.* **60**, 17328–17347 (2021). doi: 10.1021/acs.iecr.1c03074
24. S. Dogné, B. Flamion, Endothelial glycocalyx impairment in disease: Focus on hyaluronan shedding. *Am. J. Pathol.* **190**, 768–780 (2020). doi: 10.1016/j.ajpath.2019.11.016; pmid: 32035885
25. K. M. Yamada, M. Sixt, Mechanisms of 3D cell migration. *Nat. Rev. Mol. Cell Biol.* **20**, 738–752 (2019). doi: 10.1038/s41580-019-0172-9; pmid: 31582855
26. J. Renkawitz *et al*., Nuclear positioning facilitates amoeboid migration along the path of least resistance. *Nature* **568**, 546–550 (2019). doi: 10.1038/s41586-019-1087-5; pmid: 30944468
27. M.-B. Voisin, D. Pröbstl, S. Nourshargh, Venular basement membranes ubiquitously express matrix protein low-expression regions: Characterization in multiple tissues and remodeling during inflammation. *Am. J. Pathol.* **176**, 482–495 (2010). doi: 10.2353/ajpath.2010.090510; pmid: 20008148
28. H. Salmon *et al*., Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J. Clin. Invest.* **122**, 899–910 (2012). doi: 10.1172/JCI45817; pmid: 22293174
29. A. Nicolas-Boluda *et al*., Tumor stiffening reversion through collagen crosslinking inhibition improves T cell migration and anti-PD-1 treatment. *eLife* **10**, e58688 (2021). doi: 10.7554/eLife.58688; pmid: 34106045
30. F. Puttur *et al*., Pulmonary environmental cues drive group 2 innate lymphoid cell dynamics in mice and humans. *Sci. Immunol.* **4**, eaav7638 (2019). doi: 10.1126/sciimmunol.aav7638; pmid: 31175176
31. N. R. J. Fernandes *et al*., CD4+ T cell interstitial migration controlled by fibronectin in the inflamed skin. *Front. Immunol.* **11**, 1501 (2020). doi: 10.3389/fimmu.2020.01501; pmid: 32793204
32. V. Vogel, Unraveling the mechanobiology of extracellular matrix. *Annu. Rev. Physiol.* **80**, 353–387 (2018). doi: 10.1146/annurev-physiol-021317-121312; pmid: 29433414
33. S. Keerthivasan *et al*., Homeostatic functions of monocytes and interstitial lung macrophages are regulated via collagen domain-binding receptor LAIR1. *Immunity* **54**, 1511–1526.e8 (2021). doi: 10.1016/j.immuni.2021.06.012; pmid: 34260887
34. M. Culty, T. E. O'Mara, C. B. Underhill, H. Yeager Jr., R. P. Swartz, Hyaluronan receptor (CD44) expression and function in human peripheral blood monocytes and alveolar macrophages. *J. Leukoc. Biol.* **56**, 605–611 (1994). doi: 10.1002/jlb.56.5.605; pmid: 7525818
35. J. R. Fraser, T. C. Laurent, U. B. Laurent, Hyaluronan: Its nature, distribution, functions and turnover. *J. Intern. Med.* **242**, 27–33 (1997). doi: 10.1046/j.1365-2796.1997.00170.x; pmid: 9260563
36. S. E. Armstrong, D. R. Bell, Relationship between lymph and tissue hyaluronan in skin and skeletal muscle. *Am. J. Physiol. Heart Circ. Physiol.* **283**, H2485–H2494 (2002). doi: 10.1152/ajpheart.00385.2002; pmid: 12388305
37. M. Erickson, R. Stern, Chain gangs: New aspects of hyaluronan metabolism. *Biochem. Res. Int.* **2012**, 893947 (2012). doi: 10.1155/2012/893947; pmid: 22216413
38. B. McDonald *et al*., Interaction of CD44 and hyaluronan is the dominant mechanism for neutrophil sequestration in inflamed liver sinusoids. *J. Exp. Med.* **205**, 915–927 (2008). doi: 10.1084/jem.20071765; pmid: 18362172
39. H. C. DeGrendele, M. Kosfiszer, P. Estess, M. H. Siegelman, CD44 activation and associated primary adhesion is inducible via T cell receptor stimulation. *J. Immunol.* **159**, 2549–2553 (1997). doi: 10.4049/jimmunol.159.6.2549; pmid: 9300670
40. C. S. Bonder, S. R. Clark, M. U. Norman, P. Johnson, P. Kubes, Use of CD44 by CD4+ Th1 and Th2 lymphocytes to roll and adhere. *Blood* **107**, 4798–4806 (2006). doi: 10.1182/blood-2005-09-3581; pmid: 16497973
41. H. Y. Lim *et al*., Hyaluronan receptor LYVE-1-expressing macrophages maintain arterial tone through hyaluronan-mediated regulation of smooth muscle cell collagen. *Immunity* **49**, 326–341.e7 (2018). doi: 10.1016/j.immuni.2018.06.008; pmid: 30054204
42. L. A. Johnson *et al*., Dendritic cells enter lymph vessels by hyaluronan-mediated docking to the endothelial receptor LYVE-1. *Nat. Immunol.* **18**, 762–770 (2017). doi: 10.1038/ni.3750; pmid: 28504698
43. W. Lawrance, S. Banerji, A. J. Day, S. Bhattacharjee, D. G. Jackson, Binding of hyaluronan to the native lymphatic vessel endothelial receptor LYVE-1 is critically dependent on receptor clustering and hyaluronan organization. *J. Biol. Chem.* **291**, 8014–8030 (2016). doi: 10.1074/jbc.M115.708305; pmid: 26823460
44. T. N. Wight, Provisional matrix: A role for versican and hyaluronan. *Matrix Biol.* **60-61**, 38–56 (2017). doi: 10.1016/j.matbio.2016.12.001; pmid: 27932299
45. J. Lesley *et al*., TSG-6 modulates the interaction between hyaluronan and cell surface CD44. *J. Biol. Chem.* **279**, 25745–25754 (2004). doi: 10.1074/jbc.M313319200; pmid: 15060082
46. T. N. Wight, I. Kang, M. J. Merrilees, Versican and the control of inflammation. *Matrix Biol.* **35**, 152–161 (2014). doi: 10.1016/j.matbio.2014.01.015; pmid: 24513039
47. L. A. Johnson, S. Banerji, B. C. Lagerholm, D. G. Jackson, Dendritic cell entry to lymphatic capillaries is orchestrated by CD44 and the hyaluronan glycocalyx. *Life Sci. Alliance* **4**, e202000908 (2021). doi: 10.26508/lsa.202000908; pmid: 33687996
48. L. Wang, M. Fuster, P. Sriramarao, J. D. Esko, Endothelial heparan sulfate deficiency impairs L-selectin- and chemokine-mediated neutrophil trafficking during inflammatory responses. *Nat. Immunol.* **6**, 902–910 (2005). doi: 10.1038/ni1233; pmid: 16056228
49. L. Giuffrè *et al*., Monocyte adhesion to activated aortic endothelium: Role of L-selectin and heparan sulfate proteoglycans. *J. Cell Biol.* **136**, 945–956 (1997). doi: 10.1083/jcb.136.4.945; pmid: 9049258
50. L. Stoler-Barak *et al*., Blood vessels pattern heparan sulfate gradients between their apical and basolateral aspects. *PLOS ONE* **9**, e85699 (2014). doi: 10.1371/journal.pone.0085699; pmid: 24465652

