Exoplanet watchers find no place like home p. 1286

A history of science in 10 books p. 1292

Carbon nanotube fibers with high dynamic strength p. 1318

Science

$15
21 JUNE 2024
science.org
AAAS

GETTING CLOSER

Environmental change increases the value of social tolerance p. 1330

CONTENTS

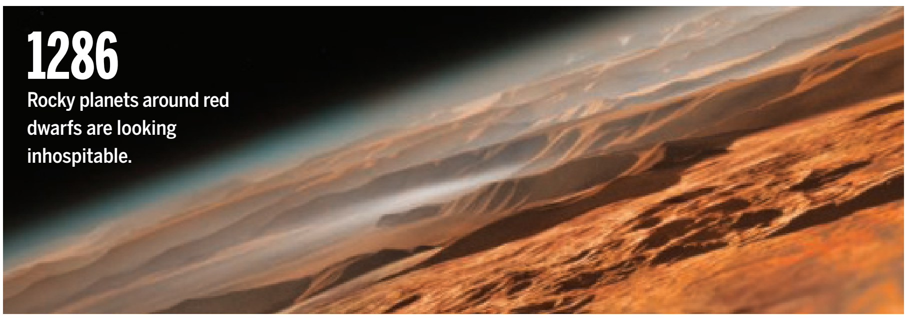
21 JUNE 2024 • VOLUME 384 • ISSUE 6702

NEWS

IN BRIEF

IN DEPTH

FEATURES



INSIGHTS

BOOKS *ET AL.*

PERSPECTIVES

ILLUSTRATIONS (TOP TO BOTTOM): ALEJANDRA PENALOZA; M. KORNMESSER/ESO

ON THE COVER

Hurricane Maria devastated a small island off the coast of Puerto Rico, leaving its population of rhesus macaques exposed to excruciating heat. In this transformed landscape, macaques that were more tolerant of others were better able to survive. This was not the case before the disaster, indicating a change in the adaptive benefits of social ties after an ecological disturbance. See page 1330. *Photo: Lauren J. N. Brent*

ILLUSTRATION: ROBERT NEUBECKER

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2024 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2627; Foreign postage extra: Air assist delivery: $107. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

EDITORIAL

Reducing nuclear dangers

Matthew Bunn is the James R. Schlesinger Professor of the Practice of Energy, National Security, and Foreign Policy, Harvard Kennedy School, Harvard University, Cambridge, MA, USA. matthew_bunn@harvard.edu

Dark clouds loom on the nuclear horizon, with threats from all directions: Russia's nuclear bombast in its war on Ukraine, China's construction of hundreds of nuclear missile silos, North Korea's missile testing, India and Pakistan's ongoing nuclear competition, and Iran's push toward nuclear weapons capability. In response, US policy-makers are discussing whether a further American nuclear arms buildup is needed. At the same time, evolving technologies, from hypersonic missiles to artificial intelligence, are straining military balances and may be making them more unstable. The risk of nuclear war has not been so high since the Cuban Missile Crisis.

These are the kinds of danger that accords for nuclear restraint were meant to mitigate by providing predictability and transparency while moderating worst-case analysis on all sides. But these agreements have been gravely weakened. After multiple Russian violations and American withdrawals, the 2010 New START Treaty is the last remaining pact limiting US and Russian nuclear forces (still over 80% of the world's total)—but it expires in February 2026 and Russia is blocking inspections that the treaty requires. No talks on arms restraints are underway. The world could soon face an unrestrained arms competition for the first time in over five decades—and a more complex one involving more countries and more technologies.

> "The risk of nuclear war has not been so high since the Cuban Missile Crisis."

But it is worth remembering that the 1983 Soviet arms control walkout—which seemed that it might be the end of such efforts—was followed 4 years later by the first treaty with real reductions and on-site inspections. Opportunities can arise unexpectedly—especially if people of good will work behind the scenes to create them. Here, the science and engineering community can play a major role. Nongovernment "Track 2" dialogues—which offer more freedom to explore ideas than is available to government representatives—can develop concepts for governments to take up later. Discussions among American and Soviet scientists and engineers were crucial in shaping early arms control ideas in the 1960s and bringing the Soviet government on board. Later Track 2 discussions helped pave the way for securing and dismantling the Cold War's deadly legacies. The US National Academies Committee on International Security and Arms Control, the Pugwash Conferences on Science and World Affairs, and groups at universities and think tanks around the world are carrying on this work today.

In-depth technical discussions involving the scientific community are essential for issues such as reducing the dangers of conflict in outer space and cyberspace and exploring how new technologies such as commercial space systems and artificial intelligence can help verify the next generation of arms restraints. New modeling of the climate catastrophe that might be caused by a nuclear war was a key element of the global movement that led to the nuclear weapons ban treaty (now in force, though no state with nuclear weapons has signed up). Concepts for verifying the dismantling of nuclear weapons without revealing classified information and for monitoring sites remotely are opening new avenues for future accords.

Initial steps forward should focus on reducing US tensions with China, Russia, and North Korea. Governments should revitalize and expand key risk reduction measures including steps to avoid and manage dangerous military accidents, to establish communication at multiple political and military levels, and to notify one another about missile tests and major military exercises. Unilateral measures—big enough to be noticed but small enough not to endanger security, like President Kennedy's announcement of a unilateral nuclear testing pause in 1963—may be an essential part of such efforts.

In addition, the United States, Russia, and China must work out ways to build predictability, reduce hostility, and avoid the dangers and costs of unrestrained competition. All three parties should end reliance on "launch on warning" policies and take missiles off alert so that decisions on life or death for millions of human beings do not have to be made in minutes.

Governments need help from scientists and engineers both in understanding the dangers that nuclear weapons continue to pose and in finding paths to reduce them. Those who have been doing much of this work are aging out. By becoming informed, joining groups working on these issues, and taking part in arms control dialogues and research efforts, a new generation of scientists and engineers, working across national boundaries, can provide that help.

–Matthew Bunn

10.1126/science.adr0532

NEWS

“The end of slavery has vast implications for everything after. To avoid [studying] that would be avoiding data.”

Neuroscientist Bianca Jones Marlin, who studies the heritable effects of trauma on neurons, reflecting on this week's U.S. holiday Juneteenth, which marks slavery's end.

IN BRIEF

Edited by **Jeffrey Brainard**

POLICY

EU approves law to restore nature

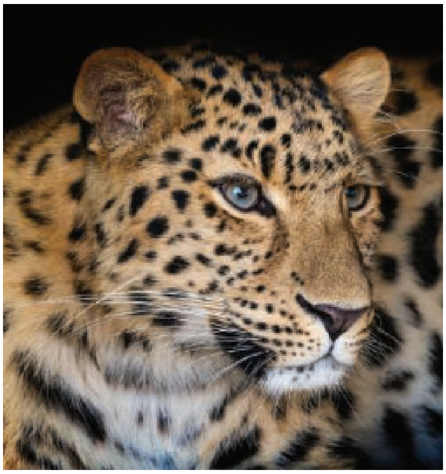
The European Union's environment ministers this week narrowly approved a controversial law to revive the continent's natural habitats, the final step in its adoption. The measure requires the EU's 27 member states to develop plans to restore the more than 80% of habitats that are rated in poor health. These plans are meant to ensure that at least 30% of these areas—which include wetlands, rivers, reefs, and forests—are returned to good condition by 2030, rising to 90% by 2050. Farmers had protested the measure, and key provisions have been weakened during its long approval process. Even so, the final ballot was close. The deciding vote came from the Austrian environment minister, Leonore Gewessler of the Greens, who went against the wishes of the Austrian People's Party (ÖVP), the senior partner in Austria's governing coalition. ÖVP officials said after the vote they would file criminal charges against Gewessler for “abuse of office.”

Argentina's science spending cut

FUNDING | Argentine President Javier Milei is making good on promises before his election last year to slash government spending on science, a report says. From January to May, the government gave research agencies $376 million Argentine pesos ($416 million), or 24% less than during the same span in 2023, according to the Ibero-American Center for Research in Science, Technology, and Innovation, a group of university-affiliated researchers in Argentina. The cuts pose a particular threat to 18 agencies that pay researcher salaries, the report suggests: Since Milei took office in December 2023, 14 of these agencies have spent 75% of their budgets for the 2024 fiscal year that ends on 31 December. If such trends continue, the drop in research spending will rival the sharpest in Argentina's history, recorded in 2018. Milei has said his austerity plans are needed to reduce severe inflation and resurrect Argentina's battered economy.

Russia, North Korea to aid big cat

CONSERVATION | North Korea and Russia are rekindling an old flame: science cooperation. In an initial step, the military allies and global pariahs are banding together to conserve the rare Amur leopard (*Panthera pardus orientalis*), the news outlet RIA Novosti reported last week. During the Cold War, the two nations were steadfast science and technology partners, and this week they were expected to announce new science collaborations during a visit by Russian President Vladimir Putin to Pyongyang, North Korea. U.N. sanctions ban most science cooperation with North Korea, but the project to save the critically endangered leopard is unlikely to draw international ire. Camera traps confirm only about 100 roam forests in East Asia; the last confirmed sightings on the Korean Peninsula date to the 1970s. Last month, Russia and China made a separate pact establishing the Land of Big Cats, a 1.75-million-hectare transboundary reserve for Amur leopards and Siberian tigers.



The Amur leopard, a subspecies, is threatened by poaching, deforestation, and habitat loss.

PHOTO: ERNIE JANES/NPL/MINDEN PICTURES

Voyage ends for probes' leader

PLANETARY SCIENCE | Edward Stone, who led NASA's landmark Voyager mission for 50 years, charting the farthest reaches of the Solar System, died last week. He was 88. Taking advantage of a rare alignment of the outer planets, Voyager 1 and 2, launched in 1977, explored Jupiter and Saturn, and Voyager 2 continued on to Uranus and Neptune. As the Voyagers flew deeper into space, Stone continued as project scientist. He also became head of NASA's Jet Propulsion Laboratory (JPL) in 1991, overseeing other planetary science missions. Now well into interstellar space, the Voyager spacecraft endure: Last week JPL restored full operation of Voyager 1's remaining instruments, which had been offline since November 2023, Linda Spilker, a JPL planetary scientist, reported at a meeting of NASA's Outer Planets Assessment Group. Stone was “just a brilliant scientist and wonderful leader,” she said, before leading the group—begun because of Voyager—in a moment of silence in tribute.

FDA picks fall COVID-19 vaccine

PUBLIC HEALTH | The U.S. Food and Drug Administration (FDA) last week endorsed a

IN FOCUS An image of giant kelp (*Macrocystis pyrifera*) off California's Channel Islands was honored this week as a finalist in this year's BigPicture Natural World Photography Competition, run by the California Academy of Sciences. Researchers are studying how warming ocean waters have reduced populations of kelp, a foundation of one of Earth's most productive ecosystems.

strategy for a COVID-19 vaccine to be used this fall that favors a composition different from one recently recommended by an agency advisory committee. The continuing evolution of SARS-CoV-2 makes choosing the most effective vaccine challenging. This month, the panel of outside scientists suggested FDA greenlight a vaccine design matched to the variant JN.1; most current infections stem from viruses descended from it. JN.1 is also the version the World Health Organization recommended for vaccines in April, when it was more prevalent. But a drawback of this approach is that JN.1 has been overtaken by two offspring, KP.2 and KP.3, implicated in up to 50% of current U.S. cases. On 14 June, FDA announced it will permit a vaccine aimed at JN.1 but favors targeting KP.2 "if feasible"—a suggestion that vaccinemakers Pfizer and Moderna will likely follow. Another maker, Novavax, is already producing the JN.1 vaccine; the company began earlier because its vaccine takes longer to make.

Galaxy's central black hole ignites

ASTRONOMY | For the first time, researchers have reported glimpsing what appears to be the "switching on" of a supermassive black hole in the center of a galaxy. In 2019, the galaxy, SDSS1335+0728, suddenly brightened and soon displayed what resembled an active galactic nucleus, a central black hole that voraciously consumes gas from its surroundings. The change in the previously unheralded galaxy, 300 million light-years away, was detected by the Zwicky Transient Facility, a telescope designed to spot changes in the night sky. Some galaxies temporarily brighten because of supernovae or black holes ripping up a nearby star, but SDSS1335+0728 has continued to get brighter in every wavelength along the electromagnetic spectrum, a team reports this week in *Astronomy & Astrophysics*. The group hopes future observations will rule out other possibilities—including a slow star-ripping event or an entirely new process—and shed light on what ignites these cosmic beacons, the universe's most luminous steady light sources.

U.S. secretly knocked China vax

MISINFORMATION | In 2020 and 2021, the U.S. military ran a secret campaign to discourage people in the Philippines and Muslim-majority countries from using a COVID-19 vaccine made in China, Reuters reported in an investigation published last week. The strategy may have caused preventable deaths. U.S. military officials told the news service the campaign was meant to counter China's growing political influence and push back against online claims it made that SARS-CoV-2 originated in the United States. Public health officials blasted the U.S. effort, in which a Department of Defense unit used phony online accounts to post messages criticizing China's Sinovac vaccine—the first against COVID-19 to become available in the Phillipines after the U.S. and other wealthy countries scooped up doses of other vaccine types. Studies indicated Sinovac was less effective than other vaccines, but the World Health Organization endorsed it. The U.S. messages targeted at Muslims suggested Sinovac contained a pork derivative, which its maker has denied.

Study backs diabetes drug as safe

BIOMEDICINE | Paternal use of the widely used diabetes drug metformin—without other diabetes medications and in the 90 days before conception—was not associated with major birth defects in the resulting babies, a study has found. The findings, based on 384,000 births in Israel, contradict a large Danish analysis that found a 40% increased risk of major congenital malformations (MCMs) in babies of men who took metformin. The authors of the new report, published on 17 June in the *Annals of Internal Medicine*, found a 14% decrease in the risk of MCMs in the offspring of men who took metformin alone. The researchers write that the Danish analysis may have been confounded by its possible failure to exclude all mothers with diabetes, correct for poorer heart health in fathers taking metformin, and separately analyze men taking a variety of other diabetes drugs along with metformin—factors that may be linked to MCMs.

PHOTO: FRESSIA JUNQI PENG

AGRICULTURAL SCIENCE

Wheat scientists unveil historic 'gold mine'

Genome sequencing of century-old collection could help breeders restore long-lost traits

By **Erik Stokstad**

An antique collection of wheat from around the world could breathe new vigor into the staple. When plant breeders created modern wheat during the 19th and 20th centuries, they focused on crossing and selectively breeding a few key varieties, creating a finicky racehorse of a crop: high yielding but vulnerable to disease, heat, and drought and reliant on a liberal application of fertilizer. Part of the solution, according to a study published this week by *Nature*, may lie in the genetic diversity in 827 kinds of wheat, many of them long vanished from farms.

The research is the culmination of a massive, decadelong effort to characterize those crop populations, or landraces—sequencing their genomes, planting them in fields, and scrutinizing their traits. "It is a herculean work," says geneticist Jorge Dubcovsky of the University of California, Davis, who wasn't involved in the study. "This will be a fantastic new resource for the global wheat research community." Already scientists have identified genes that, if bred into modern wheat, could reduce the crop's need for nitrogen fertilizer and increase its resistance to wheat blast, a disease now threatening harvests in much of the world.

The landrace collection was assembled in England starting in 1924, when Arthur Ernest Watkins joined the University of Cambridge's Plant Breeding Institute. Watkins was studying wheat anatomy, examining variation in traits such as the leaflike structures at the top of the stalk. He realized these traits might help with differentiating landraces. So, he convinced the London Board of Trade to collect wheat samples on his behalf. Over 2 decades, consuls and business agents across the British Empire and beyond visited local markets and bought grain that had been grown in as many environments as possible, acquiring 7000 samples of wheat from 32 countries.

Ever since, curators kept the collection viable by sowing and collecting seeds every few years—a practice interrupted only by World War II, when some landraces were lost. Meanwhile, in the wider world, farmers stopped growing many of these landraces as new, high-yielding wheat arrived. Others lost their uniqueness as curious farmers interbred them with the modern varieties. This erosion of genetic diversity makes the Watkins collection—now maintained by the John Innes Centre (JIC), a plant science institute in Norwich, England—a valuable "snapshot of time," says Alison Bentley, a wheat geneticist at Australian National University.

To find out how valuable it is, Simon Griffiths, a geneticist at JIC, wanted to sequence the genomes of the landraces and chart the novel genetic diversity they contained. Given the size of the wheat genome—40 times larger than rice's—the task was outside his grasp. Help came from evolutionary geneticist Shifeng Cheng of the Agricultural Genomics Institute at Shenzhen, which specializes in unraveling crop genomes. In just 3 months Cheng's team sequenced the 827 historic landraces, along with 208 modern varieties for comparison. The genome data arrived at JIC in a suitcase full of hard drives, Griffiths recalls.

The team first took a broad look at the genetic diversity contained in the wheat genomes—a "jaw dropping moment," Griffiths says. The historic landraces contained twice as much diversity as all the modern varieties. "I could see that we were exploring a gold mine," Cheng says.

For breeders to make use of that diversity, however, they need to know which landraces could endow wheat with potentially desirable traits. So the team created crosses from JIC's seed collection, resulting in 6762 unique populations. After growing them in greenhouses and fields in the United Kingdom and China, the team measured 137 traits including yield, nutritional content, and stress tolerance. Then they developed algorithms to connect those traits to the genome data and pinpoint the sources of genes responsible for useful traits. Cheng says the hardest part was finding computational resources large enough to handle the petabyte of data. "If you've never experienced that, you cannot imagine it."

In their *Nature* paper, Griffiths and Cheng's team highlight genes and genetic markers for

PHOTO: KGPA LTD/ALAMY

Some of the historic wheat in a century-old collection was gathered from as far as Australia.

a variety of important traits, such as nutritional content and stress tolerance. Catherine Feuillet, chief scientist at Inari Agriculture, a plant biotechnology company headquartered in Cambridge, Massachusetts, says it's not clear yet how important those particular genes will be but that the new resources have "massive" value. "You can basically identify genes for any characteristic that matters to wheat," she says.

Mark Sorrells, a wheat breeder at Cornell University, suspects the landraces harbor genes that could allow wheat to thrive with less nitrogen fertilizer. The plants in the Watkins collection date from before cheap fertilizer became widely available, he notes, so they may be adapted to less nutrient rich soils. After growing the landraces in fields with either high or low levels of fertilizer, the researchers note in the *Nature* paper that they found a "very promising gene cluster" related to efficient nitrogen use.

Already the landraces have yielded a potentially invaluable gene called *Pm4*, which imparts resistance to wheat blast. Discovered in the Watkins collection and described in *Nature Plants* this week, *Pm4* counters a particularly virulent strain of the fungus that causes the disease. Breeders at the International Maize and Wheat Improvement Center in Mexico are now working to add *Pm4* to wheat varieties grown in Southeast Asia, where harvests are in jeopardy.

Cheng says breeders in China are already adding other traits from the Watkins collection to modern wheat varieties grown in the country. A collaborator in northern China, for example, is looking to enhance the ability of wheat to grow in salty soil. And in India, the source of 128 landraces in the collection, researchers at Punjab Agricultural University are working to improve disease resistance with seeds they acquired from JIC. Parveen Chhuneja, a geneticist at the university, says their efforts will now "gain momentum" because of the information in the new paper.

Sorrells cautions that "plant breeders have to be patient long term." Crossing adds undesirable characteristics that must be removed with further crosses, he explains, so it could take a decade or longer for traditional breeders to improve modern varieties with genes from the old landraces.

Breeders will have plenty of freedom to experiment with the new traits, because the genomic resources described in the *Nature* paper as well as the seeds are available cost free to researchers around the world. "We are the custodians of the material," Griffiths says. "We're not the owners." ■

UNITED KINGDOM

As U.K. election nears, major parties reveal their science plans

All pledge support for research but would restrict immigration, dismaying researchers

By **Holly Else**

The United Kingdom's Labour Party, the likely winner of national elections set for 4 July, has unveiled ambitious science-related plans. It promises to develop decadelong science spending plans, cut net carbon emissions from electricity production to zero by the end of the decade, and reduce the use of animals in some kinds of research if it prevails.

The promises are among the science-related provisions included in manifestos, or policy statements, released last week by the U.K.'s three largest parties: Labour, the Conservatives, and the Liberal Democrats.

The election is widely seen as a referendum on 14 years of Conservative rule, and polls suggest Labour is favored to form a new government, which would be led by veteran politician Keir Starmer.

Economic issues, health care, and immigration have dominated the campaign. But the manifestos all express support for R&D, which is "fantastic," says Daniel Rathbone, head of the Campaign for Science and Engineering, an advocacy group. "That cross-party support for research … is critical for the U.K.'s future and will allow us to tackle many of the challenges that we face as a society."

On funding, Labour is pledging to have research agencies—including UK Research and Innovation, the nation's main public science funder—set spending plans that cover 10 years. That is longer than the 3-year blueprints traditionally used in the government's budgeting process. The longer cycle will "allow meaningful partnerships with industry to keep the UK at the forefront of global innovation," according to Labour's manifesto.

Many researchers have been calling for such long-term planning. Kieron Flanagan, a research policy specialist at the University of Manchester, says, "It a very good idea to be thinking like this," because it allows for continuity. But he notes that lawmakers serve just 5-year terms in Parliament. Because the government could change hands within the 10-year window, the plans could end up becoming a "vague promise or commitment of a direction of travel that might change."

Labour didn't put figures to those spending plans, leaving researchers guessing whether a Labour win would mean more or less money for science. But one of Labour's shadow science ministers has suggested the party will maintain current levels of spending on science, which amounts to about 3% of the U.K.'s gross domestic product (GDP).

The Conservative manifesto says the party will uphold its existing promise to boost science investment to £22 billion per year by 2026–27. (The percent of GDP that sum would represent is unclear.)

Labour Party leader Keir Starmer (left) debated U.K. Conservative Party leader Rishi Sunak in early June.

PHOTO: JONATHAN HORDLE/ITV VIA GETTY IMAGES

The Liberal Democrats offer the most ambitious science spending plan. The party, which has seen its popularity wane since it formed a coalition with the Conservatives in 2010, says it will invest 3.5% of GDP in research and innovation by 2034, but offers no other details. The party also says it will reinstate a goal of spending 0.7% of GDP on foreign aid, a funding stream that has previously helped support research partnerships with other countries.

On immigration, the manifestos may do little to tamp down fears that the next government will make it more difficult for foreign scientists and students to come to the U.K. Wellcome and the U.K.'s Royal Society have called for reducing the up-front fees paid by scientists seeking work visas. But earlier this year the Conservative government hiked migrant health care costs, making it more expensive to bring dependents, and boosted the minimum earnings required for a visa.

Both Labour and the Conservatives have said they want to reduce immigration. And the Conservatives have floated further visa price rises. "That would be very worrying in terms of our ability to attract people to the U.K.," says Martin Smith, head of Wellcome's policy lab.

On climate policy, the Conservative manifesto makes no major new promises. But the party remains committed to achieving net zero emissions by 2050, in part by tripling offshore wind capacity and scaling up nuclear power. But the Conservatives also say they will build new gas-powered power stations to help keep the lights on when the weather hampers wind and solar.

Labour says it will "follow the science" by delivering almost all U.K. electricity from zero-carbon sources by 2030. The party wants to double onshore wind production, triple solar power, quadruple offshore wind, and get Hinkley Point C, a new nuclear reactor besieged by delays, over the finish line.

Meeting this zero-carbon target will be "very difficult," the think tank Public First concluded in an October 2023 report analyzing Labour's plan. One big challenge, it noted, will be modernizing the U.K.'s power grid to handle new renewable power sources.

Still, Labour's "detailed, more holistic" climate plan could benefit the research community, says Penny Clark, a social scientist at University College London. Ambitious, high-level targets often catalyze funding that filters down to researchers, she says.

Labour also pledges to work with researchers and industry to phase out "animal testing," although it is not clear whether that means ending the use of animals in biomedical research. ■

Holly Else is a journalist based in Kenya.

The U.S. government dropped charges against a researcher but held the Cleveland Clinic responsible.

PHOTO: DUSTIN FRANZ/BLOOMBERG VIA GETTY IMAGES

SCIENTIFIC INTEGRITY

Settlement over China funding puts institutions on notice

Cleveland Clinic fined $7.6 million for alleged mismanagement of NIH grants

By **Jeffrey Mervis**

Over the past 5 years, the U.S. Department of Justice (DOJ) has won only a handful of criminal cases in which it prosecuted scientists alleged to have defrauded the government by not disclosing research support they received from China. But last month DOJ sent a clear message that, despite that poor track record, research institutions will be held accountable for mistakes in monitoring outside support to their faculty.

A 17 May settlement with the Cleveland Clinic Foundation (CCF) requires the medical colossus to pay the government $7.6 million to resolve allegations it mismanaged three grants from the National Institutes of Health (NIH). It's a record amount for a case involving foreign research support, a mechanism U.S. policymakers believe China has used to steal U.S. technology. "Today's settlement illustrates the importance of being truthful at every stage of the grants process," U.S. Attorney Rebecca Lutzko said in a statement announcing the settlement.

In addition to the fine, the settlement requires a top CCF administrator "to personally attest" to the accuracy of all information it submits to NIH. "That is a very meaningful sanction or burden on an institution," says attorney Michael Vernick of Akin Gump Strauss Hauer & Feld, who has represented research institutions in similar settlements but was not involved in the CCF case.

The CCF settlement was brought under a Civil War–era law called the False Claims Act (FCA), designed to prevent price gouging by government contractors. Since 2019, the law has been applied in a handful of cases to collect money from institutions accused of failing to properly monitor their researchers' foreign ties. Experts on research compliance say the government's use of the FCA to combat alleged attempts by China to obtain U.S. technology should be a wake-up call to every university receiving federal research dollars.

"When NIH and the Department of Justice looked at what the Cleveland Clinic had done, they saw no messaging or system in place that would cause people to change how they were doing things," suggests Mary Millsaps, director of research compliance at North Carolina State University. "And having a high-level official be on the hook for all reporting is because they didn't see leadership buy-in."

CCF itself wasn't the government's initial target. In 2018, NIH began to investigate CCF cardiovascular geneticist Qing Wang after getting an FBI list of some 3000 scientists the agency believed had received funding through one of the many Chinese programs aimed at attracting world-class scientists, many of Chinese ancestry. A team in NIH's office of extramural research crosschecked the FBI list with a roster of NIH-funded scientists and, when Wang's name popped out, began to look at his publications for evidence he had not disclosed research support from China or that his work there was interfering with his research commitments to NIH.

In January 2019, NIH asked CCF to investigate those allegations and report back. In April 2020, CCF sent NIH a 17-page summary of its findings. Within days NIH suspended Wang's $2.8 million grant and CCF terminated his employment. He was arrested on 13 May 2020, charged with making false claims in connection with the grant.

However, in July 2021 DOJ abruptly dropped its effort to prosecute him without explaining its decision. Wang, who grew up in China but trained in the United States and became a U.S. citizen in 2005, has restarted his research career at China's Huazhong University of Science and Technology.

To Wang's lawyer, Peter Zeidenberg, the reason the government threw in the towel is clear: His client had followed the rules and voluntarily disclosed his Chinese support. "He told them everything—both NIH and CCF," says Zeidenberg, a partner with ArentFox Schiff LLP. The CCF settlement, he adds, simply confirms Wang's innocence.

But DOJ's lawyers may have had a different reason for abandoning its criminal prosecution of Wang. As cited in the settlement, DOJ had turned up evidence that CCF allowed employees to share passwords to gain access to NIH's reporting system during the period covered by the targeted grant. Such unauthorized access, sources tell *Science*, meant it would have been impossible for the government to prove Wang had personally falsified any reports.

NIH's evolving policy on reporting outside sources of funding also appears to have nudged the government to shift its focus from Wang to CCF. Institutions whose researchers apply for grants are obliged to tell NIH about "current and pending sources" of research, but until recently they believed those requirements only covered other federal grants that they managed. (Grants are technically awarded not to an individual, but to their institution, which is responsible for meeting all attached obligations.)

But in July 2019, NIH announced it needed to be told about all research support going to a grant applicant—not just federal grants but also in-kind support, grants from foreign governments, and funding for visiting scientists in their labs. NIH called it a reminder, but universities saw it as a major change. "It substantially raised the bar in terms of what grant-making agencies expected to be disclosed," Vernick says.

Whether the 2019 notice was simply a reminder or broke new ground, Millsaps and others speculate that DOJ pursued a settlement because it felt CCF hadn't revised its practices to fit the new reality. "There's probably additional egregious behavior that occurred," Millsaps says. "Maybe the Cleveland Clinic didn't cooperate [with DOJ] or take steps to make sure that their investigators understood their obligation to disclose."

CCF denies wrongdoing and blames Wang for any violations of federal rules. "Any forms submitted to NIH by the Cleveland Clinic relating to Dr. Wang's grants were based on information supplied by Dr. Wang, information the Cleveland Clinic did not know at the time was false," it said in a statement when the settlement was announced. CCF declined further comment.

The CCF settlement eclipses two earlier civil settlements involving alleged nondisclosures between the government and the Van Andel Institute. In those 2019 and 2021 cases, totaling $6.6 million, the scientists involved in the grants were never criminally charged with any wrongdoing. But as with CCF, the second Van Andel settlement required a senior executive to sign off on all NIH submissions relating to "other support."

Such personal scrutiny is not sustainable, Millsaps and other research administrators say, because of the large number of grants managed by a large institution such as CCF, the tight deadlines for most grant submissions, and the amount of information that the senior official would need to vet to ensure compliance. Rather, they think the requirement is NIH's way of telling institutions to shape up.

"Universities should not be confused about what is expected of them," Millsaps says. "But whether they have communicated that to the researchers is another question. Some are probably doing it better than others. But everybody needs to improve. Or they may be held liable." ■

SCIENCE POLICY

Republicans float plan to overhaul NIH

Lawmakers propose reducing 27 institutes to 15

By **Jocelyn Kaiser**

The Republican leaders of two committees in the House of Representatives have floated a proposal to restructure the U.S. National Institutes of Health (NIH) that includes a call to curb risky pathogen research.

The plan was released last week by Representative Cathy McMorris Rodgers (R–WA), chair of the commerce committee, and endorsed by Representative Robert Aderholt (R–AL), who chairs the appropriations panel that funds NIH. It is one of several efforts taking shape in Congress to formally approve, or reauthorize, the $47.1 billion agency's policies and programs for the first time since 2006.

The proposal contains a mix of familiar and new ideas. It would shrink NIH's 27 institutes to 15, a concept some research leaders have supported in the past. For instance, the neurological, eye, and dental and craniofacial research institutes would become a single neuroscience and brain institute.

The National Institute of Allergy and Infectious Diseases (NIAID) would be split into two parts: one focusing on infectious disease and the other on the immune system and arthritis. The move appears to be a result of Republican concerns that Anthony Fauci, who led NIAID for 38 years, held too much power. All institute directors would be limited to two 5-year terms, an idea previously floated by the U.S. National Academies of Sciences, Engineering, and Medicine.

The plan would "pause" NIH funding of so-called gain-of-function research—studies that manipulate viruses in ways that might make them a pandemic threat—until stronger regulations are set. And it calls for strengthening NIH's sexual harassment policies and disclosing royalty payments made to NIH scientists.

NIH does not comment on pending legislation. But organizations representing patients and other groups are likely to resist any rejiggering of NIH's institutes.

Congress isn't expected to focus on this and other House and Senate proposals for reauthorizing NIH until next year. ■

Polio workers conducted a vaccination campaign in early June in Karachi, Pakistan.

INFECTIOUS DISEASES

Wild poliovirus makes comeback in Afghanistan and Pakistan

2024 target of ending all transmission will likely be missed

By **Leslie Roberts**

The Global Polio Eradication Initiative (GPEI) has been struggling in Africa with large outbreaks spawned by the polio vaccine itself, which is made of live but weakened poliovirus. But the program at least had the wild virus on the run, driving down the cases it causes in its last strongholds, Pakistan and Afghanistan. Now, GPEI is facing a devastating setback on that front, too.

In 2021 and 2022, through intensive vaccination campaigns, GPEI pulled off what its independent monitoring board (IMB) called a "remarkable achievement." It cleared the wild virus from the teeming megacity of Karachi, Pakistan, which has long been a source of reinfection for the rest of the country. For more than 2 years, Karachi saw no paralytic polio cases, and surveillance of wastewater turned up no sign of the wild virus.

The virus also disappeared from Pakistan's two other historic "reservoirs": areas around Quetta in the southwestern province of Balochistan, and Peshawar in the northern province of Khyber Pakhtunkhwa (KP), both near the long, porous border with Afghanistan. The virus was cornered in just seven KP districts and a few in the East region of northern Afghanistan—the smallest geographic area in the history of the eradication program, according to a GPEI advisory group.

But now, wastewater samples reveal the virus is back in Karachi and around Quetta and Peshawar, likely brought by people from those holdout districts. As of early June, the virus had spread to about 40 districts in Pakistan. It had also returned to Kandahar in the South region of Afghanistan. The number of paralytic polio cases remains relatively low because much of the population has been immunized—so far this year there have been five in each country. But plenty of virus is circulating, ready to strike any unvaccinated or undervaccinated child. GPEI's goal of stopping all transmission of the wild virus this year is looking increasingly elusive.

The current epidemiology is "alarming," says Natalia Molodecky, who worked with GPEI in Pakistan's polio emergency operations center and is now a consultant for the Task Force for Global Health in Atlanta. "We are seeing continuous detections [of the wild virus] in Karachi, and we are seeing a lot more transmission across the country."

The first signs of trouble came in August and September 2023, when a couple of wastewater samples in the historic reservoirs tested positive for the virus. Over the past 4 months, it has really taken off, says GPEI's Hamid Jafari, who directs the eradication program in the region that includes Afghanistan and Pakistan. In just 1 week in May, the wild virus was detected in 26 environmental samples in Pakistan, and that was before the high season for its transmission began this summer.

"The picture looks horrific," Jafari concedes, "but actually the situation is not so bad." For one, there are now more environmental surveillance sites in Pakistan than ever before, which explains some of the uptick, he says. Just two genetic clusters or lineages of the wild poliovirus are now circulating in the two countries, down from 12 in 2020, a loss of diversity that is often a sign the virus is on the way out, Jafari says.

The virus now spreading across Pakistan originated in Afghanistan and traveled out of the border area with the hundreds of thousands, if not millions, of people, constantly on the move, Jafari says. The recent surge coincided with the Pakistani government's "repatriation" of about half a million Afghan refugees. Stopping virus transmission will mean targeting this "moving reservoir," which is repeatedly missed during standard door-to-door vaccination campaigns, he adds. "It is not possible for one country to be virus-free without the other," says Aziz Memon, who chairs Rotary International's PolioPlus Committee in Pakistan.

Karachi contains all the conditions the virus needs to thrive: very high population density and constant movement—nearly 1 million people move in or out of the city every day, Molodecky says—along with poor sanitation and hygiene. Many children are infected with multiple pathogens, which lowers vaccine efficacy. Karachi also harbors pockets of strong resistance to polio vaccination. Polio workers, often from the same community, sometimes agree to fake finger marking—a sign a child has been vaccinated—or fail to report so-called "refusals," Jafari says.

Michael Galway, who leads the Bill & Melinda Gates Foundation's eradication efforts in Afghanistan and Pakistan, says, "It will be a heavy lift to clear out the virus from all of these historic reservoirs, but the proof of concept is there. We have done it before." Southern KP is "a different reality," he says. There, insecurity, tribal conflict, and community boycotts make it almost impossible for the polio program to operate.

Memon thinks that with redoubled efforts, GPEI can still stop wild poliovirus transmission this year. Galway and Jafari are less optimistic. Jafari says that if the program can confirm what's driving the circulation of wild virus and find innovative ways to reach missed children in the newly reinfected areas and in the entrenched border area, "then we have a strong chance to stop transmission during the next low season. I think it will take until mid-2025." ■

Leslie Roberts is a journalist based in Washington, D.C.

EARTH SCIENCES

Ancient earthquake likely rerouted the Ganges

Discovery of new seismic concern stokes flooding fear for densely populated delta region

By **Eli Kintisch**

Feeding the world's largest delta, the mighty Ganges River passes through lowlands in India and Bangladesh where 156 million people live today. The delta's central river channel hasn't moved in centuries. But it shifted dramatically 2500 years ago when a massive earthquake struck the region, a new study suggests. Its authors say a similar quake-induced shift in the Ganges now could mean a "devastating modern occurrence" of flooding.

Rivers occasionally flow over their banks to create a new path, a phenomenon known as an avulsion. Within river deltas, the main path of the water, the river channel, can move every millennium or so, says geoscientist Elizabeth Chamberlain of Wageningen University & Research, lead author of the study this week in *Nature Communications*.

But avulsions generally "take place over decades" as a river gradually shifts its banks, says co-author Michael Steckler, a geophysicist at Columbia University's Lamont-Doherty Earth Observatory. And usually it's smaller river channels that shift, he adds. "A big river like [the Ganges] doesn't change so easily," he says. On rare occasions, earthquakes can trigger instantaneous avulsions, but none has been noted in a delta, most of which are seismically inactive.

Chamberlain, Vanderbilt University geologist Steven Goodbred, and colleagues stumbled on a counterexample by chance. They were in the region studying one of the Ganges's previous river channels, indicated by satellite images, about 100 kilometers south of Dhaka, the capital of Bangladesh. Late in the day, they serendipitously came across a large hole dug for a future pond and saw vertical columns of sand jutting up through the mud—structures called seismites that form when earthquakes churn sediment layers. Given the seismic activity in the region, the researchers had long suspected seismites lurking below. "Steve said, 'I've been looking for these for 20 years!'" Chamberlain says. "It was a needle in a haystack." The team told the workers "Don't fill it yet—we'll come back tomorrow and work on it," Chamberlain says. "They were going to fill it that night." Had that happened, Steckler says, his colleagues wouldn't have seen anything."

The seismite samples they took ultimately revealed the major river's secret. With a dating technique that uses light to determine when sand or mud grains were last at the surface, the scientists determined that the layers the seismites intruded had been buried in an earthquake 2500 years ago. Similar analysis showed the last sediment deposited in the old Ganges channel had the exact same age. Sand grains in a second remnant channel 85 kilometers downstream also dated back 2500 years. The conclusion? The earthquake had caused the avulsion, the first ever recognized in a delta. "We realized this spectacular archive recorded a major event that reorganized the delta's water pathways," Chamberlain says, adding that the delta's river channel probably hasn't avulsed since then.

Vamsi Ganti, a geomorphologist at the University of California, Santa Barbara, calls the paper a "sound study," saying the team did "a commendable job of bringing together large amounts of detailed field stratigraphic data and geochronological measurements."

Bangladesh, where the Ganges is known as Padma, is prone to large earthquakes, and the authors fear what a repeat of the past event could bring given the many millions now living in the delta region. In 2008, an avulsion of the Kosi River, a tributary of the Ganges, occurred after heavy monsoon rains—not an earthquake. The incident killed more than 500 people in the Indian state of Bihar and was considered at the time the worst Indian flood in 50 years. One million people were evacuated.

A Ganges avulsion caused by an earthquake could be far worse. The group calls such a danger a "cascading risk" because it adds to the toll of a quake in the region. "A big earthquake by itself is damaging. A big earthquake suddenly moving an immense river is even more severe, especially considering the population density of the Bengal," Chamberlain says.

The team hopes its research can help policymakers plan for and mitigate flood hazards in the Ganges delta and beyond. "It is crucial that we understand the links between past earthquakes and as this provides a blueprint for how lowland rivers may respond to major earthquakes across the globe," Ganti says, "especially in tectonically active, populous regions." ■

People living in Dhaka (top) and elsewhere on the Ganges River delta could suffer extreme flooding if an earthquake suddenly changed the river's course there. Vertical sand layers (bottom) suggest that happened 2500 years ago.

PHOTOS: (TOP TO BOTTOM) JANA CAVOJSKA/SOPA IMAGES/LIGHTROCKET VIA GETTY IMAGES; ELIZABETH CHAMBERLAIN

NO PLACE LIKE HOME

The hunt for Earth-like planets has run into headwinds. Some astronomers are looking for signs of habitability on bigger worlds

By **Daniel Clery**

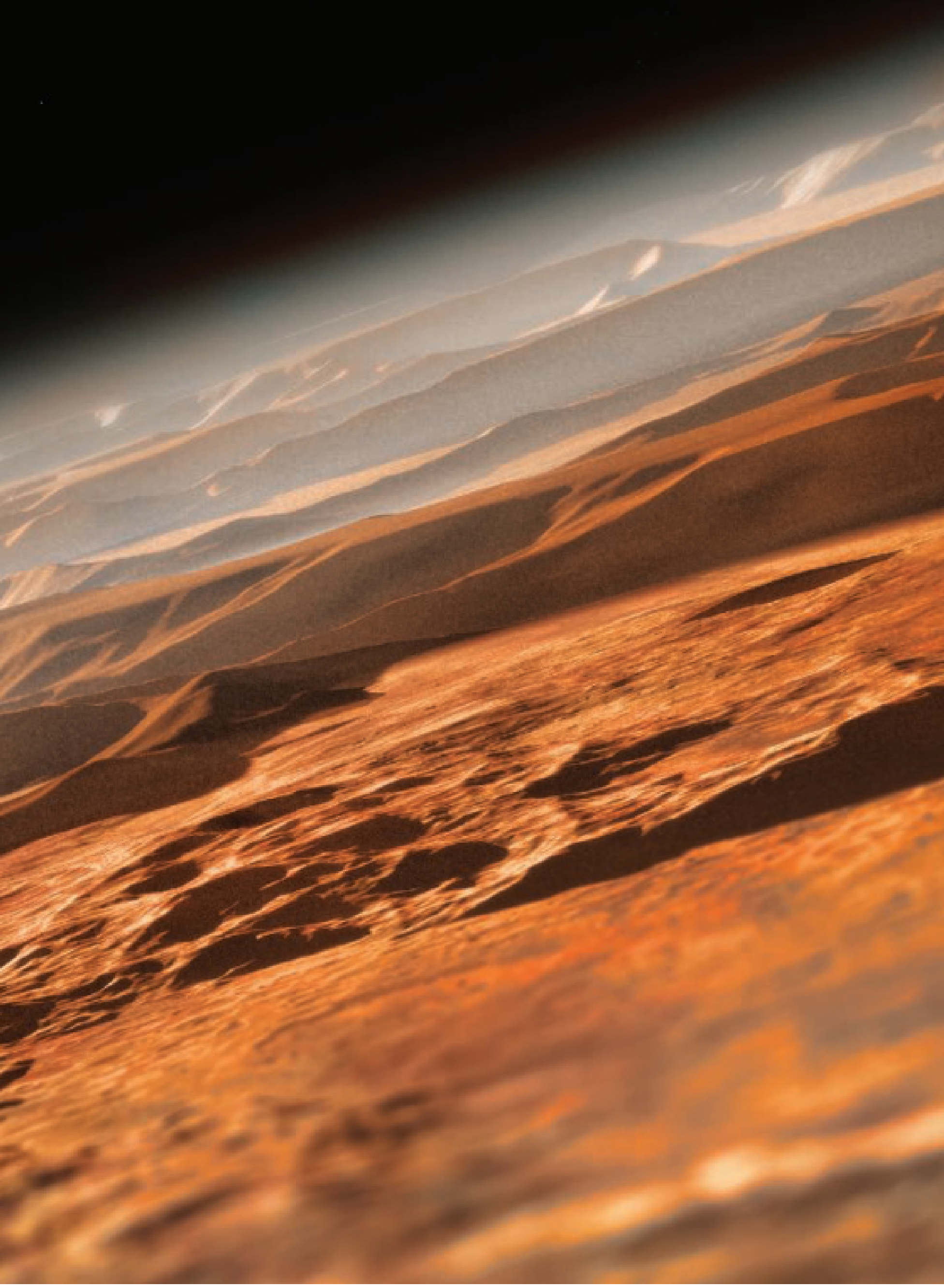
Living on one of the seven Earth-size planets in the TRAPPIST-1 system would be strange indeed. Looming ominously in the sky is an enormous red star, prone to fiery outbursts and appearing several times bigger than the Sun. Hours of the day don't exist; each planet is tidally locked to the star so that one side is forever scorchingly hot, the other eternally frozen. Along the margin dividing the day- and nightsides—the only place with a tolerable climate—a ceaseless wind blows and the star hangs on the horizon, in perpetual sunset.

A short stroll into the dark side brings your planetary companions into view. Every few days one or more passes overhead like a floating lantern, larger than the Moon. Keen observers of the night sky might also notice a bright yellow star, one of the system's close neighbors, and wonder what life might be like near what humans call the Sun.

But for those of us who live 41 light-years away on a world warmed by that benign yellow star, the quest to learn whether one of the TRAPPIST-1 planets could make a comfortable home for our imaginary observer has been an exercise in frustration.

When the seven known planets around TRAPPIST-1 were revealed in 2017, they were a gift not only for science fiction writers, but also for astronomers, who viewed the system as the best place to look for a habitable planet with JWST, NASA's 2-year-old orbiting observatory. It seemed a perfect match of instrument and target. JWST's ability to

IMAGE: M. KORNMESSER/ESO



The red dwarf TRAPPIST-1 looms over one of its seven planets. The star may have stripped away the planets' atmospheres.

peer into exoplanets' atmospheres in the infrared, where life-friendly molecules such as water and carbon dioxide leave their fingerprints, is unique. And TRAPPIST-1, besides being relatively nearby, is a red dwarf only slightly larger than Jupiter, so cool and dim that its light does not swamp that of its planets. The rocky worlds whirl around in orbits much tighter than Mercury's, which means that despite the feeble star, four of the seven are reckoned to sit in a habitable zone where liquid water could exist on their surfaces. The fast orbits, of 3 weeks or less, also mean the planets regularly cross the face of the star from the vantage of Earth—and JWST.

Those transits are a boon to observers, because if there is an atmosphere, some of the starlight will pass through it. Its chemical constituents can selectively absorb the light at specific wavelengths, creating troughs in the starlight's spectrum. Astronomers have used the technique to find evidence for carbon dioxide, methane, and water in the atmospheres of large, hot, uninhabitable planets. Detecting these gases around the TRAPPIST-1 planets would help build the case that they could make suitable homes for life.

That was never going to be easy, because red dwarfs like TRAPPIST-1 tend to be tempestuous, erupting and flaring in ways that can blast away atmospheres and also confound the weak atmospheric signal JWST is trying to detect. By the end of its third year, JWST will have lavished 175 hours of observing time on TRAPPIST-1. However, JWST has yet to find any firm evidence of an atmosphere around TRAPPIST-1's planets.

So, some astronomers want to widen the net. A working group advising NASA and the Space Telescope Science Institute (STScI), which operates JWST, is calling for a broad 500-hour study of 15 to 20 small rocky planets around various red dwarfs, to settle once and for all whether such planets can host atmospheres. "If we find nothing, it will be a disappointment, but it would feel good to have a definitive answer," says working group chair Seth Redfield of Wesleyan University.

Others are beginning to think the search for habitability needs to expand to include other kinds of planets. Must they be rocky and Earth-size? Perhaps larger super-Earths, which could be wrapped in global oceans, should also be considered. Or maybe an even larger body, a mini-Neptune, could hold a water ocean under a thick hydrogen atmosphere that lets in just enough life-giving light. "These are more speculative, and very different from the planetary body we do know has life," says Charles Cockell, director of the Centre for Astrobiology at the University of Edinburgh. "But any planet with the right conditions should be investigated."

Whatever the size, only a handful are within JWST's reach at the moment. René Doyon, an astronomer at the University of Montreal, counts just six: four temperate TRAPPIST-1 worlds that have not yet been exhaustively searched, one potentially watery super-Earth called LHS 1140b, and one hydrogen-wrapped Neptune-like world called K2-18b. "We only have a handful of objects," Doyon says. Michaël Gillon of the University of Liège, whose team discovered TRAPPIST-1's planets, is similarly pessimistic. "We could have good surprises, but habitable planets are a bit out of reach for now," he says.

Time, too, is limited. JWST is now expected to last for up to 20 years, twice what was predicted at time of launch, but finding and studying habitable atmospheres, which is proving harder than expected, may push against that time limit. "Small exoplanets are really hard. It'll be messy," Redfield says. "We're on the precipice, but how close is hard to ascertain."

ASTRONOMERS HAVE ALWAYS JUDGED that the most likely home for life near another star is a rocky planet just like Earth. "We can't look everywhere," Cockell says, "and the best way to narrow down is to follow the constraints of the life that we know." Atmospheric oxygen is not a necessity because life on Earth has existed without it. But at a bare minimum, most astrobiologists agree you need water, a source of energy, and a place for prebiotic molecules to concentrate and react. That could be in rock pools along a sunny shoreline, in hot springs, or around ocean hydrothermal vents. And you need a temperate climate to keep the water liquid, which greenhouse gases such as carbon dioxide help maintain.

But true Earth twins—similar-size planets in Earth-like orbits around Sun-like stars—are not easy to study. It's impossible to snap a picture of such a planet with today's telescopes, which can directly image only the biggest, hottest planets in wide orbits that take them away from the star's dazzle. Even gaining transit data from an Earth twin would be difficult: It would transit its star just once every 365 days and, when it does, the dip in brightness it would cause in a star as large and bright as the Sun would be almost too tiny to measure.

Attention has instead turned to small, red dwarf stars, or M-dwarfs, which are between 10% and 60% of the size of the Sun and less than 7% as bright. Exoplanet surveys have found M-dwarfs to be awash with small rocky planets in tight orbits, some making transits every few hours. And, in the galactic neighborhood, there are lots of them. Of the 60 stars closest to Earth, 50 are M-dwarfs. "M-dwarfs are so attractive to us from an observational standpoint," Redfield says.

But transit results for planets around TRAPPIST-1, the M-dwarf poster child, have shown no clear signs of atmospheres. Results from a technique called eclipse photometry—measuring brightness just before and just after a planet disappears behind its star—have also been discouraging. By subtracting the brightness of the star alone from that of the star and the planet together, researchers can gauge how brightly the planet's dayside glows—an indicator of temperature. A planet that is cooler than expected when illuminated by the star likely has an atmosphere carrying heat away to its nightside.

A March 2023 study of TRAPPIST-1b, the innermost planet, which receives four times as much radiation as Earth, showed no evidence of a moderating atmosphere. In July 2023, researchers turned to its neighbor TRAPPIST-1c, whose similarity to Venus in size and radiation received from its star led to hopes for a thick atmosphere. They found it, too, was boiling hot, suggesting little to no air to sweep heat around to its nightside. However, the team could not rule out a thin gas layer, and a later modeling effort suggested a wisp of oxygen or water vapor could fit the data.

Finding atmospheres around TRAPPIST-1b and c was always a long shot, considering their closeness to the star. But there's a more fundamental concern: that TRAPPIST-1 planets are prone to atmospheric loss because of their star, says Laura Kreidberg of the Max Planck Institute for Astronomy. M-dwarfs are particularly violent in their youth, blasting nearby planets with bursts of ultraviolet and x-ray radiation that can strip away an atmosphere. Even if a planet's atmosphere survives its star's turbulent youth, M-dwarfs continue to flare throughout their long lives. These slow-burning stars are expected to last for trillions of years, hundreds of times longer than the Sun, and TRAPPIST-1 is already nearly twice the Sun's age.

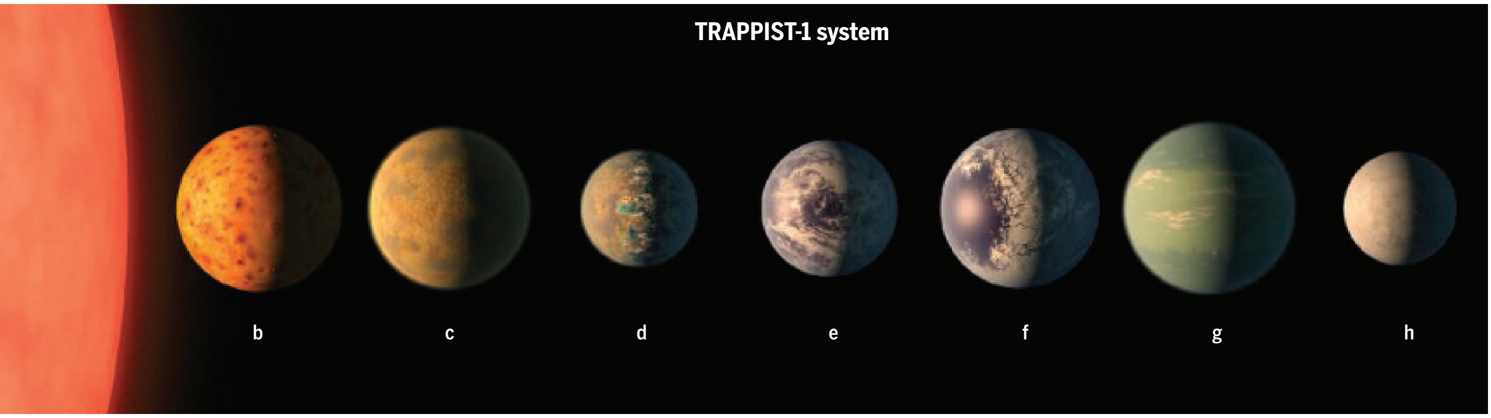
Despite the setbacks, researchers are keen to push out from TRAPPIST-1b and c to the cooler planets in the system, farther from the star and more likely to hold on to any atmosphere. But that comes at a cost in terms of telescope time, because cooler planets emit less infrared, making eclipse photometry more difficult. Moreover, a cooler atmosphere will shrink closer to the planet, weakening any transit signal. For example, it's estimated that JWST will need more than 100 hours of observing time to detect carbon dioxide around one of the cooler TRAPPIST-1 planets. "There is some sticker shock over how many hours you need per planet," Kreidberg says.

TIME ON JWST is precious. Each year, STScI apportions more than 10,000 hours among hundreds of observing teams—and only about 30% of that time goes to exoplanet studies. In 2023, acting STScI Director Nancy Levenson, aware of how long it takes to scrutinize cool rocky exoplanets, offered 500 hours of special "director's discretionary time" to exoplanet astronomers in cycle 3, the third year of observing that begins on 1 July. "This is something that would not happen through our normal time allocation processes," she says.

STScI set up a working group, led by Redfield, to recommend how to use the time. Over the past year, the group held three town hall meetings and received 42 white papers suggesting programs. On 1 April, the group recommended the 500 hours be devoted to a survey of 15 to 20 rocky planets around M-dwarfs to definitively answer the question of whether such worlds can have atmospheres. "We will try to probe a broad enough sample so we will know one way or another," Redfield says.

The proposal does not call for characterizing the atmospheres by laboriously gathering transit spectra. Instead, it says astronomers should use eclipse photometry to simply detect whether an atmosphere exists. Some see the lack of spectroscopy—and the move away from TRAPPIST-1—as a missed opportunity. "I'd like a more balanced program," Doyon says.

An added concern for advocates of TRAPPIST-1 is that only one program looking at the system was approved for cycle 3.



Four of the seven TRAPPIST-1 planets—d through g—sit in the habitable zone of their star and could have liquid water on their surfaces.

IMAGES: R. HURT AND T. PYLE/NASA

That study, 129 hours spread across cycles 3 and 4, will test a new strategy to remove the effects of stellar activity, which can vary day by day. On 15 occasions, JWST will observe the star while TRAPPIST-1b and e both transit in quick succession. The transit of b will give observers a spectrum with no atmosphere present. When e transits, shortly before or after, the star should be in a very similar state, so if there are any differences between the two spectra, they should be due to absorption by e's atmosphere alone.

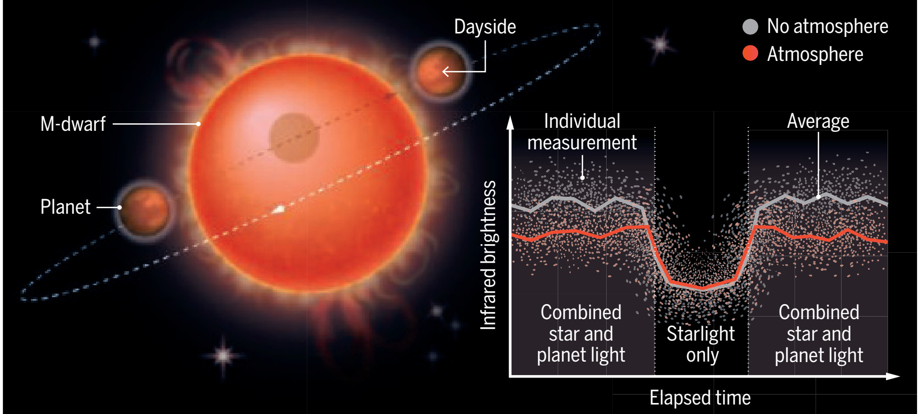
Although the tactic will be immensely valuable if successful, some had been hoping for a more systematic examination of the TRAPPIST-1 planets, says Julien de Wit, a planetary scientist at the Massachusetts Institute of Technology. "Our hope is that the exploration of the system is not put on hold until this program is completed."

Reading starlight

In all but a handful of cases, today's telescopes cannot directly image planets outside the Solar System. Instead, they monitor how the whole system changes, both in brightness and in its spectrum, as the planets orbit their stars. Astronomers get the most information when systems are aligned edge-on to Earth, because they can see changes when a planet goes behind the star (an eclipse) and passes in front (a transit).

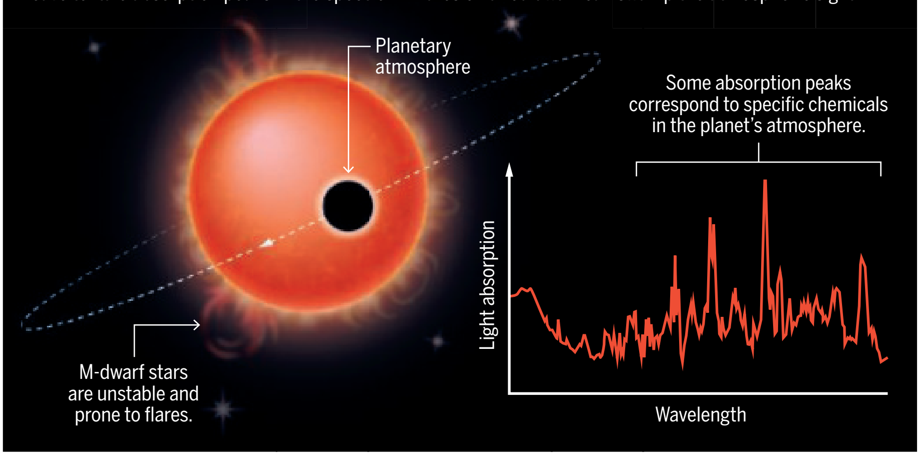
Eclipse photometry

Many planets that orbit close to their star have a permanent, hot dayside. The difference in infrared glow just before and during an eclipse gives astronomers the temperature of the planet's dayside. An airless planet will be brighter and hotter than one with atmospheric winds that carry heat to the nightside.



Transit spectroscopy

During a transit, some starlight passes through the planet's atmosphere—if it exists—and gas molecules leave telltale absorption peaks in the spectrum. Flares on a red dwarf can swamp the atmospheric signal.



GRAPHIC: N. BURGESS/*SCIENCE*

ALTHOUGH M-DWARFS may be the most common kind of star, rocky worlds like Earth are not the most common type of planet. Of the more than 5600 exoplanets confirmed to date, the vast majority fall somewhere in size between rocky Earths and gas giants—a range not seen in the Solar System. Exactly what a planet halfway between rocky and gaseous is like—and whether it, too, might be habitable—remains largely a mystery, although its density might offer a clue.

If a planet's density is high, it probably has an Earth-like composition: an iron core surrounded by rock and a thin atmosphere. If the density is lower, all sorts of possibilities open up: It could have a rocky core surrounded by a deep and thick hydrogen atmosphere, or it could have a somewhat thinner atmosphere and a global water ocean or a layer of ice. To unravel those possibilities, researchers need to probe the planet's atmosphere and see whether their findings match climate models based on the planet's composition.

One such planet JWST has been examining is LHS 1140b, a world squarely in the habitable zone of its M-dwarf star, 49 light-years from Earth. In January, Doyon's team reported precise measurements of its mass (5.6 times that of Earth) and size (1.73 times the size of Earth), making it less dense than a scaled-up Earth.

In March, other researchers described the first JWST transit spectrum of LHS 1140b, taken in July 2023. The data, reported in a preprint on arXiv, are preliminary, but they show hints of what could be a nitrogen-dominated atmosphere like Earth's, and no signs at all of hydrogen, leaving water or ice as the favored explanation for its low density. "That's why we're excited about this planet," says team member Renyu Hu of the California Institute of Technology.

In unpublished work based on two transit observations of LHS 1140b taken in December 2023, Doyon and his colleagues confirmed the lack of hydrogen. That leaves two likely scenarios, according to their climate modeling: that the planet may be an airless rock wrapped in water ice, like Jupiter's moon Europa; or, if its atmosphere is thicker and warmed by carbon dioxide, that ice could be punctuated by a comfortably warm "bull's-eye ocean" in the middle of its star-facing side, half as big as the Atlantic Ocean, he says.

Of his shortlist of six potentially habitable worlds, Doyon says LHS 1140b is the most likely to have retained an atmosphere. "We need more observations with JWST!" But that will be a challenge. LHS 1140b's position in the sky means it is often out of view for JWST, which can only see it transiting its star four times a year. Doyon says as many as a dozen transits will be needed to confirm the atmosphere and see whether it contains carbon dioxide—an indicator of a watery surface.

More speculative is the idea that something even closer to Neptune in size could host life beneath a thick hydrogen-rich atmosphere. One candidate is the planet K2-18b, 8.6 times Earth's mass and 124 light-years away. A team led by Nikku

The comforts of home?

Astrobiologists think a planet needs an atmosphere and liquid water to have a chance at hosting life. Detecting their existence has been difficult, even for JWST, NASA's giant infrared space telescope. Some astronomers think just six potentially habitable worlds are within reach of the telescope.

TRAPPIST-1d–g

Four Earth-size rocky worlds

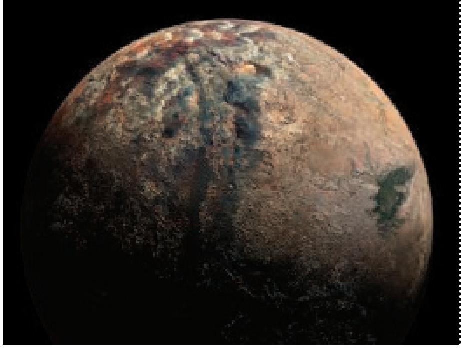
Distance 41 light-years

Size 0.79 to 1.13 Earth radii

Temperature 13°C to 76°C

Density 0.79 to 0.92 Earth's density. Lower densities suggest some gas or liquid layers

Atmosphere Not detected on innermost planets



LHS 1140b

An ice- or ocean-wrapped super-Earth

Distance 49 light-years

Size 1.73 Earth radii

Temperature ~30°C

Density 1.07 Earth's density. Models suggest a makeup of 10% to 20% water or ice

Atmosphere Data suggest nitrogen, with water and carbon dioxide

K2-18b

A watery mini-Neptune wreathed in hydrogen

Distance 124 light-years

Size 2.61 Earth radii

Temperature –8°C

Density 0.48 Earth's density. A thick atmosphere, but thinner than other gas giants

Atmosphere Hydrogen with methane and carbon dioxide

Madhusudhan of the University of Cambridge looked at a spectrum of the planet taken by the Hubble Space Telescope in 2019 and saw evidence of hydrogen, expected of large gaseous planets. The team modeled what sort of climate might exist around a mini-Neptune below such an atmosphere. They found a small number of solutions in which a liquid water ocean could exist, the temperature kept at the right level by greenhouse warming from its moderately thick atmosphere.

The team took another look at K2-18b last year with JWST. In the Solar System, planets with hydrogen-rich atmospheres, such as Saturn, also contain small amounts of ammonia and methane. Sunlight breaks up those compounds in the upper atmosphere but they get regenerated lower down. On K2-18b, a thinner atmosphere and water ocean might interfere with that recycling, particularly for ammonia, according to modeling by Madhusudhan's team. The JWST transit spectrum matched that prediction, showing evidence of methane and carbon dioxide but no ammonia, a sign, Madhusudhan says, that the water ocean exists. He calls K2-18b a "hycean" (a mashup of hydrogen and ocean) world. "No other hypotheses match the data, so far," he says. The team is currently studying JWST data taken from several other candidate hycean worlds.

Not everyone is ready to buy into the idea of habitable hycean worlds. "Most modelers are pretty pessimistic about liquid water on the surface of a sub-Neptune," Gillon says. The level of greenhouse warming would have to be finely tuned to achieve temperate conditions, because thick hydrogen atmospheres around such planets can quickly generate runaway effects, heating the surface above the boiling point. They can also be so thick that they are opaque to light. "Could you have life? You need ultraviolet photons to get through to create complex chemistry."

More generally, Kreidberg doubts life could emerge on worlds with deep global oceans, because any molecular building blocks would be too dilute. "You're never going to meet" other complex molecules, she says. Cockell isn't as quick to dismiss such unfamiliar locations for life. "It's different but has physical and chemical conditions similar to places we know," he says. "It's an exciting hypothesis to test."

Despite the slow progress, those seeking habitable worlds take heart from the surprises that JWST is quietly and steadily finding on larger, hotter worlds. It has found snowlike flakes of quartz filling the skies of WASP-17 b. It has identified the supersonic winds that transfer heat to a cloud-filled nightside on WASP-43 b. And it has found gritty clouds of silt and sand high up in the atmosphere of VHS 1256 b. "It's already a miracle what we are learning," Gillon says.

The hunt for cooler, wetter, habitable planets is sure to take surprising turns as well. Just last month, NASA's Transiting Exoplanet Survey Satellite, designed to discover exoplanets by monitoring stars for transits, found a planet that could become another promising target for JWST: an Earth-size temperate world around the red dwarf Gliese 12, even closer to Earth than TRAPPIST-1. "These are our first baby steps in learning about rocky planets," Kreidberg says. "It's an extraordinary position to be in." Gillon agrees. "It's a very exciting time. Life is the goal, obviously, but we will have to be patient." ■

IMAGES: (LEFT TO RIGHT) IANM35/GETTY IMAGES; BRIGHTSTARS/GETTY IMAGES; PABLO CARLOS BUDASSI/WIKIMEDIA COMMONS

INSIGHTS

BOOKS *et al.*

ILLUSTRATION: ALEJANDRA PENALOZA

REVIEW ROUNDUP

Ten must-read science histories

2024 marks the centennial of the History of Science Society (HSS). For 100 years, our members have crafted accounts of science as human endeavor and cultural phenomenon. A now-vast bibliography reflects the diversity of the sciences, the local forces and global exchanges that shape their development, and the evolving methods and questions that drive historical inquiry. For this year's summer reading list, we asked 10 of our book prize winners to choose the books they would recommend to *Science* readers. Their list includes groundbreaking works covering a variety of disciplines, new perspectives on the origins of standard measurements and modern science, and explorations of stories that have gone untold. **—Matthew Shindell, HSS Secretary**

The Experiential Caribbean

Reviewed by **Tiago Saraiva**[1]

Science came into being during the long 17th century, replacing dogmatic beliefs of ultimate causes and scholastic norms inherited from tradition with matters of fact established through sensorial experience and the handling of natural phenomena. Making excellent use of Spanish Inquisition archives documenting accusations against more than 100 practitioners of this new form of knowledge-making, in *The Experiential Caribbean*, historian Pablo F. Gómez details the many challenges involved in producing authoritative claims about nature that will be accepted by broad audiences. The knowledge-makers in question were not Galileo, Boyle, Hernández, or Newton, working in European universities and courts and in newly inaugurated institutions such as the Royal Society or the French Royal Academy of Sciences. They were Francisco Mandinga, Paula de Eguiluz, Bernardo Macaya, and Antonio Congo—Black men and women of the Caribbean.

Gómez unveils a coherent set of practices to "understand, classify, and manipulate the natural world" that emerged in the early modern period among people of African origin throughout the main port cities of the Spanish Empire in the Caribbean. He asserts the leading position of Black specialists who mobilized ideas, materials, and techniques from Africa, Europe, and the Americas to become authorities in the healing of free Black urban dwellers in the poor neighborhoods of Cartagena or Havana, of enslaved workers in mines and plantations, and even of well-off agents of empire. To make his case, Gómez rejects epistemological prejudice against "black ways of knowing the world," associated in the imperial archive with categories such as witches or shamans, and focuses instead on "Mohanes"—his favored term denoting "master of sorcerers"—as producers of specialized forms of knowledge about bodies, plants, rocks, or the skies.

Gómez's narrative transforms Bernardo Macaya's command of tropical storms through power-objects with techniques of

West Central African origin into a pivotal event for the history of science. And he does the same with Antonio Congo's practice of collecting plants and roots that he called "his people," which Congo believed talked to him in dreams, explaining for which cures they should be used. Gómez's efforts expand the history of science, rendering it an important field beyond European parochial geographies. *The Experiential Caribbean* is a must-read for anyone interested in considering how experiential knowledge historically enabled engagement with the material world.

Saraiva is the author of *Fascist Pigs: Technoscientific Organisms and the History of Fascism* (MIT Press, 2016), for which he received the HSS's Pfizer Award in 2017.

The Experiential Caribbean: Creating Knowledge and Healing in the Early Modern Atlantic, *Pablo F. Gómez*, University of North Carolina Press, 2017, 314 pp.

Wild by Design

Reviewed by **Emily Pawley**[2]

Historians have long described two strains guiding US environmentalism and shaping US nature: "preservationism," which aimed to protect an imagined

Ecological restoration, such as coral cultivation, involves choices that reflect human values and power.

"pristine wilderness," and "conservationism," which cast all nature as a stockpile of "natural resources." In this magnificent book, historian and former wetlands ecologist Laura Martin traces the century-long sweep of an equally consequential third strain of science and action: ecological restoration.

Whereas conservation and preservation attempted to slow or stop change, ecological restoration emerged where catastrophic damage had already occurred: forests stripped of native plants and predators, the whirling clouds of the Dust Bowl, or the radioactive devastation in the wake of nuclear tests in the Pacific. The ecological sciences, Martin demonstrates, forged their central insights through processes of damage and repair. Thus, Eugene and Howard Odum developed theories of "ecosystem disturbance" as they irradiated and poisoned swathes of land, part of an Atomic Energy Commission (AEC) project bracing for World War III.

Restoration efforts were often just as drastic. Martin describes wetlands dug, coral farmed, and "invasive" species shot, burned, or poisoned. Ubiquitous "natural" presences prove to be recent constructions: Rainbow trout from the Pacific coast now strike at flies from Vermont to Japan, their range expanded and their maturation speed doubled by scientists allied with the AEC and General Mills. Such consequences ballooned as restoration ecology itself grew, moving from botanists building plant communities at women's colleges, to state agencies pitting themselves against invasive species, and ultimately out to a web of foundations and corporation-funded "off-site mitigation," an ancestor of today's carbon offset markets.

One thing is clear throughout all these stories: Designing the wild requires choices, and choices inevitably embody values and involve power. Martin investigates who benefits and who loses from restoration, as when the American Bison Society placed its reserves on Native American reservations, intentionally eroding Tribal sovereignty. She invites us to question what "wildness" restorations are expected to restore: the precolonial "baselines" that emerged in the 1980s? "Novel ecosystems" that can shift as the climate shifts?

These issues matter because Martin wants restoration to work. "The narrative that environmental harm is irreversible," she writes, "is, by definition, disempowering." By showing how much wildness has already been designed over the past century, and under what assumptions, Martin enables us to plan better designs of our own, to conceive of restoration as "an optimistic collaboration" among humans and other species. Those acting on the ecological emergencies we face should look back on the stories told by this important book.

Pawley is the author of *The Nature of the Future: Agriculture, Science, and Capitalism in the Antebellum North* (Univ. of Chicago Press, 2020), for which she received the HSS's Philip J. Pauly Prize in 2021.

Wild by Design: The Rise of Ecological Restoration, *Laura J. Martin*, Harvard University Press, 2022, 336 pp.

The Making of Mr. Gray's Anatomy

Reviewed by **Anita Guerrini**[3]

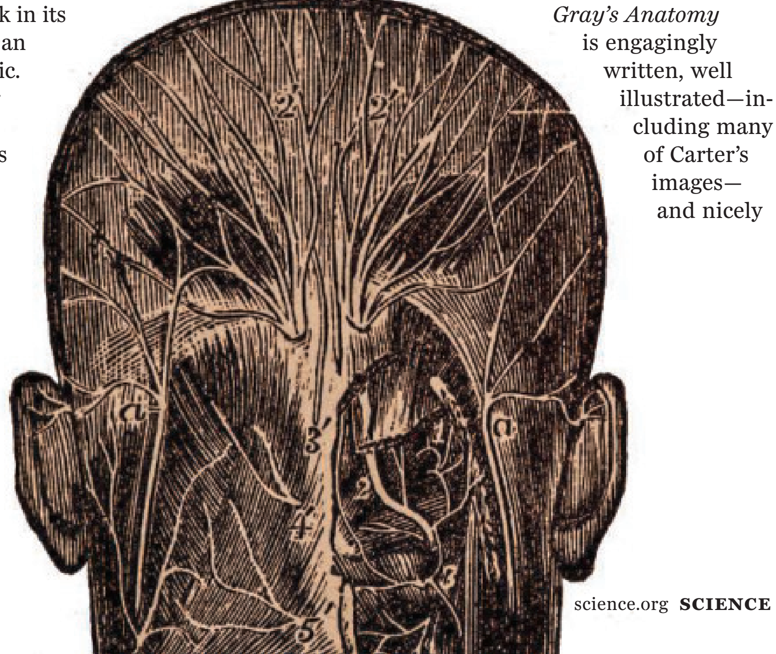
The 42nd edition of *Gray's Anatomy*, the renowned illustrated reference manual for physicians, appeared in 2020. Ruth Richardson's 2008 book, *The Making of Mr. Gray's Anatomy*, commemorated the 150th anniversary of that publication, which was originally titled *Anatomy, Descriptive and Surgical*. In it, she explores in fascinating detail what made *Gray's Anatomy* such an important book in its time and such an enduring classic.

The Making of Mr. Gray's Anatomy takes place in "the London of Charles Dickens—between *A Christmas Carol* (1843) and *Great Expectations* (1860–61)," a period of enormous social and political change mired in enduring inequality. Richardson's book encompasses not only the composition of *Gray's Anatomy* by the surgeon Henry Gray and the physician-artist Henry Vandyke Carter but also the publication process, the dissections of the bodies of the workhouse poor that underpinned it, and even the women who sewed its bindings. Her quest to uncover these details in the face of little documentary evidence makes the book, in addition, a meditation on the historical art.

Much of the success of *Gray's Anatomy* is owed to its distinctive symbiotic relationship between text and illustrations. Gray (1827–1861) began his surgical studies at St. George's Hospital in London at the age of 15 and remained there for the rest of his life. His evident talents as a dissector and surgeon gained assistance from the patronage of the powerful Sir Benjamin Brodie. Richardson contrasts the ambitious Gray with the decidedly unconnected Carter (1831–1897), the book's illustrator. The son of a provincial artist, Carter came to London as an apprentice surgeon-apothecary in 1848.

While Gray represented the old world of patronage, Carter represented the new meritocracy, and a decade later, he could place "MD" after his name thanks to the still-new University of London. He nonetheless struggled with depression and self-doubt. Richardson describes how Gray repeatedly exploited his younger colleague's talents and deprived him of equal credit on the title page. Carter moved to India shortly before the book was published.

The Making of Mr. Gray's Anatomy is engagingly written, well illustrated—including many of Carter's images—and nicely



Its vivid illustrations helped to make *Gray's Anatomy* a valuable reference.

PHOTOS (TOP TO BOTTOM): PCLARCK2/ISTOCK PHOTO; VINTAGEMEDSTOCK/ALAMY STOCK PHOTO produced. Richardson tells a compelling story of the intellectual, economic, social, and physical circumstances surrounding this important book and does not let readers forget the human capital, living and dead, that made it possible.

Guerrini is the author of *The Courtiers' Anatomists: Animals and Humans in Louis XIV's Paris* (Univ. of Chicago Press, 2015), for which she received the HSS's Pfizer Award in 2018.

The Making of Mr. Gray's Anatomy: Bodies, Books, Fortune, Fame, *Ruth Richardson,* Oxford University Press, 2008, 322 pp.

Procedures such as pulse taking were historically conducted differently in Chinese and Greek medicine.

The Expressiveness of the Body and the Divergence of Greek and Chinese Medicine

Reviewed by **Leah DeVun**[4]

Shigehisa Kuriyama's landmark 1999 study begins with a murder mystery. Each witness gives such a strikingly different version of the events that, as Kuriyama writes, "the very idea of truth becomes suspect." So, too, with the history of medicine. As Kuriyama shows through his comparative history of Greek and Chinese medicine, our truths about the human body are, in part, a result of where we stand in the world. For instance, Kuriyama places an illustration from Andreas Vesalius's *De humani corporis fabrica* alongside one from Hua Shou's *Shisijing fahui*. The reader cannot help but notice that the two images reveal vastly different visions of human anatomy. Next, Kuriyama compares the history of Greek and Chinese pulse taking. What appears at first to be a shared diagnostic method turns out to be two wildly divergent practices based in fundamentally dissimilar understandings of the signs of disease. As we learn, Greek and Chinese writers also differed in their descriptions of the body's organs and their functions.

How could two cultures look at and touch the same parts of the body and perceive such different things? Furthermore, if the body is always and forever the same, how can historians, who study change over time, write its history? As Kuriyama argues, how we feel pulses, organs, blood, or breath is remarkably variable across time and geography. By comparing how we gather information from our senses and how we communicate it through language, Kuriyama traces how perception both reflects and creates our sense of the body's reality, which in turn shapes trajectories of health and treatment.

Two decades after its publication, Kuriyama's study is now considered a classic in multiple fields. It is sophisticated enough to be of interest to experts but is written in such fantastically unpretentious prose that students and generalists alike will also read it with pleasure. Although my research is in a different field, Kuriyama's book spurred me to shed some of my own assumptions about the stability of medical categories of sex and gender over time, which proved crucial for my own work. *The Expressiveness of the Body* is not only a valuable contribution to our understanding of Greek and Chinese medicine but also a work of methodological brilliance that invites readers to question where we ourselves stand in the world and how that might prompt us to reevaluate what we take for granted in the body.

DeVun is the author of *The Shape of Sex: Nonbinary Gender from Genesis to the Renaissance* (Columbia Univ. Press, 2021), for which she received the HSS's Margaret W. Rossiter History of Women in Science Prize in 2023.

The Expressiveness of the Body and the Divergence of Greek and Chinese Medicine, *Shigehisa Kuriyama,* Zone Books, 1999, 344 pp.

Networked Sovereignty

Reviewed by **Allison Margaret Bigelow**[5]

As an information scientist and Pascua Yaqui woman, Marisa Elena Duarte does what few historians of science can in her 2017 book, *Networked Sovereignty*: She weaves together Native and non-Native knowledge to highlight the best problem-solving capacities of each tradition. Her case study is the building of high-speed internet in reservations throughout the United States, from Idaho to Minnesota and especially along the US-Mexico border, where transnational Indigenous peoples have lived for thousands of years and now navigate information systems informed by US border control.

Some connections Duarte makes between Native and non-Native thought are philosophical. For example, she presents network theory as a framework for understanding relationships between human, plant, river, and animal nations following Indigenous ontologies and as the human-designed bundles of cables, towers, satellites, devices, and bytes that allow for long-range communication in Western approaches.

Other connections are more literal, as when she describes the work of University of California, San Diego, physicist Hans-Werner Braun, who in the late 1990s—together with a team of industry engineers—determined that the best location for a new satellite transmission tower for the National Science Foundation Network was on Tribal lands. Braun and the engineers worked with Native leaders to identify a Tribal citizen with IT expertise, Michael Peralta, and a professor of ethnic studies, Ross Frank, who could plan, source grants, and implement a network that served both groups. Community members participated at every stage. Once the network was built, citizens from the 19 Nations who consulted on the project designed an archive for community and government use, and Tribal enterprises sprang up to support training in IT, graphic design, and audiovisual recording.