51. M. Weber *et al*., Interstitial dendritic cell guidance by haptotactic chemokine gradients. *Science* **339**, 328–332 (2013). doi: 10.1126/science.1228456; pmid: 23329049
52. A. J. L. Ridley *et al*., Chemokines form complex signals during inflammation and disease that can be decoded by extracellular matrix proteoglycans. bioRxiv 508420 [Preprint] (2022); https://doi.org/10.1101/2022.09.20.508420.
53. H. B. Schiller *et al*., Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol. Syst. Biol.* **11**, 819 (2015). doi: 10.15252/msb.20156123; pmid: 26174933
54. A. C. McQuattie-Pimentel *et al*., The lung microenvironment shapes a dysfunctional response of alveolar macrophages in aging. *J. Clin. Invest.* **131**, e140299 (2021). doi: 10.1172/JCI140299; pmid: 33586677
55. T. J. Bell *et al*., Defective lung function following influenza virus is due to prolonged, reversible hyaluronan synthesis. *Matrix Biol.* **80**, 14–28 (2019). doi: 10.1016/j.matbio.2018.06.006; pmid: 29933044
56. A. J. Day, C. M. Milner, TSG-6: A multifunctional protein with anti-inflammatory and tissue-protective properties. *Matrix Biol.* **78-79**, 60–83 (2019). doi: 10.1016/j.matbio.2018.01.011; pmid: 29362135
57. N. S. Baranova *et al*., Incorporation of pentraxin 3 into hyaluronan matrices is tightly regulated and promotes matrix cross-linking. *J. Biol. Chem.* **289**, 30481–30498 (2014). doi: 10.1074/jbc.M114.568154; pmid: 25190808
58. D. C. Briggs *et al*., Inter-α-inhibitor heavy chain-1 has an integrin-like 3D structure mediating immune regulatory activities and matrix stabilization during ovulation. *J. Biol. Chem.* **295**, 5278–5291 (2020). doi: 10.1074/jbc.RA119.011916; pmid: 32144206
59. S. R. Reeves *et al*., Asthmatic bronchial epithelial cells promote the establishment of a Hyaluronan-enriched, leukocyte-adhesive extracellular matrix by lung fibroblasts. *Respir. Res.* **19**, 146 (2018). doi: 10.1186/s12931-018-0849-1; pmid: 30071849
60. L. Zhuo *et al*., SHAP potentiates the CD44-mediated leukocyte adhesion to the hyaluronan substratum. *J. Biol. Chem.* **281**, 20303–20314 (2006). doi: 10.1074/jbc.M506703200; pmid: 16702221
61. B. Matuska *et al*., Pathological hyaluronan matrices in cystic fibrosis airways and secretions. *Am. J. Respir. Cell Mol. Biol.* **55**, 576–585 (2016). doi: 10.1165/rcmb.2015-0358OC; pmid: 27243106
62. D. F. Boyd *et al*., Exuberant fibroblast activity compromises lung function via ADAMTS4. *Nature* **587**, 466–471 (2020). doi: 10.1038/s41586-020-2877-5; pmid: 33116313
63. M. McMahon *et al*., ADAMTS5 is a critical regulator of virus-specific T cell immunity. *PLOS Biol.* **14**, e1002580 (2016). doi: 10.1371/journal.pbio.1002580; pmid: 27855162
64. C. Yang *et al*., The high and low molecular weight forms of hyaluronan have distinct effects on CD44 clustering. *J. Biol. Chem.* **287**, 43094–43107 (2012). doi: 10.1074/jbc.M112.349209; pmid: 23118219
65. P. Teder *et al*., Resolution of lung inflammation by CD44. *Science* **296**, 155–158 (2002). doi: 10.1126/science.1069659; pmid: 11935029
66. P. L. Bollyky *et al*., ECM components guide IL-10 producing regulatory T-cell (TR1) induction from effector memory T-cell precursors. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7938–7943 (2011). doi: 10.1073/pnas.1017360108; pmid: 21518860
67. P. L. Bollyky *et al*., Intact extracellular matrix and the maintenance of immune tolerance: High molecular weight hyaluronan promotes persistence of induced CD4+CD25+ regulatory T cells. *J. Leukoc. Biol.* **86**, 567–572 (2009). doi: 10.1189/jlb.0109001; pmid: 19401397
68. L. Schaefer, Extracellular matrix molecules: Endogenous danger signals as new drug targets in kidney diseases. *Curr. Opin. Pharmacol.* **10**, 185–190 (2010). doi: 10.1016/j.coph.2009.11.007; pmid: 20045380
69. C. W. Frevert, J. Felgenhauer, M. Wygrecka, M. V. Nastase, L. Schaefer, Danger-associated molecular patterns derived from the extracellular matrix provide temporal control of innate immunity. *J. Histochem. Cytochem.* **66**, 213–227 (2018). doi: 10.1369/0022155417740880; pmid: 29290139
70. A. M. Piccinini, K. S. Midwood, DAMPening inflammation by modulating TLR signalling. *Mediators Inflamm.* **2010**, 1–21 (2010). doi: 10.1155/2010/672395; pmid: 20706656
71. M. Blich *et al*., Macrophage activation by heparanase is mediated by TLR-2 and TLR-4 and associates with plaque progression. *Arterioscler. Thromb. Vasc. Biol.* **33**, e56–e65 (2013). doi: 10.1161/ATVBAHA.112.254961; pmid: 23162016
72. A. J. Mayfosh, N. Baschuk, M. D. Hulett, Leukocyte heparanase: A double-edged sword in tumor progression. *Front. Oncol.* **9**, 331 (2019). doi: 10.3389/fonc.2019.00331; pmid: 31110966
73. C. E. McQuitty, R. Williams, S. Chokshi, L. Urbani, Immunomodulatory role of the extracellular matrix within the liver disease microenvironment. *Front. Immunol.* **11**, 574276 (2020). doi: 10.3389/fimmu.2020.574276; pmid: 33262757
74. B. Hinz, The extracellular matrix and transforming growth factor-β1: Tale of a strained relationship. *Matrix Biol.* **47**, 54–65 (2015). doi: 10.1016/j.matbio.2015.05.006; pmid: 25960420
75. M. C. Lampi, C. A. Reinhart-King, Targeting extracellular matrix stiffness to attenuate disease: From molecular mechanisms to clinical trials. *Sci. Transl. Med.* **10**, eaao0475 (2018). doi: 10.1126/scitranslmed.aao0475; pmid: 29298864
76. C. Zhu, W. Chen, J. Lou, W. Rittase, K. Li, Mechanosensing through immunoreceptors. *Nat. Immunol.* **20**, 1269–1278 (2019). doi: 10.1038/s41590-019-0491-1; pmid: 31534240
77. V. S. Meli *et al*., Biophysical regulation of macrophages in health and disease. *J. Leukoc. Biol.* **106**, 283–299 (2019). doi: 10.1002/JLB.MR0318-126R; pmid: 30861205
78. J.-C. Escolano *et al*., Compliant substrates enhance macrophage cytokine release and NLRP3 inflammasome formation during their pro-inflammatory response. *Front. Cell Dev. Biol.* **9**, 639815 (2021). doi: 10.3389/fcell.2021.639815; pmid: 33855019
79. J. Herrera, C. A. Henke, P. B. Bitterman, Extracellular matrix as a driver of progressive fibrosis. *J. Clin. Invest.* **128**, 45–53 (2018). doi: 10.1172/JCI93557; pmid: 29293088
80. H. L. Sladitschek-Martens *et al*., YAP/TAZ activity in stromal cells prevents ageing by controlling cGAS-STING. *Nature* **607**, 790–798 (2022). doi: 10.1038/s41586-022-04924-6; pmid: 35768505
81. S. Ma *et al*., A role of PIEZO1 in iron metabolism in mice and humans. *Cell* **184**, 969–982.e13 (2021). doi: 10.1016/j.cell.2021.01.024; pmid: 33571427
82. A. G. Solis *et al*., Mechanosensation of cyclical force by PIEZO1 is essential for innate immunity. *Nature* **573**, 69–74 (2019). doi: 10.1038/s41586-019-1485-8; pmid: 31435009