Frameworks such as the "digital divide," which suggest a permanent gap between those who have access to technology and those who do not, fail to explain the complex, creative, and hopeful ways in which Native Nations engage with the backbone of the internet. Some Nations that Duarte describes created for-profit data processing centers, powered by renewable energy, for long-term economic and environmental well-being; others focused on building databases to improve patient care in government-run Indian Health Service clinics. Oftentimes their experiments failed, and they tried again in iterative learning processes that responded to the specific needs of their peoples. As expressions of Native sovereignty, their efforts were place-based and grounded in community. Duarte asks readers to consider what it would look like if more projects began by asking what community members want to learn from scientific research.

Bigelow is the author of *Mining Language: Racial Thinking, Indigenous Knowledge, and Colonial Metallurgy in the Early Modern Iberian World* (Omohundro Institute of Early American History and Culture and Univ. of North Carolina Press, 2020), for which she received the HSS's Philip J. Pauly Prize in 2022.

Networked Sovereignty: Building the Internet Across Indian Country, *Marisa Elena Duarte*, University of Washington Press, 2017, 208 pp.

Secrets of Women

Reviewed by **Tara Nummedal[6]**

The publication of Belgian physician Andreas Vesalius's 1543 book, *On the Fabric of the Human Body,* is widely recognized as a pivotal moment of the Scientific Revolution. Dedicated to Holy Roman Emperor Charles V, the tome opens with a rowdy frontispiece centering Vesalius himself dissecting a female body and contains more than 200 spectacular woodcuts of flayed bodies and open torsos that still circulate widely today, making it a virtuoso example of what the printed book could do in the 16th century. It was also a call for reform, an argument for grounding medicine in direct experience with the human body rather than in ancient Greek and Roman texts. Taking up Vesalius's call, so the story goes, European medical men from the 16th century onward bravely risked social and religious taboos about opening the human body to dissect the cadavers of criminals and transform the Western understanding of medicine as a result.

Vesalius's access to print and patronage enshrined his reputation as a decisive figure in the history of human dissection. However, as Katharine Park demonstrates in her accessible and well-illustrated volume *Secrets of Women*, Vesalius's fame obscured a much longer and far more interesting history of human dissection. Park locates its origins in mid-13th-century Northern Italy and centers spaces of religion, family, and kinship, relegating the context of medical education to "a distant third."

Late medieval Italians opened bodies for all kinds of reasons: to embalm them, to seek physical signs of sanctity, to seek the origins of life in the uterus, to determine cause of death, or to remove infants from mothers who died during childbirth. In all of these contexts, female bodies were "a privileged object of dissection" because they were understood to hold internal secrets of generation, life, and sanctity that only dissection could reveal.

Park begins each chapter of her book with the opening of a female body: an abbess who had visions of Christ planting his cross in her heart, a patrician wife who died after birthing her seventh child, and a married virgin who reported that Jesus had removed her heart and caused her to lactate. Eventually, she returns to the anonymous female criminal who appeared at the center of Vesalius's frontispiece, allowing readers to see this famous image in a new light—as an emblem that contains three centuries of religious, social, and medical practice of human dissection in which the secrets of the female body were always central.

Nummedal is the author of *Anna Zieglerin and the Lion's Blood: Alchemy and End Times in Reformation Germany* (Univ. of Pennsylvania Press, 2019), for which she received the HSS's Pfizer Award in 2022.

Secrets of Women: Gender, Generation, and the Origins of Human Dissection, *Katharine Park*, Zone Books, 2006, 424 pp.

Bitter Roots

Reviewed by **Robyn d'Avignon[7]**

Abena Dove Osseo-Asare's pathbreaking first book, *Bitter Roots*, is the crowning achievement of a generation of scholarship that challenged the near-omission of Africa from the history of science. This was a legacy of the Euro-American origins of the field, and of the racist depiction of Africans, since the era of the transatlantic slave trade, as primitive and irrational. Osseo-Asare rewrites this bitter history through the biographies of six healing plants, all indigenous to Africa, that expatriate and African scientists sought to transform into pharmaceuticals. But this is not a simple story of "biopiracy" or the appropriation of Indigenous plant knowledge by outsiders. As Osseo-Asare shows, histories of medicinal uses of plants within and beyond Africa, and the unruly distribution of plants themselves, did not respect the boundaries of empires, nation-states, or ethnolinguistic "groups" in or beyond the continent.

Drawing on archives and oral histories from Ghana, Madagascar, and South Africa, Osseo-Asare contrasts the "fragmented, synchronous stories of shared creation" of plant knowledge produced through the interactions of African healers, urban plant vendors, and scientists with narratives of "priority" that insist that "we know the time of each discovery." Colonialism created a forced "open source" economy of plant knowledge in Africa, but the benefits of this exchange were uneven, as colonial-era scientists relied on patents to exclude the claims of Africans to plant-based poisons and remedies. We encounter British soldiers and arrow poisons in the Gold Coast alongside Dutch settlers experimenting with the hoodia root, historically used by the Indigenous Khoisan to abate hunger in southern Africa.

The book soars in its treatment of African scientists whom decolonizing states charged with finding plant-based formulas for national pharmaceutical industries. African scientists, most of them men, relied on the practices and knowledge of healers and plant vendors, many of whom were women. But like their colonial predecessors, African scientists—struggling to forge careers in conditions of extreme inequality—relied on patents to restrict the benefactors of diffuse plant knowledge creation.

What is fair use? Can plants be owned? Is there such a thing as an African science? Or a Ghanaian science? In concise and inviting prose, Osseo-Asare pries open these legal, moral, and political questions, introducing

Hoodia gordonii is used in traditional medicine practices in parts of Africa.

PHOTO: KIM TAYLOR/NPL/MINDEN PICTURES

readers to the varied ways in which African men and women have long shaped our global histories of science, even when we have failed to recognize their contributions.

d'Avignon is the author of *A Ritual Geology: Gold and Subterranean Knowledge in Savanna West Africa* (Duke Univ. Press, 2022), for which she received the HSS's Pfizer Award in 2023.

Bitter Roots: The Search for Healing Plants in Africa, *Abena Dove Osseo-Asare*, University of Chicago Press, 2014, 288 pp.

Errors belatedly discovered in the calculation of the meter's length are still present today.

PHOTO: HJBC/ALAMY STOCK PHOTO

The Measure of All Things

Reviewed by **Joshua Nall**[8]

When John Quincy Adams asked Thomas Jefferson for his views on the metric system, Jefferson highlighted the essential dilemma faced by any state that seeks to change how people measure the world around them: "Shall we mould our citizens to the law, or the law to our citizens?" People accustomed to measures consonant with their local lived experience are typically unenthusiastic about changing the entire basis of their economic life on the imposed whim of a distant government. Yet almost the entire world has acceded to precisely this change. How?

This story, as told by Ken Alder in his 2003 book, *The Measure of All Things*, is much wilder and considerably more fascinating than the history of measurement has any right to be. Determining a new basis for metric standards—specifically, a single unit of measure, the meter—turns out to have been one of the great adventure stories in the history of science.

Its origins are found in that most radical of moments—the French Revolution. Here, in the shadow of the storming of the Bastille, populists and learned *savans* collaborated on a project fit for the new age of reason: the overthrow of myriad old measures by a new, universal length derived neither from regional customs nor the fiat of kings but rather from nature itself. Intended to be a system "for all people, for all time," the meter would use the globe as a common standard, being one ten-millionth of the distance from the North Pole to the equator. All they had to do was measure it.

This task took two eminent astronomers, Jean-Baptiste-Joseph Delambre and Pierre-François-André Méchain, more than 7 years of exacting labor and proved so taxing that it drove the latter to an early grave. The pair's charge required them to triangulate thousands of distant waypoints from the north of France down into Spain during a time of political tumult and almost constant war. Their expeditions were hampered by extreme weather, inhospitable terrain, hyperinflation, epidemics, and the constant threat of violence from locals ever-skeptical of the outsiders and their curious instruments.

Alder discovers at the heart of this story a very human tragedy in the figure of Méchain, a man who grows increasingly wracked with guilt over his cover-up of confounding errors belatedly discovered within his data. This error still resides in the meter that we use today. But, as Alder shows us, the validity of this unit has proven to come not from its purported natural perfection but rather from the epic unrepeatability of its derivation. A measure for all mankind, but one that is singularly man-made.

Nall is the author of *News from Mars: Mass Media and the Forging of a New Astronomy, 1860–1910* (Univ. of Pittsburgh Press, 2019), for which he received the HSS's Philip J. Pauly *Prize in 2020.*

The Measure of All Things: The Seven-Year Odyssey and Hidden Error That Transformed the World, *Ken Alder*, Free Press, 2003, 448 pp.

Mind Fixers

Reviewed by **Michael Robinson**[9]

In 1978, the social critic Martin Gross rejoiced in psychiatry's turn away from psychoanalysis, with its focus on "mind cures," and its new embrace of the biology of the brain. In 10 years, he explained, "psychiatry will come out of the dark ages." Many professionals felt the same. They rejoiced over the new "biological revolution" that seemed—with new pharmaceuticals and diagnostic techniques such as CAT scans and MRIs—to be transforming psychiatry. In their view, psychiatry was returning to its 19th-century roots, when early pioneers such as Theodor Meynert and Emil Kraepelin had placed the biology of the brain at the center of their research programs. The long dominance of Freudian psychiatrists, who vaulted into prominence after World War II, seemed to be over.

Yet this story, Anne Harrington argues in *Mind Fixers*, her superb history of American psychiatry, was a myth. Most of the drugs at the center of this "biological revolution" were not new (Xanax, rockstar drug of the 1990s, had been developed by the pharmaceutical company Upjohn in the 1960s), and the imaging tools of the human brain, while spectacular, were not conclusive about the origins of mental illness. Many of the drugs at the center of this biological revolution, promoted relentlessly by Big Pharma, scarcely worked better than placebos.

This is not the only myth-busting moment in *Mind Fixers*. The early work of biological psychiatrists, Harrington shows, did not yield great results. Patients suffering from schizophrenia and manic depression were not cured by insulin-induced comas, electroconvulsive therapy, or psychosurgeries such as lobotomies. In fact, they usually fared worse. It was the failure of such techniques that led to the decline of biological psychiatry after World War I. And it was World War II that gave new life to Freudian psychiatrists. As millions of American soldiers returned from Europe, and later Korea,

Widespread screening for phenylketonuria implies a triumph over the rare genetic disorder, but the reality is more complicated.

suffering from "war neurosis," "shell shock," and "battle fatigue," how could traumatic experience and its effect on the mind not be taken seriously?

Of course, Freudian psychologists managed to bungle things as well. Analysts were hostage to their own prejudices. Focused on formative experiences, they frequently attached their patients' problems to bad parenting, or more specifically, bad mothering. Homosexuality was the result of overprotective mothers whereas autism was blamed on disengaged ones.

No group has all the answers, Harrington shows. The history of American psychiatry, like the brain itself, is strange, complex, and filled with convolutions.

Robinson is the author of *The Lost White Tribe: Explorers, Scientists, and the Theory That Changed a Continent* (Oxford Univ. Press, 2016), for which he received the HSS's Watson, Helen, Miles, and Audrey Davis Prize in 2019.

Mind Fixers: Psychiatry's Troubled Search for the Biology of Mental Illness, *Anne Harrington,* Norton, 2020, 384 pp.

The PKU Paradox

Reviewed by **Marga Vicedo**[10]

Phenylketonuria (PKU) is a rare genetic disorder. Yet most industrialized countries have mandatory programs to screen newborns for it. In *The PKU Paradox*, historian of science Diane B. Paul and pediatrician Jeffrey P. Brosco ask: What generated such interest in this rare disease? Their engaging and insightful examination of the medical and social history of PKU shows how the disorder was constructed as a model of medical success for genetic testing.

Because the severe cognitive impairments caused by PKU can be prevented when affected infants are placed on a diet low in phenylalanine, the disease became a paradigmatic example to prove that biology is not destiny; that "genetic" does not always mean "unchangeable." But Paul and Brosco argue that therapeutic optimism led to underestimating the difficulties of "fixing" a genetic disorder.

After the development of the Guthrie test—a pinprick puncture of a newborn's heel to collect blood—the push for mandatory PKU screening gained momentum in the mid-1960s, although randomized clinical trials to establish the test's efficacy were lacking. As a result of false positives, some infants suffered ill effects from being put on low-phenylalanine diets. False negatives prevented others from being treated. In addition, it was unclear how long PKU patients needed to stay on the unpalatable and restrictive diet. Children and adolescents found the diet difficult to adhere to, and women with PKU who abandoned the diet and became pregnant discovered that their infants were affected with problems in utero. The expense burdened PKU patients' families as well. Yet no counseling or support was available.

For Paul and Brosco, the PKU story reveals the risks of following the technological imperative. Should we mandate routine screening for a genetic condition just because we have the technology to do so? The answer is not straightforward. Scientific promises sometimes clash with social realities.

The *PKU Paradox* is an important book. Written in 2013, before the discovery of CRISPR and mRNA-based technologies that augur great advances in genomics research and therapeutics, it raises profound questions that concern all of us. History alone cannot tell us how to act in the future, but it can provide valuable guidance as we enter uncharted territories.

Vicedo is the author of *Intelligent Love: The Story of Clara Park, Her Autistic Daughter, and the Myth of the Refrigerator Mother* (Beacon Press, 2021), for which she received the HSS's Watson, *Helen, Miles, and Audrey Davis Prize in 2022.*

The PKU Paradox: A Short History of a Genetic Disease, *Diane B. Paul and Jeffrey P. Brosco*, Johns Hopkins University Press, 2013, 320 pp.

10.1126/science.adq1754

[1]Department of History, Drexel University, Philadelphia, PA, USA. Email: tsaraiva@drexel.edu [2]Department of History, Dickinson College, Carlisle, PA, USA. Email: pawleye@dickinson.edu [3]School of History, Philosophy, and Religion (emerita), Oregon State University, Corvallis, OR, USA. Email: anita.guerrini@oregonstate.edu [4]Department of History, Rutgers University, New Brunswick, NJ, USA. Email: leah.devun@rutgers.edu [5]Department of Spanish, Italian, and Portuguese, University of Virginia, Charlottesville, VA, USA. Email: amb8fk@virginia.edu [6]Department of History, Brown University, Providence, RI, USA. Email: nummedal@brown.edu [7]Department of History, New York University, New York, NY, USA. Email: robyn.davignon@nyu.edu. [8]Department of History and Philosophy of Science, University of Cambridge, Cambridge, UK. Email: jfkn2@cam.ac.uk [9]Department of Humanities, University of Hartford, West Hartford, CT, USA. Email: microbins@hartford.edu [10]Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, Ontario, Canada. Email: marga.vicedo@utoronto.ca

PERSPECTIVES

ATMOSPHERIC CHEMISTRY

Mapping urban haze

Direct airborne measurements of emission fluxes reveal major sources over California

By **Thomas Karl**

Volatile organic compounds (VOCs) drive chemistry in Earth's atmosphere and play an important role in air pollution management. Globally, vegetation is the largest source of atmospheric VOCs (*1*). However, regionally, this biogenic input can vary considerably, and with large differences between the contributions of anthropogenic, pyrogenic, and biogenic emissions. The regional apportionment of VOCs therefore remains challenging. On page 1324 of this issue, Pfannerstill *et al.* (*2*) report on a wide range of VOC emissions through direct airborne observations over California. They show that up to 60% of the VOC reactivity is associated with biogenic emissions. Their observations also allow VOC emissions to be constrained along transects over southern California, providing a much clearer picture of the anthropogenic and biogenic VOC fluxes. The findings have important implications for air quality regulations aimed at managing ozone and secondary aerosol formation in the South Coast Air Basin region and beyond.

Naturally occurring ozone (a molecular oxygen species) in the upper atmosphere (stratosphere) prevents the Sun's ultraviolet radiation from harming life on Earth. However, ozone in the atmospheric layer closest to Earth's surface (troposphere) is formed through the photochemical reaction of nitrogen oxides (NO_x) and VOCs. It has been 60 years since tropospheric ozone formation was closely linked to NO_x and VOC emissions (*3*). The debate as to whether anthropogenic or biogenic VOC emissions dominate ozone formation in different geographical settings has remained unsettled. From a chemical perspective, ozone can be controlled depending on whether an air mass is NO_x-limited or VOC-limited (*4*). In general, the formation of ozone in urban areas tends to be more limited by VOCs, whereas remote areas are more sensitive to NO_x. Ozone abatement strategies are further complicated by regional VOC emissions that may be dominated by either anthropogenic or biogenic sources. Sixty years ago, excessive ozone production in California could clearly be attributed to VOC and NO_x emissions from cars. Strict emission controls have since helped to reduce these emissions from anthropogenic combustion sources. In the 1960s, it was reported that plants can also emit vast amounts of reactive terpenes, fueling tropospheric ozone formation (*5*). This sparked a provocative regulatory question—do trees pollute more than cars? It was the starting point of rethinking the VOC component of air pollution management.

With the decline of anthropogenic VOCs released into the atmosphere (*6*), attention shifted toward managing NO_x emissions, which tends to be technologically more difficult in VOC-sensitive regions (*7*). Recent observations in California have suggested that a new class of anthropogenic VOCs called volatile chemical products (VCPs) may play an important role in generating ozone and secondary organic aerosols (SOAs), particulate matter that forms through the oxidation of VOCs. This could outweigh the importance of biogenic VOCs in populated areas (*8*), a possibility that has triggered debate on the origin of important reactive VOC classes (such as monoterpenes and sesquiterpenes). These hydrocarbons have been mostly associated with biogenic emissions.

Airborne flux measurements by Pfannerstill *et al.* over southern California allow previous local-scale observations to be extended to scales relevant for regional air quality modeling and have settled this debate. Previous studies on urban VOC emissions have relied on indirect methods (such as emission inventories), whereas Pfannerstill *et al.* performed airborne eddy covariance measurements to directly observe atmospheric fluxes. Mass spectrometry was used to identify more than 400 VOC species. The authors found that despite the semi-arid land cover and the emergence of VCPs, VOCs emitted from vegetation still dominate the reactivity of VOCs over the entire South Coast Air Basin region. Pfannerstill *et al.* also estimated that up to 60% of the ozone and SOA formation potentially originates from biogenic VOCs. Moreover, they showed that warmer temperatures lead to an increase in air mass reactivity of ~30% per degree Celsius, largely caused by the increase in VOC emissions from vegetation.

The earliest descriptions of vegetation in southern California date back to those of Spanish explorers in the 1500s. Such explorations described shrubs and giant cactus in some areas, but also lush grass and green mountains (*9*). Some 480 years later, land-surface characteristics can be mapped from space at decameter resolution, and there is no longer any ambiguity about land cover. However, from an air quality perspective, the diversity and genetic differences between plants species still make it difficult to assign standardized emission potentials of VOCs to different land-cover types and plant functional types. Furthermore, the emergence of urban VCP emissions (*10, 11*) suggests previously underestimated anthropogenic emission sources. The mapping of VOC emissions from airborne platforms, as performed by Pfannerstill *et al.*, opens a new frontier for the improvement and verification of spatially resolved emission maps, which form the backbone of any meaningful air quality prediction model system. Constraining the total emission flux of VOCs to the atmosphere remains a major challenge. Technological improvements in mass spectrometry have made it possible to show that thousands of VOCs are emitted to the atmosphere across the urban-rural interface. Although controlled studies on vehicles, smokestacks, and some industrial processes have been quite successful in providing VOC emission factors for the extrapolation of emissions in bottom-up approaches, the emergence of VCPs together with VOC emissions from vegetation complicates the issue considerably. This may require additional complementary approaches, similar to the flux observations presented by Pfannerstill *et al.*, who assessed VOC emission processes on a regional scale, helping to improve future air chemistry and climate models. ■

Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria. Email: thomas.karl@uibk.ac.at

REFERENCES AND NOTES

1. A. B. Guenther *et al.*, *Geosci. Model Dev.* **5**, 1471 (2012).
2. E. Y. Pfannerstill *et al.*, *Science* **384**, 1324 (2024).
3. A. J. Haagen-Smit, *Ind. Eng. Chem.* **6**, 1342 (1952).
4. S. Silmann *et al.*, *Annu. Rev. Energy Environ.* **18**, 31 (1993).
5. R. A. Rasmussen, F. W. Went, *Proc. Natl. Acad. Sci. U.S.A.* **53**, 215 (1965).
6. C. Warneke *et al.*, *J. Geophys. Res.* **117**, D00V17 (2012).
7. A. Neftel *et al.*, *J. Geophys. Res.* **117**, 8188 (2002).
8. M. M. Coggon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026653118 (2021).
9. H. R. Wagner, F. de Ulloa, *Calif. Hist. Soc. Q.* **3**, 307 (1924).
10. T. Karl *et al. Proc. Natl. Acad. Sci. U.S.A.* **115**, 1186 (2018).
11. B. C. McDonald *et al.*, *Science* **359**, 760 (2018).

10.1126/science.adq1423

CELL BIOLOGY

Cellular senescence in normal physiology

Long associated with aging, senescent cells can promote health and have physiological roles

By **João Pedro de Magalhães**

The aging of the world's population has intensified interest in understanding the aging process and devising strategies and interventions to prolong a healthy life span. Cellular senescence, when cells become irreversibly growth arrested after a period of in vitro cell proliferation or in response to sublethal stress or oncogene expression (*1*, *2*), plays a role in aging phenotypes and age-associated diseases (*1*). Increasing evidence shows that senescent cells also have essential physiological functions, such as in tumor suppression, development, wound healing, tissue remodeling, regeneration, and vasculature. This raises important questions about the similarities and differences between senescent cell types and how they function in homeostasis and pathology, and it creates additional challenges in targeting them therapeutically.

Despite the importance of cellular senescence in tumor suppression and aging phenotypes, its precise definition is still debated. Furthermore, there is no single biomarker of senescence but rather several markers—including growth arrest, senescence-associated β-galactosidase activity, telomere-associated DNA damage, and expression of the cell cycle inhibitors p16 and p21—that have been used to identify and sometimes quantify senescent cells both in vitro and in vivo (*1*, *2*). Senescent cells have been found to secrete proinflammatory cytokines and other factors, a characteristic called the senescence-associated secretory phenotype (SASP), that may disrupt tissue homeostasis and contribute to a proinflammatory state (*1*). Indeed, a seminal study demonstrated that genetically eliminating senescent cells, using an inducible system that triggers apoptosis in p16-expressing cells, ameliorated signs of aging and extended the median life span of mice by 24 to 27% (*3*). Thus, the study and development of senolytics—drugs that selectively kill senescent cells—has gained traction in recent years (*1*).

Although several studies in mouse models support the hypothesis that senescent cells can trigger or contribute to age-associated phenotypes, more recent studies have revealed additional roles for senescent cells in nonharmful and even physiological processes. Indeed, eliminating senescent cells in mice can be detrimental to health, highlighting the importance of these cells in mammalian homeostasis and physiology. For example, senescent cells become more prevalent with age, particularly in the liver, and are often vascular endothelial cells. The continuous or acute removal of these senescent cells in mice disrupted blood-tissue barriers and led to the buildup of blood-borne macromolecular waste, resulting in perivascular fibrosis in a variety of tissues and subsequent health deterioration (*4*). These results underscore the functional and structural roles that senescent cells play in tissues that are important for organismal health.

Another recent study using p16 as a marker identified in young mice a population of senescent fibroblasts in the basement membrane adjacent to epithelial stem cells of the lungs that monitor barrier integrity and respond to tissue inflammation to promote epithelial regeneration (*5*). A different study detected senescence in multiple cell types of the lungs of neonatal mice and demonstrated that reducing senescent cells either pharmacologically or genetically (by p21 deletion) disrupted lung development (*6*). These findings suggest that programmed senescence, which is apparently not induced as a damage response but rather associated with developmental processes, is crucial for lung development in mice. However, reduction of senescent cells once lungs were developed in 7-day-old mice attenuated hyperoxia-induced lung injury, suggesting that senescent cells can also limit tissue repair (*6*). Paradoxically, eliminating senescent cells might exacerbate pulmonary hypertension in mice, despite the increased presence of senescence markers in the lungs of both mice and patients with hypertension (*7*). Notably, a study using p16 transgenic mice also found a role for p16 and cellular senescence in promoting insulin secretion by pancreatic beta cells, which was also observed in human cells, suggesting that cellular senescence can in some contexts enhance cellular function (*8*).

Studies of senescence in vivo often use senescence-ablator mice, which are genetically engineered to allow for the selective elimination of senescent cells, typically by targeting cells expressing p16. One important consideration is that many of these mouse models use different constructs to target p16 (*3*, *4*). Different constructs can lead to differences in the rate and effectiveness of senescent cell elimination. As such, technical disparities might explain inconsistencies in some results and further suggest that certain proportions or types of senescent cells within a tissue could be healthy whereas other senescent cell types and amounts might be pathologic. It is also important to note that even though p16 is the most widely used marker in these experiments, it is not a universal marker of cell senescence, and senescent cells can be observed independently of p16 expression (*9*).

Although cellular senescence has been observed during development in mice and other organisms (*1*, *2*), its functions in development are still being unveiled. Senescent cells and activation of senescence signaling have been observed in human placentas and were down-regulated in placentas from pregnancies with intrauterine growth restriction (*9*). Interestingly, senescence-deficient mice, owing to deletion of genes encoding key senescence-associated signaling proteins (p16, p21, and p53), exhibited morphological aberrations in the placenta, suggesting that cellular senescence and the underlying signaling pathways regulate placental structure and function (*9*). Additionally, senescent cells in mice promote hair growth, and clusters of senescent cells appear to enhance the activity of adjacent stem cells to stimulate tissue renewal (*10*). Thus, damage-independent programmed senescence serves various functions during the development of different mammalian tissues, and the precise functions of these senescent cells require further investigation.

The role of senescence in wound healing and repair has been recognized for several years, and it has been suggested that senescent cells, probably by means of the SASP, recruit immune cells to promote tissue repair (*2*). However, a broader picture is emerging regarding the role of cellular senescence in regeneration and tissue repair across different animals. In an invertebrate cnidarian (*Hydractinia symbiolongicarpus*), senescent cells can induce the reprogramming of neighboring somatic cells into stem cells, driving whole-body regeneration (*11*). Genetic or pharmacological inhibition of cellular senescence prevented reprogramming and regeneration, suggesting that senescent cells signal

Genomics of Ageing and Rejuvenation Lab, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK. Email: jp@senescence.info

to nearby cells at an injury site to prepare for regeneration (*11*). Similarly, in newts (*Notophthalmus viridescens*), cellular senescence enhances limb regeneration, particularly by promoting muscle dedifferentiation and generating regenerative progenitor cells (*12*). In zebrafish, senescent cells appear after fin amputation, and their pharmacological removal impairs tissue regeneration (*13*). Although the relevance of these observations to mammals is unclear, several studies in mice have also demonstrated that senescent cells are important for tissue regeneration. For example, senescence of hepatic stellate cells is induced after liver injury, and pharmacological or genetic ablation of these senescent cells impairs liver regeneration (*14*).

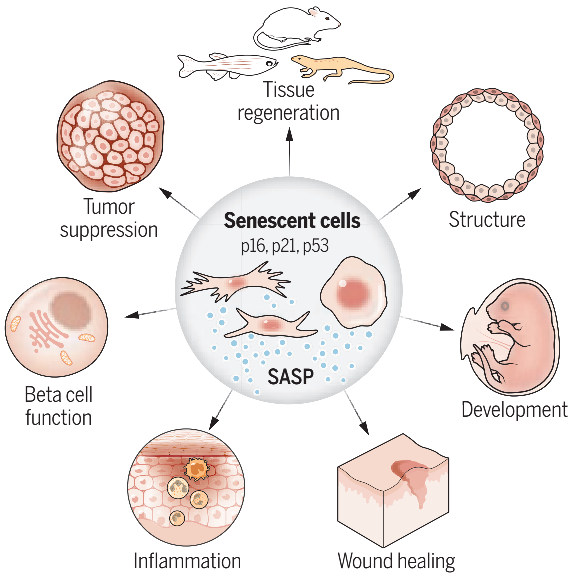
Cellular senescence is an adaptive process; senescent cells are not aging, dysfunctional cells, they are functional cells that are important for several physiological processes (see the figure). Unfortunately, the word "senescence" implies a detrimental or nonfunctional role that is no longer accurate (*15*). Although the lack of universal senescence markers and of a clear definition of cellular senescence poses challenges, it is unquestionable that cells exhibiting senescence-associated markers play crucial roles in various normal physiological functions and in maintaining tissue homeostasis. Although damage and stress can induce cellular senescence, perhaps to recruit immune cells through the SASP and promote tissue repair and remodeling, cellular senescence can also arise independently of molecular damage or injury, for example, during development. Furthermore, senescence induced by injury can encourage regeneration and wound healing, and the degree of senescent cell involvement in the regeneration of different tissues is an exciting avenue for future research. Although a role for senescent cells in aging has been suggested by many studies (*1*–*3*), the recent findings that demonstrate normal physiological functions of senescent cells reveal a more complicated picture of the potential role of cellular senescence in mammalian aging.

A few limitations should be acknowledged. A growing number of pharmacological interventions, namely senolytics, have been shown to have health benefits in aging mice (*1*). But these drugs have off-target effects, and hence genetic manipulations, as has been the focus here, might offer a more precise way to test the role of senescent cells in health, aging, and disease. Nonetheless, several studies in mice have used both genetic and pharmacological methods to target senescent cells and shown that their elimination can impair normal physiology, such as organ development and tissue regeneration (*6*, *14*).

Another important caveat is that most in vivo studies of cellular senescence have been conducted in mice, and its role in humans remains poorly understood. Senescent cells have been observed in the context of human age-associated diseases (*2*), often as part of inflammatory responses. However, whether senescent cells are protective or harmful (or both) in human pathologies remains to be established. Just as inflammation promotes tissue repair but may inadvertently contribute to tissue dysfunction, senescence is also likely pleiotropic and may be detrimental or harmful in humans, depending on context. Perhaps some types of senescent cells prevent tissue degeneration, by promoting regeneration as well as maintaining structure and function, thereby contributing to health maintenance, whereas other types—or an excessive amount—contribute to degeneration. Notably, not all senescent cells are the same. They can have different inducers, express different markers, and originate from different tissues and cell types; our understanding of the different types of senescent cells is still very limited. Many questions remain regarding the temporal dynamics of different types of senescent cells, their functions, and their clearance by the immune system.

Functions of senescent cells

Recent evidence from mice and other model organisms has revealed that cellular senescence is important during the development of several tissues and organs, tissue regeneration in several animals, inflammation and wound healing, tumor suppression, insulin secretion in pancreatic beta cells, and has structural roles in the vascular system and placenta. There can be different types of senescent cells, which have a senescence-associated secretory phenotype (SASP), and are associated with the expression of markers such as p16 and p21.



Maybe short-term, transient cellular senescence is beneficial whereas long-term, chronic senescence becomes detrimental. Indeed, senescent cells are not only heterogeneous but perhaps dynamic, shifting with time and with changes in the tissue microenvironment. More research is needed to discriminate between healthy and pathologic senescent cells, and there are ongoing projects, such as the Cellular Senescence Network (SenNet; https://sennetconsortium.org/), that aim to identify, classify, and characterize senescent cells in humans and model organisms to improve understanding.

The growing number of physiological roles of senescent cells raises challenges concerning the development of senolytics and other senotherapeutics that target them. Perhaps removing certain types of senescent cells in some tissues is beneficial, whereas removing other types of senescent cells or too many of them is detrimental. As such, effective therapies will need the precision—at a certain dosage and administration—to eliminate pathologic senescent cells while sparing healthy senescent cells. Or perhaps modulators of cellular senescence, which alter the function of senescent cells without necessarily eliminating them, need to be explored. Thus, the development of aging therapies centered on senescent cell targeting remains a promising avenue but is likely to be more complex than initially anticipated. ■

GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. J. P. de Magalhães, J. F. Passos, *Mech. Ageing Dev*. **170**, 2 (2018).
2. D. Muñoz-Espín, M. Serrano, *Nat. Rev. Mol. Cell Biol*. **15**, 482 (2014).
3. D. J. Baker *et al*., *Nature* **530**, 184 (2016).
4. L. Grosse *et al*., *Cell Metab*. **32**, 87 (2020).
5. N. S. Reyes *et al*., *Science* **378**, 192 (2022).
6. H. Yao *et al*., *Nat. Commun*. **14**, 273 (2023).
7. E. Born *et al*., *Circulation* **147**, 650 (2023).
8. A. Helman *et al*., *Nat. Med*. **22**, 412 (2016).
9. H. Gal *et al*., *EMBO J*. **38**, e100849 (2019).
10. X. Wang *et al*., *Nature* **618**, 808 (2023).
11. M. Salinas-Saavedra *et al*., *Cell Rep*. **42**, 112687 (2023).
12. H. E. Walters, K. E. Troyanovskiy, A. M. Graf, M. H. Yun, *Aging Cell* **22**, e13826 (2023).
13. S. Da Silva-Álvarez *et al*., *Aging Cell* **19**, e13052 (2020).
14. N. Cheng, K.-H. Kim, L. F. Lau, *JCI Insight* **7**, e158207 (2022).
15. D. Gems, C. C. Kern, *Geroscience* **44**, 2461 (2022).

ACKNOWLEDGMENTS

J.P.d.M. thanks lab members, particularly D. Ribeiro and R. Avelar. J.P.d.M. used AI-assisted tools to prepare this Perspective (see supplementary materials). J.P.d.M. is chief scientific officer of YouthBio Therapeutics; an adviser and/or consultant for the Longevity Vision Fund, 199 Biotechnologies, and NOVOS; and the founder of Magellan Science Ltd.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj7050

10.1126/science.adj7050

CANCER

An unexpected corridor to brain metastasis

Breast cancer cells migrate from the bone marrow to the leptomeninges

By **Lea Monteran** and **Neta Erez**

The brain is a main site of breast cancer metastasis, which inflicts a grim prognosis on patients (*1*). Leptomeningeal metastases are a fatal form of brain metastasis, with a median survival of less than 6 months after diagnosis (*2*). This kind of metastasis results from the spread of cancer cells to the membranous layers filled with cerebrospinal fluid (CSF) that surround the brain and spinal cord. However, the mechanisms involved in cancer spread to the leptomeninges remain unclear. On page 1317 of this issue, Whiteley *et al.* (*3*) report the characterization of a previously unidentified route by which breast cancer cells (BCCs) metastasize to the leptomeninges in mice, migrating from vertebral metastasis along the external surface of blood vessels. The findings elucidate how BCCs hijack migratory pathways normally used by neuronal cells and could inform future therapies.

Leptomeningeal metastases are diagnosed in up to 10% of patients with solid tumors, including breast cancer (*4*), but the anatomical routes used by cancer cells to invade this harsh microenvironment remain largely unknown. The two main paths for cancer cell dissemination are via the bloodstream or the lymphatic system. In the brain, it is generally accepted that metastatic dissemination involves breaching the blood-brain barrier (BBB) or the blood-CSF barrier to colonize the leptomeninges.

To better understand leptomeningeal metastasis, Whiteley *et al.* used mouse models to generate a variant of BCCs that target the leptomeninges and found that these cells, but not the parental BCCs, highly express the cell adhesion protein integrin α6. This expression was important for BCC migration in vivo, which was mediated by the engagement of integrin α6 with laminin present on the external surface of emissary blood vessels that connect the bone marrow to the leptomeninges. This suggests that metastases in the central nervous system (CNS) follow two distinct routes: via blood vessels, crossing the BBB to the parenchyma, or through vertebral metastatic lesions in the bone, crawling into the spinal canal and leptomeninges by adhesion-mediated migration on emissary vasculature, directly connecting the bone marrow to the leptomeninges (see the figure).

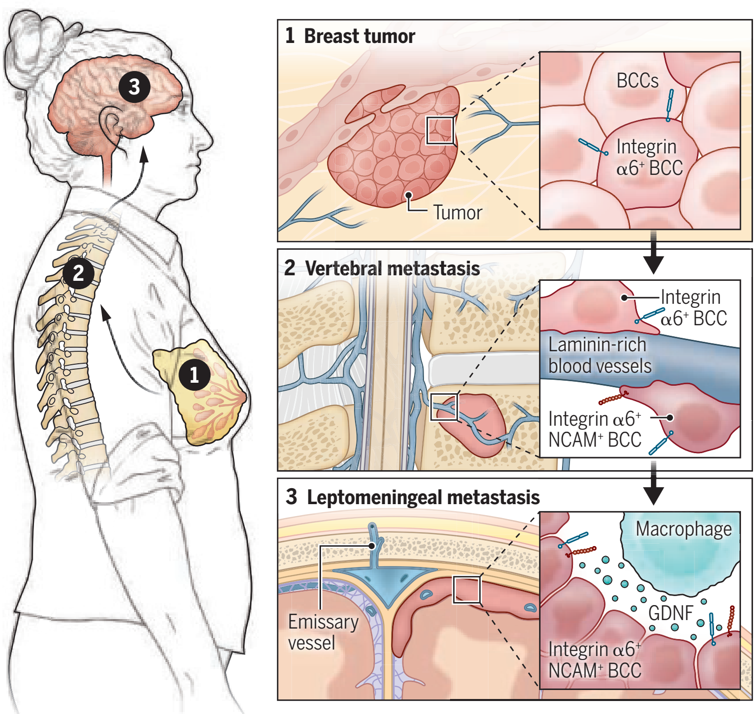
The integrin α6–laminin axis was previously implicated in the migration of neural stem cells (*5*) and in the invasion of acute lymphoblastic leukemia (ALL) cells from the bone marrow into the leptomeninges (*6*). Furthermore, recent studies reported lymphatic vasculature in the brain and connections of bone marrow in the skull with the meninges, which highlights the interactions between the CSF and brain immune surveillance (*7*–*9*). The skull and vertebral bone marrow were also recently described as reservoirs of monocytes and neutrophils for the CNS, which can be mobilized upon spinal cord injury or neuroinflammation (*10*). The study of Whiteley *et al.* suggests that BCCs hijack these physiological pathways that enable migration of neuronal and immune cells from the bone marrow into the leptomeninges.

Once in the leptomeninges, cancer cells must survive in the nutrient-poor CSF. Several cancer cell–derived factors were previously implicated in this process. For example, expression of complement component 3 by BCCs and lung cancer cells promoted disruption of the blood-CSF barrier to allow growth-promoting factors from the plasma to enter the CNS and support cancer cell growth (*11*). Leptomeningeal metastatic subpopulations of lung and breast carcinomas can also express lipocalin 2, an iron chelator, enabling them to collect essential iron in this hypoxic microenvironment (*12*). Using gene expression analyses, Whiteley *et al.* found that growth of BCCs in the leptomeninges involves increased expression of neural cell adhesion molecule (NCAM).

NCAM is normally expressed by neurons, and its ligand, glial cell line–derived neurotrophic factor (GDNF), is secreted by reactive microglia and macrophages in response to brain tissue injury, blocking apoptotic stress responses (*13*). Whiteley *et al.* show that GDNF-expressing macrophages are present next to BCCs in the leptomeninges and that blocking the NCAM-GDNF axis inhibits cancer cell growth. Notably, analyses of biopsies from breast cancer

Cancer spread to the leptomeninges

Breast cancer cells (BCCs) hijack neural and innate immunity signaling to reach the leptomeninges. BCCs expressing integrin α6 exit bone metastatic lesions and enter the leptomeninges by migrating on the external side of laminin-rich emissary blood vessels. Upon entering the leptomeningeal space, a subpopulation of BCCs expressing NCAM survive and thrive in this hostile microenvironment by hijacking growth-promoting signals from microglia and macrophages that express GDNF.



GDNF, glial cell line–derived neurotrophic factor; NCAM, neural cell adhesion molecule.

GRAPHIC: N. BURGESS/*SCIENCE*

Department of Pathology, Faculty of Medical & Health Sciences, Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel.
Email: netaerez@tauex.tau.ac.il

patients revealed that integrin α6 and NCAM were expressed in BCCs in leptomeningeal metastatic lesions, which suggests that these mechanisms are important for human leptomeninges metastasis. Previous studies have found increased amounts of GDNF in the CSF of patients with leptomeningeal metastasis (*11*), suggesting that BCCs that express NCAM exploit a physiological neuroprotective signaling pathway to enable their survival in the harsh leptomeningeal niche.

An intriguing hypothesis is that BCC dissemination through migration from the bone along the external surface of blood vessels may be generalized to other metastatic scenarios. The Batson venous plexus is a network of veins that drain several internal organs to the vertebral venous plexus, connecting the breasts, prostate, vertebrae, and brain. The Batson plexus has long been proposed to be a metastatic conduit (*14*), and the findings of Whiteley *et al.* suggest that BCCs may use similar mechanisms along this route to metastasize from the breasts to bone. Prostate cancer also leads to bone metastasis, further supporting this possibility.

The study by Whiteley *et al.* highlights the importance of choosing adequate models because some mouse models are useful to study distinct brain metastatic destinations, but not others. Indeed, the tropism to leptomeninges or parenchyma was shown to be an inherent feature of specific lung, breast, and melanoma cancer cell lines (*15*). The findings also underscore the importance of accurately characterizing niche-specific microenvironments within the brain and the specific pathways used by cancer cells to manipulate them, which may inform accurate treatment strategies. Future studies should explore whether these findings can be used for the treatment of patients with leptomeningeal metastasis, providing potential new targets to constrain the entry and survival of BCCs in the leptomeninges. ■

REFERENCES AND NOTES

1. M. Valiente *et al.*, *Trends Cancer* **4**, 176 (2018).
2. M. E. Freret, A. Boire, *J. Exp. Med.* **221**, e20212121 (2024).
3. A. E. Whiteley *et al.*, *Science* **384**, eadh5548 (2024).
4. J. V. Brower *et al.*, *J. Clin. Neurosci.* **27**, 130 (2016).
5. L. A. Flanagan, L. M. Rebaza, S. Derzic, P. H. Schwartz, E. S. Monuki, *J. Neurosci. Res.* **83**, 845 (2006).
6. H. Yao *et al.*, *Nature* **560**, 55 (2018).
7. F. Herisson *et al.*, *Nat. Neurosci.* **21**, 1209 (2018).
8. J. A. Mazzitelli *et al.*, *Nat. Neurosci.* **26**, 2052 (2023).
9. J. Rustenhoven, J. Kipnis, *Nature* **612**, 417 (2022).
10. A. Cugurra *et al.*, *Science* **373**, eabf7844 (2021).
11. A. Boire *et al.*, *Cell* **168**, 1101 (2017).
12. Y. Chi *et al.*, *Science* **369**, 276 (2020).
13. E. P. Duarte *et al.*, *Growth Factors* **30**, 242 (2012).
14. O. V. Batson, *Ann. Surg.* **112**, 138 (1940).
15. M. Valiente *et al.*, *Cancer Res.* **80**, 4314 (2020).

ACKNOWLEDGMENTS

N.E. is supported by the Israel Science Foundation, the Department of Defense Breast Cancer Research Program Breakthrough Award, the Israel Cancer Research Fund (ICRF project grant), and the Melanoma Research Alliance.

10.1126/science.adq4019

CANCER

JAKing up immunity

Janus kinase (JAK) inhibitors improve antitumor responses

By **Massimo Gadina**[1] and **John J. O'Shea**[2]

The understanding of the immune system has increased considerably in the past 30 years, leading to the development of better therapies for cancer as well as immune-mediated disorders. These advances include unleashing T cell antitumor activity with monoclonal antibodies targeting immune checkpoints and Janus kinase inhibitors (JAKis or jakinibs) that block cytokine signaling for autoimmune, atopic, and inflammatory pathologies (*1*, *2*). On pages 1315 and 1314 of this issue, Zak *et al.* (*3*) and Mathew *et al.* (*4*), respectively, report that combining these therapeutic approaches results in improved clinical responses compared with immune checkpoint inhibition alone in patients with relapsed or refractory Hodgkin lymphoma as well as metastatic non–small cell lung cancer (NSCLC). These are exciting advances, but at the same time the strategy might seem paradoxical. Why would jakinibs enhance immune-mediated elimination of cancer cells given that they limit lymphocyte activation and proliferation?

Along with the discovery of the numerous factors that mediate host defense and control cancer came the realization that the immune system comprises many drivers and brakes to fine-tune the response. Countering lymphocyte activation are immune checkpoints, which are intrinsic inhibitory factors that limit immune-mediated damage. Immune checkpoint inhibitors (ICIs), including antibodies against programmed cell death protein 1 (PD-1), PD-1 ligand 1 (PD-L1), and cytotoxic T lymphocyte–associated antigen 4 (CTLA4), release T cell inhibition so they can recognize and be activated by tumor-specific antigens. ICIs have substantially improved survival and permit long-lasting responses for multiple malignancies.

However, not all patients respond to ICIs. Resistance can be mediated by a variety of factors, including an immunosuppressive environment often driven by T regulatory cells or myeloid-derived suppressor cells (MDSCs) that prevent T cell activation (*5*). Interferons (IFNs) are cytokines with important antiviral and antitumor activity, which activate JAK signaling (*6*). One mechanism by which tumors escape immunosurveillance is JAK deletion (*7*, *8*). Despite their positive role in controlling tumors, IFNs also limit immune responses and induce the expression of immune checkpoints (*9*). Paradoxically, the chronic inflammatory status of the tumor microenvironment results in increased expression of IFN-stimulated genes, which also contributes to immunotherapy resistance. Specifically, persistent IFN signaling sustains progressive $CD8^+$ T cell differentiation toward a terminal-exhausted status, which is associated with ICI failure. Similar to tumors, chronic viral infections also result in T cell exhaustion driven by IFN signaling. Indeed, blocking type I IFNs improved antiviral responses in chronic lymphocytic choriomeningitis virus (LCMV) infection (*10*–*12*).

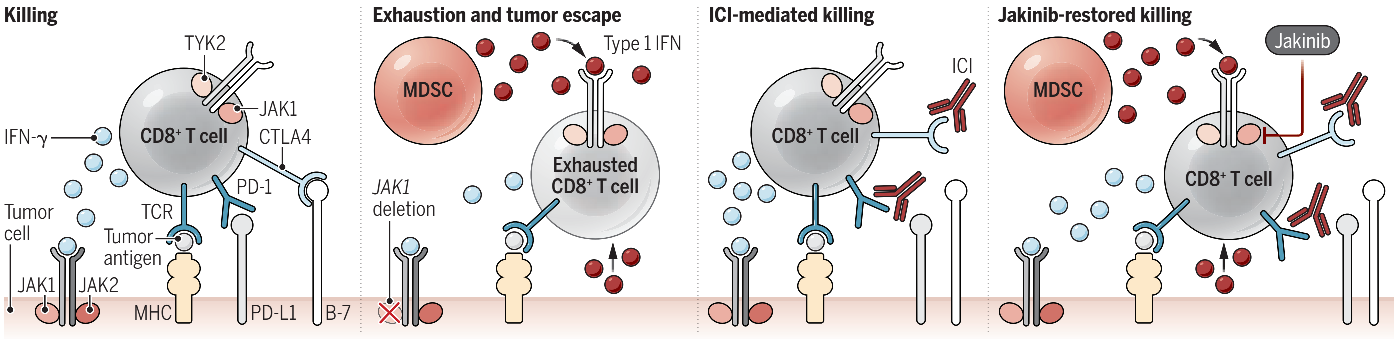
Zak *et al.* searched a collection of small molecules that rescue T cell exhaustion associated with chronic LCMV infection. Unbiased drug screening highlighted a role for jakinibs. After in vitro validation, ruxolitinib, a drug approved for the treatment of myeloproliferative neoplasms, was selected for in vivo experiments in mice infected with LCMV; ruxolitinib administration resulted in increased numbers of antigen-specific $CD8^+$ T cells. Notably, jakinib treatment did not adversely affect viral load, which was instead seen if an IFN-α/β receptor (IFNAR)–specific antibody was given, and which also increased the numbers of $CD8^+$ T cells.

Zak *et al.* and Mathew *et al.* used a jakinib plus an ICI to treat tumor-bearing mouse models, and the combination significantly reduced tumor growth. They found that jakinib addition expanded $CD8^+$ T cells with expression of Ki67 (a marker of proliferation), features of memory precursors or progenitor cells, and reduced markers of exhaustion (see the figure). Notably, the infiltration of granulocytes (such as neutrophils and MDSCs) was reduced in tumors of mice receiving ICI plus jakinib. The myeloid cells present in the tumors exhibited reduced expression of MDSC-associated markers and increased expression of major histocompatibility complex class II (MHCII), promoting antigen presentation to T cells. Consistent with this evidence of improved antitumor immune responses, tumor-infiltrating T cells showed increased

[1]Translational Immunology Section, Office of Science and Technology, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), National Institutes of Health (NIH), Bethesda, MD, USA. [2]Lymphocyte Cell Biology Section, Molecular Immunology and Inflammation Branch, NIAMS, NIH, Bethesda, MD, USA. Email: gadinama@nih.gov; john.oshea@nih.gov

Restoring responses to immune checkpoint inhibitors

Tumor-specific CD8+ T cells recognize tumor antigens and secrete IFN-γ, which activates JAKs. IFNs and T cell activation also induce the expression of immune checkpoints, PD-L1 and CTLA4. Chronic stimulation by tumor antigens results in CD8+ T cell exhaustion, which is enhanced by type I IFNs secreted by MDSCs and the tumor. ICIs promote tumor cell killing, but some patients do not respond. The addition of a jakinib with an ICI limits the immunosuppressive effects of IFNs and enhances antitumor responses.



CTLA4, cytotoxic T lymphocyte–associated antigen 4; ICIs, immune checkpoint inhibitors; IFN-γ, interferon-γ; jakinib, JAK inhibitor; JAKs, Janus kinases; MDSCs, myeloid-derived suppressor cells; MHC, major histocompatibility complex; PD-1, programmed cell death protein 1; PD-L1, programmed cell death 1 ligand 1; TCR, T cell receptor.

degranulation and increased expression of CD44 (a marker of lymphocyte activation) when a jakinib was added to ICI treatment.

These preclinical results catalyzed the design of clinical trials in patients that were either treatment-naïve or refractory to ICI therapy. Zak *et al.* assessed the efficacy of ruxolitinib, a JAK1 and JAK2 inhibitor, plus the ICI nivolumab, which targets PD-1, in a phase 1 open-label trial of 19 patients with relapsing-refractory Hodgkin lymphoma who had previously received ICI and were unresponsive or showed a mixed response. In this trial, overall survival was 87% at 2 years compared to previous reports of 23.8% with ICI alone (*13*). Mathew *et al.* undertook a phase 2 clinical trial of a different jakinib (itacitinib, which targets JAK1) with a different ICI (pembrolizumab, which also targets PD-1) in 21 patients with treatment-naïve NSCLC. In this trial, median progression-free survival was nearly 2 years, compared to the 6.5 to 10.3 months observed in other trials with only ICI (*14*). In both studies, the addition of the jakinib recapitulated the positive preclinical findings.

An important difference between the two trials was the timing of jakinib administration. Zak *et al.* gave ruxolitinib for 8 days before the start of ICI therapy, whereas Mathew *et al.* started itacitinib plus pembrolizumab after two 3-week cycles of pembrolizumab alone. This was limited to two cycles, after which patients were maintained on pembrolizumab only. Despite the differences in trial design and the choices of jakinib and ICI, the results were positive. The addition of ruxolitinib decreased the numbers of myeloid progenitors and monocytes and hence reduced the number of MDSCs, and increased the numbers of cytokine-producing CD8+ T cells in Hodgkin lymphomas. Notably, complete responders had greater reduction of neutrophil/lymphocyte ratio and reduced percentage of mononuclear cells in the periphery.

Similarly, the addition of itacitinib caused a proliferative burst in CD8+ T cells that did not become terminally exhausted. Other important findings reported by Mathew *et al.* were increased T cell plasticity and decreased CD8 clonal diversity, suggesting that the expansion of these T cells was antigen-driven rather than simply homeostatic proliferation. The effects of itacitinib addition were noteworthy in terms of decreased expression of type I IFN-stimulated genes leading to improved immune functions and antitumor responses.

Back to the paradox. Tumors in which *JAK1* is deleted highlight the importance of IFN-γ in eliminating tumors (*7*). Thus, reversing T cell exhaustion is critical, but blockade of IFN-γ antitumor activity needs to be avoided—timing matters. Additionally, the choice of jakinib and its selectivity could be an important factor. There are four JAKs, which are responsible for signaling by 57 cytokines (*2*). Ruxolitinib inhibits JAK1 and JAK2 and targets the signaling from a large number of cytokines. Itacitinib is relatively JAK1 selective and therefore inhibits the signaling from a more restricted number of cytokines but would block type 1 and type 2 IFNs. In principle, extended use of ruxolitinib or itacitinib could limit lymphocyte expansion and activation. If the goal is to inhibit just type 1 IFNs, deucravacitinib might be desirable as it is a selective inhibitor of TYK2, a member of the JAK family that mediates type 1 IFN signaling. However, in the studies of Zak *et al.* and Mathew *et al.*, the beneficial effects of jakinibs seem to be independent of duration and selectivity of jakinib. Moreover, both trials showed that temporary JAK inhibition allowed immune responses to reset, so the paradox remains.

Beyond jakinibs, targeting JAKs with small interfering RNAs (siRNAs) has now been reported (*15*). It would be interesting to know whether these modified therapeutic oligonucleotides will hold an advantage over small molecules in augmenting ICI efficacy.

Aside from the exciting findings of the early phase trials reported by Zak *et al.* and Mathew *et al.*, they provide a great deal of data with complex analyses of immune responses. The trials that led to the approval of jakinibs for autoimmune or atopic disease in the early 2010s were not based on complex biomarkers, largely because the technology used in the current studies was not widely available. Will this bounty of new information permit even better adjustment of dose, context, and timing? Such a deep assessment of the immune response may also allow the mechanisms underlying jakinib treatment failure in cancer as well as autoimmune pathologies to be deciphered. It will be exciting to see how such sophisticated data might be used in the clinic and to inform research. ■

REFERENCES AND NOTES

1. K. Chamoto, T. Yaguchi, M. Tajima, T. Honjo, *Nat. Rev. Immunol.* **23**, 682 (2023).
2. R. L. Philips *et al.*, *Cell* **185**, 3857 (2022).
3. J. Zak *et al.*, *Science* **384**, eade8520 (2024).
4. D. Mathew *et al.*, *Science* **384**, eadf1329 (2024).
5. C. Meyer *et al.*, *Cancer Immunol. Immunother.* **63**, 247 (2014).
6. A. V. Villarino, M. Gadina, J. J. O'Shea, Y. Kanno, *Cell* **181**, 1696 (2020).
7. J. M. Zaretsky *et al.*, *N. Engl. J. Med.* **375**, 819 (2016).
8. E. Alspach, D. M. Lussier, R. D. Schreiber, *Cold Spring Harb. Perspect. Biol.* **11**, a028480 (2019).
9. L. M. McLane, M. S. Abdel-Hakeem, E. J. Wherry, *Annu. Rev. Immunol.* **37**, 457 (2019).
10. E. B. Wilson *et al.*, *Science* **340**, 202 (2013).
11. R. Baccala *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8925 (2014).
12. J. R. Teijaro *et al.*, *Science* **340**, 207 (2013).
13. R. D. Schouten *et al.*, *Transl. Lung Cancer Res.* **9**, 1736 (2020).
14. T. S. K. Mok *et al.*, *Lancet* **393**, 1819 (2019).
15. Q. Tang *et al.*, *Mol. Ther. Nucleic Acids* **35**, 102117 (2024).

ACKNOWLEDGMENTS

The authors acknowledge Y. Kanno and J. Shayne for discussion and helpful comments. J.J.O'S. receives royalty income from US Patent for JAK inhibitors.

10.1126/science.adq1717

RETROSPECTIVE

Daniel C. Dennett (1942–2024)

Philosopher and science advocate

By **Rafael Yuste**[1] *and* **Michael Levin**[2]

Daniel Clement Dennett, who merged philosophical thinking with scientific evidence from many fields, died on 19 April at age 82. In the central question of the nature of the mind, Dennett took a neuroscience-anchored position, arguing that consciousness and all higher cognitive abilities could be understood as a direct consequence of the physiology of the brain. He expanded his ideas in best-selling books about religion and evolution and became one of the strongest voices publicly defending atheism and Darwinism.

Born in Boston on 28 March 1942 to a diplomat and a professor, Dennett spent his childhood years in Beirut. He graduated with a BA in philosophy in 1963 from Harvard University, where he studied with Willard Quine. In 1965, Dennett received a DPhil in philosophy at the University of Oxford, under Gilbert Ryle, with a groundbreaking thesis, "Content and conscience," which has since been translated into many languages. He taught at the University of California, Irvine, for 6 years and then moved to Tufts University in 1971, where he remained for the rest of his career. At the time of his death, Dennett was Fletcher Professor of Philosophy and co-director of the Center for Cognitive Studies at Tufts. He was also a prestigious University Professor, which allowed him to teach in any department, an appropriate honor for an interdisciplinary scholar like Dennett.

Starting with his doctoral thesis, Dennett's main body of work focused on the philosophical examination of both basal and higher mental abilities, which he grounded in scientific research. In his book *Kinds of Minds*, he laid out ways of thinking about the different types of cognition, starting from its humble evolutionary origins. In *Consciousness Explained*, he proposed a unified theory of consciousness, whereby sensory systems work in parallel to generate perceptions, which are then "edited" and integrated into a unified "Cartesian theater" with a common "narrative." He argued for strict materialism, without room for subjective elements of consciousness, which he thought were confusing and unscientific. His position was prescient, as recent brain recording and imaging research has demonstrated how the parallel processing of sensory information is then integrated into a unified framework in brain-wide functional signals.

Dennett was also deeply interested in evolution as the major force for the creation of the mind, and he argued for a central role for natural selection, which inevitably leads to more sophisticated and complex organisms. His strong Darwinian stand led him to publicly defend atheism and oppose religion. He also argued that morality was the natural outcome of a selection process.

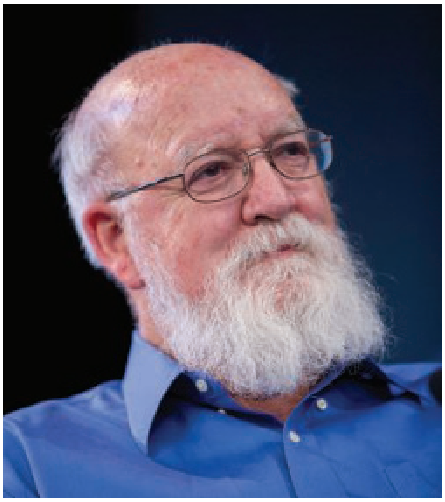


PHOTO: DAVID LEVENSON/GETTY IMAGES

In his later work, Dennett explored the origins of intelligence and argued for a more general definition of intelligent systems, beyond biological ones. As long as there are Darwinian-style selection rules for emergent properties (properties of a complex system that arise from interactions among its parts), he envisioned that systems could acquire more complex phenotypes with more sophisticated functions and inevitably progress, like the steps in a ratchet, toward a more intelligent behavior. He was comfortable extending intelligent and goal-directed behavior to systems as simple as viruses and molecular networks and those as complex as artificial intelligence (AI) and the internet. His position anticipated many of the current arguments and discussions about consciousness in AI long before the technology existed. His views meshed seamlessly across disciplines, merging ideas from computer science, evolutionary biology, and cognitive neuroscience to form a rich understanding of embodied minds.

Dan was a generous thinker, sharing inspiration and wisdom with his colleagues and students. M.L. was a student in Dan's philosophy of mind course when they met in 1993. He had the good fortune to later work with Dan as a fellow faculty member. Dan was equally generous and inspiring in both capacities, offering the best of critical analysis but also freedom to explore wild hypotheses.

Dan loved coming to M.L.'s lab to personally look through microscopes and directly see the new biology that could inspire his thinking. He had an infectious excitement about new ideas and new ways to experimentally test his and others' theories. His cunningly designed thought experiments, which he referred to as "intuition pumps," could be used to powerfully transmit counterintuitive concepts, clearly and often with great humor.

R.Y. met Dan in his later years and explored with him a "modern synthesis" of philosophy and neuroscience, aiming to help explain how cognitive properties can be mechanistically implemented by the dynamics and emergent properties of neural circuits. Echoing the mid-20th-century synthesis of evolutionary biology and genetics, this goal remains one of the most formidable intellectual challenges of our generation.

Immensely curious from a very young age, Dennett was a sketch artist, sculptor, jazz pianist, skilled navigator, computer engineer, and inspiring lecturer. He managed a farm in Maine with his wife, Susan, for many years. He and Susan adopted two children and later enjoyed their roles as grandparents.

Dennett held honorary doctorates from McGill University and the universities of Connecticut, Edinburgh, Bucharest, and Amsterdam. He became a AAAS fellow in 2009, and his many awards included the American Humanist Association's Humanist of the Year in 2004. In 2012, Dennett was awarded the Erasmus Prize, one of Europe's most distinguished recognitions for exceptional contributions to culture, society, or social science.

Dennett's combination of intellectual humility and brilliance is an uncommon one. A prolific writer and engaging speaker, he was scrupulously honest and always more interested in clarity and insight than in cheap rhetorical wins. His frequently used strategy of presenting an opponent's strongest-possible argument and directly engaging the reader or audience in the debate, an approach he called "steel-manning," is a typical example of the profound legacy he leaves us. The fields of consciousness studies, philosophy, and neuroscience will deeply miss his timelessly classic yet always youthful intellectual curiosity. ■

10.1126/science.adq5873

[1]Department of Biological Sciences, Columbia University, New York, NY, USA. [2]Department of Biology, Tufts University, Boston, MA, USA. Email: rmy5@columbia.edu; michael.levin@tufts.edu

POLICY FORUM

ARTIFICIAL INTELLIGENCE

GPTs are GPTs: Labor market impact potential of LLMs

Research is needed to estimate how jobs may be affected

By **Tyna Eloundou[1], Sam Manning[2], Pamela Mishkin[1], Daniel Rock[3]**

We propose a framework for evaluating the potential impacts of large-language models (LLMs) and associated technologies on work by considering their relevance to the tasks workers perform in their jobs. By applying this framework (with both humans and using an LLM), we estimate that roughly 1.8% of jobs could have over half their tasks affected by LLMs with simple interfaces and general training. When accounting for current and likely future software developments that complement LLM capabilities, this share jumps to just over 46% of jobs. The collective attributes of LLMs such as generative pretrained transformers (GPTs) strongly suggest that they possess key characteristics of other "GPTs," general-purpose technologies (*1, 2*). Our research highlights the need for robust societal evaluations and policy measures to address potential effects of LLMs and complementary technologies on labor markets.

We consider the progress of LLMs' capabilities and the potential breadth of the complementary technologies that they spawn, underscoring that maximizing the impact of LLMs requires their integration within broader systems (*3–5*). The rubric that we develop [see supplementary materials (SM) section 3.1.2] defines task exposure to LLMs, in the spirit of prior work on quantifying exposure to machine learning (*6–8*). Following prior literature, we use the concept of exposure as a proxy for potential economic impact, irrespective of labor-augmentation or displacement effects (see section 2 of the SM for further discussion of the literature). Our approach differs from the broader scope of complementary innovations in general-purpose technology discussions, focusing more narrowly on advanced software capabilities than on the potential for business process reengineering, new intangible assets creation, or workforce retraining. General-purpose technologies such as electricity or computing historically have had far-reaching effects that took decades to fully materialize. With evidence of the general-purpose technology potential of LLMs, we urge caution in making long-term predictions while offering an outline of where work might change.

[1]OpenAI, San Francisco, CA, USA. [2]Centre for the Governance of AI, Oxford, UK. [3]Wharton School, University of Pennsylvania, Philadelphia, PA, USA. Email: pamela@openai.com

RATING TASKS AND OCCUPATIONS

We use the O*NET 27.2 database (*9*), which covers 1016 occupations and their Detailed Work Activities (DWAs) and tasks, focusing on the 923 occupations with available task information.

DWAs describe broader activities that represent the work performed within an occupation, e.g., a DWA of "Execute sales or other financial transactions" might be associated with a range of occupations. Tasks are more specific work units linked to one or more DWAs, such as the task "Deliver email confirmation of completed transactions and shipment." O*NET categorizes tasks as "Core" or "Supplemental," with the latter weighted half as much as the former in our analysis (or noted otherwise). Results are robust to alternative weightings (SM section 9.3). We sourced employment and wage data from the 2020–2022 Occupational Employment series by the US Bureau of Labor Statistics (BLS).

Although the term "LLM" can refer to a broad array of data modalities, we confine our definition of LLM-powered software to principally apply to text, code, and images. We do not tie this analysis to a particular model or model provider but rather to a paradigm of release (through chatbot and application programming interfaces) that was dominant in mid-2023. Progress has been made integrating LLMs with robotics, but the uncertainty around their reliability and adoption, as well as the difficulty conveying the potential of these integrations in our rubric format, led us to exclude assumptions about future robotics integrations from our analysis.

We collect results by applying our exposure rubric (SM section 3.1), where we define exposure as the capacity of an LLM or LLM-powered system to reduce the time required for a human to complete a task by at least 50% while preserving or improving quality. The chosen 50% time reduction threshold, though arbitrary, was used for practical purposes to facilitate interpretation by human annotators and GPT-4. Our estimates of task-level potential impacts are likely to be greater than real-world productivity gains, which will be influenced by factors other than the capabilities of LLMs and complementary technologies (*10*).

Our analysis uses four key measures to gauge exposure levels at the occupation level: (i) E1, the share of an occupation's tasks where access to an LLM alone or with a simple interface would lead to 50% time savings; (ii) E2, the share of an occupation's tasks where additional software is needed on top of an LLM to realize 50% time savings; (iii) E1 + E2, the share of tasks exposed assuming full LLM and software integration; and (iv) E1 + 0.5 × E2, an intermediate measure of exposure while complementary technologies are not fully implemented.

The authors, along with a group of contractors experienced in evaluating LLM outputs, labeled the exposure of DWAs and tasks in the O*NET dataset according to the rubric. We ask human annotators to use their best judgment in labeling DWAs and tasks as E0, E1, or E2 accord-

Human and GPT-4 ratings

Mean and standard deviation of ratings from human and GPT-4 labels are highly similar. Occupation level represents the average share of tasks within an occupation exposed. Task level represents the share of tasks across all occupations exposed.

Occupation-level exposure

	Human		GPT-4	
	Mean	SD	Mean	SD
E1	0.14	0.14	0.14	0.16
E1 + 0.5 × E2	0.30	0.21	0.34	0.22
E1 + E2	0.46	0.30	0.55	0.34

Task-level exposure

	Human		GPT-4	
	Mean	SD	Mean	SD
E1	0.15	0.36	0.14	0.35
E1 + 0.5 × E2	0.31	0.37	0.35	0.35
E1 + E2	0.47	0.50	0.56	0.50

ing to the rubric. Although these human raters had experience working with contemporary LLMs, they do not have diverse occupational experience. This limits their understanding of O*NET work tasks, occupational workflows, and occupation-specific software tools. Future research aims to broaden the pool of human raters across various occupations to enhance the validation of these exposure scores.

We then prompted an early GPT-4 model to apply a slightly modified version of the rubric given to human labelers. We complement the human ratings with GPT-4 ratings to explore the feasibility of using LLMs to aid in social science research, particularly to scale up processes that might otherwise be time-consuming or resource intensive. The version given to GPT-4 was edited to enhance agreement with human ratings across a sample of O*NET task/occupation pairs. Indeed, small changes to the rubric can substantially affect GPT-4's labels, highlighting the fragility of this method in the absence of an alternative validation set (see SM section 3.3.2 for additional discussion of the limitations of this approach). Final agreement rates across human and GPT-4 ratings are visualized at the occupation level in fig. S4 and at the task level in fig. S7.

ESTIMATING EXPOSURE

Our results show LLMs' relevance to approximately 14% of tasks per occupation on average (see the table). When considering the partial implementation of complementary software (E1 + 0.5 × E2), this figure doubles to 30 to 34%, and with full implementation (E1 + E2), it climbs to 46 to 55%. Using the National Employment Matrix data from the BLS, we estimate that ~80% of workers are in occupations with at least 10% of tasks exposed assuming partial implementation of complementary software (E1 + 0.5 × E2), whereas 18.5% of workers have more than 50% of their tasks exposed (fig. S2).

Exposure is prevalent in occupations that involve generating written text or code, as well as those with routine information processing tasks. In fig. S8, we plot exposure by Job Zone—a measure in the O*NET database that groups occupations that are similar in (i) the level of education needed to get a job in the occupation, (ii) the amount of related experience required to do the work, and (iii) the extent of on-the-job training needed to do the work. The amount of preparation required ranges from 3 months (Job Zone 1) to 4 or more years (Job Zone 5). Job Zones 4 and 5 are the most exposed, indicating that occupations needing "extensive preparation" (e.g., lawyers, pharmacists, database administrators) are more exposed than those with lower entry barriers (e.g., dishwashers, floor sanders).

Tasks marked E2 indicate that they are exposed only by software leveraging LLMs and domain-specific scaffolding, not by generally available LLMs. Within an occupation, on average, the addition of domain-specific software exposes an additional 32 to 41% of tasks according to human and GPT-4 ratings, respectively (see occupation-level means in the table). This suggests that LLM-enhanced software could more than double the share of tasks exposed relative to the baseline of LLMs alone (mean E1 of 14% of tasks for both annotation types).

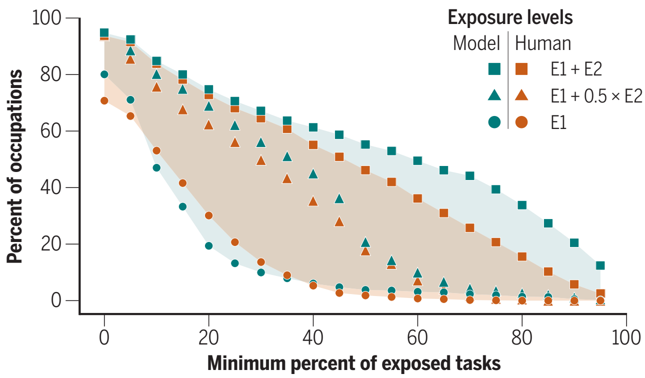
Occupational exposure to LLMs are shown in terms of the percentage of tasks affected (occupational exposure) with specific exposure levels (E1, E1 + 0.5 × E2, E1 + E2) (see the first figure). For instance, human annotators determined that 1.8, 17.8, and 46.2% of occupations are $E1_{50}$, $(E1 + 0.5 \times E2)_{50}$, and $(E1 + E2)_{50}$ exposed (e.g., have at least 50% of tasks exposed at each level), respectively. The vertical gap between E1 and E1 + E2 on this figure indicates the potential additional exposure from artificial intelligence (AI) tools beyond direct LLM interactions alone. In general, higher-wage occupations are more exposed to LLMs than lower-wage occupations (see the second figure).

In exploratory analysis, we used GPT-4 to evaluate the automation potential of tasks using an automation-focused rubric. These GPT-4 ratings were not validated with human ratings. This automation rubric (see SM section 3.2) rates tasks on the basis of the proportion of task components that LLM-based systems could autonomously complete with high quality and reliability. It spans five automation categories, from Full Automation—tasks fully manageable by LLMs without human intervention—to No Automation, where LLMs are unable to reliably perform any part of the task independently with high quality.

Our exploratory analysis estimates only 1.86% of tasks could be fully automated by LLMs plus additional software integrations without human oversight. Still, more than 71% of tasks have at least some component that an LLM plus additional software could plausibly complete with high quality (table S1). High LLM exposure in occupations correlates with higher automation potential, with automation scores explaining 55.6% of the task-level exposure variance (table S2). This suggests that occupations with greater automation risk may also have higher augmentation potential (fig. S3). This challenges the notion that it may be possible to predetermine whether LLMs will

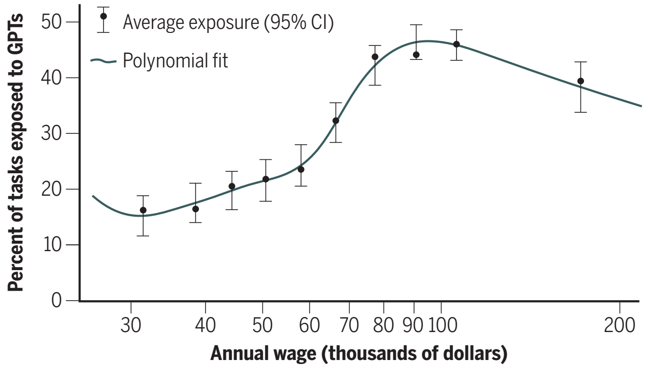
Exposure of occupations to GPTs

Exposure intensity across the economy is shown in terms of percent of affected occupations. The horizontal axis represents the share of tasks within an occupation exposed at the different exposure levels denoted in the figure legend. Shaded areas denote the range of exposure levels.



Exposure to GPTs by income

The share of all tasks within occupations that are exposed to LLMs and partial LLM-powered software (E1 + 0.5 × E2) is shown against the median annual wage for the occupation. Data reflect human rating.



CI, confidence interval; GPT, generative pretrained transformer; LLM, large language model.

GRAPHIC: M. HERSHER/*SCIENCE*

augment or automate jobs, despite recommendations for AI to complement rather than replace human work (*11*).

One specific criterion for general-purpose technologies is that they exhibit innovational complementarities: The productivity of research and development (R&D) efforts in downstream industries improves owing to their development. This could lead not only to capital deepening (an increase in the capital-to-labor ratio), but also a change in the direction of innovation (*3*). Our data can offer some suggestions as to whether work will change in R&D roles as a result of LLMs. We construct a network of occupations using detailed work activities that they share in common (fig. S13). Nodes in this network are O*NET occupational varieties, and nodes share an edge if they share a DWA in common. The network is highly modular (Louvain modularity of 0.709), and we detect 11 distinct occupational clusters using the Louvain method. The two job groups (clusters) that are most exposed to LLMs are "Scientists and Researchers," then "Technologists," such as software engineers and data scientists (fig. S16). The R&D-intensive roles in the top two categories are in general more heavily exposed than many other kinds of work. This suggests that when LLMs improve, they have the potential to cause downstream improvements in R&D productivity for workers in the sectors deploying them (see SM section 8.2.).

IMPLICATIONS AND LIMITATIONS

As LLMs are nascent, public investment in tracking LLM adoption and ensuing labor market impacts is urgently needed and will be vital for shaping effective policy responses. Although this paper does not test specific policies, recent work has outlined US policy options on education, training, tax, and safety net reforms that could help ensure that the economic gains from LLMs are large and broadly distributed (*12*). Transitional policy tools are essential to preventing harms (e.g., from concentrated labor demand shocks) for negatively affected workers. Initiatives such as wage insurance and unemployment insurance reform can ease transitions and bolster economic security. Furthermore, collaborations between AI labs and governments to extend AI's economic gains equitably could help to direct the development of LLMs in a more socially beneficial direction.

Despite a growing literature that aims to understand the labor market impacts of these systems, there is still no clear understanding of how "exposure" to AI systems will translate to real-world impacts on labor demand, wages, inequality, job quality, and other key outcomes. Public investments in measurement and tracking of LLM adoption, complementary technology development, and the ensuing labor market impacts are a critical input to informing optimal public policy responses.

Using a rubric-based approach to measure exposure has limitations (SM section 3.3). It does not capture the creation of potential new tasks and potential changes to returns to technological capital (*13*). Furthermore, we cannot conclude that there will be any impacts related to the equilibrium demand for human labor across occupations using this method. This method also misses shifts in innovation direction rather than just capital deepening (*3*), even if it can suggest where LLMs might influence labor task intermediates to innovation. Despite these shortcomings, this measurement is useful for identifying tasks and occupations likely to be affected by LLMs. However, predicting long-term labor market changes due to LLMs' general-purpose abilities remains a complex challenge.

There are specific limitations in using a LLM to assist with this rubric-based methodology. Task classification by GPT-4 is affected by changes in rubric language, prompt structure, example inclusion, detail level, key term definitions, and model version. Refining prompts with a small validation set can better align model outputs with the rubric's intended meaning. We made slight adjustments between GPT-4 and human annotator rubrics to guide the model toward accurate labels without biasing human judgments. Although we use multiple annotation sources, none is considered the definitive ground truth. There is room for improvement and innovation in creating effective rubrics for LLM classification of complex information.

We do not systematically examine sources of disagreement across labeling sources, but there are a few categories of tasks that qualitatively appear to have greater variation in exposure ratings. These include tasks where using an LLM or domain-specific software would necessitate substantial individual or group behavior or norm changes (such as expecting to meet with humans to conduct a meeting or negotiation); tasks where human oversight, judgment, or empathy are expected or required (such as certain kinds of decision-making or counseling); and tasks already potentially exposed by technology where it is difficult to judge the added impact of using an LLM (such as making reservations).

Predicting the trajectory of LLM applications is fraught with uncertainties owing to emergent capabilities, shifts in human biases, and technological evolution. Our forward-looking projections, rooted in contemporary trends and perceived technological potential, are subject to change with new breakthroughs. Tasks currently deemed beyond the reach of LLMs may become viable with future innovations, whereas those seeming ripe for disruption could encounter unexpected barriers. Through continued effort to better understand LLM capabilities and their possible effects on the workforce, policymakers and other key stakeholders can make better-informed decisions, efficiently navigating the complex landscape of AI and its implications for the future of work. ■

REFERENCES AND NOTES

1. T. F. Bresnahan, M. Trajtenberg, *J. Econom*. **65**, 83 (1995).
2. R. G. Lipsey, K. I. Carlaw, C. T. Bekar, *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth* (Oxford Univ. Press, 2005).
3. T. Bresnahan, "Artificial intelligence technologies and aggregate growth prospects" in *Prospects for Economic Growth in the United States*, J. W. Diamond, G. R. Zodrow, Eds. (Cambridge Univ. Press, 2019), pp. 132–170.
4. A. K. Agrawal, J. S. Gans, A. Goldfarb, "AI adoption and system-wide change," *Tech. Rep.*, National Bureau of Economic Research (2021).
5. A. Goldfarb, B. Taska, F. Teodoridis, *Res. Policy* **52**, 104653 (2023).
6. E. Brynjolfsson, T. Mitchell, D. Rock, *AEA Pap. Proc*. **108**, 43 (2018).
7. E. W. Felten, M. Raj, R. Seamans, *AEA Pap. Proc*. **108**, 54 (2018).
8. M. Webb, "The impact of artificial intelligence on the labor market," Working paper, Stanford University, 2020.
9. O*NET, O*NET 27.2 Database (2023).
10. M.S. Svanberg *et al*., "Beyond AI Exposure: Which Tasks are Cost-Effective to Automate with Computer Vision?" Working paper; https://futuretech-site.s3.us-east-2.amazonaws.com/2024-01-18+Beyond_AI_Exposure.pdf.
11. E. Brynjolfsson, *Daedalus* **151**, 272 (2022).
12. D. H. Autor, D. A. Mindell, E. Reynolds, *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines* (MIT Press, 2022).
13. D. Acemoglu, P. Restrepo, *Am. Econ. Rev*. **108**, 1488 (2018).

ACKNOWLEDGMENTS

The authors thank those who helped annotate task exposure, including M. A. Saeed, B. Zitha, M.Ö. Şenen, J.J., and P. Hoeschele. The authors thank L. Fuld, A. Glat, M. Lampe, and J. Susser for excellent research assistance, and M. Brundage for feedback on this paper. The authors thank T. Markov and V. Goel for setting up the infrastructure used to run rubrics with GPT-4. The authors thank L. Ahmad, D. Bakong, M. Beane, S. Benzell, E. Brynjolfsson, P. Eloundou-Enyegue, M. Frank, C. Frey, S. Giroux, G. Hadfield, J. Heidecke, J. Hersh, A. Hickey, L. Hitt, E. Horvitz, K. Hosanagar, S. Hu, A. Kachra, C. Kim, K. Klinova, D. Kokotajlo, G. Krueger, M. Lampe, J. Milin, A. Mehta, D. Robinson, L. Schiavo, D. Selsam, S. Shoker, P. Tambe, and J. Wu for feedback and edits at various stages of the project. S.M. performed the majority of work while under contract with OpenAI. GPT-4 and ChatGPT were used for writing, coding, and formatting assistance in this project (see SM). D.R. is a cofounder and board member of Workhelix, Inc..

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj0998

10.1126/science.adj0998

LETTERS

The border between the the US and Mexico faces challenges in meeting global sustainability goals.