83. B. Aykut *et al*., Targeting Piezo1 unleashes innate immunity against cancer and infectious disease. *Sci. Immunol.* **5**, eabb5168 (2020). doi: 10.1126/sciimmunol.abb5168; pmid: 32826342
84. R. L. Gieseck 3rd, M. S. Wilson, T. A. Wynn, Type 2 immunity in tissue repair and fibrosis. *Nat. Rev. Immunol.* **18**, 62–76 (2018). doi: 10.1038/nri.2017.90; pmid: 28853443
85. T. A. Wynn, K. M. Vannella, Macrophages in tissue repair, regeneration, and fibrosis. *Immunity* **44**, 450–462 (2016). doi: 10.1016/j.immuni.2016.02.015; pmid: 26982353
86. M. Schnoor *et al*., Production of type VI collagen by human macrophages: A new dimension in macrophage functional heterogeneity. *J. Immunol.* **180**, 5707–5719 (2008). doi: 10.4049/jimmunol.180.8.5707; pmid: 18390756
87. D. H. Madsen *et al*., M2-like macrophages are responsible for collagen degradation through a mannose receptor-mediated pathway. *J. Cell Biol.* **202**, 951–966 (2013). doi: 10.1083/jcb.201301081; pmid: 24019537
88. S. O'Reilly, M. Ciechomska, N. Fullard, S. Przyborski, J. M. van Laar, IL-13 mediates collagen deposition via STAT6 and microRNA-135b: A role for epigenetics. *Sci. Rep.* **6**, 25066 (2016). doi: 10.1038/srep25066; pmid: 27113293
89. M. Jinnin, H. Ihn, K. Yamane, K. Tamaki, Interleukin-13 stimulates the transcription of the human alpha2(I) collagen gene in human dermal fibroblasts. *J. Biol. Chem.* **279**, 41783–41791 (2004). doi: 10.1074/jbc.M406951200; pmid: 15271999
90. A. Oriente *et al*., Interleukin-13 modulates collagen homeostasis in human skin and keloid fibroblasts. *J. Pharmacol. Exp. Ther.* **292**, 988–994 (2000). pmid: 10688614
91. J. E. Kolodsick *et al*., Protection from fluorescein isothiocyanate-induced fibrosis in IL-13-deficient, but not IL-4-deficient, mice results from impaired collagen synthesis by fibroblasts. *J. Immunol.* **172**, 4068–4076 (2004). doi: 10.4049/jimmunol.172.7.4068; pmid: 15034018
92. A. E. Postlethwaite, M. A. Holness, H. Katai, R. Raghow, Human fibroblasts synthesize elevated levels of extracellular matrix proteins in response to interleukin 4. *J. Clin. Invest.* **90**, 1479–1485 (1992). doi: 10.1172/JCI116015; pmid: 1401080
93. R. Sugimoto *et al*., Effect of IL-4 and IL-13 on collagen production in cultured LI90 human hepatic stellate cells. *Liver Int.* **25**, 420–428 (2005). doi: 10.1111/j.1478-3231.2005.01087.x; pmid: 15780068
94. R. L. Gieseck 3rd *et al*., Interleukin-13 activates distinct cellular pathways leading to ductular reaction, steatosis, and fibrosis. *Immunity* **45**, 145–158 (2016). doi: 10.1016/j.immuni.2016.06.009; pmid: 27421703
95. J. A. Knipper *et al*., Interleukin-4 receptor α signaling in myeloid cells controls collagen fibril assembly in skin repair. *Immunity* **43**, 803–816 (2015). doi: 10.1016/j.immuni.2015.09.005; pmid: 26474656
96. A. N. Donlan *et al*., IL-13 is a driver of COVID-19 severity. *JCI Insight* **6**, e15010 (2021). doi: 10.1172/jci.insight.150107; pmid: 34185704
97. C. Sharpe, D. J. Thornton, R. K. Grencis, A sticky end for gastrointestinal helminths; the role of the mucus barrier. *Parasite Immunol.* **40**, e12517 (2018). doi: 10.1111/pim.12517; pmid: 29355990
98. C. M. Evans *et al*., The polymeric mucin Muc5ac is required for allergic airway hyperreactivity. *Nat. Commun.* **6**, 6281 (2015). doi: 10.1038/ncomms7281; pmid: 25687754
99. S. Z. Hasnain *et al*., Muc5ac: A critical component mediating the rejection of enteric nematodes. *J. Exp. Med.* **208**, 893–900 (2011). doi: 10.1084/jem.20102057; pmid: 21502330
100. S. Z. Hasnain *et al*., Immune-driven alterations in mucin sulphation is an important mediator of Trichuris muris helminth expulsion. *PLOS Pathog.* **13**, e1006218 (2017). doi: 10.1371/journal.ppat.1006218; pmid: 28192541
101. N. M. Koropatkin, E. A. Cameron, E. C. Martens, How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012). doi: 10.1038/nrmicro2746; pmid: 22491358
102. A. S. Luis *et al*., A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **598**, 332–337 (2021). doi: 10.1038/s41586-021-03967-5; pmid: 34616040
103. D. Ndeh *et al*., Metabolism of multiple glycosaminoglycans by Bacteroides thetaiotaomicron is orchestrated by a versatile core genetic locus. *Nat. Commun.* **11**, 646 (2020). doi: 10.1038/s41467-020-14509-4; pmid: 32005816
104. K. Agaronyan *et al*., Tissue remodeling by an opportunistic pathogen triggers allergic inflammation. *Immunity* **55**, 895–911.e10 (2022). doi: 10.1016/j.immuni.2022.04.001; pmid: 35483356
105. D. P. Dyer *et al*., Differential structural remodelling of heparan sulfate by chemokines: The role of chemokine oligomerization. *Open Biol.* **7**, 160286 (2017). doi: 10.1098/rsob.160286; pmid: 28123055
106. D. P. Dyer, C. L. Salanga, B. F. Volkman, T. Kawamura, T. M. Handel, The dependence of chemokine-glycosaminoglycan interactions on chemokine oligomerization. *Glycobiology* **26**, 312–326 (2016). pmid: 26582609
107. A. L. Gray *et al*., Chemokine CXCL4 interactions with extracellular matrix proteoglycans mediate widespread immune cell recruitment independent of chemokine receptors. *Cell Rep.* **42**, 111930 (2023). doi: 10.1098/j.celrep.2022.111930; pmid: 36640356
108. T. M. Handel, D. P. Dyer, Perspectives on the biological role of chemokine:glycosaminoglycan interactions. *J. Histochem. Cytochem.* **69**, 87–91 (2021). doi: 10.1369/0022155420977971; pmid: 33285085
109. E. Migliorini *et al*., Cytokines and growth factors cross-link heparan sulfate. *Open Biol.* **5**, 150046 (2015). doi: 10.1098/rsob.150046; pmid: 26269427
110. J. C. Bahr, X.-Y. Li, T. Y. Feinberg, L. Jiang, S. J. Weiss, Divergent regulation of basement membrane trafficking by human macrophages and cancer cells. *Nat. Commun.* **13**, 6409 (2022). doi: 10.1038/s41467-022-34087-x; pmid: 36302921
111. M.-B. Voisin, A. Woodfin, S. Nourshargh, Monocytes and neutrophils exhibit both distinct and common mechanisms in penetrating the vascular basement membrane in vivo. *Arterioscler. Thromb. Vasc. Biol.* **29**, 1193–1199 (2009). doi: 10.1161/ATVBAHA.109.187450; pmid: 19498176
112. M.-B. Voisin *et al*., Neutrophil elastase plays a non-redundant role in remodeling the venular basement membrane and neutrophil diapedesis post-ischemia/reperfusion injury. *J. Pathol.* **248**, 88–102 (2019). doi: 10.1002/path.5234; pmid: 30632166
113. H. Koga *et al*., Inhibition of neutrophil elastase attenuates airway hyperresponsiveness and inflammation in a mouse model of secondary allergen challenge: Neutrophil elastase