Edited by Jennifer Sills

Sustainable development in global border regions

The likelihood of achieving Sustainable Development Goals (SDGs) by the 2030 deadline looks dim (*1*), especially in border regions. More than one-sixth of the world's population lives in regions within a 60-km buffer zone of an international border. Most border communities face poverty, remoteness from governance centers, economic underdevelopment, and infrastructure deficits (*2*). Often situated in ecologically vulnerable zones, these regions are prone to natural disasters and biodiversity crises (*3*). They are also the sites of 45% of global armed conflicts, and at least 40 countries have territorial disputes (*4*). Given increasing regional conflicts, surging energy and food costs, and geopolitical strains, border zones require urgent support to achieve SDGs.

More than 30 countries are embroiled in wars or regional conflicts (*5*), which severely affects the ability of border regions to make progress on SDGs. Conflict leads to heightened humanitarian crises (SDGs 1, 2, 3, 6, and 16), socioeconomic downturns (SDGs 8, 9, 11, and 12), and intensified ecological and climate crises (SDGs 13, 14, and 15) in border regions. The COVID-19 pandemic dealt a major blow to global trade, affecting the prosperity of regions that depend on border ports, and economic recovery has been slow (SDG 8) (*6*). Water disputes in transboundary rivers such as the Nile, Mekong, and Jordan exacerbate water security challenges (SDG 6) (*7*). The erection of barriers in many border regions compromises migrant rights (SDG 10.7), labor market security (SDG 8.8), and the protection of women and children (SDG 16.2) (*8*), and it also disrupts ecosystems and wildlife migration (SDG 15) (*9*). Institutional and cultural diversity in border areas hinders cross-border cooperation (SDG 17) (*10*), and the influx of refugees strains public services and exacerbates social polarization (SDGs 1.3, 8.8, 10) (*11*).

Developing border areas, which is key to global success, requires collaboration among neighboring countries, local governments, nongovernmental organizations, businesses, communities, and scientists. The United Nations (UN) should establish an agency dedicated to sustainable border development, managing complex relations, and solving mutual challenges. With UN support, stakeholders can work together to restore stability in conflict zones through diplomacy, enhance humanitarian aid, and manage refugees effectively. The UN and relevant countries should boost cross-border infrastructure; China's Belt and Road Initiative in Central and Southeast Asia (*12*) can serve as a guide. Cooperative frameworks to resolve cross-border resource use and ecosystem protection can also facilitate these goals. Communities should be encouraged to participate and build capacity toward achieving global targets. To promptly address challenges, the UN should lead affected countries to enhance monitoring of SDGs and promote information, resource, and knowledge sharing. Collaborative efforts can foster sustainable development in border areas.

Yi Cheng[1,2], Zhenci Xu[3], Haimeng Liu[2]*
[1]Beijing Academy of Social Sciences, Beijing 100101, China. [2]Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. [3]Department of Geography, The University of Hong Kong, Hong Kong 999077, China.
*Corresponding author. Email: liuhm@igsnrr.ac.cn

REFERENCES AND NOTES

1. United Nations, "The sustainable development goals report 2023: Special edition" (UN, 2023); https://desapublications.un.org/publications/sustainable-development-goals-report-2023-special-edition.
2. C. Xiao, Z. Feng, Z. You, F. Zheng, *J. Cleaner Prod.* **371**, 133685 (2022).
3. M. A. Titley, S. H. Butchart, V. R. Jones, M. J. Whittingham, S. G. Willis, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011204118 (2021).
4. Y. Cheng, thesis, University of Chinese Academy of Sciences, Beijing, China (2023) [in Chinese].
5. World Population Review, "Countries currently at war/Countries at war 2024" (2024); https://worldpopulationreview.com/country-rankings/countries-currently-at-war.
6. D. Egger *et al.*, *Sci. Adv.* **7**, eabe0997 (2021).
7. T. R. Albrecht, A. K. Gerlak, *Water Secur.* **17**, 100124 (2022).
8. S. Nurhasanah, in "Fourth Asia-Pacific research in social sciences and humanities, arts and humanities stream (AHS-APRISH 2019)" (Atlantis Press, 2023), pp. 540–551.
9. J. D. Linnell *et al.*, *PLOS Biol.* **14**, e1002483 (2016).
10. C. McCall, *Coop. Confl.* **46**, 201 (2011).
11. W. Arrocha, *Crime Law Soc. Change* **71**, 245 (2019).
12. D. Normile, *Science* **374**, 674 (2021).

10.1126/science.adp0639

Sex-based differences affect conservation

The world's ecosystems, and the organisms that inhabit them, face unprecedented threats, including a changing climate and anthropogenic disturbances. Understanding the baseline habitat requirements of living organisms is essential to developing effective conservation strategies (*1*). Nonetheless, studies often fail to consider the biological sex of organisms, an oversight that could substantially impede conservation efforts (*2*).

Sex-based differences can affect morphology, physiology, behavior, and distribution in a wide variety of taxa (*2*–*4*). For instance, heat-related mortality was higher in male Magellanic penguins

(*Spheniscus magellanicus*) (*5*) and in female Australian flying foxes (*Pteropus* spp.) (*6*) compared with the opposite sex because of sex-specific body conditions and physiology. The males and females of some species, such as shortfin mako sharks (*Isurus oxyrinchus*) (*7*) and golden-winged warblers (*Vermivora chrysoptera*) (*8*), spatially segregate, likely because of sex-specific foraging behavior or social factors in courtship and mating.

Variations have also been documented in the phenologies and life history strategies across males and females of many species (*9*, *10*). For example, the timing of spawning migration differs by sex in yellowfin soles (*Limanda aspera*), with males staying longer on the spawning grounds compared with females (*11*), and in bighorn sheep (*Ovis canadensis*), with males and females moving between seasonal ranges at different times (*12*).

Conservation and management strategies should be tailored to each sex in the case of differences, as if they were separate species (*2*). Yet, most conservation initiatives do not take sex into account when addressing the consequences of habitat loss and thus protect areas that favor the behavior and needs of only one sex (*8*). Researchers and conservation managers should monitor and report sex ratios across breeding and nonbreeding periods to facilitate sex-specific vulnerability assessments for each species. Each report should provide the factors and variables that the authors used to classify males, females, and hermaphrodites, including the range of sexual phenotypes. Conservation strategies focused on the specific needs of sexes will more effectively protect the full population.

Elena Gissi[1,2,3]*, R. Terry Bowyer[4], Vernon C. Bleich[5]
[1]National Research Council, Institute of Marine Science (CNR ISMAR), 30122 Venice, Italy. [2]Oceans Department, Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA. [3]National Biodiversity Future Center, 90133 Palermo, Italy. [4]Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. [5]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Reno, NV 89557, USA.
*Corresponding author.
Email: elena.gissi@cnr.it

REFERENCES AND NOTES

1. G. Ceballos, P. R. Ehrlich, R. Dirzo, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6089 (2017).
2. R. T. Bowyer, *Sexual Segregation in Ungulates: Ecology, Behavior, and Conservation* (Johns Hopkins Univ. Press, 2022).
3. V. J. Wearmouth, D. W. Sims, in *Advances in Marine Biology*, vol. 54, D. W. Sims, Ed. (Academic Press, 2008), pp. 107–170.
4. E. Gissi, L. Schiebinger, R. Santoleri, F. Micheli, *Front. Ecol. Environ.* **21**, 324 (2023).
5. K. A. Holt, P. D. Boersma, *Ornithol. Appl.* **124**, duab052 (2022).
6. J. A. Welbergen, S. M. Klose, N. Markus, P. Eby, *Proc. R. Soc. B* **275**, 419 (2008).
7. G. R. Mucientes *et al.*, *Biol. Lett.* **5**, 156 (2009).
8. R. E. Bennett, A. D. Rodewald, K. V. Rosenberg, *Biol. Conserv.* **240**, 108266 (2019).
9. J.-F. Lemaître *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8546 (2020).
10. C. T. Williams, H. E. Chmura, C. K. Deal, K. Wilsterman, *Integr. Comp. Biol.* **62**, 980 (2022).
11. M. Olmos *et al.*, *Prog. Oceanogr.* **215**, 103035 (2023).
12. V. Geist, *Mountain Sheep: A Study in Behavior and Evolution* (Univ. Chicago Press, 1971).

10.1126/science.adp1088

Plastic foam pollution from Chinese fisheries

In 2024, China issued its most recent government work report, which includes a requirement for enhanced control of plastic pollution in water and soil (*1*). However, the policy does not address contamination caused by the plastic foam used in fishing gear.

Plastic foam pollution by fisheries is serious and widespread. Fishermen in coastal provinces of China, such as Guangdong, Guangxi, Fujian, and Hainan, commonly use closed-cell foam materials, such as expanded polypropylene and polystyrene, which are known for their cost-effectiveness and low density (*2*). The plastic foam serves as fishing rafts and as floating balls—which can reach up to 2 m in diameter—that hold up fishing nets and aquaculture cages (*3*, *4*). In the 135-km^2 Maowei Sea alone, located in the Beibu Bay of the South China Sea, an estimated 399 metric tonnes of fishing plastic foam enters the ocean each year (*5*).

Floating foam has a limited life span and steadily deteriorates (*5*), leaving substantial amounts of debris in the ocean and potentially contributing to pollution (*6*). The foam debris floats on the surface of the water and drifts with the wind and currents (*7*). Accidental ingestion of foam debris by marine organisms can lead to intestinal obstruction and damage to the digestive tract, imposing long-term health problems (*8*). When plastic foam debris degrades into microplastics, its chemical composition and the pollutants that it absorbs can compound marine pollution (*9*). These foam microplastics can also enter the human body through the food web, causing potential chronic damage (*10*, *11*).

Reducing the harm caused to oceans by plastic foam necessitates a comprehensive approach. The government should implement additional laws and regulations related to fishery foam waste and manage fishery projects accordingly. Long-term monitoring mechanisms are essential to quantitatively and systematically track the abundance of foam debris and foam microplastics in seawater. Enterprises and research institutions should collaborate to promote upgraded fishing gear materials and foam recycling technology. Additionally, workers engaged in marine production should be incentivized to gradually replace plastic foam with stable plastic, such as high-density polyethylene, through publicity and subsidies.

Hui Wang, Peng Liu, Yihui Yuan, Ning Wang*
State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou 570228, China.
*Corresponding author.
Email: wangn02@foxmail.com

REFERENCES AND NOTES

1. Central Government of the People's Republic of China, "Government work report" (2024); https://www.gov.cn/yaowen/liebiao/202403/content_6939153.htm [in Chinese].
2. M. Kuroda *et al.*, *Sci. Total Environ.* **925**, 171421 (2024).
3. GRT News, "The substitution of aquaculture waters for the environmental protection float has unexpectedly resulted in a more environmentally detrimental outcome" (2020); https://mp.weixin.qq.com/s/6eN5RMFv7sVCvD1QUPj9dQ [in Chinese].
4. "The increase in white pollution necessitates a specific protocol for managing marine debris along the southern shore of Xiamen Bay," *Chuanmeiplus* (2023); https://mp.weixin.qq.com/s/4vv3HZACs7n3qQe_ci0UZw [in Chinese].
5. Y. Tian *et al.*, *Water Res.* **219**, 118551 (2022).
6. L. Feld, V. H. da Silva, J. Strand, *Mar. Pollut. Bull.* **180**, 113774 (2022).
7. J. Lee *et al.*, *Mar. Policy* **51**, 182 (2015).
8. A. Turner, *Environ. Sci. Technol.* **54**, 10411 (2020).
9. K. Tanaka *et al.*, *Mar. Pollut. Bull.* **69**, 219 (2013).
10. Z. Huang *et al.*, *J. Hazard. Mater.* **446**, 130673 (2023).
11. A. D. Vethaak, J. Legler, *Science* **371**, 672 (2021).

10.1126/science.adp4110

ERRATA

Erratum for the Report "SCF[Fbxl3] controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins" by L. Busino *et al.*, *Science* **384**, eadq4777 (2024).

Published 23 May 2024
10.1126/science.adq4777

Erratum for the Research Article "Size, distribution, and vulnerability of the global soil inorganic carbon" by Y. Huang *et al.*, *Science* **384**, eadq4518 (2024).

Published 23 May 2024
10.1126/science.adq4518

Erratum for the Research Article "Ciliopathy patient variants reveal organelle-specific functions for TUBB4B in axonemal microtubules" by Dodd and Mechaussier *et al.*, *Science* **384**, eadq2178 (2024).

Published 9 May 2024
10.1126/science.adq2178

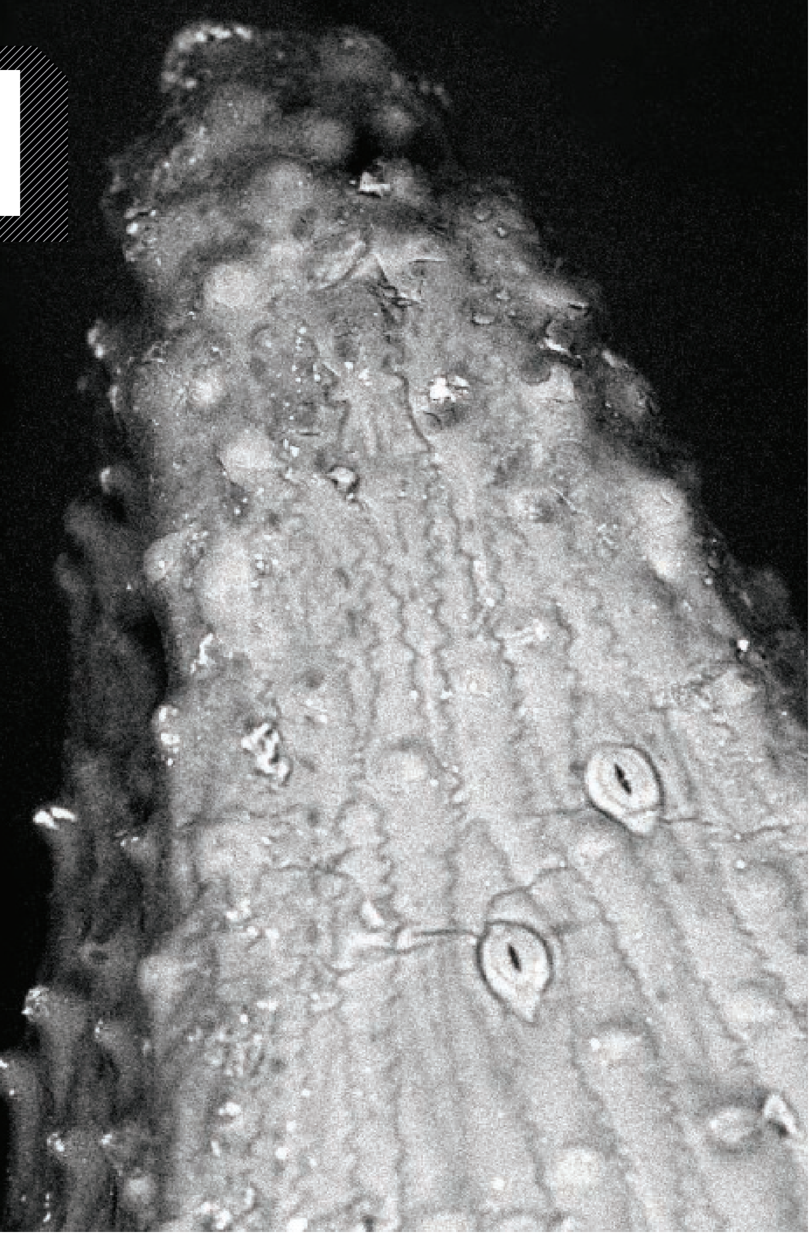
Erratum for the Report "Transmembrane molecular pump activity of Niemann-Pick C1 protein" by J. P. Davies *et al.*, *Science* **384**, eadq2125 (2024).

Published 9 May 2024
10.1126/science.adq2125

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



FLUID DYNAMICS

Bidirectional fluid flow inspired by nature

Both engineered and natural structures can exhibit directional liquid transport that is driven by the specific patterning of the surface. Yang *et al*. drew inspiration from the *Crassula muscosa* plant, in which the direction of fluid flow is determined by the shape and orientation of the structured fins, which can vary from stem to stem, and by the surface tension of the fluid. The authors constructed biomimetic counterparts that can include magnetic particles, which allows for changes in the flow directions simply through small changes in the angle and geometry of the fins. —Marc S. Lavine

Science p. 1344, 10.1126/science.adk4180

Scanning electron microscopy image of a *C. muscosa* leaf showing knobby structures that help to control fluid transport between leaves

PHOTO: YANG *ET AL.*

ADAPTIVE SOCIALITY

In times of change, make friends

Changing environments generally lead to shifts in the availability of essential resources. Such shifts can be major selective forces and will happen especially rapidly when changes are caused by extreme events such as hurricanes. Testard *et al*. looked at the response of an isolated population of rhesus macaques in Puerto Rico to the loss of tree cover caused by the category 4 Hurricane Maria. Reduced tree cover led to a reduction in shade, which is critical for the monkeys. In response, there was a general increase in social tolerance among the animals in the population. The most tolerant animals had the highest survival. —Sacha Vignieri

Science p. 1330, 10.1126/science.adk0606

CARBON CYCLING

New estimates from old bombs

Net primary productivity (NPP), the storage of carbon within plant tissues resulting from photosynthesis, is a major carbon sink that we rely on for slowing climate change. Global NPP estimates are variable, leading to uncertainty in modeling current and future carbon cycling. Graven *et al*. updated NPP estimates using radiocarbon data from nuclear bomb testing in the 1960s. This analysis of radiocarbon uptake into vegetation suggested that current models underestimate NPP, likely by underestimating the carbon stored in short-lived, nonwoody tissues. This work suggests that plants store more carbon but for a shorter time frame than is currently recognized. —Bianca Lopez

Science p. 1335, 10.1126/science.adl4443

PHOTOSYNTHESIS

Hydrogen highlights

The protein complex photosystem II brings together a large number of cofactors and pigments within a scaffold that allows for movement of water and protons, which are key to the water oxidation chemistry performed at the oxygen-evolving center during photosynthesis. However, it is often challenging to identify water molecules, especially protons, in x-ray crystal structures. Hussein *et al*. instead used cryo–electron microscopy, which has different benefits and drawbacks, to visualize photosystem II at even higher resolution than existing x-ray structures. Hydrogen atoms are more visible with this approach, and the authors identified more than half of those expected to be present, providing key insights into the functioning of this crucial enzyme. —Michael A. Funk

Science p. 1349, 10.1126/science.adn6541

ELECTROCHEMISTRY

Tungsten leaves so cobalt can stay

Currently, large-scale water electrolysis is conducted in basic conditions. There are efficiency advantages to using acidic conditions instead, but nearly all anodic catalysts other than the rare and expensive iridium dissolve too rapidly to be practical. Ram *et al*. report a strategy for stabilizing a more Earth-abundant cobalt catalyst by deliberately introducing tungsten that leaches out. Water and hydroxide fill the resulting vacancies to form a structure that keeps the cobalt in place for at least 600 hours of catalytic activity at a current density of one amp per square centimeter. —Jake S. Yeston

Science p. 1373, 10.1126/science.adk9849

NEUROSCIENCE

How the brain controls the heart rate

Voluntary control of one's heart rhythm can be achieved through feedback training. This technique is practiced in areas such as free diving and meditation, and offers the promise of future applications in therapy for arrhythmias, pain, and depression. However, the neural circuits underlying this biofeedback remain poorly understood. Using biofeedback-based operant learning, Yoshimoto *et al*. developed a rat model of self-regulated heart rate control. Neuronal activity in the anterior cingulate cortex was central to the development of bradycardia. Electrophysiology, calcium imaging, and synaptic tracing techniques revealed the complete neuronal pathway from the anterior cingulate cortex to the heart through several relay stations. —Peter Stern

Science p. 1361, 10.1126/science.adl3353

QUANTUM OPTICS

Protecting photon interference

The use of topological structures in optics has provided a route to controlling and protecting the behavior of propagating light beams. Extending that concept to quantum light sources has shown that quantum states of photons can be similarly protected. Ehrhardt *et al*. have demonstrated interference of propagating photons, an essential feature for quantum information processing. With the interference of topological origin, the results illustrate a way forward toward next-generation photonic quantum circuitry and scalable quantum computing protected by topologically robust quantum gates. —Ian S. Osborne

Science p. 1340, 10.1126/science.ado8192

VASCULAR DISEASE

Cell treatment for limb revascularization

Peripheral vascular disease can cause chronic limb-threatening ischemia (CLTI), which decreases quality of life and can lead to amputation. Patel *et al*. found that a specific population of pro-angiogenic/arteriogenic monocytes (PAMs) are increased in the circulation in response to CLTI. PAMs from patients with CLTI secreted growth factors that led to increased vascular cell proliferation in vitro. Markers of angiogenesis and arteriogenesis were increased in mice with hindlimb ischemia treated with PAMs from CLTI patients. The PAMs adhered to ischemic human muscle more than other monocytes. In a clinical study, PAMs injected into human ischemic limb muscle were partially retained around the injection site for up to 72 hours. These results suggest that using patient-derived PAMs could be a therapeutic approach for treating CLTI. —Brandon Berry

Sci. Transl. Med. (2024) 10.1126/scitranslmed.adf0555

IN OTHER JOURNALS

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

CANCER RISK

The risk of ink

Tattoos have exploded in popularity across Europe and the US in recent decades despite possible health risks tied to carcinogenic chemicals in the ink. Nielsen *et al.* ran a correlational study using data between 2007 and 2017 from the Swedish National Cancer Register to investigate the potential health risks of tattooing. Compared with nontattooed individuals, Swedish people with tattoos showed a 21% greater risk of malignant lymphoma, diffuse large B-cell lymphoma, and follicular lymphoma. —Ekeoma E. Uzogawa

EClinicalMedicine (2024) 10.1016/j.eclinm.2024.102649

SIGNAL TRANSDUCTION

Protection from cryptic RNA splicing

RNA splicing is pivotal in post-transcriptional gene regulation. In humans, there are challenges to accurate splicing caused by long interspersed nuclear elements (LINEs), which can generate long segments of double-stranded DNA that can trigger immune responses. Zheng *et al.* have identified heterogeneous nuclear ribonucleoprotein M (hnRNPM) as an RNA-binding protein that preferentially binds cryptic splice sites found in these large introns. In tumors lacking hnRNPM, such enhanced signaling promotes the infiltration by immune cells. Thus, although hnRNPM appears to protect the integrity of the transcriptome by preventing cryptic splicing, it could also be a target for enhancing inflammatory immune responses against tumors. —L. Bryan Ray

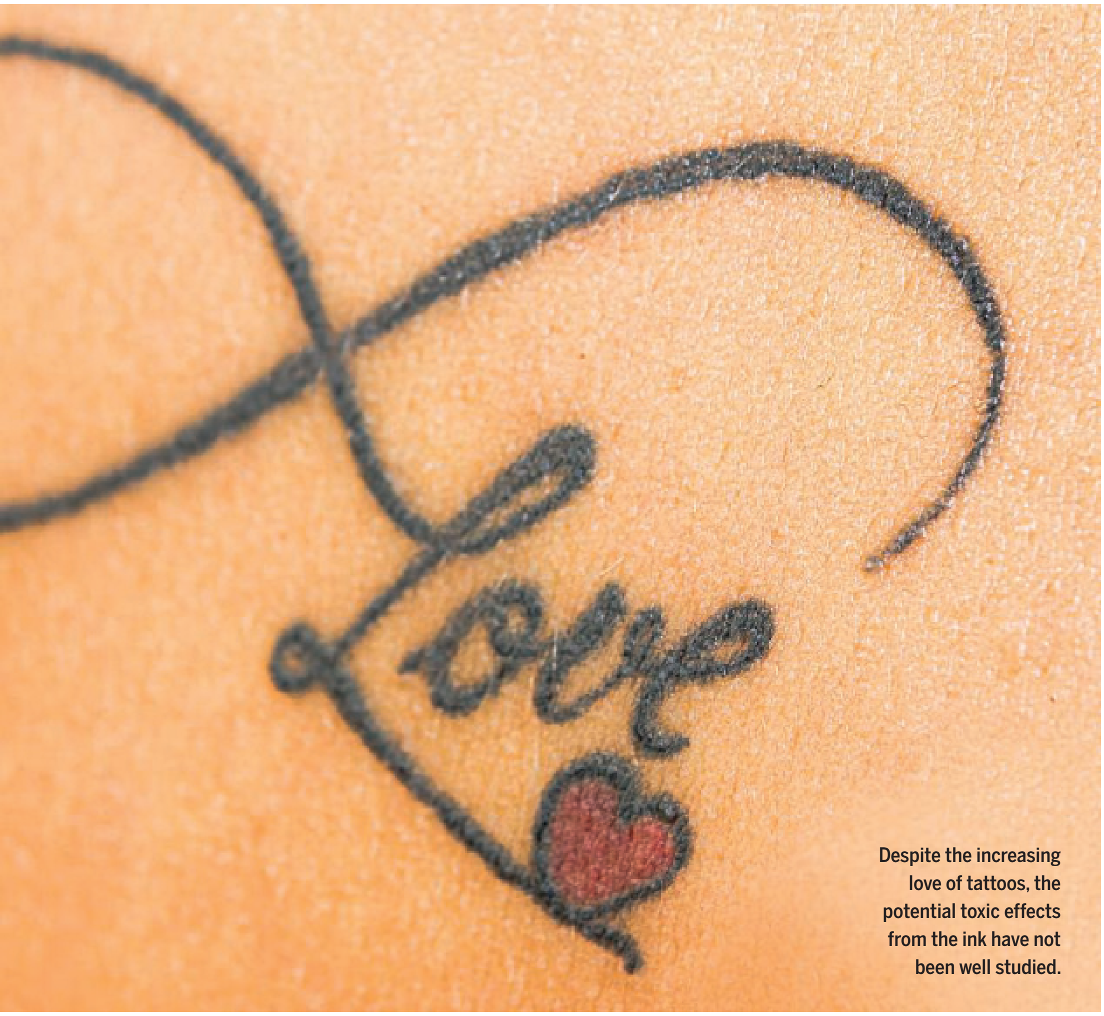
Mol. Cell (2024) 10.1016/j.molcel.2024.05.004

NEUROSCIENCE

Stick to what we know

Everyday life requires deciding between the uncertainties of exploring for new information by learning and experimentation and the exploitation of acquired knowledge for more predictable rewards. Shifting between exploration and exploitation depends on changes in attentional tuning, which are mediated by norepinephrine projections originating from the locus coeruleus. Turner *et al.* used quantitative magnetic resonance imaging to scan this region of the brain. The authors found that older age was

PHOTO: ZORAN MLADENOVIC/ALAMY STOCK PHOTO



Despite the increasing love of tattoos, the potential toxic effects from the ink have not been well studied.

associated with an exploitation bias and lower microstructural integrity of the locus coeruleus. Larger stores of prior knowledge, declining cognitive control, and a focus on close social bonds converge in later life. This combination creates a tendency to exploit options with known rewards while avoiding the ambiguity and risk associated with exploration. —Peter Stern

Proc. Natl. Acad. Sci. U.S.A. (2024)
10.1073/pnas.2322617121

NANOMATERIALS

Patterning wires

Alternating the composition along a nanowire can be useful for applications such as catalysis and thermoelectrics. Chen *et al.* developed a method for changing the composition along tellurium nanowires and nanotubes, creating segmented heteronanostructures. The authors demonstrate the synthesis of 25 segmented heteronanostructures using 13 different elements, and a silver tellurium/lead tellurium pair demonstrated interesting thermoelectric behavior. The strategy could be used to produce other interesting combinations with a variety of applications. —Brent Grocholski

Nat. Commun. (2024)
10.1038/s41467-024-47446-7

X-RAY OPTICS

Adaptive mirrors for x-ray optics

Penetrating x-rays, those found in the short-wavelength region of the electromagnetic spectrum, are a powerful probe of matter in the physical and biological sciences. Although optical lenses manipulate light using refraction, x-rays typically are guided by reflection from high-quality crystalline silicon or quartz glass mirrors. This reflection hinders the focusing ability of the x-ray beam and thus limits the spatial resolution for imaging. Inoue *et al.* developed a deformable mirror based on a single-crystal lithium niobate substrate coated with a highly reflecting layer of platinum. Application of voltages to a series of electrodes attached to the underside of the lithium niobate substrate can be used to deform the mirror surface. The ability to fine-tune the x-ray beams using adaptive optics should provide a route to higher-resolution x-ray microscopy. —Ian S. Osborne

Optica (2024)
10.1364/OPTICA.516909

EDUCATION

Public intervention, private improvements

Intervention to improve public schools can also improve private schools, multiplying the effects of the program. Working with the provincial government in Punjab, Pakistan, Andrabi *et al.* conducted a randomized trial spanning 80 villages. A government intervention in public schools improved test scores among fourth-grade students in both public and private schools. Private school improvements gained from hiring better educated and better trained teachers were driven by competition, being most pronounced when private schools were physically closer to public schools and the public schools were higher performing. Accounting for these private sector impacts increased the cost effectiveness of the public investment by 85%. —Brad Wible

Q. J. Econ. (2024)
doi.org/10.1093/qje/qjae014

STRUCTURAL BIOLOGY

Exciting competition

Excitatory neurotransmission is controlled in the brain by glutamate receptors called AMPARs. Dysregulation of AMPARs mediates many neurological disorders. Understanding how allosteric modulators affect the function of these receptors is paramount for developing treatments for related disorders. Hale *et al.* used cryo–electron microscopy to visualize AMPARs bound to glutamate in the presence of positive and/or negative allosteric modulators. In the presence of the negative modulator, glutamate binding was decoupled from channel opening. When both positive and negative modulators were present, the negative modulator outcompeted the positive one by modifying its binding domain, resulting in receptor inhibition. This allosteric site could thus be a target for drug development. —Mattia Maroso

Nat. Struct. Mol. Biol. (2024)
10.1038/s41594-024-01328-0

RESEARCH ARTICLE SUMMARY

CLINICAL TRIALS

Combined JAK inhibition and PD-1 immunotherapy for non–small cell lung cancer patients

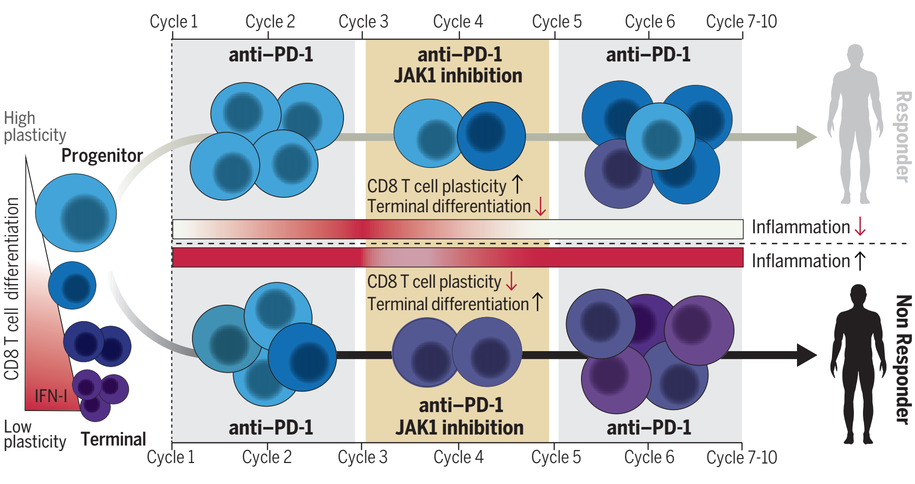
Divij Mathew†, Melina E. Marmarelis†, Caitlin Foley, Joshua M. Bauml, Darwin Ye, Reem Ghinnagow, Shin Foong Ngiow, Max Klapholz, Soyeong Jun, Zhaojun Zhang, Robert Zorc, Christiana W. Davis, Maximillian Diehn, Josephine R. Giles, Alexander C. Huang, Wei-Ting Hwang, Nancy R. Zhang, Adam J. Schoenfeld, Erica L. Carpenter, Corey J. Langer, E. John Wherry*, Andy J. Minn*

INTRODUCTION: Inflammation is a hallmark of cancer but is also required to generate optimal antitumor immune responses. Although short exposure to cytokines such as interferon (IFN) can enhance tumor immunity, prolonged exposure can result in immunosuppression. The dual nature of inflammation makes it challenging to harness the beneficial effects of cytokine activation during cancer immunotherapy while avoiding detrimental consequences. Thus, a therapeutical strategy to effectively modulate these often-opposing functions of cytokine signaling could improve immunotherapy efficacy and mitigate development of resistance.

RATIONALE: Preclinical studies in mice have demonstrated that blockade of type-one IFN (IFN-I) signaling can improve immune function during chronic viral infections and enhance the efficacy of immunotherapy for cancer. Moreover, tumors from patients with lung and other cancer types that relapse after—or are resistant to—immune checkpoint blockade (ICB) can have high expression of IFN-stimulated genes. Thus, we conducted a phase 2 clinical trial for first-line metastatic non–small cell lung cancer (NSCLC) with programmed death ligand 1 (PD-L1) expression ≥50%. Patients were administered 6 weeks of anti–PD-1 (PD-1, programmed cell death protein 1) immunotherapy, followed by a combination of anti–PD-1 and itacitinib [a selective Janus kinase 1 (JAK1) inhibitor] for 6 weeks before continuing with anti–PD-1 alone. To understand the immunological effects of transient JAK inhibition aimed at interfering with persistent inflammation occurring after anti–PD-1, we evaluated the association between clinical response and the evolution of CD8 T cell differentiation, immune signaling, and inflammatory markers.

RESULTS: In mice, the addition of itacitinib after the start of ICB improved response of interferon-stimulated gene (ISG)–high resistant tumors and increased the proportion of proliferating precursor-like CD8 T cells in the periphery. Use of an IFN-I–receptor blocking antibody similarly improved ICB efficacy, suggesting that blockade of IFN-I signaling is sufficient to phenocopy the effects of delayed JAK inhibition. In humans, the combination of anti–PD-1 with delayed, transient, itactinib treatment in patients with NSCLC led to an overall response rate of 67% and median progression-free survival of 23.8 months. Patients were categorized into three response groups based on the timing of radiographic response: patients with clinical response within the first two cycles of anti–PD-1 monotherapy (αPD1.R), patients who responded only after a 6-week course of concurrent itacitinib was added at the start of the third cycle of anti–PD-1 (JAKi.R), or patients who did not respond regardless of treatment (NR). Each clinical response group had distinctive immunological changes coupled to inflammatory features. The αPD1.R patients had low baseline inflammation and CD8 T cell responses after anti–PD-1 alone. JAKi.R patients had elevated inflammatory markers, poor CD8 T cell responses, and blunted immune signaling after anti–PD-1 alone. However, after addition of itacitinib, subsequent clinical responses in JAKi.R patients were associated with decreased inflammatory signaling accompanied by an increase in a "fate-flexible" CD8 T cell progenitor-like population. This fate-flexible population was linked to features of greater CD8 T cell plasticity. By contrast, NR patients had high baseline inflammation refractory to JAK inhibition. This persistent inflammatory and IFN-I signaling in NR patients was associated with CD8 T cell terminal differentiation and treatment failure.

CONCLUSION: As a therapeutic strategy to block the immunosuppressive effects of persistent IFN and/or chronic inflammation, JAK inhibition after initial anti–PD-1 is safe, feasible, and associated with durable and high response rates in NSCLC. JAK inhibition may be particularly beneficial in patients with elevated inflammation who have poor CD8 T cell responses to anti–PD-1 alone. In this study, JAK inhibition enhanced CD8 T cell plasticity by decreasing the inflammatory and IFN-I signals that drive terminal differentiation of CD8 T cells. However, some patients with the highest baseline levels of inflammation were largely unaffected by JAK1 inhibition and had progressive terminal CD8 T cell differentiation and disease progression. Our findings suggest that JAK inhibition can target chronic immunoregulatory functions of cytokine signaling that contribute to relapse during cancer immunotherapy and is a strategy warranting further preclinical and clinical investigation. ■



JAK inhibitors to improve clinical and immune response to anti–PD-1 for lung cancer patients. In a clinical trial of patients with NSCLC treated with anti–PD-1, mitigating persistent IFN-I and inflammation with a JAK inhibitor decreased terminal differentiation of CD8 T cells and expanded fate-flexible CD8 T cell progenitors in responding patients. Patients with progressive disease had high baseline inflammation that minimally changed with JAK1 inhibition.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: andyminn@upenn.edu (A.J.M.); wherry@pennmedicine.upenn.edu (E.J.W.)
†These authors contributed equally to this work.
Cite this article as D. Mathew *et al.*, *Science* **384**, eadf1329 (2024). DOI: 10.1126/science.adf1329

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf1329

RESEARCH ARTICLE SUMMARY

CLINICAL TRIALS

JAK inhibition enhances checkpoint blockade immunotherapy in patients with Hodgkin lymphoma

Jaroslav Zak*†, Isaraphorn Pratumchai†, Brett S. Marro†, Kristi L. Marquardt, Reza Beheshti Zavareh, Luke L. Lairson, Michael B. A. Oldstone, Judith A. Varner, Livia Hegerova, Qing Cao, Umar Farooq, Vaishalee P. Kenkre, Veronika Bachanova*‡, John R. Teijaro*‡

INTRODUCTION: Checkpoint immunotherapy has revolutionized cancer treatment in the past decade for many cancer patients. Despite durable responses in some patients, many either fail to respond or develop resistance to the current immunotherapy regimens. Thus, a substantial unmet clinical need remains in nonresponsive patients to reinvigorate checkpoint immunotherapy responses.

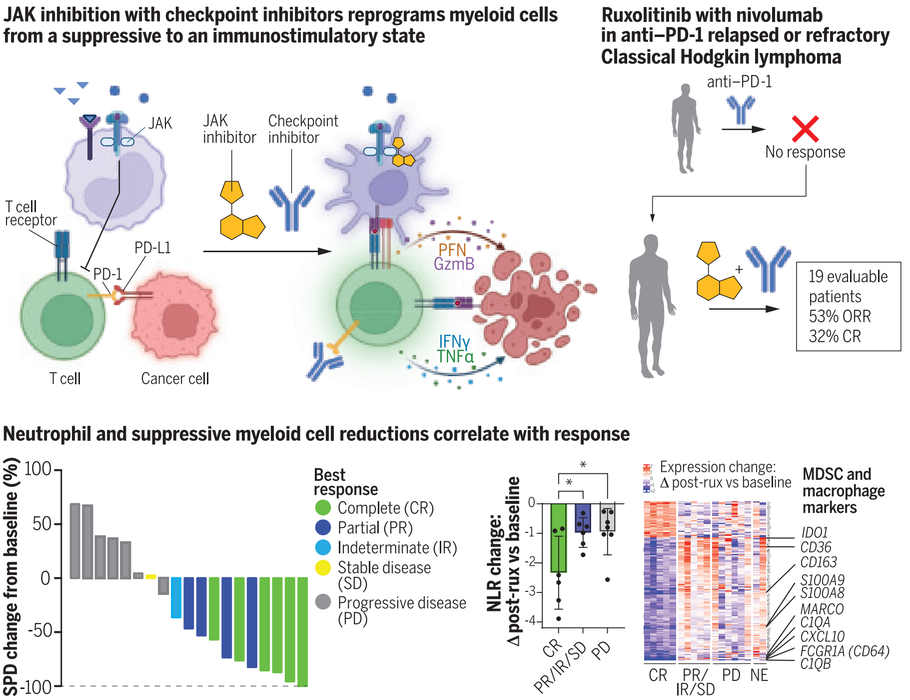
RATIONALE: T cell exhaustion limits current responses to immunotherapy. Using the results of a high-throughput screen for small molecules reversing T cell exhaustion, we explored the hypothesis that a small molecule could enhance antitumor T cell responses and improve the efficacy of immune checkpoint inhibitor (ICI) immunotherapy in cancer.

RESULTS: Following the screening of the ReFrame library in the lymphocytic choriomeningitis virus (LCMV) Clone 13 (Cl13) model of immune suppression, we identified a preponderance of Janus kinase inhibitors (JAKis), which could dramatically rescue the function of exhausted T cells. Using the first clinically approved JAKi, ruxolitinib, we demonstrated that, rather than suppress the immune response, JAK inhibition could enhance T and NK cell numbers and function in Cl13 as well as syngeneic tumor models and enhance tumor control in combination with ICI immunotherapy. Notably, this combination was effective in multiple solid tumor and lymphoma models, including those without prior ICI resistance and those in which ruxolitinib monotherapy had no effect on tumor growth.

Treatment of both Cl13-infected and tumor-bearing mice with JAKis also modulated the transcriptomic and functional properties of myeloid cells from an immune suppressive state to one with immune-stimulatory potential. The ability of ruxolitinib to enhance checkpoint inhibitor efficacy was conditional on the presence of myeloid cells, and myeloid cells in ruxolitinib and ICI–treated mice showed up-regulation of antigen presentation molecules, including major histocompatibility complex class II (MHC-II). These results demonstrate that ruxolitinib's ability to enhance antitumor T cell responses is partly T cell extrinsic and dependent on myeloid cell modulation.

We report the results of a phase I clinical trial in patients with Hodgkin lymphoma who relapsed or were refractory after prior checkpoint inhibitor immunotherapy. Ruxolitinib was administered in combination with nivolumab. The combination yielded a best overall response rate of 53% with 6/19 patients achieving a complete metabolic response to therapy. Clinical response correlated with reductions in neutrophil-to-lymphocyte ratios (NLRs), percentages of suppressive myeloid cells, and increases in the numbers of cytokine-producing T cells after ruxolitinib treatment. In both preclinical models and patient samples, the transcriptomic signature of granulocyte colony-stimulating factor (G-CSF) (a JAK-dependent cytokine) signaling was significantly down-regulated by ruxolitinib, suggesting that ruxolitinib may prevent suppressive programming of myeloid cells owing to excessive JAK–signal transducer and activator of transcription (STAT) signaling by cytokines such as G-CSF.

CONCLUSION: Our results support the therapeutic potential of small-molecule JAK inhibition in combination with checkpoint inhibitors in cancer. We demonstrate that JAK inhibition is not immune suppressive but rather stimulatory, suggesting that the context of JAK inhibitor usage dictates the ultimate treatment outcome. Given that the immunomodulatory properties and clinical efficacy were not exclusive to Hodgkin lymphoma, it will be important to explore the clinical potential of JAK inhibitors with ICI in solid tumors, particularly those in which suppressive myeloid cells correlate with poor response to ICI monotherapy.



JAK inhibition reprograms myeloid cells to enhance checkpoint inhibitor immunotherapy. Combination of the JAKi ruxolitinib with anti–PD-1 antibody nivolumab promotes MHC-II^{+} immunostimulatory and reduces suppressive myeloid cells, enhancing T cell division and function. In a clinical trial of patients with Hodgkin lymphoma relapsed or refractory following previous anti–PD-1 therapy, ruxolitinib with nivolumab achieved clinical responses that correlated with reduction in suppressive myeloid cells. ORR, overall response rate; CR, complete response; NE, nonevaluable; SPD, sum of the product of diameters; MDSC, myeloid-derived suppressor cell; rux, ruxolitinib. The upper left image was created with BioRender.com.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jaroslav@scripps.edu (J.Z.); bach0173@umn.edu (V.B.); teijaro@scripps.edu (J.R.T.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as J. Zak *et al.*, *Science* 384, eade8520 (2024). DOI: 10.1126/science.ade8520

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.ade8520

RESEARCH ARTICLE SUMMARY

DRUG DISCOVERY

AlphaFold2 structures guide prospective ligand discovery

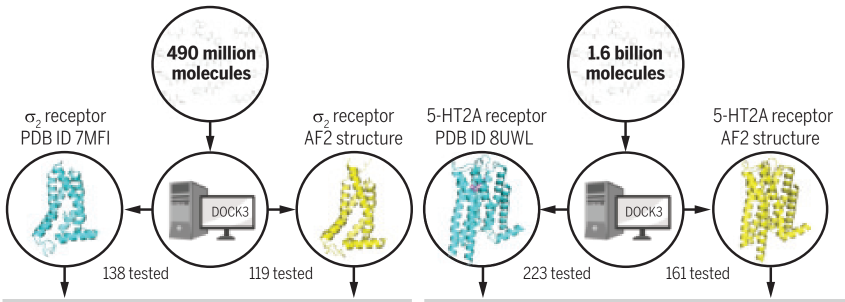
Jiankun Lyu*†, Nicholas Kapolka†, Ryan Gumpper†, Assaf Alon†, Liang Wang†, Manish K. Jain, Ximena Barros-Álvarez, Kensuke Sakamoto, Yoojoong Kim, Jeffrey DiBerto, Kuglae Kim, Isabella S. Glenn, Tia A. Tummino, Sijie Huang, John J. Irwin, Olga O. Tarkhanova, Yurii Moroz, Georgios Skiniotis*, Andrew C. Kruse*, Brian K. Shoichet*, Bryan L. Roth*

INTRODUCTION: Deep-learning methods to predict protein structures, like AlphaFold2 (AF2) and RosettaFold, have had great impact on structural biology, but their influence on drug discovery is less clear. Recent retrospective docking studies suggest that AF2 models struggle to recapitulate ligand binding modes and to distinguish active from decoy molecules in ligand discovery simulations, compared with the same calculations on experimental structures. Still, these studies are retrospective; how AF2 models perform prospectively for predicting new ligands has not, to our knowledge, been explored.

RATIONALE: To address this prospective gap, we selected two therapeutic targets for which the AF2 models appeared before the experimental structures were released: the σ_2 and serotonin 2A (5-HT2A) receptors. Whereas the AF2 model resembled experimental structures overall, there were meaningful conformational differences in residues at the ligand binding sites, particularly for the 5-HT2A receptor. To prospectively test the relative ability of the AF2 models to guide the discovery of new ligands, we docked libraries of hundreds of millions to billions of molecules against both the modeled and the experimental structures, prioritizing for synthesis and testing hundreds of high-ranking molecules for each model and structure. We assessed the performance of the AF2 structures versus the experimental structures by hit rate (number experimentally active per number tested) and by hit potency.

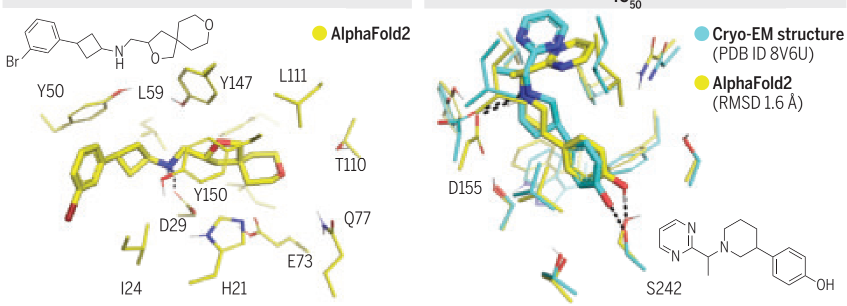
RESULTS: Surprisingly, prospective docking against the AF2 models was as effective as it was for docking against the experimental structures. For the σ_2 receptor, 55% of the AF2-derived docking hits were active at 1 μM, whereas docking against the crystal structure led to a 51% hit rate. For the 5-HT2A receptor, 26% of the molecules from the AF2-derived model bound at 10 μM, whereas for the cryo–electron microscopy (cryo-EM) structure, 23% bound. Comparing the affinities of these hits yielded similar conclusions. Against the σ_2 receptor, the top 18 hits from the AF2 docking had inhibition constant (K_i) values between 1.6 and 84 nM, similar to the distribution from docking against the receptor crystal structure. Against the 5-HT2A receptor, the AF2 model led, if anything, to more potent and selective compounds compared with docking against the experimental structure. The most potent AF2-derived agonists had median effective concentration (EC_{50}) values ranging from 42 nM to 1.6 μM, whereas the cryo-EM–derived docking hits had EC_{50} values ranging from 246 nM to 3 μM. Three of the AF2-derived docking hits were subtype selective, whereas the cryo-EM–derived docking hits were not. A cryo-EM structure of an AF2-derived agonist bound to the 5-HT2A receptor superposed well with the docking prediction and supported several residue conformations anticipated by the AF2 model that differed from those observed in the original experimental structure.

CONCLUSION: Differences in the ligand binding sites between AF2 models and experimental structures may reduce the ability of the models to recognize known ligands. For a subset of AF2 models, however, these differences may represent low-energy, alternate receptor conformations that can guide the discovery of new ligands just as well as experimental structures can, potentially expanding the range of proteins that may be targeted for structure-based drug discovery. ■



51%	Hit rate	54%
6	Number of hits K_i < 5 nM	2
19	Number of hits K_i < 50 nM	13
1.8 nM	Best affinity	1.6 nM

23%	Hit rate	26%
71 to 114 nM	Affinity K_i	15 to 24 nM
246 nM to 3 μM	Agonist EC_{50}	42 nM to 1.6 μM
NA	Antagonist IC_{50}	907 nM to 6.2 μM



AF2 structures template prospective ligand discovery. More than 490 million and 1.6 billion molecules were docked against the AF2 and experimental structures for the σ_2 and 5-HT2A receptors, respectively. More than 100 high-ranking molecules were prioritized for synthesis and testing for each model and structure (about 500 total molecules). Judging by hit rate and hit affinity, docking against the AF2 models was as effective as docking against the experimental structures. The cryo-EM structure of the agonist Z7757 in complex with the 5-HT2A receptor superposed closely with the computational docking prediction on the AF2 model. IC_{50}, median inhibitory concentration; NA, not applicable; PDB, Protein Data Bank; RMSD, root mean square deviation. Single-letter abbreviations for the amino acid residues are as follows: D, Asp; E, Glu; H, His; I, Ile; L, Leu; Q, Gln; S, Ser; T, Thr; and Y, Tyr.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jiankunlyu@gmail.com (J.L.); yiorgo@stanford.edu (G.S.); andrew_kruse@hms.harvard.edu (A.C.K.); bshoichet@gmail.com (B.K.S.); bryan_roth@med.unc.edu (B.L.R.)
†These authors contributed equally to this work.
Cite this article as J. Lyu *et al.*, *Science* **384**, eadn6354 (2024). DOI: 10.1126/science.adn6354

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adn6354

RESEARCH ARTICLE SUMMARY

CANCER

Breast cancer exploits neural signaling pathways for bone-to-meninges metastasis

Andrew E. Whiteley†, Danhui Ma†, Lihua Wang, Seok-Yeong Yu, Claire Yin, Trevor T. Price, Brennan G. Simon, Katie R. Xu, Kathleen A. Marsh, Maegan L. Brockman, Tatiana M. Prioleau, Katherine I. Zhou, Xiuyu Cui, Peter E. Fecci, William R. Jeck, Chad M. McCall, Jadee L. Neff, Dorothy A. Sipkins*

INTRODUCTION: Our group previously discovered that hematopoietic-lineage cells can traffic from the vertebral or skull bone marrow (BM) to the leptomeninges (LM) by migrating along the laminin-rich external surface of emissary veins (EVs). These EVs, part of the vertebral-calvarial BM vasculature, pass through foramina in the bone surface and emerge as LM blood vessels, thereby directly connecting the BM compartment to the central nervous system (CNS). This abluminal vascular migration route bypasses the need for cells to enter the circulation, cross the CNS blood-brain or blood–cerebrospinal fluid (CSF) barriers, or disrupt tissue boundaries.

The BM is the most frequent site of breast cancer (BC) metastasis, and most BC patients diagnosed with leptomeningeal disease (LMD) have established vertebral bone metastases. This observation suggests that EV trafficking could be an efficient mechanism of entry for BC cells (BCCs) into the LM; however, EV trafficking by carcinoma cells has not been previously demonstrated.

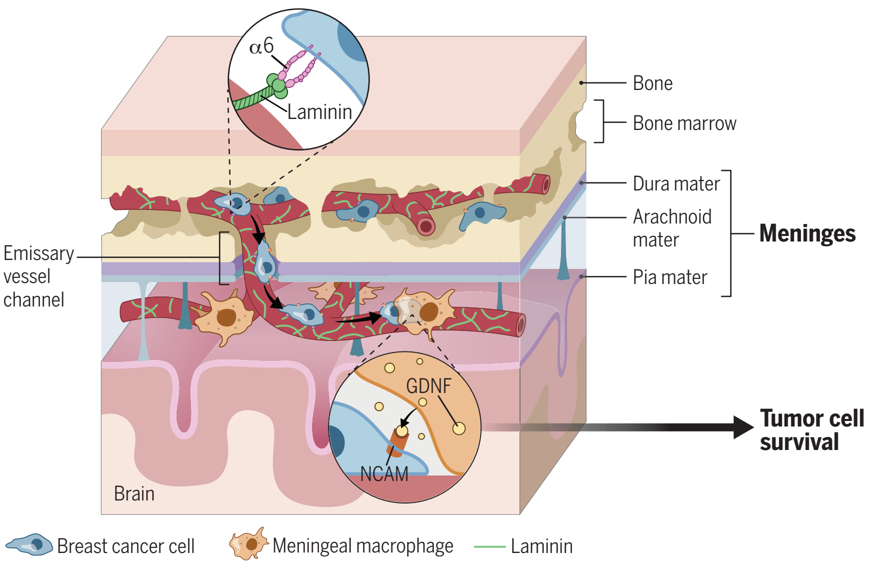
In contrast to the nutrient-rich BM milieu, the nutrient-poor LM are a harsh environment for survival. The adaptive mechanisms that allow tumors to thrive under cellular stressors within this niche are little understood. Furthermore, although the LM harbor a relative paucity of immune cells, they contain numerous resident macrophage populations. Macrophages are well-described to play protumoral roles in many tissues; yet, how BCCs might subvert meningeal macrophages to enhance their survival in the LM is only beginning to be investigated.

RATIONALE: Metastases to the LM—the CSF-containing membranes surrounding the brain and spinal cord—occur in a wide variety of hematologic and solid malignancies, including leukemia, lymphoma, BC, lung cancer, and melanoma. When LM metastases arise, they are often rapidly fatal. The molecular mechanisms that enable LM metastasis are poorly understood, and there are limited interventions to prevent or treat this deadly disease complication.

RESULTS: By applying a combination of intravital and ex vivo three-dimensional confocal microscopy, micro–computed tomography (micro-CT), and histologic analyses to mouse models of bone-metastatic BC and LM metastasis, we demonstrated that BCCs could traffic to the LM from the BM through abluminal EV migration. We also found that BC cell surface expression of integrin α6, a laminin receptor, was essential for this process. Engraftment of mice with BC cells with CRISPR-mediated deletion of α6 inhibited LM colonization, decreased LMD development, and prolonged survival. Conversely, induced expression of α6 in BCCs increased LMD.

Imaging also demonstrated that the majority of BCCs colocalized with macrophages after entering the LM and that their presence stimulated macrophage secretion of the neuronal prosurvival molecule, glial-derived neurotrophic factor (GDNF). GDNF is minimally expressed in the healthy adult brain and LM but is secreted by reactive CNS microglia and macrophages in response to brain injury, where it is deposited in the extracellular matrix and serves to block apoptotic neuronal stress responses. Echoing this role in neurons, we found that BCCs that express the GDNF receptor, neural cell adhesion molecule (NCAM), can transduce antiapoptotic signaling that enhances their survival amid nutrient deprivation. Intrathecal GDNF blockade, macrophage-specific GDNF ablation, or deletion of NCAM from BCCs inhibited BC growth within the LM. Lastly, immunohistochemical analysis of patient samples showed that α6 expression was associated with meningeal-based metastases and that these metastases were highly enriched for BC-NCAM and stromal-GDNF expression.

CONCLUSION: Our data provide evidence that BCCs hijack a hematopoietic migration pathway to enter the LM via the BM. We also show that BCCs, mirroring neurons under stress, co-opt meningeal macrophages to aid their survival. Our findings describe previously unrecognized roles for GDNF and integrin α6 signaling in promoting BC LMD and thus provide the foundation for predictive, preventative, and therapeutic approaches to BC LM metastasis management. ■



BCC metastasis and prosurvival interactions in the meninges. BCCs that express the integrin α6 subunit, part of a cell surface receptor enabling adhesion to the matrix molecule laminin, can migrate along the laminin-rich exterior surface of BM-LM bridging emissary vessels to invade the LM. BCCs then colocalize with meningeal macrophages to stimulate secretion of GDNF, which binds BC NCAM receptors and activates survival signaling pathways.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dorothy.sipkins@duke.edu
†These authors contributed equally to this work.
Cite this article as A. E. Whiteley *et al.*, *Science* **384**, eadh5548 (2024). DOI: 10.1126/science.adh5548

READ THE FULL ARTICLE AT https://doi.org/10.1126/science.adh5548

RESEARCH ARTICLES

NANOMATERIALS

Carbon nanotube fibers with dynamic strength up to 14 GPa

Xinshi Zhang[1,2]†, Xudong Lei[3,4]†, Xiangzheng Jia[5]†, Tongzhao Sun[2,6], Jiajun Luo[1,2], Shichen Xu[1,2], Lijun Li[2], Dan Yan[1,2], Yuanlong Shao[1,2], Zhenzhong Yong[7,8,9], Yongyi Zhang[7,8,9]*, Xianqian Wu[3,4]*, Enlai Gao[5]*, Muqiang Jian[2]*, Jin Zhang[1,2]*

High dynamic strength is of fundamental importance for fibrous materials that are used in high–strain rate environments. Carbon nanotube fibers are one of the most promising candidates. Using a strategy to optimize hierarchical structures, we fabricated carbon nanotube fibers with a dynamic strength of 14 gigapascals (GPa) and excellent energy absorption. The dynamic performance of the fibers is attributed to the simultaneous breakage of individual nanotubes and delocalization of impact energy that occurs during the high–strain rate loading process; these behaviors are due to improvements in interfacial interactions, nanotube alignment, and densification therein. This work presents an effective strategy to utilize the strength of individual carbon nanotubes at the macroscale and provides fresh mechanism insights.

Ultrahigh dynamic strength and energy-absorbing fibrous materials are needed in high–strain rate applications, such as ballistic impact and untraceable debris impact on aircraft and spacecraft (*1*, *2*). The impact resistance of fibrous materials is closely related to the hierarchical structures, starting from individual building blocks at the nanoscale, then moving to microfibrils, and finally to the macroscopic ensembles, which provide various methods to dissipate the mechanical energy. A few high-performance fibers, such as ultrahigh–molecular weight polyethylene (UHMWPE), poly(*p*-phenylene-2,6-benzobisoxazole) (PBO), and aramid fibers (*3*, *4*), have been developed and play vital roles in modern industries. Despite these achievements, the fabrication of higher-performance fibers is still driven by increasing industrial demands. One promising route is the assembly of ultrahigh-performance nanomaterials into macroscopic architectures (*5*, *6*).

Carbon nanotubes (CNTs) with an intrinsic strength of more than 100 GPa (*7*, *8*) have been considered as promising building blocks for constructing high-performance and multifunctional fibers for applications in both quasi-static and dynamic environments (*2*, *9*–*13*). For example, Xie *et al*. (*2*) demonstrated the potential applications of CNT fibers (CNTFs) for high–strain rate environments using a stroboscopic quantification method. Three spinning methods are used to fabricate CNTFs, including wet spinning (*14*, *15*), vertically aligned array spinning (*16*), and direct aerogel spinning (*17*). Among them, direct aerogel spinning is promising for the continuous and scalable fabrication of ultrastrong fibers consisting of high–aspect ratio CNTs (*18*–*20*). However, the quasi-static and dynamic mechanical properties of CNTFs are limited by the poor interfacial interactions, low nanotube alignment, and high porosity formed in the spinning process (*21*–*24*). To solve these issues, various posttreatment approaches have been developed to modify the hierarchical structures of CNTFs (*25*, *26*), including solution densification (*27*, *28*), mechanical treatment (*20*), and thermal annealing (*29*). However, the tensile strength of CNTFs (<10 GPa), especially the dynamic strength, is far lower than that of individual CNTs (>100 GPa), indicating that there is still plenty of room for improving the strength of CNTFs.

We developed a strategy that includes progressive stretching, infusion with PBO nanofibers and molecular chains (hereafter PBOs), and mechanical rolling to improve the interfacial interactions, nanotube alignment, and densification of CNTFs (Fig. 1A and fig. S1). Briefly, functionalized CNTFs (F-CNTFs) are first immersed in chlorosulfonic acid (CSA) solution containing PBOs and swell visibly owing to the protonation effect (*30*). Afterward, PBOs are infused into F-CNTFs during the progressive stretching treatment (PBO-CNTFs). Finally, mechanical rolling is used to densify PBO-CNTFs (D-PBO-CNTFs) (*31*). This approach can be used to continuously produce high-performance tows (Fig. 1B and fig. S2). Our experimental characterizations demonstrate that the as-obtained fibers have a highly ordered and densely packed structure with strong interfacial interactions (Fig. 1, C and D). Consequently, D-PBO-CNTFs exhibit a high quasi-static tensile strength of 8.2 ± 0.2 GPa and a toughness of 170.3 ± 17.9 MJ m^{-3}, yielding a 355% increase in the tensile strength and a 106% increase in the toughness compared with F-CNTFs (Fig. 1E). The resultant fibers exhibit a dynamic strength of 14.0 ± 0.7 GPa and a toughness of 462.6 ± 102.1 MJ m^{-3} under a high strain rate of about 1400 s^{-1}, values that are higher than those of commercial fibers (e.g., 7.2 ± 0.8 GPa and 295.0 ± 59.5 MJ m^{-3} for PBO fibers, 4.6 ± 0.2 GPa and 174.6 ± 11.6 MJ m^{-3} for Kevlar 29 fibers tested under the same conditions as D-PBO-CNTFs; Fig. 1E).

Fabrication and structure of CNTFs

We first produced continuous CNTFs by floating catalyst chemical vapor deposition (fig. S3A) (*17*). The as-spun fibers contained impurities such as metallic catalyst particles and amorphous carbon (fig. S3B). To purify and functionalize these fibers with oxygen-containing functional groups (e.g., hydroxy groups), we adopted the posttreatments of weak oxidation and acid washing (fig. S1A). Through these treatments, thermogravimetric analyses demonstrated that most of the residual impurities in F-CNTFs had been removed (fig. S4), and Raman spectra and x-ray photoelectron spectroscopy (XPS) of F-CNTFs confirmed the functionalization of CNTs (figs. S5 and S6 and table S1) (*31*). Compared with the pristine fibers, F-CNTFs produced by the purification process exhibited a more compact structure (fig. S7), and the as-obtained fibers subjected to a 12-hour oxidation treatment exhibited an improved tensile strength (fig. S8 and table S1); these fibers were chosen as the raw fibers for further optimization in the following investigation.

When immersed in a CSA bath, F-CNTFs swell by side-wall protonation (fig. S9), which offers plenty of room to rearrange the entangled and porous networks for the improvements of alignment and densification of fibers. By subjecting F-CNTFs to CSA treatment for 10 min with a stretching ratio of 20%, the as-obtained CSA-CNTFs exhibited a substantial

[1]Beijing National Laboratory for Molecular Sciences, School of Materials Science and Engineering, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. [2]Beijing Graphene Institute (BGI), Beijing 100095, China. [3]Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China. [4]School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China. [5]Department of Engineering Mechanics, School of Civil Engineering, Wuhan University, Wuhan 430072, China. [6]State Key Laboratory of High-efficiency Coal Utilization and Green Chemical Engineering, College of Chemistry and Chemical Engineering, Ningxia University, Yinchuan 750021, China. [7]Key Laboratory of Multifunctional Nanomaterials and Smart Systems, Advanced Materials Division, Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Suzhou 215123, China. [8]Division of Nanomaterials and Jiangxi Key Lab of Carbonene Materials, Jiangxi Insitute of Nanotechnology, Nanchang 330200, China. [9]School of Nano-Tech and Nano-Bionics, University of Science and Technology of China, Hefei 230026, China.
*Corresponding author. Email: yyzhang2011@sinano.ac.cn (Y.Z.); wuxianqian@imech.ac.cn (X.W.); enlaigao@whu.edu.cn (E.G.); jianmq-cnc@pku.edu.cn (M.J.); jinzhang@pku.edu.cn (J.Z.)
†These authors contributed equally to this work.

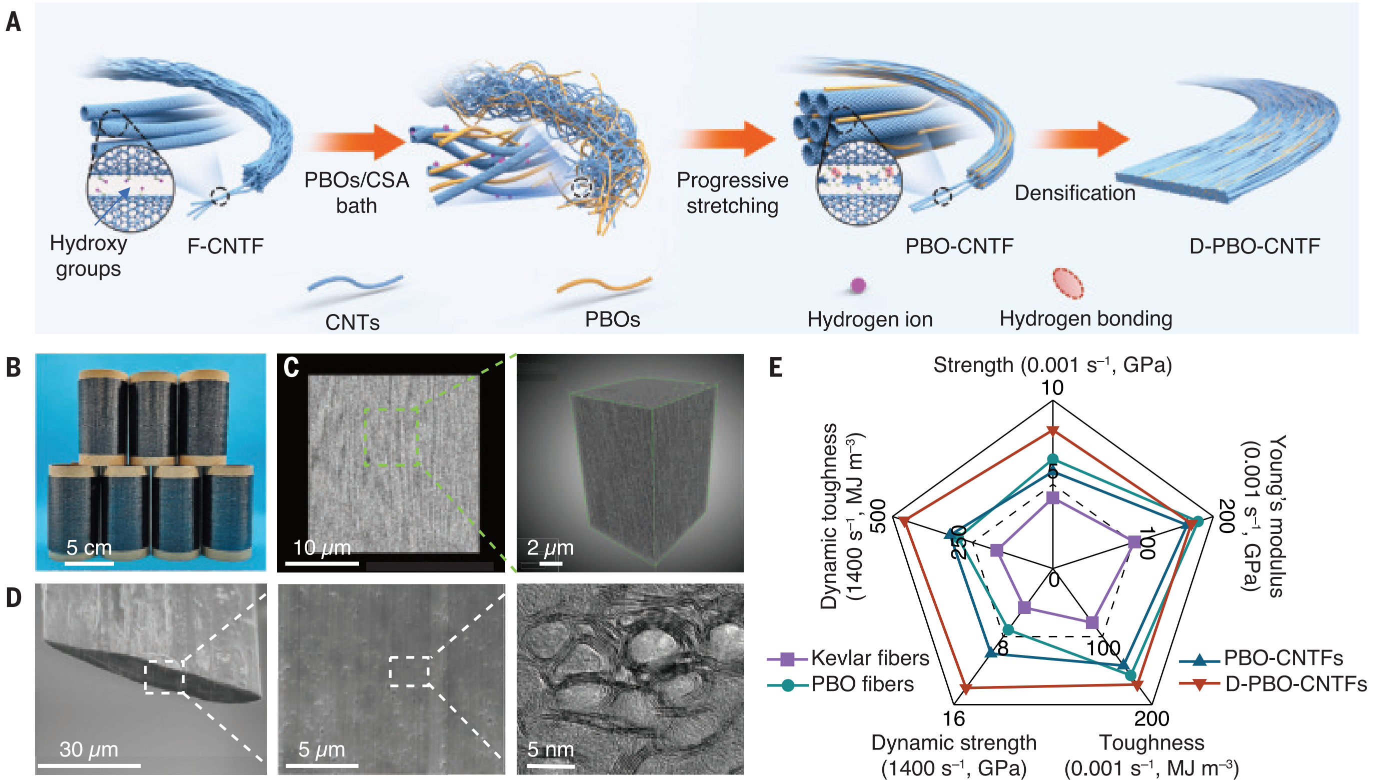


Fig. 1. Preparation, morphology, and mechanical properties of CNTFs. (**A**) Strategy to develop highly packed and well-aligned CNTFs. (**B**) Digital photograph of PBO-CNTF tows. (**C**) Three-dimensionally reconstructed void microstructure (right) derived from nano-CT results (left) for D-PBO-CNTFs. (**D**) SEM (left and middle) and TEM (right) images of the radial cross section of D-PBO-CNTF cut by a focused ion beam. (**E**) Radar chart for comparing the mechanical performance of different CNTFs and commercial fibers (PBO and Kevlar 29 fibers).

improvement in tensile strength (figs. S10 and S11 and tables S2 and S3). Furthermore, a progressive stretching treatment was introduced. After optimizing the drawing speeds under a certain stretching process, longitudinal scanning electron microscopy (SEM) and transmission electron microscopy (TEM) images showed that the CNTs yield an increase in the alignment along the fiber axial direction (fig. S12). Herman's orientation factor (*f*) values measured by wide-angle x-ray scattering (WAXS) also support an improvement in the alignment of CNTs. The *f* for CSA-CNTFs obtained at an optimized stretching rate of 3% per min was 0.90, higher than that of F-CNTFs (0.66) and other fibers prepared at higher drawing speeds (fig. S13 and table S4) (*31*). This might be because the mechanical treatment at a low drawing speed provides adequate time to disentangle and reorient the CNTs, resulting in the enhancement of tensile strength, Young's modulus, and toughness (fig. S14 and table S5). Solely progressive stretching treatment cannot fully eliminate the voids, as shown in the CSA-CNTFs (fig. S16A). To address this issue, PBOs were infused into CNTFs by uniformly dissolving PBO fibers in a CSA bath at different weight percentages (wt %) (fig. S15) during the progressive stretching treatment. Both the elemental analyses and thermogravimetric analyses demonstrated that PBO-CNTFs treated in the PBOs-CSA solution with a PBO concentration of 0.05 wt % exhibited the highest PBO content (fig. S17 and table S6). Such PBO-CNTFs showed high densification (fig. S16), alignment (Fig. 2A and fig. S18), and mechanical properties (fig. S19 and table S7). Optimal D-PBO-CNTFs were fabricated by applying mechanical rolling on these PBO-CNTFs.

We compared the alignment, densification, and interfacial interactions of F-CNTFs, CSA-CNTFs, PBO-CNTFs, and D-PBO-CNTFs. First, we investigated the alignment of these fibers. The orientation factors measured by WAXS for CSA-CNTFs (0.90), PBO-CNTFs (0.94), and D-PBO-CNTFs (0.92) were higher than for F-CNTFs (0.66), which is consistent with results of polarized Raman spectra (fig. S20). This indicates that the PBOs-assisted progressive stretching treatment can effectively improve the alignment of CNTs along the fiber axis (Fig. 2, B to D, and table S8). Second, the densification of these fibers was evaluated by small-angle x-ray scattering (SAXS) (*32*), which indicated that D-PBO-CNTFs have the densest structure (fig. S21). We also reconstructed the microstructures of fibers using nanoscale x-ray computed tomography (nano-CT) and measured the porosities accordingly (Fig. 2, F and H, fig. S22, and movies S1 to S4). The porosities of F-CNTFs (4.8%), CSA-CNTFs (2.7%), and PBO-CNTFs (1.9%) are higher than that of D-PBO-CNTFs (1.0%). These results are consistent with the cross-sectional SEM images, SAXS characterization, and density measurements (Fig. 2, E and G, and table S9). Third, we explored the interfacial interactions by Fourier transform infrared spectroscopy (FTIR) spectra. The redshift of the hydroxy group peak indicated the hydrogen-bond interactions between CNTs and PBOs (fig. S23A).

Performance of CNTFs

The structural optimization endows the fibers with improvements in quasi-static mechanical properties and electrical conductivities. D-PBO-CNTFs have a tensile strength of 8.2 ± 0.2 GPa, a Young's modulus of 172.7 ± 9.6 GPa, a toughness of 170.3 ± 17.9 MJ m^{-3}, and an electrical conductivity of 2.9×10^6 S m^{-1}, which are 4.6, 1.7, 2.1, and 5.8 times those of F-CNTFs (1.8 ± 0.2 GPa, 99.8 ± 7.3 GPa, 82.5 ± 6.4 MJ m^{-3}, and 0.5×10^6 S m^{-1}; Fig. 3, A and B, and table S9), respectively. F-CNTFs show an intertube slippage failure morphology, whereas D-PBO-CNTFs exhibit a failure morphology with much fewer pull-out bundles (fig. S24).

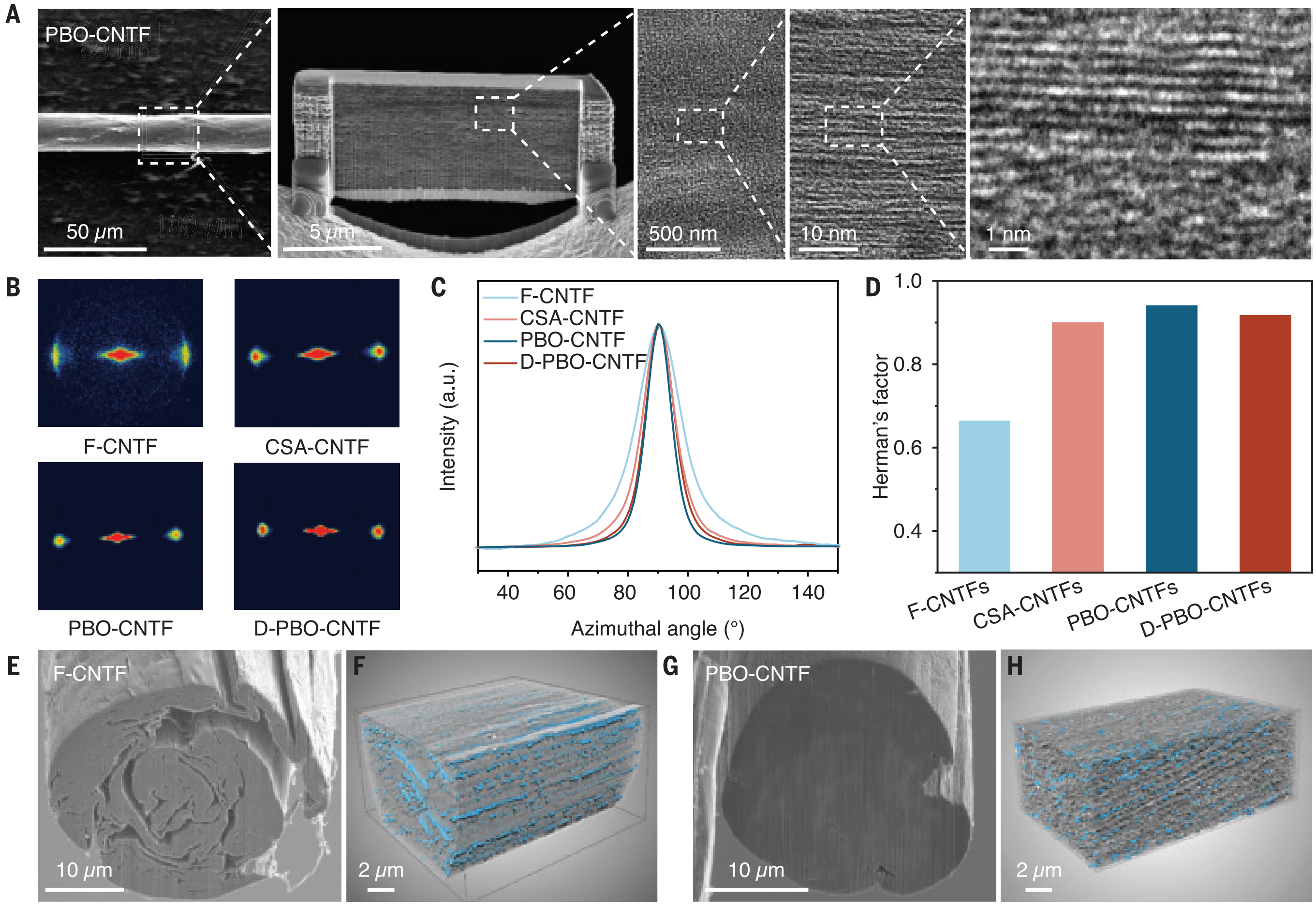


Fig. 2. Structural characterization of CNTFs. (**A**) SEM (two images on the left) and TEM (remaining) images of the axial cross section of PBO-CNTFs cut by a focused ion beam. (**B**) WAXS patterns of F-CNTFs, CSA-CNTFs, PBO-CNTFs, and D-PBO-CNTFs. (**C**) Azimuthal intensity profile of different fibers. a.u., arbitrary units. (**D**) Comparison of the alignment of CNTs within fibers. (**E** and **G**) SEM images of cross sections cut by a focused ion beam for F-CNTFs (E) and PBO-CNTFs (G). (**F** and **H**) Three-dimensional void microstructures reconstructed by nano-CT (blue represents the internal voids) for F-CNTFs (F) and PBO-CNTFs (H).

From the ultrahigh quasi-static mechanical properties, we can further estimate the ballistic resistance of fibers via Cunniff velocity (c^*) = $[(\sigma\varepsilon/2\rho)\,(E/\rho)^{1/2}]^{1/3}$ (*4*), where σ, ε, E, and ρ are tensile strength, elongation at break, Young's modulus, and density of fibers, respectively. These data show that D-PBO-CNTFs have the best potential for the application of ballistic-resistant materials (Fig. 3C and table S11). It should be noted that c^* is a rough estimation because it depends on the loading rates and local plastic deformation (*2*). More accurately, the high–strain rate performance of fibers was investigated using a mini-split Hopkinson tension bar (figs. S25 to S27) (*3*, *31*). The shape of stress-strain curves of CNTFs at the high strain rates (Fig. 3D), that is, the stress after the peak value does not drop suddenly, is different from that at quasi-static loading (Fig. 3A). This is because CNTFs have insufficient time to regulate rate-sensitive conformations, such as disentanglement, reorientation, and slippage of nanotubes, thus exhibiting a "cascade-like" breaking of individual CNTs (*22*, *33*). Compared with the quasi-static strength, D-PBO-CNTFs exhibit a dynamic strength of 9.2 ± 0.8, 11.0 ± 0.7, and 14.0 ± 0.7 GPa at strain rates of about 500, 950, and 1400 s^{-1}, which increase by 12.2, 34.1, and 70.7%, respectively. This indicates an improvement in strengthening efficiency as the strain rate increases (table S12). The dynamic strength of D-PBO-CNTFs at a strain rate of about 1400 s^{-1} is 6.1, 2.3, and 1.4 times that of F-CNTFs, CSA-CNTFs, and PBO-CNTFs, respectively (Fig. 3, D and E), and substantially surpasses those of all other high-performance fibers (Fig. 3F and table S12). Meanwhile, the dynamic toughness of D-PBO-CNTFs reaches 462.6 ± 102.1 MJ m^{-3} (Fig. 3G), which exceeds that of other high-performance fibers (*34*, *35*). Furthermore, the fracture morphologies show that the intertube slippage of CNTFs at the high strain rates is inhibited, and PBO-CNTFs and D-PBO-CNTFs exhibit a ductile-to-brittle transition in the fracture mode (fig. S28) (*22*, *31*).