inhibition attenuates allergic airway responses. *Respir. Res.* **14**, 8 (2013). doi: 10.1186/1465-9921-14-8; pmid: 23347423
114. L. M. Bradley, M. F. Douglass, D. Chatterjee, S. Akira, B. J. G. Baaten, Matrix metalloprotease 9 mediates neutrophil migration into the airways in response to influenza virus-induced toll-like receptor signaling. *PLOS Pathog.* **8**, e1002641 (2012). doi: 10.1371/journal.ppat.1002641; pmid: 22496659
115. D. Talmi-Frank *et al*., Extracellular matrix proteolysis by MT1-MMP contributes to influenza-related tissue damage and mortality. *Cell Host Microbe* **20**, 458–470 (2016). doi: 10.1016/j.chom.2016.09.005; pmid: 27736644
116. K. Lim *et al*., Neutrophil trails guide influenza-specific CD8$^+$ T cells in the airways. *Science* **349**, aaa4352 (2015). doi: 10.1126/science.aaa4352; pmid: 26339033
117. M. D. Prakash *et al*., Granzyme B promotes cytotoxic lymphocyte transmigration via basement membrane remodeling. *Immunity* **41**, 960–972 (2014). doi: 10.1016/j.immuni.2014.11.012; pmid: 25526309
118. A. Fischer *et al*., Neutrophils direct preexisting matrix to initiate repair in damaged tissues. *Nat. Immunol.* **23**, 518–531 (2022). doi: 10.1038/s41590-022-01166-6; pmid: 35354953
119. M. M. LaRue *et al*., Metabolic reprogramming of tumor-associated macrophages by collagen turnover promotes fibrosis in pancreatic cancer. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2119168119 (2022). doi: 10.1073/pnas.2119168119; pmid: 35412885
120. F. C. Simões *et al*., Macrophages directly contribute collagen to scar formation during zebrafish heart regeneration and mouse heart repair. *Nat. Commun.* **11**, 600 (2020). doi: 10.1038/s41467-019-14263-2; pmid: 32001677
121. Y. Matsubayashi *et al*., A Moving Source of Matrix Components Is Essential for De Novo Basement Membrane Formation. *Curr. Biol.* **27**, 3526–3534.e4 (2017). doi: 10.1016/j.cub.2017.10.001; pmid: 29129537
122. M. Cescon, F. Gattazzo, P. Chen, P. Bonaldo, Collagen VI at a glance. *J. Cell Sci.* **128**, 3525–3531 (2015). pmid: 26377767
123. Y. Zhu *et al*., Tissue-resident macrophages in pancreatic ductal adenocarcinoma originate from embryonic hematopoiesis and promote tumor progression. *Immunity* **47**, 323–338.e6 (2017). doi: 10.1016/j.immuni.2017.07.014; pmid: 28813661
124. R. Afik *et al*., Tumor macrophages are pivotal constructors of tumor collagenous matrix. *J. Exp. Med.* **213**, 2315–2331 (2016). doi: 10.1084/jem.20151193; pmid: 27697834
125. M. Guilliams, G. R. Thierry, J. Bonnardel, M. Bajenoff, Establishment and maintenance of the macrophage niche. *Immunity* **52**, 434–451 (2020). doi: 10.1016/j.immuni.2020.02.015; pmid: 32187515
126. Y. Wang *et al*., Tissue-resident macrophages promote extracellular matrix homeostasis in the mammary gland stroma of nulliparous mice. *eLife* **9**, e57438 (2020). doi: 10.7554/eLife.57438; pmid: 32479261
127. K. Sadtler *et al*., Developing a pro-regenerative biomaterial scaffold microenvironment requires T helper 2 cells. *Science* **352**, 366–370 (2016). doi: 10.1126/science.aad9272; pmid: 27081073
128. X. Wang *et al*., Type 2 immunity induced by bladder extracellular matrix enhances corneal wound healing. *Sci. Adv.* **7**, eabe2635 (2021). doi: 10.1126/sciadv.abe2635; pmid: 33863719
129. M. Guilliams *et al*., Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022). doi: 10.1016/j.cell.2021.12.018; pmid: 35021063
130. L. A. Dworkin, H. Clausen, H. J. Joshi, Applying transcriptomics to studyglycosylation at the cell type level. *iScience* **25**, 104419 (2022). doi: 10.1016/j.isci.2022.104419; pmid: 35663018