To directly assess the impact resistance of these fibers, we performed laser-induced high-velocity transverse impact testing (Fig. 3H and figs. S31 and S32). The specific energy dissipation power (SEDP) of a single fiber, which is equal to 7.5 × 10^{10} kg^{-1} multiplied by its transverse velocity (c_T), is a figure of merit to evaluate its dynamic energy absorption capacity (figs. S29 and S30) (*2*, *31*), where c_T is the transverse velocity. It is challenging to accurately measure the SEDP of D-PBO-CNTFs with a narrow ribbon-like cross section because it depends on the bending resistance along the impact direction (fig. S33 and table S13). Hence, we only compared the SEDP values of the other three fibers. Among these fibers, PBO-CNTFs have the highest SEDP value [(8.7 ± 1.0) × 10^{13} m kg^{-1} s^{-1}; Fig. 3I and table S14] (*2*). The impact resistance of PBO-CNTFs is attributed to the high longitudinal wave speed. Finite element simulations demonstrated that the high longitudinal wave speed [$c_L = (E/\rho)^{1/2}$ = 11.4 km s^{-1}] helps to delocalize the impact energy (fig. S34, A and B), which is consistent with experimental results (fig. S32 and table S14). After the high-velocity transverse impact, the craters on the surface of PBO-CNTFs are

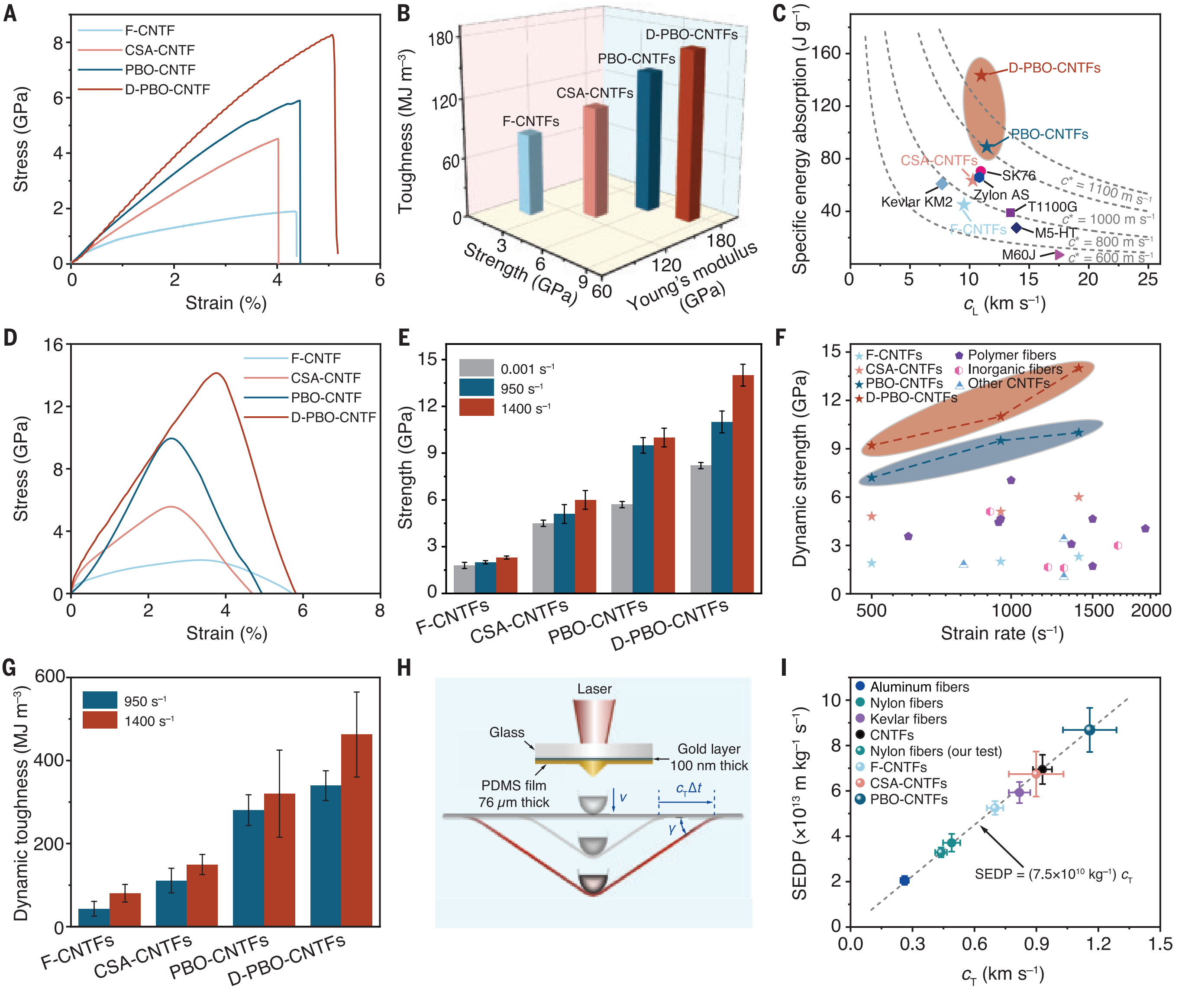


Fig. 3. Mechanical properties of CNTFs. (**A**) Quasi-static stress-strain curves of F-CNTFs, CSA-CNTFs, PBO-CNTFs, and D-PBO-CNTFs. (**B**) Comparison of quasi-static tensile strength, Young's modulus, and toughness of different CNTFs. (**C**) Comparison of specific energy absorption and longitudinal wave velocity of our fibers (indicated by stars) and other high-performance fibers. (**D**) Stress-strain curves of F-CNTFs, CSA-CNTFs, PBO-CNTFs, and D-PBO-CNTFs at high strain rates of about 1400 s^{-1}. (**E**) Comparison of the strength of CNTFs at different strain rates. Error bars indicate the standard deviation of the means of independent measurements. (**F**) Comparison of the dynamic strength of our fibers and other high-performance fibers at high strain rates. (**G**) Comparison of the dynamic toughness of CNTFs at different strain rates. Error bars indicate the standard deviation of the means of independent measurements. (**H**) Schematic diagram of laser-induced high-velocity transverse impact on a single fiber. Here, PDMS, v, Δt, and γ represent polydimethylsiloxane, impact velocity, interval time, and deflection angle, respectively. (**I**) SEDP values of different fibers. Error bars indicate the standard deviation.

not visible compared with those on CSA-CNTFs (fig. S34, C and D), suggesting that the high dynamic strength of PBO-CNTFs helps to maintain the structural integrity.

Strengthening mechanism of CNTFs

To further determine whether covalent bonds are broken, we measured Raman mapping spectra on the impact region of the fibers and observed an increase in the I_D/I_G value (fig. S37), where I_D is the intensity of D-band and I_G is the intensity of G-band, which is a signal of bond breaking or atomic rearrangement that also contributes to dissipating energy (*2*, *36*–*38*). The analyses we describe next show that the simultaneous breakage of CNTs is of central importance to achieving ultrahigh dynamic performance, which results from improvements in interfacial interactions, nanotube alignment, and densification of CNTFs. Regarding the interfacial interactions, we performed experiments using in situ Raman spectroscopy and stress relaxation. When the applied strain of F-CNTFs is less than 1.3%, the G-band frequency in the Raman spectra shows a small downshift. Afterward, a plateau region extends from 1.3 to 2.5% (Fig. 4A), indicating that the further increase of applied strain does not transfer the stress into CNTs. For CSA-CNTFs, because both the alignment and densification are improved compared with F-CNTFs, the plateau region extends from 2.5 to 3.3% (fig. S38A). By contrast, there is a continuous strengthening stress

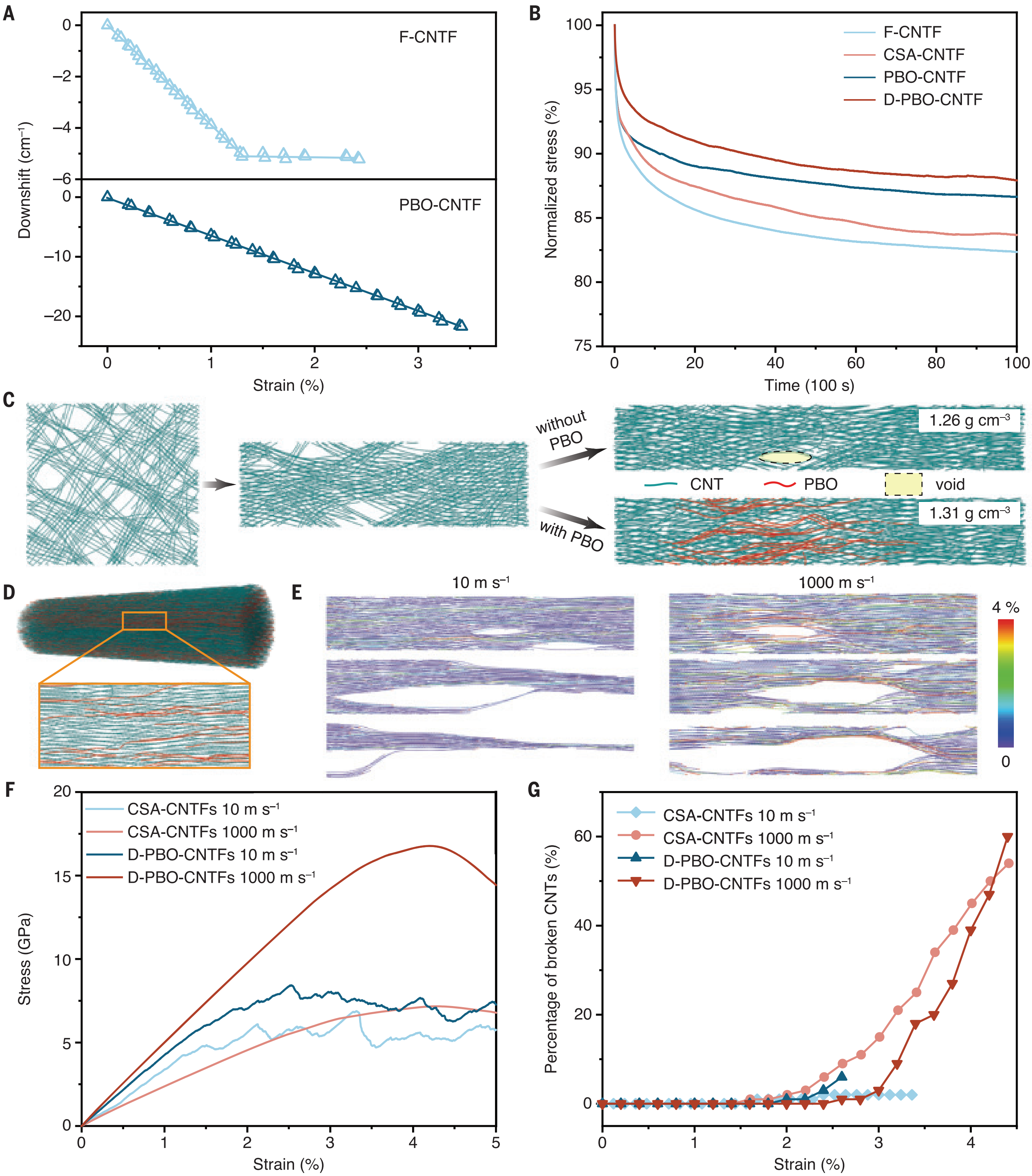


Fig. 4. Mechanistic analyses of the dynamic performance of CNTFs. (**A**) Dependence of Raman frequency downshifts on the applied strains for F-CNTFs and PBO-CNTFs. (**B**) Stress-relaxation curves of different fibers at 1.5% strain. (**C**) Simulation snapshots of progressive stretching treatment for CNTFs with and without PBO. (**D**) Structures and cross-sectional morphology of the simulated model. (**E**) Snapshots of deformed D-PBO-CNTFs under low (left) and high (right) loading velocities. The atoms are colored according to the bond strain. (**F**) Stress-strain curves of CSA-CNTFs and D-PBO-CNTFs under low and high loading velocities. (**G**) Percentage of broken CNTs for CSA-CNTFs and D-PBO-CNTFs during the tensile process, which is counted up to the peak stress in the stress-strain curves.

transfer to CNTs over the whole strain range for PBO-CNTFs and D-PBO-CNTFs, and the downshift per unit strain is about twice as high in PBO-CNTFs and D-PBO-CNTFs compared with F-CNTFs and CSA-CNTFs (Fig. 4A and fig. S38). Meanwhile, D-PBO-CNTFs show the highest resistance to stress relaxation (Fig. 4B). The increased load transfer efficiency suggests that there are strong interfacial interactions between PBOs and CNTs. We also investigated the interfacial interactions by atomistic simulations. At the nanoscale, our calculations show that the persistence lengths of multiwalled CNTs and PBOs are 1.9×10^6 and 38 nm, respectively, which are generally consistent with prior reports (*39*, *40*). These results indicate that PBOs are much more flexible than CNTs. The discrepancy in flexibility between PBOs and CNTs implies that PBOs would preferentially adsorb onto the rigid CNTs (figs. S39 and S40). The combination of relatively flexible PBOs and relatively rigid CNTs would result in an increased effective contact area. Meanwhile, the interfacial energy between PBOs and CNTs (71.2 meV per atom) is also higher than that between two CNTs (55.5 meV per atom). Consequently, the increased contact area and interfacial energy account for the improvements in interfacial interactions and load transfer efficiency (fig. S41).

We explored the effect of nanotube alignment and densification on the behavior of CNTFs using coarse-grained molecular dynamics (CGMD) simulations. The simulations of progressive stretching treatment show that the fibers undergo a reorganization of microstructures—including curving, reorientation, straightening, and sliding of nanotubes and bundles—toward a more uniform, compact, and ordered structure (Fig. 4C and fig. S42). As shown in Fig. 4C, the voids are observed in CNTFs without the addition of PBOs, whereas these voids are filled in CNTFs with the addition of PBOs. This leads to an increase in simulated densities from 1.26 to 1.31 g cm^{-3}, values that are generally consistent with experimental characterizations of CSA-CNTFs (1.24 g cm^{-3}) and PBO-CNTFs (1.28 g cm^{-3}), as summarized in table S9. Meanwhile, the alignment of CNTs in CGMD simulations is also improved from 0.91 for CSA-CNTFs to 0.93 for PBO-CNTFs, which is attributed to an increased load transfer efficiency due to the addition of PBOs. These simulations are generally consistent with our above experimental characterizations of the alignment and densification. Finally, we did tensile tests of fibers under low and high loading velocities (Fig. 4, D to F). The strain distributions show that the strain primarily localizes at the slippage region under the low loading velocity, whereas CNTs deform more uniformly under the high loading velocity (Fig. 4E). Our simulations demonstrate that the tensile strength of D-PBO-CNTFs under low and high loading velocities is 8.4 and 16.3 GPa, respectively (Fig. 4F). To understand the macroscopic performance of CNTFs under low and high loading velocities, we performed analyses by tracking the microscopic behaviors of each CNT within the fibers. Figure 4G shows that under the low loading velocity, only about 2 and 6% of CNTs within CSA-CNTFs and D-PBO-CNTFs, respectively, break (*31*), indicating that the slippage of nanotubes dominates the failure. By comparison, under the high loading velocity, CNTFs have insufficient time to fully regulate rate-sensitive conformations, such as disentanglement, reorientation, and slippage of nanotubes, which results in a high percentage of CNTs breaking. Up to the peak stress, about 54 and 60% of CNTs break for CSA-CNTFs and D-PBO-CNTFs, respectively. The percentage of broken CNTs within CSA-CNTFs is higher than that broken within D-PBO-CNTFs before a strain of about 4%. This is because the inferior structure of CSA-CNTFs leads to stress concentration as well as premature breakage of CNTs. The breakage of CNTs within the D-PBO-CNTFs is much more simultaneous than that within CSA-CNTFs, and the percentage of broken CNTs within the D-PBO-CNTFs becomes larger after a strain of about 4% (Fig. 4G). This simultaneous breakage of CNTs within the D-PBO-CNTFs under a high loading velocity originates from the improvements in interfacial interactions, alignment, and densification, which accounts for the high dynamic strength of these fibers.

Summary

We developed a strategy to fabricate CNTFs with a dynamic strength of 14 GPa. This strategy leads to improvements in interfacial interactions, nanotube alignment, and densification within the fibers. Multiscale analyses combined with experimental evidence revealed that the dynamic performance of CNTFs is primarily due to the simultaneous breakage of individual nanotubes and the exceptional impact-energy delocalization that occurs during the high–strain rate loading process. Our work provides a feasible route to harness the intrinsic strength of individual CNTs at the macroscale to fabricate impact-resistant fibrous materials.

REFERENCES AND NOTES

1. S. Eswarappa Prameela *et al*., *Nat. Rev. Mater.* **8**, 81–88 (2023).
2. W. Xie *et al*., *Nano Lett.* **19**, 3519–3526 (2019).
3. J. Luo *et al*., *Nat. Commun.* **14**, 3019 (2023).
4. Y. Q. Li, H. L. Fan, X. L. Gao, *Compos. Part B* **238**, 109890 (2022).
5. B. Fang, D. Chang, Z. Xu, C. Gao, *Adv. Mater.* **32**, e1902664 (2020).
6. I. A. Kinloch, J. Suhr, J. Lou, R. J. Young, P. M. Ajayan, *Science* **362**, 547–553 (2018).
7. B. Peng *et al*., *Nat. Nanotechnol.* **3**, 626–631 (2008).
8. Y. Bai *et al*., *Science* **369**, 1104–1106 (2020).
9. Y. Wen *et al*., *Nano Lett.* **22**, 6035–6047 (2022).
10. J. Mu *et al*., *Science* **365**, 150–155 (2019).
11. H. Chu *et al*., *Science* **371**, 494–498 (2021).
12. C. Chen *et al*., *Chem. Rev.* **123**, 613–662 (2023).
13. L. W. Taylor *et al*., *Nano Lett.* **21**, 7093–7099 (2021).
14. N. Behabtu *et al*., *Science* **339**, 182–186 (2013).
15. S. G. Kim *et al*., *Carbon* **196**, 59–69 (2022).
16. M. Zhang, K. R. Atkinson, R. H. Baughman, *Science* **306**, 1358–1361 (2004).
17. Y. L. Li, I. A. Kinloch, A. H. Windle, *Science* **304**, 276–278 (2004).
18. H. G. Chae, S. Kumar, *Science* **319**, 908–909 (2008).
19. K. Koziol *et al*., *Science* **318**, 1892–1895 (2007).
20. W. Xu, Y. Chen, H. Zhan, J. N. Wang, *Nano Lett.* **16**, 946–952 (2016).
21. J. J. Vilatela, J. A. Elliott, A. H. Windle, *ACS Nano* **5**, 1921–1927 (2011).
22. D. Y. Wang *et al*., *J. Mech. Phys. Solids* **173**, 105241–105267 (2023).
23. P. F. Wang *et al*., *Carbon* **102**, 18–31 (2016).
24. X. P. Hu *et al*., *Compos. Commun.* **28**, 100933 (2021).
25. Y. Jung, Y. S. Cho, J. W. Lee, J. Y. Oh, C. R. Park, *Compos. Sci. Technol.* **166**, 95–108 (2018).
26. J. S. Bulmer, A. Kaniyoor, J. A. Elliott, *Adv. Mater.* **33**, e2008432 (2021).
27. J. Lee *et al*., *Nat. Commun.* **10**, 2962 (2019).
28. X. Zhang *et al*., *Sci. Adv.* **8**, eabq3515 (2022).
29. D. Lee *et al*., *Sci. Adv.* **8**, eabn0939 (2022).
30. Y. J. Wang, S. Xia, H. Li, J. F. Wang, *Adv. Funct. Mater.* **29**, 1903876 (2019).
31. Materials and methods are available as supplementary materials.
32. J. Qiu *et al*., *ACS Nano* **7**, 8412–8422 (2013).
33. Y. N. Zhang, L. X. Zheng, G. Z. Sun, Z. Y. Zhan, K. Liao, *Carbon* **50**, 2887–2893 (2012).
34. J. Lim, W. W. Chen, J. Q. Zheng, *Polym. Test.* **29**, 701–705 (2010).
35. M. Cheng, W. N. Chen, T. Weerasooriya, *J. Eng. Mater. Technol.* **127**, 197–203 (2005).
36. W. B. Lu, T. Chou, B. Kim, *Phys. Rev. B* **83**, 134113 (2011).
37. N. G. Chopra *et al*., *Nature* **377**, 135–138 (1995).
38. S. Ozden *et al*., *Nano Lett.* **14**, 4131–4137 (2014).
39. N. Fakhri, D. A. Tsyboulski, L. Cognet, R. B. Weisman, M. Pasquali, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14219–14223 (2009).
40. D. B. Roitman, M. McAdon, *Macromolecules* **26**, 4381–4383 (1993).

ACKNOWLEDGMENTS

We thank T. Zhou and S. Qu for providing CNTF samples and participating in discussions. We thank H. Lu for helpful discussions. We are grateful for technical support for Nano-X from Suzhou Institute of Nano-Tech and Nano-Bionics (SINANO), Chinese Academy of Sciences, and Carl Zeiss (Shanghai) Management Co., Ltd. The numerical calculations in this work were done on the supercomputing system in the Supercomputing Center of Wuhan University. **Funding:** This work was supported by Ministry of Science and Technology of China grants 2022YFA1203302 and 2022YFA1203304; National Natural Science Foundation of China grants 52202032, 52021006, T2188101, 52102035, 12272391, and 12232020; Beijing Natural Science Foundation grant 2222094; Strategic Priority Research Program of CAS grant XDB36030100; and Beijing National Laboratory for Molecular Sciences grant BNLMS-CXTD-202001. **Author contributions:** Conceptualization: J.Z., M.J., X.Z.; Methodology: X.Z., X.L., X.J., T.S., J.L., S.X., L.L., D.Y.; Investigation: X.Z., X.L., X.J.; Visualization: X.Z., X.L., and X.J.; Supervision: J.Z., M.J., E.G., X.W., Y.Z.; Writing – original draft: X.Z., X.L., X.J.; Writing – review and editing: J.Z., M.J., E.G., X.W., Y.Z., Y.S., Z.Y. **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.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj1082
Materials and Methods
Supplementary Text
Figs. S1 to S42
Tables S1 to S17
References (*41*–*64*)
Movies S1 to S4

Submitted 7 June 2023; resubmitted 29 February 2024
Accepted 10 May 2024
10.1126/science.adj1082

AIR POLLUTION

Temperature-dependent emissions dominate aerosol and ozone formation in Los Angeles

Eva Y. Pfannerstill[1]*†, Caleb Arata[1], Qindan Zhu[2,3,4]‡, Benjamin C. Schulze[4], Ryan Ward[4], Roy Woods[5], Colin Harkins[3,6], Rebecca H. Schwantes[6], John H. Seinfeld[4], Anthony Bucholtz[5], Ronald C. Cohen[2,7], Allen H. Goldstein[1,8]*

Despite declines in transportation emissions, urban North America and Europe still face unhealthy air pollution levels. This has challenged conventional understanding of the sources of their volatile organic compound (VOC) precursors. Using airborne flux measurements to map emissions of a wide range of VOCs, we demonstrate that biogenic terpenoid emissions contribute ~60% of emitted VOC OH reactivity, ozone, and secondary organic aerosol formation potential in summertime Los Angeles and that this contribution strongly increases with temperature. This implies that control of nitrogen oxides is key to reducing ozone formation in Los Angeles. We also show some anthropogenic VOC emissions increase with temperature, which is an effect not represented in current inventories. Air pollution mitigation efforts must consider that climate warming will strongly change emission amounts and composition.

Ambient air pollution is the fourth-ranking human health risk factor globally (*1*), leading to an estimated 4.2 million premature deaths per year (*2*). Important pollutants causing cardiovascular and respiratory diseases are fine particulate matter ($PM_{2.5}$) and tropospheric ozone (*2*). Volatile organic compounds (VOCs) are precursors to both: A large fraction of $PM_{2.5}$ is secondary organic aerosol (SOA) that forms through the oxidation of VOCs (*3*). In the presence of nitrogen oxides (NO_x) and sunlight, VOC oxidation leads to ozone formation.

Ninety-nine percent of the world's population lives in places where the World Health Organization air quality guidelines are not met (*2*). This includes the US megacity of Los Angeles, where ozone and $PM_{2.5}$ are frequently at unhealthy levels, especially in the summer (*4*, *5*). As in many industrialized cities, technologies such as efficient three-way catalytic converters and efforts spurred by regulation led to a steep decrease in automotive VOC emissions and thus to a decades-long decrease of air pollutant concentrations (*6*). However, concentrations of ozone and $PM_{2.5}$ particle pollution have stopped decreasing since ~2010 (*7*, *8*). Recent studies indicate the increasing relative importance of volatile chemical products, which now contribute as much as half of urban fossil fuel VOC emissions in industrialized cities (*9*, *10*). The relative contribution of the biogenic VOC fraction must also have increased with declining transportation emissions. As a result of these changes in the emitted VOC mixture over the course of a few decades, the sources of secondary air pollution have been called into question. For example, model-observation comparisons have raised doubts on whether the models correctly reproduce the emission source mixture contributing to SOA (*11*, *12*). The origin of SOA in Los Angeles is under debate, with some studies reporting a predominantly vehicular source (*11*, *13*), whereas a temperature-dependent analysis of $PM_{2.5}$ concentrations and isoprene concentrations indicated a major biogenic origin (*8*). The percentage of days on which $PM_{2.5}$ exceeds 12 $\mu g/m^3$ is <10% at 20°C and >40% at 30°C and reaches 70% at 40°C (*8*). Similarly, the likelihood of ozone exceedances (≥70 parts per billion) is close to 0% at 20°C and >70% at temperatures ≥30°C, with temperature-dependent biogenic emissions of reactive terpenoids suggested as one of the driving factors (*14*). However, concentration-based temperature dependencies may be influenced by meteorology instead of emissions because hotter days tend to be more stagnant (*15*). This shows the need for direct emission observations. With climate change, an increase in the number of days with high temperatures is expected (*16*, *17*). Thus, it is important to understand how increasing temperatures affect VOC emission amounts and mixture and what this means for secondary air pollutant formation, as well as regulation strategies.

[1]Department of Environmental Science, Policy and Management, University of California at Berkeley, Berkeley, CA, USA. [2]Department of Earth and Planetary Science, University of California at Berkeley, Berkeley, CA, USA. [3]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA. [4]NOAA Chemical Sciences Laboratory, Boulder, CO, USA. [5]Department of Environmental Science and Engineering, California Institute of Technology, Pasadena, CA, USA. [6]Department of Meteorology, Naval Postgraduate School, Monterey, CA, USA. [7]Department of Chemistry, University of California at Berkeley, Berkeley, CA, USA. [8]Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley, CA, USA.
***Corresponding author. Email: ahg@berkeley.edu (A.H.G.); e.pfannerstill@fz-juelich.de (E.Y.P.)**
†Present address: Institute for Energy and Climate Research 8: Troposphere, Forschungszentrum Jülich, Jülich, Germany.
‡Present address: Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

Spatially resolved direct airborne measurements of VOC emission fluxes

Previous efforts to understand the magnitude and composition of VOC emissions in Los Angeles, as in other megacities, have relied on indirect methods—either by using traditional bottom-up emission inventories (*18*) or by inferring emissions top down from concentration measurements with chemical transport models (*10*). Both approaches are indirect and rely on a range of assumptions and thus are subject to large uncertainties. To overcome these limitations, we performed airborne eddy covariance measurements to provide the first direct observations of spatially resolved VOC emissions in Los Angeles. Emission and deposition fluxes were calculated from 10-Hz concentration and vertical wind speed measurements by using continuous wavelet transformation (*19*). State-of-the-art instrumentation [proton transfer reaction (PTR)–time-of-flight mass spectrometry–time-of-flight mass spectrometry] provided a comprehensive range of VOC species for which spatially resolved urban fluxes were observed, including source-specific tracers for biogenics, vehicle emissions, personal care products, and solvents, among others. Nine flights were conducted in June 2021 between 11:00 and 17:00, with flight days selected to cover a temperature range as wide as possible (for maps of flight tracks, see figs. S1 and S5). Median flight temperatures ranged from 23° to 31°C, with minima and maxima stretching from 15° to 37°C, respectively.

Ozone formation in Los Angeles is still sensitive to VOCs, with recent analyses suggesting that current NO_x emissions need to be reduced substantially (>50%) to move to a NO_x-sensitive ozone formation regime (*7*, *14*). The contribution of VOCs to ozone formation depends on the reaction frequency of each VOC species with the hydroxyl (OH) radical, the primary oxidant in the daytime troposphere. This reaction frequency is referred to as VOC OH reactivity (hereafter: OH reactivity) and is calculated for emissions as

$$\mathrm{OHR}_F = k_{\mathrm{OH,VOC}} \times F_{\mathrm{VOC}} \tag{1}$$

where OHR_F is the OH reactivity of the flux in meters per second squared, $k_{\mathrm{OH,VOC}}$ is the reaction rate constant of a VOC with the OH radical in meter cubed per molecule per second, and F_{VOC} is the flux of the VOC in molecules per meter squared per second. Because the daytime boundary layer over Los Angeles can be considered as a box to which emissions are continuously added, emitted OH reactivity can be considered a proxy for in situ ozone formation potential within the Los Angeles basin. The contribution of VOCs to SOA formation is more complex to quantify. We estimated aerosol yields using the statistical oxidation model (*20*) combined with a one-dimensional volatility

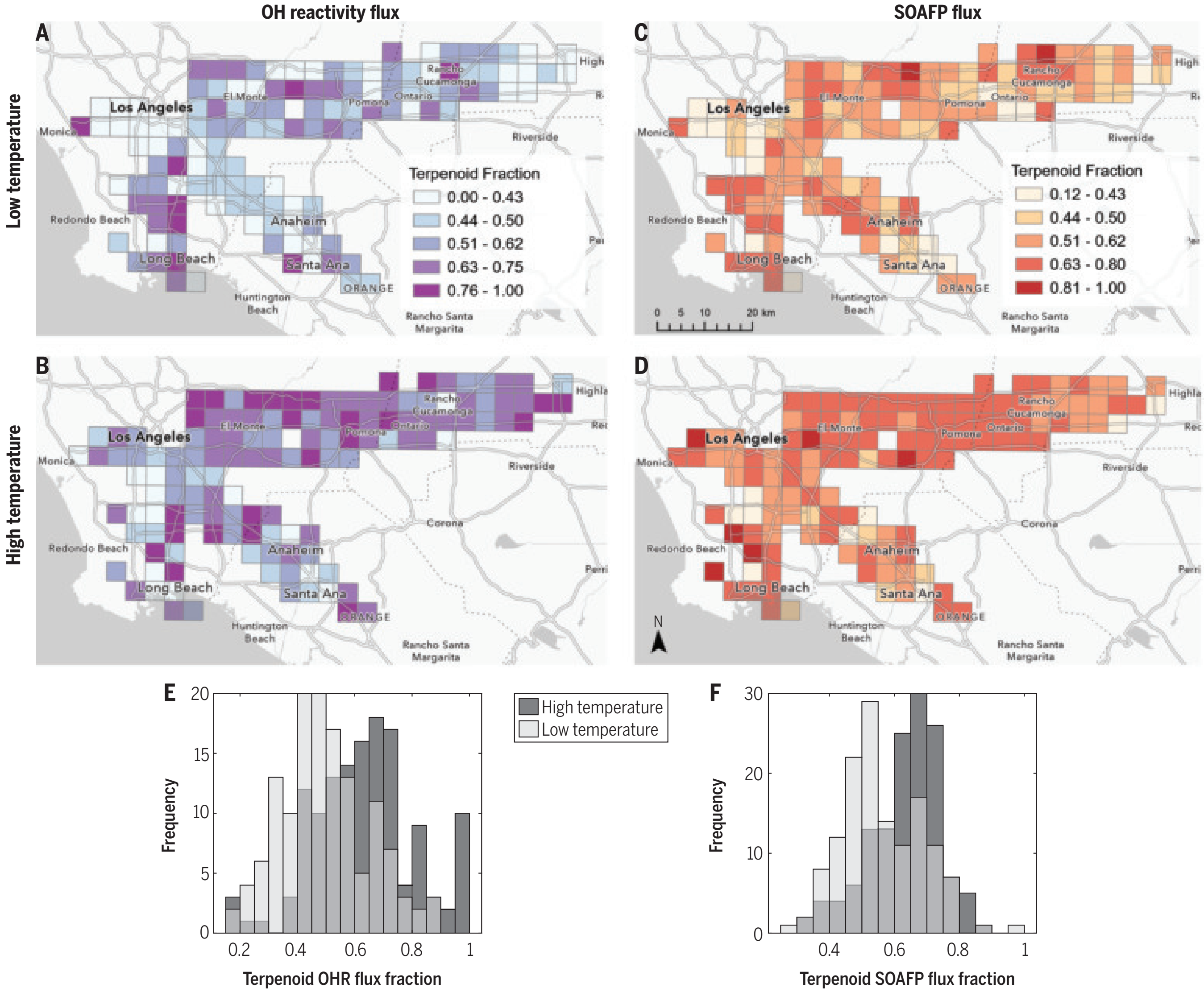


Fig. 1. Terpenoid fraction of OH reactivity and SOA formation potential of measured emissions at low versus high temperatures in Los Angeles. (**A** to **D**) Maps show the terpenoid [isoprene (C_5H_8) + monoterpenes ($C_{10}H_{16}$, $C_{10}H_{16}O$, $C_{10}H_{18}O$) + sesquiterpenes ($C_{15}H_{24}$)] fraction in the summed VOC OH reactivity flux at low temperature (A), high temperature (B), and in the summed SOA formation potential (SOAFP) of VOC emissions at low (C) and high (D) temperatures. For each 4-km by 4-km grid cell, "low temperature" was defined as an ambient temperature (2 m above ground) in the lowest 25% of all flux measurements ($n \geq 6$) conducted over that grid cell, and "high temperature" was defined as an ambient temperature in the upper 25% of all flux measurements conducted over that grid cell. (**E** and **F**) Frequency of grid cells' terpenoid fractions at low (E) and high (F) temperatures is shown. OHR, OH reactivity of the flux; SOAP, SOA potential.

basis set for oxygenated VOCs (*21*). Ambient particle formation also depends on factors such as ambient humidity (*22*), preexisting particle surface area (*23*), and the mixture of VOCs present (*24*), which we have not considered here but are unlikely to substantially change the relative contribution of individual precursor classes to the SOA budget.

Temperature dependence of the VOC mixture contributing to secondary air pollutants

Mapped VOC emissions (fig. S5) were attributed to footprint areas (fig. S1) and averaged to 4-km by 4-km grid cells, and the VOC OH reactivity and SOA formation potential were calculated for each of the observed 410 VOC species. The fraction of terpenoids (isoprene, monoterpenes, and sesquiterpenes, typically attributed to biogenic emissions) both in the summed OH reactivity and SOA formation potential was large even in the lowest temperature bin and grew substantially with temperature in most areas of Los Angeles (Fig. 1). In the lower 25th percentile of temperatures, OH reactivity in 45% of the grid cells and SOA formation potential in 67% of the grid cells were dominated by terpenoids. This especially included regions at the less urbanized hillsides on the outskirts of the metropolitan area, for example, north of El Monte. In the top 25th percentile, these values increased to 78 and 88%, respectively. Thus, at high temperatures, terpenoids became the main drivers for the formation potential of ozone and particles even in the downtown area (in Fig. 1 below and southeast of the label "Los Angeles"), where traffic and consumer product emissions are expected to be largest on the basis of current emission inventories.

The composition of the observed average VOC emission mixture is shown in Fig. 2 (for high and low temperature composition, see fig. S6). Terpenoids contributed ~16% of the measured VOC mass flux. Owing to the high reactivity of the terpenoids and the low volatility of their oxidation products, terpenoids accounted for 56 ± 26% of the emitted OH reactivity and 56 ± 33% of the SOA formation potential. The OH reactivity contribution here

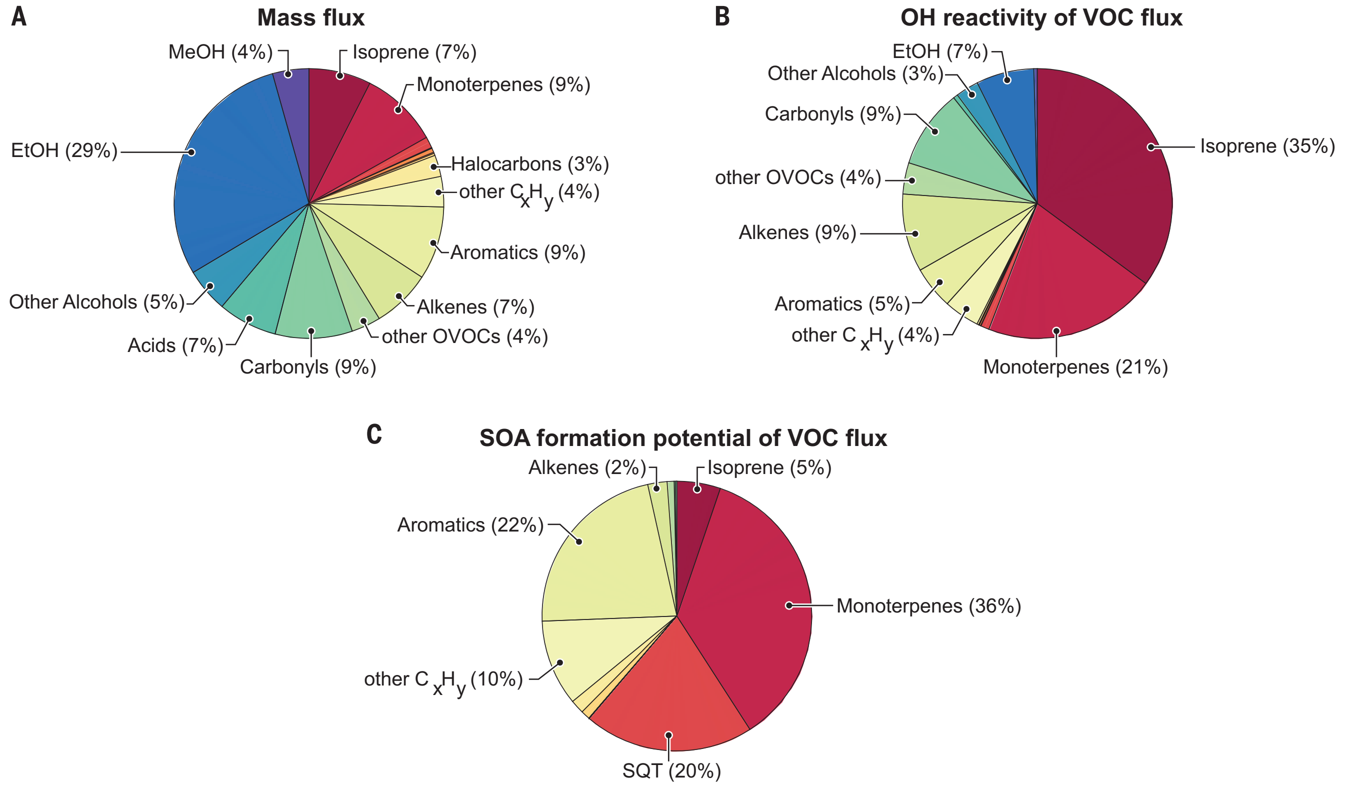


Fig. 2. Contribution of different VOC classes. (**A** to **C**) Contributions of VOC classes to (A) mass flux, (B) OH reactivity of VOC flux, and (C) SOAFP of VOC surface flux of all VOCs measured during the RECAP-CA flights in June 2021. Alkanes (contributing up to 4% of emitted OH reactivity and up to 24% of SOAFP) were not measured. The pie charts show the median composition over all flights. EtOH, ethanol; MeOH, methanol; OVOC, oxygenated VOC; SQT, sesquiterpenes.

is similar to the ~60% biogenic contribution to ozone in Berlin during a heatwave (*25*) and is supported by a modeling analysis for Los Angeles that showed that terpenoids and NO_x combined contribute >55% of the MDA8 ozone formed in Los Angeles (*26*). An inventory-based study also found ~50% contribution of biogenic terpenoids to ozone formation potential in Los Angeles, although with a higher isoprene and a lower monoterpene contribution (*18*). The same study predicted only 23% of the (annual average) SOA formation potential from terpenoids, but our value agrees well with a radiocarbon analysis of organic aerosol in summertime Los Angeles that attributed 58% ± 15% to nonfossil sources (*27*). Ethanol, which was a major constituent of the VOC mass flux, is a low-reactivity VOC and thus only contributed ~5% of the OH reactivity. The terpenoid fraction here is higher than in concentration-based Los Angeles observations (*28*). Because these VOCs react rapidly, the magnitude of their contribution to secondary products may be underestimated on the basis of ambient concentrations alone.

We note that the VOC emission mixture observed here is incomplete, because the PTR method is not sensitive to alkanes. However, alkanes are a comparably minor and relatively unreactive fraction of total VOC, contributing only 4% of the measured mid-day VOC OH reactivity in Los Angeles in 2010 (*28*). Alkanes, which are longer-lived than terpenoids, would become more relevant contributors to ozone formed downwind of the Los Angeles basin. For SOA formation, long-chain alkanes are of higher relevance. We estimate that including long-chain alkanes would reduce the relative terpenoid contribution to average summertime SOA formation by 24% of the total (*18*). Seasonality and diel variability in emissions composition are not captured by our measurements, which are limited to daytime in June, a time period that is, however, representative for air quality standard exceedance conditions.

The VOCs included in Fig. 3 were among the largest contributors to OH reactivity and/or SOA formation potential of the emissions, and they all increased with temperature. The observed increase of isoprene, monoterpene and sesquiterpene emissions with temperature is in accordance with the expectations for biogenic emissions (*29*). Reflecting this state of knowledge, two current inventories (CARB/MEGAN and the combination of BEIS + FIVE-VCP) also predicted increases in emissions of these VOCs with temperature. However, the measured amount of monoterpene and sesquiterpene emissions was a factor of 2 to 3 higher than predicted by the inventories, and Fig. 3B indicates that temperature sensitivity of monoterpene emissions in the inventories may be underestimated, as shown in a recent evaluation of the MEGAN model (*30*). The measured isoprene emission matched the inventories within the uncertainty of the fluxes.