ACKNOWLEDGMENTS

We are very grateful to our colleagues at the Wellcome Centre for Cell-Matrix Research and the Lydia Becker Institute for Immunology and Inflammation for many stimulating discussions. We would especially like to thank A. Day, D. Thornton, R. Lennon, A. MacDonald, and T. Hardingham and the anonymous referees for critical review of the manuscript. Figures have been drawn in BioRender. **Funding**: This work was supported by MRC-UK grant MR/V011235/1 (J.E.A.) and Wellcome Trust grants 106898/A/15/Z (J.E.A.), 218570/Z/19/Z (D.P.D.), and 203128/A/16/Z (T.E.S., D.P.D., and J.E.A.). **Competing interests**: The authors declare no competing interests. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 24 May 2022; accepted 22 December 2022
10.1126/science.abp8964

TECHNICAL COMMENT

PALEONTOLOGY

Comment on "Sexual selection promotes giraffoid head-neck evolution and ecological adaptation"

Sukuan Hou[1,2,3]*†, Qinqin Shi[1,2]†, Michael J. Benton[4], Nikos Solounias[5,6]

Wang *et al*. (Research Articles, 3 June 2022, eabl8316) reported an early Miocene giraffoid that exhibited fierce head-butting behavior and concluded that sexual selection promoted head-neck evolution in giraffoids. However, we argue that this ruminant is not a giraffoid and thus that the hypothesis that sexual selection promoted giraffoid head-neck evolution is not sufficiently supported.

The newly described mammal *Discokeryx xiezhi*, from the early Miocene of northern China, has disk-shaped headgear and complicated head-neck joints, supposedly used in fierce head-butting behavior (*1*). This was seen as an unusual precursor of the "necking" combat of modern giraffes, and the switch in combat mode was suggested as a key reason why giraffes evolved their long necks. Further, it is suggested (*1*) that these early giraffoids avoided competition with coexisting bovids and cervids by taking marginal niches, which promoted intensive sexual competition.

Wang *et al*. (*1*) highlight six key cranial characteristics supporting the giraffoid affinity of *Discokeryx*, all of which we regard as problematic. First, neither the parietal participation in headgear support nor the median position of frontal or parietal headgear are diagnostic features of Giraffoidea, as these features are also present in bovids (*2*). Second, dermal origin of headgear does not support giraffoid affinity as bovid horncores also have a dermal origin (*3*, *4*); moreover, it is impossible to recognize dermal bone based on the developed bone of *Discokeryx* because the microscopic structure of the developed bone is not diagnostic of its embryonic origin (*5*). Third, it is not clear that giraffid bone histology shows lamellar structure or large-scale osteons (*6*); figures in Wang *et al*. (*1*) show only thick lamellar bone (with several layers of lamellae) surrounding vascular canals in the thin sections of the headgear of *Discokeryx*, *Honanotherium*, and the bovid *Turcocerus* sp., which is a common character in a typical developed secondary osteon. Therefore, osteons cannot demonstrate "horn" differentiation between giraffoids and bovids. Fourth, in giraffids there is only one foramen on the parietal bone beneath the ossicone, indicating the existence of the cornual vein that joins the superficial temporal, as in other horned ruminants (*7*), which is mistaken for the foramen far behind an ossicone (*1*). Finally, similar semicircular canals in *Discokeryx*, extant giraffes, and the bovid *Tsaidamotherium* suggest that this feature most likely does not have phylogenetic significance (*1*). In addition, the bilobed lower canine, which is considered one of the most diagnostic features of Giraffidae (*8*), has not yet been found in *Discokeryx* or other Pecora. The headgear of *Discokeryx* was described as covered with keratinous integument—which is not found in giraffids.

Wang *et al*.'s phylogenetic analysis (*1*) reached the result that *Discokeryx* was a sister group of the bovid *Tsaidamotherium*. We checked their data matrix and found that some important characters and taxa were omitted, in addition to some problematic scoring. The upper dentition and some skull characters of

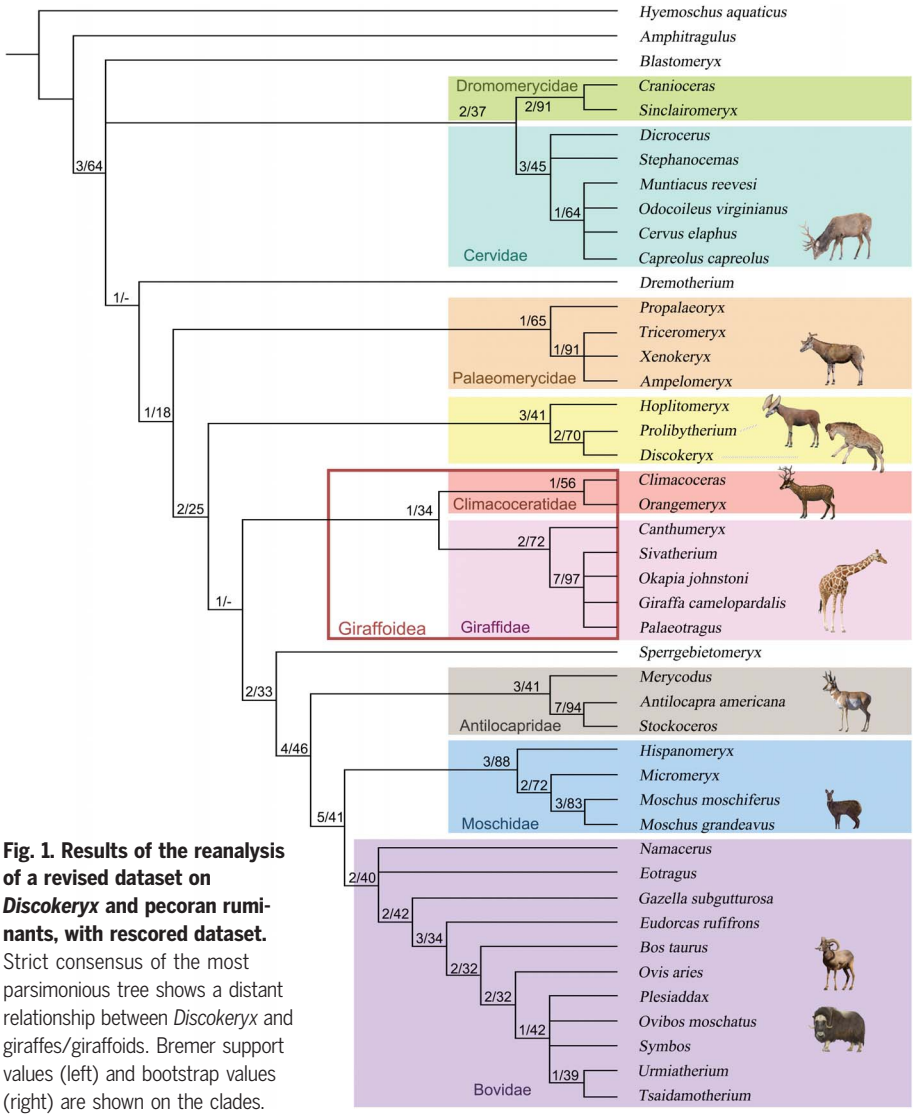


Fig. 1. Results of the reanalysis of a revised dataset on *Discokeryx* and pecoran ruminants, with rescored dataset. Strict consensus of the most parsimonious tree shows a distant relationship between *Discokeryx* and giraffes/giraffoids. Bremer support values (left) and bootstrap values (right) are shown on the clades.

[1]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. [2]CAS Center for Excellence in Life and Paleoenvironment, Beijing 100044, China. [3]University of the Chinese Academy of Sciences, Beijing 100049, China. [4]School of Earth Sciences, University of Bristol, Bristol, UK. [5]Department of Anatomy, New York Institute of Technology College of Osteopathic Medicine, Old Westbury, NY, USA. [6]Department of Paleontology, American Museum of Natural History, New York, NY, USA.
***Corresponding author. Email: housukuan@ivpp.ac.cn**
†These authors contributed equally to this work.

Tsaidamotherium are coded with a question mark even though the data are available (*9*), whereas the astragalus was coded in all fossil taxa, a character that had not been reported in some extinct Pecora including *Discokeryx*. We evaluated the phylogenetic data without adding taxa or characters, clarified some ambiguous statements of the headgear characters, rescored the inaccurate coding, and produced a strict consensus parsimony tree based on morphological data (Fig. 1). Our reanalysis shows that *Discokeryx* is not in the clade of giraffoids but rather is a more primitive ruminant with a stem phylogenetic position, whereas *Tsaidamotherium* is certainly in the clade of bovids. Our results indicate that character coding is very important in phylogenetic analyses and should be considered carefully.

Aside from the systematic position of *Discokeryx*, we have concerns about the argument concerning head-neck evolution in giraffes. Wang *et al.* (*1*) claimed "intensive sexual combats" in the evolution of giraffoids based on more diverse headgear and head-neck morphologies than in other ruminant groups. Further, they argued (*1*) that "the ecological positioning on the marginal niches promoted the intensive sexual competition, and the fierce sexual combats fostered extreme morphologies to occupy the special niches in giraffoids". We query these claims for four reasons. First, as we argue, *Discokeryx* is not a giraffoid. Second, there is no evidence that 'necking' caused neck elongation (*10*). Third, the headgear diversity of giraffoids has been overweighted by artificially using different criteria in different pecoran groups, thus it cannot be used to infer "intensive sexual combats" in giraffoids. Fourth, the reference to *Discokeryx* living in marginal niches in which sexual selection was intense is belied by the fact that it is part of a fossil fauna including many other mammals, and that there were dietary overlaps between *Discokeryx* and other mammals. Similarly, giraffoids were also claimed as having lived in marginal niches (*1*), which were in fact shown to be a dominant part of late Miocene faunas (*11*) and spanned the entire dietary spectrum of browsing, mixed feeding, and grazing (*12*).

REFERENCES AND NOTES

1. S.-Q. Wang *et al.*, *Science* **376**, eabl8316 (2022).
2. B. Bohlin, *Cavicornier der Hipparion-Fauna Nord-China*, Paleont Sin. Ser C, Vol. 9, pp. 1–166 (Geological Survey of China, 1935).
3. E. B. Davis, K. A. Brakora, A. H. Lee, *Proc. Biol. Sci.* **278**, 2857–2865 (2011).
4. A. Nasoori, *Biol. Rev. Camb. Philos. Soc.* **95**, 986–1019 (2020).
5. W. K. Ovalle, P. C. Nahirney, *Netter's Essential Histology: with Correlated Histopathology* (Elsevier Health Sciences, 2020).
6. T. Ganey, J. Ogden, J. Olsen, *Anat. Rec.* **227**, 497–507 (1990).
7. C. A. Spinage, *Afr. J. Ecol.* **6**, 53–61 (1968).
8. W. R. Hamilton, *Philos. Trans. R. Soc. B.* **283**, 165–229 (1978).
9. Q. Shi, *Sci. China Earth Sci.* **57**, 258–266 (2014).
10. G. Mitchell, D. Roberts, S. V. Sittert, J. D. Skinner, *J. Zool.* **290**, 49–57 (2013).
11. N. Solounias, "Family Giraffidae" in *The Evolution of Artiodactyls*, D. R. Prothero, S. E. Foss, Eds. (Johns Hopkins Univ. Press, 2007), chap. 21, pp. 257–277 (2007).
12. N. Solounias, F. Rivals, G. M. Semprebon, *Paleobiology* **36**, 113–136 (2010).

ACKNOWLEDGMENTS

Funding: This work was supported by the National Natural Science Foundation of China (41872005). **Competing interests:** The authors declare that they have no competing interests. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 20 July 2022; accepted 13 January 2023
10.1126/science.add9559

TECHNICAL RESPONSE

PALEONTOLOGY

Response to comment on "Sexual selection promotes giraffoid head-neck evolution and ecological adaptation"

Shi-Qi Wang[1,2]*, Jin Meng[3]*, Bastien Mennecart[4,5], Loïc Costeur[4], Jie Ye[1,2], Chunxiao Li[1,2,6], Chi Zhang[1,2], Ji Zhang[7,8], Manuela Aiglstorfer[9], Yang Wang[10,11], Yan Wu[1,2], Wen-Yu Wu[1,2], Tao Deng[1,2,6]*

Hou *et al*. challenged the giraffoid affinity of *Discokeryx* and its ecology and behavior. In our response we reiterate that *Discokeryx* is a giraffoid that, along with *Giraffa*, shows extreme evolution of head-neck morphologies that were presumably shaped by selective pressure from sexual competition and marginal environments.

Hou *et al*. (*1*) challenged our view that *Discokeryx xiezhi* is a member of giraffoids and questioned its significance on giraffoid head-neck evolution and ecological adaptation (*2*). We disagree with their arguments for the following reasons: First, Hou *et al*. mistakenly interpreted a previous study (*3*) in which the horncore proper in the bovid species *Urmiatherium intermedium* is supported by the frontal only [figures 1 and 11 in (*2*)], as reiterated by others (*4*). Bovids have neither parietal supported horncores nor median positioned horncores (*5*). Second, a giraffe's ossicone is covered by skin but the apex becomes cornified in mature males that engage in head and neck sparring (*6*). In *Discokeryx*, skin must have covered the lateral surface of the headgear whereas the dorsal surface was most likely cornified, making the headgear a highly specialized ossicone. In bovids, the supraperiosteal tissues induce horncore formation and normally there is no suture separating bony elements in horncore development (*7*). In *Discokeryx* and *Tsaidamotherium hedini*, the epiphyseal line is clear, showing a developmental ossification different from that of bovids. Third, the histological slices of *Discokeryx* and *Honanotherium* show thick lamellar bones that are clearly different from those of the bovid *Turcocerus* in which secondary osteons predominate (*2*), reflecting obvious and convincing differences in the two types. Fourth, Hou *et al*. misunderstood our interpretation of the cornual vein in giraffes and *Discokeryx*. The cornual veins do not pass through the cranial bone in bovids, in contrast to giraffes and *Discokeryx* where they pass through the large foramen on the parietal bone before joining the superficial temporal vein (*8*). Fifth, the bony labyrinth as a conservative organ remains mostly unaffected under various environmental or behavioral selective pressures compared to teeth and long bones, which is important for phylogenetic reconstruction (*9*). Giraffoid labyrinths show a distinct morphology in that the posterior ampulla insertion of the lateral semicircular canal is very close to that of the posterior canal, which is present in all giraffoids known but absent in all bovids examined, including the earliest species *Eotragus artenensis* (Fig. 1). The giraffoid condition is also present in *Tsaidamotherium hedini* (Fig. 1), a feature phylogenetically distancing the species from Bovoidea (*9*). Moreover, a lack of canines in *Discokeryx* specimens is not sufficient evidence to exclude it from giraffoids. Similarly, the canine is unknown in the unequivocal giraffid *Canthumeryx* (*10*), whereas *Orangemeryx* has one-lobed lower canines (*11*) but is still placed in Giraffoidea (*1*). Further, lack of keratinous integument is not a synapomorphy for giraffids, not to mention the cornified apex of the giraffe ossicone (*6*).

Hou *et al*. modified some characters and character codings of our dataset and, using a different method, obtained a different tree topology [figure 1 in (*1*)]. Without knowing what has been changed, we cannot respond to their result adequately. Nonetheless, their phylogeny (*1*) is similar to ours in that *Discokeryx* is grouped with *Prolibytherium* but differs in having a more basal position for Prolibytheriidae. However, in their phylogeny (*1*), Cervidae as a basal group of Pecora (more basal than *Discokeryx*) is untenable, which also broadly contradicts the molecular phylogeny in which Antilocapridae are at the base of pecorans and Cervidae pairs with Bovoidea (*12*). What surprises us is the deep insertion of *Tsaidamotherium* within bovids [figure 1 in (*1*)]. As clearly illustrated [figure S5, I to J, in (*2*)], the horncore of *Tsaidamotherium* is on the parietal bone, as in giraffoids but not bovids. Along with the epiphyseal line and bony labyrinth feature stated above, *Tsaidamotherium* is, in our view, a giraffoid and not a bovid. Hou *et al*.'s phylogeny suggests either an awkward horncore evolution in ruminants or problematic character codings in their analysis.

Hou *et al*.'s five reasons (*1*) to question our interpretation of the implications of *Discokeryx* on giraffoid evolution are largely off target. First, *Discokeryx* is a giraffoid by our phylogeny; even so, one cannot assume "an antecedent of the behavior of modern giraffids" (*1*). Second, we do not think Hou *et al*. intended to offer a scenario where head-butting could evolve to necking behavior; their second argument is an empty statement. Third, neck elongation induced by necking combat was postulated long ago (*13*) but was miscredited to us (*1*). Nonetheless, as an alternative to the traditional feeding hypothesis for elongation of giraffe necks, both need to be tested by new evidence. Our study better corroborates the necking combat hypothesis without ruling out the traditional one; both mechanisms may have worked for the neck elongation of giraffes. Fourth, the statement "the headgear diversity of giraffoids has been overweighted by artificially using different criteria in different pecoran groups" (*1*) is confusing. Strong correlation between horn and antler morphology and fighting styles in bovids and cervids has been shown (*14*). That giraffoid headgear morphologies are more diverse than those of other ruminants is partly attributable to the fibrocartilage ossification of ossicones that fuse to the cranium with greater flexibility of changes (*15*). Head-butting behavior in *Discokeryx* is supported by the most extreme head-neck structures ever known in mammals. These are by no means overweighted and artificial. Fifth, suggesting that that the mammals included in the Halamagai fossil assemblage must have dietary overlaps (*1*) could be misleading. The complex fluvial sediments, representing a high-energy tophonomic setting, have buried a wide diversity of mammals, of which some could have lived in nonoverlapping niches.

To summarize, we feel more strongly than before that *D. xiezhi* belongs to Giraffoidea.

[1]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. [2]CAS Center for Excellence in Life and Paleoenvironment, Beijing 100101, China. [3]American Museum of Natural History, New York, NY 10024, USA. [4]Naturhistorisches Museum Basel, Basel 4001, Switzerland. [5]Naturhistorisches Museum Wien, Vienna 1010, Austria. [6]University of Chinese Academy of Sciences, Beijing 100049, China. [7]School of Civil & Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430047, China. [8]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA. [9]Naturhistorisches Museum Mainz/Landessammlung für Naturkunde Rheinland-Pfalz, Mainz 55116, Germany. [10]Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, FL 32306-4100, USA. [11]National High Magnetic Field Laboratory, Tallahassee, FL 32310, USA.
*Corresponding author. Email: wangshiqi@ivpp.ac.cn (S.-Q.W.); jmeng@amnh.org (J.M.); dengtao@ivpp.ac.cn (T.D.)

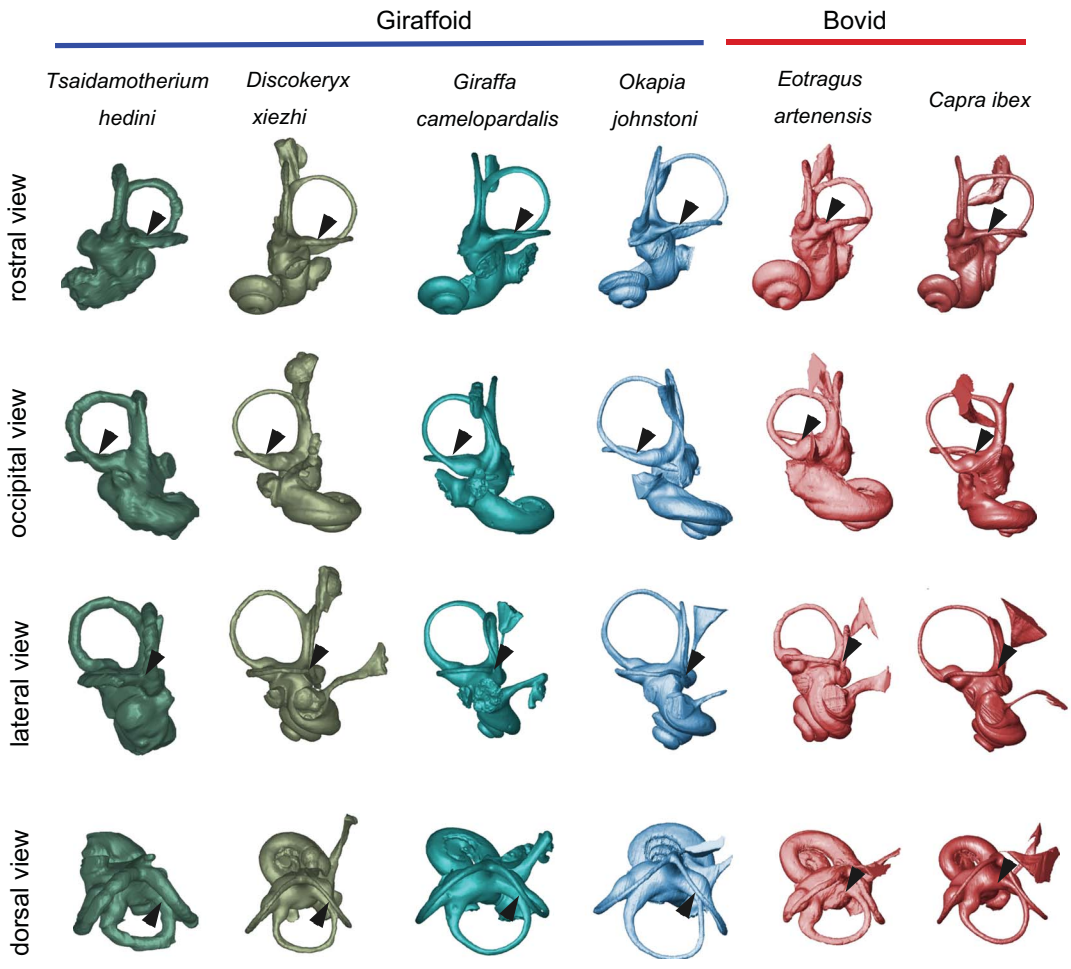


Fig. 1. Bony labyrinth of some bovids and giraffoids. The black arrowheads indicate the insertions of the lateral semicircular canal (lsc) of posterior ampulla (pa), which are different in giraffoids and bovids but are common within each group; not to scale.

Discokeryx and *Giraffa* are two examples within giraffoids that show extreme evolution of head-neck morphologies, which were most likely shaped by selective pressure from sexual competition and marginal environments.

REFERENCES

1. S. Hou, Q. Shi, M. J. Benton, N. Solounias, Comment on "Sexual selection promotes giraffoid head-neck evolution and ecological adaptation" (Science, 2023); 10.1126/science.add9559.
2. S.-Q. Wang *et al*., *Science* **376**, eabl8316 (2022).
3. B. Bohlin, *Cavicornier der Hipparion-Fauna Nord-China*, Paleont Sin. Ser C, Vol. 9, pp. 1–166 (Geological Survey of China, 1935).
4. Q.-Q. Shi, S.-Q. Wang, S.-K. Chen, Y.-K. Li, *Vert. PalAsiat.* **54**, 319–331 (2016).
5. N. Solounias, in *The Evolution of Artiodactyls*, D. R. Prothero, S. E. Foss, Eds. (Johns Hopkins Univ. Press, 2007), chap. 22.
6. C. S. Churcher, in *Horns, Pronghorns, and Antlers: Evolution, Morphology, Physiology, and Social Significance*, G. A. Bubenik, A. B. Bubenik, Eds. (Springer, 1990), chap. 5.
7. W. F. Dove, *J. Exp. Zool.* **69**, 347–405 (1935).
8. C. A. Spinage, *Afr. J. Ecol.* **6**, 53–61 (1968).
9. B. Mennecart *et al*., *Nat. Commun.* **13**, 7222 (2022).
10. W. R. Hamilton, *Philos. Trans. R. Soc. B* **283**, 165–229 (1978).
11. J. Morales, D. Soria, M. Pickford, *Geodiversitas* **21**, 229–253 (1999).
12. A. Hassanin *et al*., *C. R. Biol.* **335**, 32–50 (2012).
13. R. E. Simmons, L. Scheepers, *Am. Nat.* **148**, 771–786 (1996).
14. T. M. Caro, C. M. Graham, C. J. Stoner, M. M. Flores, *Behav. Ecol. Sociobiol.* **55**, 32–41 (2003).
15. T. Ganey, J. Ogden, J. Olsen, *Anat. Rec.* **227**, 497–507 (1990).

ACKNOWLEDGMENTS

Funding: This work was funded by the following: Strategic Priority Research Program of the Chinese Academy of Sciences XDB26000000, XDA20070203, XDA20070301, and Swiss National Science Foundation projects 200021_178853 and 200021_159854/1.
Competing interests: The authors declare that they have no competing interests. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 21 August 2022; accepted 13 January 2023
10.1126/science.ade3392

By **Jaivime Evaristo**

Hard lessons

My phone rang after I boarded a plane at the Amsterdam airport, on my way to visit family in the Philippines. It was my former Ph.D. adviser calling to tell me a preprint had just been posted that identified flaws in a paper we'd published in *Nature* looking at how forestry practices affect streamflow. My stomach dropped as he told me the authors of the critique were demanding a retraction. We couldn't talk long—the plane soon took off. I spent the 16-hour flight processing a mix of emotions—disbelief, embarrassment, frustration—and wondering what this would mean for my career.

After the plane landed, I took out my laptop and logged onto the airport WiFi so I could read the critique for myself. It was harsh and thorough, pointing out several fundamental flaws in our methods and in the underlying data, which we'd gathered from other studies.

The fallout was swift and intense. I received a flood of emails and messages. Some were from supportive colleagues, but many were harshly critical of our work. "You should not have used that metric even if it was used by earlier studies," one wrote. "Those earlier studies were also wrong!" As the first author of the paper and the person who had done all of the data analysis, I felt deeply embarrassed by the criticism.

We wrote a draft response, correcting the apparent errors in the data set and defending our methods. *Nature* sent our response out for peer review, along with the critique. We decided against publishing our response, however, after receiving feedback from peer reviewers. Our mistakes were consequential, and it was clear that the only ethical thing to do was to retract the paper.

By that point in my career, I was well past graduate school and had secured a faculty position in the Netherlands. But I was still devastated by the toxic comments that appeared on social media and in exchanges with anonymous peer reviewers. Colleagues encouraged me to have a thick skin and ignore comments that seemed to be directed at me personally. But I had a hard time doing that. I struggled with self-esteem, started to avoid meetings and conferences, and deleted my Twitter account to protect my mental health.

When it became clear that a retraction was inevitable, I formally offered my resignation to my department head. He didn't accept it, saying a resignation wasn't needed considering the errors in the paper were honest mistakes.

"Researchers should be encouraged, not punished, for ... retracting flawed work."

I decided to carry on, hopeful I could eventually recover my reputation by publishing sound science and demonstrating that I'd learned from my mistakes. But that has turned out to be a tough path. In the 3 years since the retraction came out, I've had trouble getting my work published. I've also had problems securing funding for new projects, even when the proposed work had nothing to do with the work that was retracted and external reviewers were positive about my proposal.

Despite these challenges, I don't regret our decision to retract the paper. It may have been embarrassing and humbling, but it was the right thing to do. And the experience helped me grow as a scientist. I had made my data and code for the *Nature* paper openly accessible so others could review and verify my findings, and I have a new appreciation of the value of doing so. This transparency is essential for the integrity of science, and I learned the hard way that it is better to be open and accountable, even if it means admitting mistakes. After taking a hard look at my own role in the situation, I have also realized I should have reached out to other researchers to get feedback on my methods before publishing. I can't expect myself to know everything as a scientist and my work will be stronger if I seek out diverse expertise and opinions.

In the end, the reality is that retractions are a necessary part of the scientific process—and one that shouldn't be viewed only through a negative lens. Retractions can also be an opportunity to learn and improve. Honest mistakes happen, and researchers should be encouraged, not punished, for doing the right thing and retracting flawed work. ■

Jaivime Evaristo is a tenured assistant professor at Utrecht University.

ILLUSTRATION: ROBERT NEUBECKER