The observed emissions of several anthropogenic VOCs also increased with temperature. These include toluene (Fig. 3D), ethanol (Fig. 3E), and para-chlorobenzotrifluoride (PCBTF) (Fig. 3F). The observed temperature dependences of ethanol and toluene emissions in the temperature range of 20° to 30°C approximately agree with their vapor pressure curves in this temperature range (*31*). Toluene, ethanol, and PCBTF can come from solvent-based sources, for which evaporative volatilization could explain the temperature dependence. Moreover, evaporative losses from gasoline can become an important contributor to vehicle emissions at high temperatures (*32*). The temperature dependence of these anthropogenic VOCs was not represented in either of the two inventories.

As a result of the increase of VOC emissions with temperature, the summed OH reactivity of the emissions increased two- to threefold from 20° to 30°C (Fig. 4A). This amounts to a 27% increase per 1°C (fig. S8). Although there was an increase in both the emissions of terpenoids and all other VOCs, most of the overall increase was due to the terpenoid emissions alone, which increased more than threefold. The OH reactivity of nonterpenoid VOC emissions only doubled in the

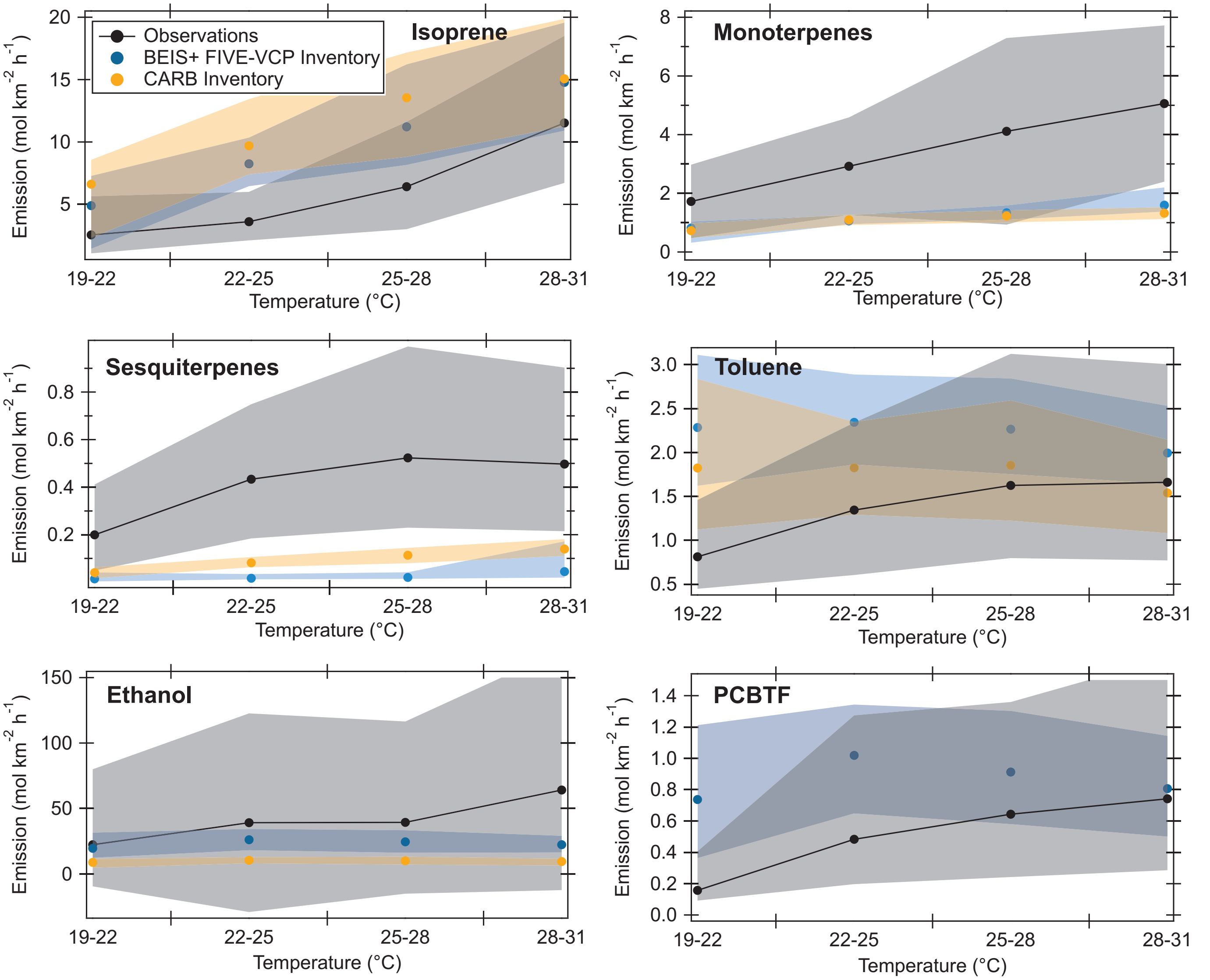


Fig. 3. Temperature dependence of emissions of individual VOC species in comparison with inventories. The VOC species shown here are among the largest contributors to SOAFP and OH reactivity in the measured VOC emissions. The circles represent medians of each temperature bin, and the shaded areas the 25th to 75th percentile of data. The temperature effect cannot be explained with regional emission differences that may co-occur with temperature differences (fig. S7). Measurement uncertainties are 76 to 96%. PCBTF is not included in the CARB inventory. Only temperature bins that include data from all regions (fig. S1) are shown. For flux measurement maps of these six species, see fig. S5.

same temperature range. The terpenoid fraction in the sum of the OH reactivity of VOC emissions increased from ~40% at 20°C to ~60% at 30°C.

Similar to OH reactivity, the SOA formation potential of emitted VOCs increased two- to threefold from 20° to 30°C, or 13% per 1°C temperature increase (Fig. 4B and fig. S8). The terpenoid fraction was enhanced from 45 to 64% of the overall SOA formation from 20° to 30°C, with the SOA-relevant terpenoid emissions approximately doubling in the same temperature range. The SOA-relevant nonterpenoid VOCs—mainly aromatics (Fig. 2)—barely increased with temperature (Fig. 4B). Although the emissions of the aromatic toluene were temperature dependent (Fig. 3) because it has a large solvent source, many other aromatics are mainly tailpipe emissions, which do not change with ambient temperature.

Monoterpene sources: Biogenic or anthropogenic?

Because monoterpenes are an especially relevant contributor to SOA formation and OH reactivity in Los Angeles, understanding whether their source is anthropogenic or biogenic is critical for air quality management. Recently, anthropogenic sources of urban monoterpenes, especially D-limonene, have been reported (*33*). Monoterpenes and sesquiterpenes are present in a variety of fragranced consumer products, for example, cleaning or personal care products (*34*).

Our data suggest a dominant, but not sole, biogenic source for the monoterpenes. The biogenic dominance is indicated both by the temperature dependence (Fig. 3B) (*35*) and by the correlation between monoterpenes and biogenic isoprene in fluxes measured in the downtown Los Angeles region (Fig. 5A). In the same region, we did not observe correlations between monoterpenes and any of the identified anthropogenic tracers, such as D5 siloxane (personal care product tracer) (Fig. 5B), toluene, or ethanol (fig. S9). Because the sesquiterpene fluxes correlated well with the monoterpene fluxes (fig. S9), a dominant biogenic source for the sesquiterpenes can be assumed, too. Monoterpene emissions did not scale with inventories of tree and shrub cover (figs. S10 and S11), likely

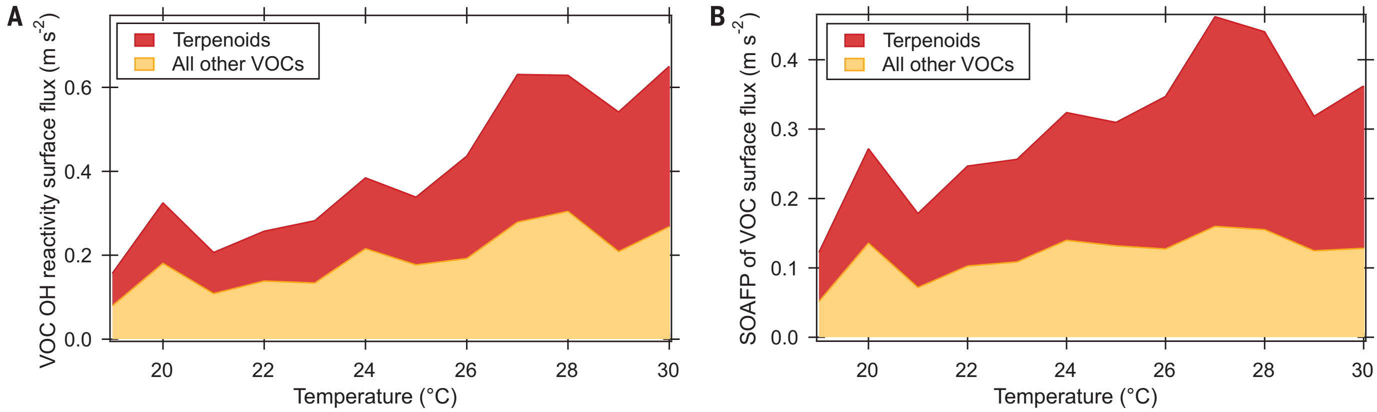


Fig. 4. Temperature dependence of the summed OH reactivity and SOAFP of the VOC emissions. (**A**) OH reactivity and (**B**) SOAFP of emitted VOCs. Primary terpenoid contribution was separated from all other VOCs and is shown as the red fraction of the stacked plot. All measured fluxes were binned into 1°C temperature bins for this plot, where "temperature" is the ambient temperature at 2 m above ground level. To ensure that the dependence is not skewed by regional emission differences, only temperature bins that cover at least three regions were included. Alkanes (contributing up to 4% of emitted OH reactivity and up to 24% of SOAFP) were not measured.

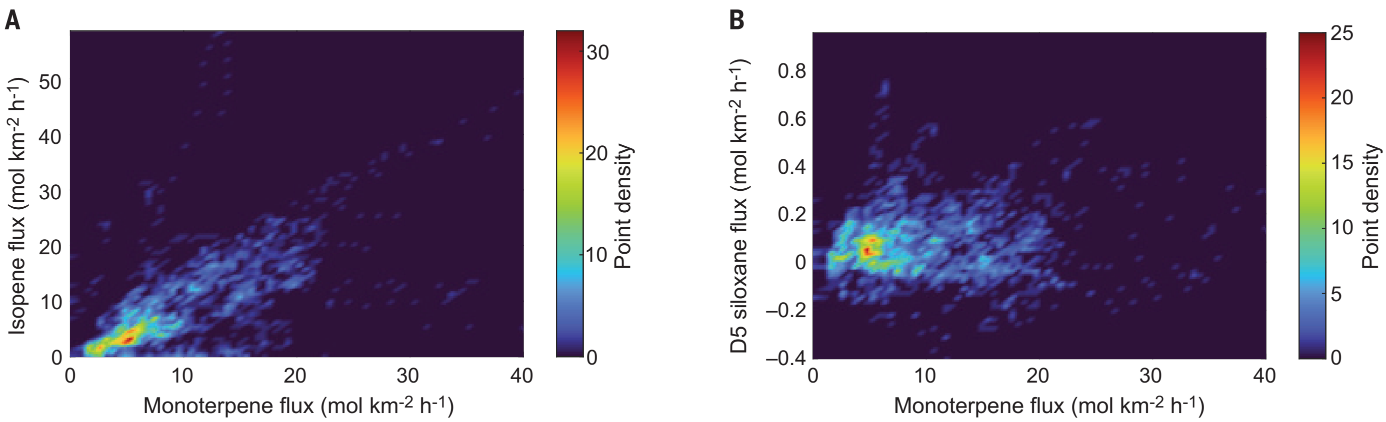


Fig. 5. Correlation of monoterpene emissions in downtown Los Angeles with other VOC fluxes, indicating biogenic origin of monoterpenes. (**A**) Density plot showing correlation (r = 0.7) between monoterpene fluxes and isoprene fluxes in the Los Angeles downtown region (for definition of regions, see fig. S1). (**B**) No correlation (r = 0.25) of monoterpene fluxes with the fluxes of an anthropogenic tracer, D5 siloxane, in the downtown Los Angeles region.

because not all trees or shrubs are monoterpene emitters. Biogenic emission inventories such as the ones included in our comparison (Fig. 3B) do not consider plant species variability within urban areas and therefore do not represent potential spatial differences in the composition and emission factors of biogenic emissions per tree. We observed regional differences in the isoprene/monoterpene emission ratio, which was ~3 in San Bernardino and ~1.5 in downtown, even though both regions have similar fractional tree coverage (fig. S12). Likely, the species composition in the San Bernardino Valley with its hillside shrublands is different.

Nonetheless, there are indications for an anthropogenic influence on measured monoterpene emissions: Some high monoterpene emissions in downtown Los Angeles (Fig. 5A) were not correlated with isoprene. In New York in winter, that is, when plant emissions are assumed to be low, monoterpenes scaled with population density (*33*). However, in Los Angeles in summer, we did not observe a correlation between monoterpene emissions and population density in the flux footprint. (This differs from the personal care product tracer D5 siloxane, which did correlate with population density; see fig. S10.) Because of this observation, we infer that a dominant anthropogenic monoterpene source is unlikely. A nonnegative matrix factorization of monoterpene and other VOC fluxes with tree cover, temperature, and population density assigned ~33% of the variability in monoterpenes fluxes to a factor dominated by population density, whereas the rest of the variability was explained by temperature and by tree and shrub cover (fig. S13). This is in good agreement with a source-apportionment study in Atlanta, Georgia, that attributed 26% of the monoterpenes to anthropogenic sources (*35*) and in which anthropogenic monoterpenes did not correlate with temperature. Coggon *et al.* suggested an anthropogenic source of monoterpenes in New York of up to 860 mg person^{-1} day^{-1} (*10*), which, by applying the population density in our flux footprints, would amount to a median of 0.74 mol km^{-2} hour^{-1} or 15% of observed mono terpenes emitted in Los Angeles.

The observed monoterpene and sesquiterpene emissions were significantly higher than predicted by the CARB and BEIS + FIVE-VCP inventories (Fig. 3). Although this may indicate an anthropogenic source not incorporated in the inventories, the size of the mismatch suggests another missing source. Because the inventories only consider an average urban tree composition, specific local monoterpene emitters may be missing, for example, eucalyptus trees, which comprise 5% of Los Angeles tree population. Moreover, flowering emissions are not included in the inventories. For example, jacaranda (*Jacaranda* sp.) trees, which are among the most abundant species in Los Angeles (*36*), were in bloom during the aircraft observations. Monoterpene emissions during bloom can be up to an order of magnitude higher than usual (*37*, *38*), and jacaranda flowers have been shown to emit a bouquet that includes monoterpenes and is rich in sesquiterpenes

(*39*, *40*). Another factor that can substantially increase biogenic terpenoid emissions and is not part of inventories is heat or drought stress (*41*, *42*). Long-term drought stress has been shown to increase sesquiterpene and monoterpene emissions while reducing isoprene emissions (*42*, *43*), which agrees with our observed inventory discrepancies. In California, 2021 was one of the driest and hottest years on record (*44*). Another contributing factor for the mismatch between observations and inventory may be the uncertainty of biogenic emission factors, their temperature dependence, and their parameterization (*30*, *45*).

Implications for urban air quality in a warming climate

Climate change is predicted to increase the number of hot days, with disproportionately large temperature increases in urban areas and in the summer (*16*, *17*). Average summer temperatures in the Los Angeles region are projected to increase by 3°C by the 2060s (*17*). Our results imply that this would lead to a doubling in emitted OH reactivity and a 40% increase in SOA formation potential of the VOCs represented in this work. This calculation assumes a linear increase in emitted OH reactivity and SOA formation potential with temperature, as observed in the limited temperature range covered in our study. Monoterpene and sesquiterpene emissions typically increase exponentially with temperature, whereas isoprene emissions are known to increase with temperature and light and eventually decrease above a temperature threshold (*46*). The biogenic VOC emission response to long-term climate change effects is subject to many uncertainties, although most studies report increased terpenoid emissions in response to warming (*47*). Our data suggest that terpenoids dominate the SOA formation potential and emitted OH reactivity above 30°C. Historically, only 10% of July days in coastal Southern California exceeded 30°C. By the 2060s, this fraction is predicted to increase to 50% (*17*). In inland Los Angeles regions, even current July temperatures exceed 30°C (*17*) nearly every day, and therefore, biogenic sources likely already dominate SOA and ozone production. This development could partly be counteracted by increased heat and drought stress that can cause irreversible damage to plants (*48*).

Even at lower summer temperatures, biogenic terpenoids dominate secondary pollutant formation potential in many parts of Los Angeles. The inventory underestimation of mono- and sesquiterpene emissions caused a factor of ~2 underestimation of VOC SOA formation potential (fig. S14) and may explain the gap between modeled and measured SOA (*49*). This, combined with the fact that inventories also did not capture the temperature dependence of anthropogenic VOCs, has implications for the representation of the emission mixture in general and for predicting how the composition and pollutant formation changes as temperatures increase.

Our findings underscore that climate change may lead to more high-ozone and high-$PM_{2.5}$ pollution events unless anthropogenic emissions are substantially reduced (*50*–*52*). Flowering and drought stress periods are expected to increase biogenic terpenoid emissions and may therefore be especially prone to high-ozone and particle pollution events. On high-temperature days, it becomes even more important to reduce NO_x emissions and anthropogenic VOC emissions, because biogenic emissions cannot be regulated and anthropogenic SOA is thought to pose a higher health risk by mass than biogenic SOA (*53*). A further reduction in NO_x emissions will help alleviate the ozone burden after transitioning to a NO_x-limited ozone formation regime.

REFERENCES AND NOTES

1. C. J. L. Murray *et al*., *Lancet* **396**, 1223–1249 (2020).
2. World Health Organization, "World Health Statistics 2022, Monitoring health for the SDGs, sustainable development goals" (2022); https://www.who.int/publications/i/item/9789240051157.
3. P. L. Hayes *et al*., *J. Geophys. Res. Atmos.* **118**, 9233–9257 (2013).
4. US Environmental Protection Agency (EPA), "Air Data - Ozone Exceedances" (2016); https://www.epa.gov/outdoor-air-quality-data/air-data-ozone-exceedances.
5. US EPA, "Air Data - Daily Air Quality Tracker" (2020); https://www.epa.gov/outdoor-air-quality-data/air-data-daily-air-quality-tracker.
6. C. Warneke *et al*., *J. Geophys. Res.* **117**, D00V17 (2012).
7. S.-W. Kim, B. C. McDonald, S. Seo, K.-M. Kim, M. Trainer, *J. Geophys. Res. Atmos.* **127**, e2021JD035606 (2022).
8. C. M. Nussbaumer, R. C. Cohen, *Environ. Sci. Technol.* **55**, 3549–3558 (2021).
9. B. C. McDonald *et al*., *Science* **359**, 760–764 (2018).
10. M. M. Coggon *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026653118 (2021).
11. P. L. Hayes *et al*., *Atmos. Chem. Phys.* **15**, 5773–5801 (2015).
12. P. K. Ma *et al*., *Atmos. Chem. Phys.* **17**, 9237–9259 (2017).
13. Y. Zhao, D. S. Tkacik, A. A. May, N. M. Donahue, A. L. Robinson, *Environ. Sci. Technol.* **56**, 15328–15336 (2022).
14. C. M. Nussbaumer, R. C. Cohen, *Environ. Sci. Technol.* **54**, 15652–15659 (2020).
15. D. E. Horton, C. B. Skinner, D. Singh, N. S. Diffenbaugh, *Nat. Clim. Chang.* **4**, 698–703 (2014).
16. G. C. Hulley, B. Dousset, B. H. Kahn, *Earths Futur.* **8**, e2020EF001480 (2020).
17. D. W. Pierce *et al*., *Clim. Dyn.* **40**, 839–856 (2013).
18. S. Gu, A. Guenther, C. Faiola, *Environ. Sci. Technol.* **55**, 12191–12201 (2021).
19. T. Karl *et al*., *J. Atmos. Sci.* **70**, 3277–3287 (2013).
20. A. L. Robinson *et al*., *Science* **315**, 1259–1262 (2007).
21. C. D. Cappa, K. R. Wilson, *Atmos. Chem. Phys.* **12**, 9505–9528 (2012).
22. G. I. Gkatzelis *et al*., *Geophys. Res. Lett.* **48**, e2020GL091351 (2021).
23. N. S. Holmes, *Atmos. Environ.* **41**, 2183–2201 (2007).
24. G. McFiggans *et al*., *Nature* **565**, 587–593 (2019).
25. G. Churkina *et al*., *Environ. Sci. Technol.* **51**, 6120–6130 (2017).
26. Q. Zhu *et al*., *Atmos. Chem. Phys.* **24**, 5265–5286 (2024).
27. P. Zotter *et al*., *J. Geophys. Res. Atmos.* **119**, 6818–6835 (2014).
28. R. F. Hansen *et al*., *J. Geophys. Res. Atmos.* **126**, e2020JD032988 (2021).
29. A. B. Guenther *et al*., *Geosci. Model Dev.* **5**, 1471–1492 (2012).
30. E. Bourtsoukidis *et al*., *Commun. Earth Environ.* **5**, 23 (2024).
31. National Institute of Standards and Technology, *NIST Webbook* (2022); https://webbook.nist.gov/chemistry/form-ser/.
32. J. I. Rubin, A. J. Kean, R. A. Harley, D. B. Millet, A. H. Goldstein, *J. Geophys. Res.* **111**, 2005JD006458 (2006).
33. G. I. Gkatzelis *et al*., *Environ. Sci. Technol.* **55**, 188–199 (2021).
34. A. C. Steinemann *et al*., *Environ. Impact Assess. Rev.* **31**, 328–333 (2011).
35. Y. Peng *et al*., *Atmos. Environ.* **288**, 119324 (2022).
36. E. G. McPherson, Q. Xiao, E. Aguaron, *Landsc. Urban Plan.* **120**, 70–84 (2013).
37. S. Fares *et al*., *Atmos. Environ.* **45**, 4557–4568 (2011).
38. R. Baghi, D. Helmig, A. Guenther, T. Duhl, R. Daly, *Biogeosciences* **9**, 3777–3785 (2012).
39. C. A. S. Pontes *et al*., *Org. Divers. Evol.* **22**, 527–541 (2022).
40. E. Guimarães, L. C. di Stasi, Rde. C. Maimoni-Rodella, *Ann. Bot.* **102**, 699–711 (2008).
41. J. K. Holopainen *et al*., *Front. Plant Sci.* **9**, 1445 (2018).
42. C. Werner *et al*., *Science* **374**, 1514–1518 (2021).
43. M. J. Potosnak *et al*., *Atmos. Environ.* **84**, 314–322 (2014).
44. California Department of Water Resources, "Water Year 2021: An Extreme Year" (2021); https://water.ca.gov/-/media/DWR-Website/Web-Pages/Water-Basics/Drought/Files/Publications-And-Reports/091521-Water-Year-2021-broch_v2.pdf.
45. L. Kaser *et al*., *Atmos. Chem. Phys.* **22**, 5603–5618 (2022).
46. A. B. Guenther, P. R. Zimmerman, P. C. Harley, R. K. Monson, R. Fall, *J. Geophys. Res.* **98**, 12609–12617 (1993).
47. J. Peñuelas, M. Staudt, *Trends Plant Sci.* **15**, 133–144 (2010).
48. E. Kleist *et al*., *Biogeosciences* **9**, 5111–5123 (2012).
49. E. A. Pennington *et al*., *Atmos. Chem. Phys.* **21**, 18247–18261 (2021).
50. D. J. Jacob, D. A. Winner, *Atmos. Environ.* **43**, 51–63 (2009).
51. E. von Schneidemesser *et al*., *Chem. Rev.* **115**, 3856–3897 (2015).
52. S. Wu *et al*., *J. Geophys. Res.* **113** (D6), 2007JD008917 (2008).
53. C. Lovett *et al*., *F1000Res.* **7**, 1031 (2018).
54. E. Y. Pfannerstill, Airborne VOC flux analysis code 2023, Zenodo (2023); https://zenodo.org/records/8411339.
55. Q. Zhu *et al*., qdzhu/FLUX: v1.0 qdzhu/FLUX: First release, Zenodo (2023); https://zenodo.org/records/8425485.

ACKNOWLEDGMENTS

The authors thank D. Baldocchi, G. Wolfe, E. Delaria, and T. Wang for insightful discussions about vertical flux divergence; A. Guenther for providing a tree cover inventory; B. McDonald for help with SOA formation potentials; M. Coggon, C. Stockwell, and C. Warneke for valuable discussions on PTR–time-of-flight mass spectrometry VOC corrections; the Regional Chemical Modeling Group of NOAA CSL for help with weather forecasting; and the Modeling and Meteorology Branch at CARB for providing their inventory. We gratefully acknowledge G. Cooper for excellent mission support; the pilots B. Kujat and G. Loudakis for their dedicated help in flight preparation, planning, and execution; and R. Weber and E. Katz for logistical support. We also thank NOAA's High Performance Computing and Communications program. **Funding:** California Air Resources Board contract nos. 20RD003 and 20AQP012 (to A.H.G. and R.C.C.); NOAA Climate Program Office's Atmospheric Chemistry, Carbon Cycle, and Climate program, grant no. NA22OAR4310540 (UCB)/ NA22OAR4310541 (AD) (to A.H.G.); Office of Naval Research Defense University Research Instrumentation Program grant no. N00014-19-1-2108 (to A.H.G.); Presidential Early Career Award for Scientists and Engineers (PECASE) (to Brian McDonald); Alexander von Humboldt Foundation Feodor Lynen Fellowship (to E.Y.P.); EPA-STAR grant no. 84001001 (to R.H.S. and Q.Z.). The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the US Environmental Protection Agency. The EPA does not endorse any products or commercial services mentioned in this publication. **Author contributions:** Conceptualization: A.H.G. and R.C.C.; Methodology: E.Y.P., C.A., Q.Z., and R.H.S.; Software: E.Y.P., C.A., and Q.Z.; Investigation: E.Y.P., C.A., B.C.S., A.B., and R. Wa., and R. Wo. Formal analysis: E.Y.P.; Visualization: E.Y.P.; Funding acquisition: A.H.G., R.C.C., and R.H.S.; Project administration: A.H.G., R.C.C., and A.B.; Supervision: A.H.G. and R.C.C.; Writing – original draft: E.Y.P.; Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data used for this manuscript are available at https://csl.noaa.gov/projects/sunvex/. The codes used for wavelet flux analysis and footprint calculation are available online (*54*, *55*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg8204
Materials and Methods
Figs. S1 to S16
Table S1 and S2
References (*56*–*103*)
Data S1

Submitted 1 March 2023; accepted 22 April 2024
10.1126/science.adg8204

ADAPTIVE SOCIALITY

Ecological disturbance alters the adaptive benefits of social ties

C. Testard[1]*†, C. Shergold[2], A. Acevedo-Ithier, J. Hart[2], A. Bernau[2], J. E. Negron-Del Valle[3,4], D. Phillips[3,4], M. M. Watowich[3,4,5], J. I. Sanguinetti-Scheck[6], M. J. Montague[1], N. Snyder-Mackler[3,4,7,8], J. P. Higham[9,10], M. L. Platt[1,11,12], L. J. N. Brent[2]

Extreme weather events radically alter ecosystems. When ecological damage persists, selective pressures on individuals can change, leading to phenotypic adjustments. For group-living animals, social relationships may be a mechanism enabling adaptation to ecosystem disturbance. Yet whether such events alter selection on sociality and whether group-living animals can, as a result, adaptively change their social relationships remain untested. We leveraged 10 years of data collected on rhesus macaques before and after a category 4 hurricane caused persistent deforestation, exacerbating monkeys' exposure to intense heat. In response, macaques demonstrated persistently increased tolerance and decreased aggression toward other monkeys, facilitating access to scarce shade critical for thermoregulation. Social tolerance predicted individual survival after the hurricane, but not before it, revealing a shift in the adaptive function of sociality.

With the intensifying climate crisis, extreme weather events are expected to become less predictable and increase in both frequency and force (*1*, *2*). These events can cause sustained destruction of an existing ecosystem, depleting resources and degrading infrastructure upon which humans and other animals depend (*3*). Persistent ecosystem modifications may then alter the selective pressures exerted on individuals living within them, which can, over time, enforce phenotypic changes (*4*). Group-living animals may cope with environmental upheavals by adaptively changing their patterns of social interaction (*5*, *6*). Yet, whether natural disturbances alter selective pressures on patterns of social interaction by changing the costs and benefits experienced by individuals and whether group-living animals can adaptively modify their social relationships as a result remain untested. The lack of empirical evidence on this topic arises from the difficulty of studying behavioral responses to extreme and unpredictable events that cannot be reproduced in the laboratory and from the rarity of long-term demographic data required to capture fitness consequences of biological traits.

On 20 September 2017, Hurricane Maria made landfall as a category 4 hurricane on Puerto Rico, causing devastating destruction and a historically tragic human death toll (*7*). A population of rhesus macaques living on Cayo Santiago, a small island off the coast of Puerto Rico, was also hit by this storm (Fig. 1A). Hurricane Maria destroyed 63% of Cayo Santiago's vegetation (Fig. 1, B and C). Five years later, the tree canopy cover remains far below prehurricane levels (Fig. 1B and fig. S1A), and temperatures on the island regularly exceed 40°C (Fig. 1D and fig. S1C). Hurricane-induced deforestation resulted in most of the island being exposed to these high temperatures. From 2018 to 2021, 48% of sample points from temperature loggers deployed on exposed (i.e., nonshaded) parts of the island detected temperatures above 40°C during the day (10 a.m. to 5 p.m.) (N = 190,240 samples from 25 temperature loggers; fig. S1C). Such extreme temperatures compromise mammals' ability to regulate internal temperature, potentially resulting in heat exhaustion and hyperthermia (*8*, *9*). Primates have limited capacity for physiological thermoregulation (*10*, *11*) and therefore often cope with extreme temperatures by changing their behavior (*12–14*). Cayo Santiago's macaques address thermoregulatory challenges primarily by seeking shade (which is, on average, 7°C cooler than in exposed areas; Fig. 1D, t test, $P < 0.001$).

Shade became a scarce and critical resource after Hurricane Maria. On the basis of evidence showing increased competition when resources are scarce or monopolizable (*15–17*), we predicted that, after the hurricane, rhesus macaques—who are widely regarded as among the most despotic of primate species (*18*)—would display lower social tolerance and more frequent aggression. Previously, we reported that instead of becoming less prosocial, macaques on Cayo Santiago drastically increased their tolerance for other monkeys in the year after Hurricane Maria (*19*). Those results led us to hypothesize that increased social tolerance was an adaptive response: Monkeys were using their social networks to access now-scarce shade critical for coping with exacerbated heat stress. In the current study, we tested this hypothesis and predicted that changes in social structure would persist and would have direct fitness consequences. We tested these predictions using 10 years of behavioral and demographic data from before and after the hurricane (Fig. 1E).

Social tolerance increased and aggression decreased up to 5 years after Hurricane Maria

To assess the long-term impacts of Hurricane Maria on social relationships, we used behavioral data collected during a 10-year period before and after the event (prehurricane years: 2013–2017; posthurricane years: 2018–2022, excluding 2020 because of the COVID-19 pandemic; Fig. 1E) in seven distinct social groups. A total of 790 different adult individuals with behavioral data before and/or after Hurricane Maria were included in this analysis. Our analyses focused on relationships in which partners spent time in close proximity to each other [within 2 m of each other (*20*)]. Whether pairs of individuals tolerate being close to each other is an active choice (*21–23*) that is repeatable and indicates an affiliative relationship between them (*24*). For these reasons, proximity is used to quantify social structure in primates (*25–28*). Indeed, proximity in close space is tightly linked to resource sharing: Macaques that do not wish to share a resource do not tolerate others in close proximity (*22*, *29*, *30*). Reduced aggression between individuals in close proximity further indicates social tolerance, so we also quantified changes in aggression.

We constructed proximity and aggression networks for each group in each year from 2013 to 2022 and compared individuals' propensity to be in proximity to or aggressive toward conspecifics before and after the hurricane. We used a linear mixed model comparing proximity and aggression in each posthurricane year to pooled prehurricane years, controlling for potential confounding variables: group size, individual age, and sex. We also conducted a year-by-year analysis, taking 2013 as a baseline for comparing all subsequent years (fig. S3). Our models accounted for uncertainty in edge weights due to unequal sampling across individuals, groups, and years (*31*).

We found that increased social tolerance persisted for 5 years after Hurricane Maria (Fig. 2, A and B). In the year immediately after the hurricane, Cayo Santiago macaques showed an acute response in which the propensity to be in proximity to another monkey was on

[1]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, USA. [2]Centre for Research in Animal Behaviour, University of Exeter, Exeter, UK. [3]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. [4]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [5]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA. [6]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [7]ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, USA. [8]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA. [9]Department of Anthropology, New York University, New York, NY, USA. [10]New York Consortium in Evolutionary Primatology (NYCEP), New York, NY, USA. [11]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA. [12]Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA.
***Corresponding author. Email: ctestard@fas.harvard.edu**
†Present address: Department of Cellular and Molecular Biology, Harvard University, Cambridge, MA, USA.

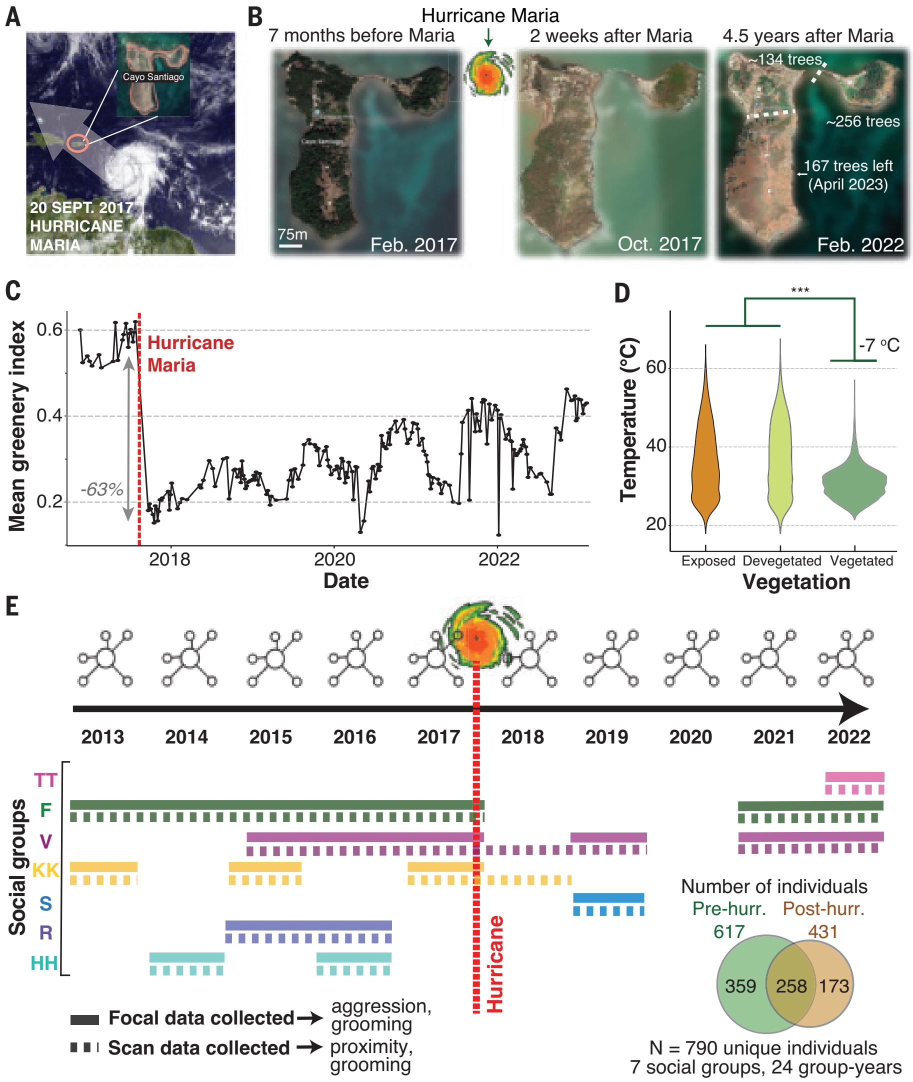


Fig. 1. Hurricane Maria caused a persistent drop in vegetation and an increase in surface temperature. (**A**) Path of Hurricane Maria making landfall on Puerto Rico, including Cayo Santiago, on 20 September 2017. VIIRS (Visible Infrared Imaging Radiometer Suite) image captured by the National Oceanic and Atmospheric Administration's Suomi NPP (National Polar-orbiting Partnership) satellite. (**B**) Google Earth images captured at three time points: 7 months before (left), 2 weeks after (middle), and 4.5 years after (right) Hurricane Maria. Tree census conducted in April 2023 revealed that only 167 trees were left on the upper part of "Big Cayo" (see fig. S1 for satellite images from more time points and tree census on the rest of Cayo Santiago). (**C**) Foliage cover from Cayo Santiago, as measured by greenness from satellite imagery [normalized difference vegetation index (NDVI)], decreased by 63% in the 3 months after Hurricane Maria. Note that the NDVI index captures all types of vegetation, including grass and small bushes. Although satellite NDVI increases steadily up until 2022, this largely reflects grass (fig. S1), which does not provide shade. Even when including all vegetation types, vegetation cover still remains far below prestorm levels. (**D**) Distribution of temperatures posthurricane (2018–2022) in three types of terrain: exposed (areas without vegetation cover both before and after Hurricane Maria), devegetated (areas that lost their vegetation cover after Hurricane Maria), and vegetated (areas with vegetation cover). Temperatures in vegetated areas are lower than in other areas (t test, $^{***}P < 0.001$). (**E**) Behavioral data (scans and focal follows) collected from up to 5 years before Hurricane Maria to 5 years after, across seven macaque social groups. Solid lines: focal follows; dotted lines: scans. (Inset) Venn diagram of the number of individuals before and/or after Hurricane Maria.

average 1.72 standard deviations higher—more than triple—compared with prehurricane years {95% confidence interval (CI) = [1.60, 1.83]; Fig. 2C and data S1; mean prehurricane was 0.18 versus 0.58 in 2018}. Two to five years after the storm, proximity among macaques remained between 0.47 and 0.84 standard deviations higher than prehurricane levels ($P < 0.001$ for all years posthurricane; Fig. 2C and data S1). Similarly, we found a higher number of proximity partners multiple years after the hurricane, peaking immediately after the event (fig. S2 and data S1). Population density across the whole island remained stable during the 10 years analyzed and therefore cannot explain these results (meaning that monkeys did not simply interact with others at higher rates because there were a greater number of available partners: mean total population size = 1498 in prehurricane years; mean = 1491.2 in posthurricane years). Changes in feeding patterns after Hurricane Maria are also unlikely to explain the observed sociality changes (see the materials and methods in the supplementary materials). The acute component of macaques' social response immediately after Hurricane Maria might reflect stronger social cohesion, which is often observed in humans in the wake of natural disasters (*32*). As the environment stabilizes around a "new normal," less-useful social connections might be pruned as a result.

Evidence of higher tolerance is corroborated by a sustained reduction in aggression after the hurricane, even as resources, particularly heat-reducing shade, were depleted and individuals were more frequently in proximity (Fig. 2, B and C). Lower aggression between members of a despotic species such as rhesus macaques in the context of scarce resources and increased proximity was unexpected given published characterizations of this species (*15*–*17*). Higher tolerance and lower aggression were not associated with consistently higher frequencies of grooming after the hurricane (fig. S2).

Persistent environmental degradation changed the adaptive value of social tolerance

Next, we investigated whether higher tolerance predicted survival after Hurricane Maria (study period: September 2017 to August 2022). We tested the effect of two individual-level network metrics from the proximity network: degree (the number of proximity partners) and strength (the sum of all edge weights, where edges were defined as the probability of observing a node in proximity to another node). This analysis included 431 different adult individuals from five social groups, of which 155 died during the study period. Because the values of network metrics changed for individuals over the course of the study period, we used time-varying mixed effects Cox proportional hazard models (*33*) to evaluate whether proximity predicted survival from one year to the next during our study period. All models implicitly accounted for age; sex was included as a fixed effect; and social group, individual ID, and year were included as random effects to

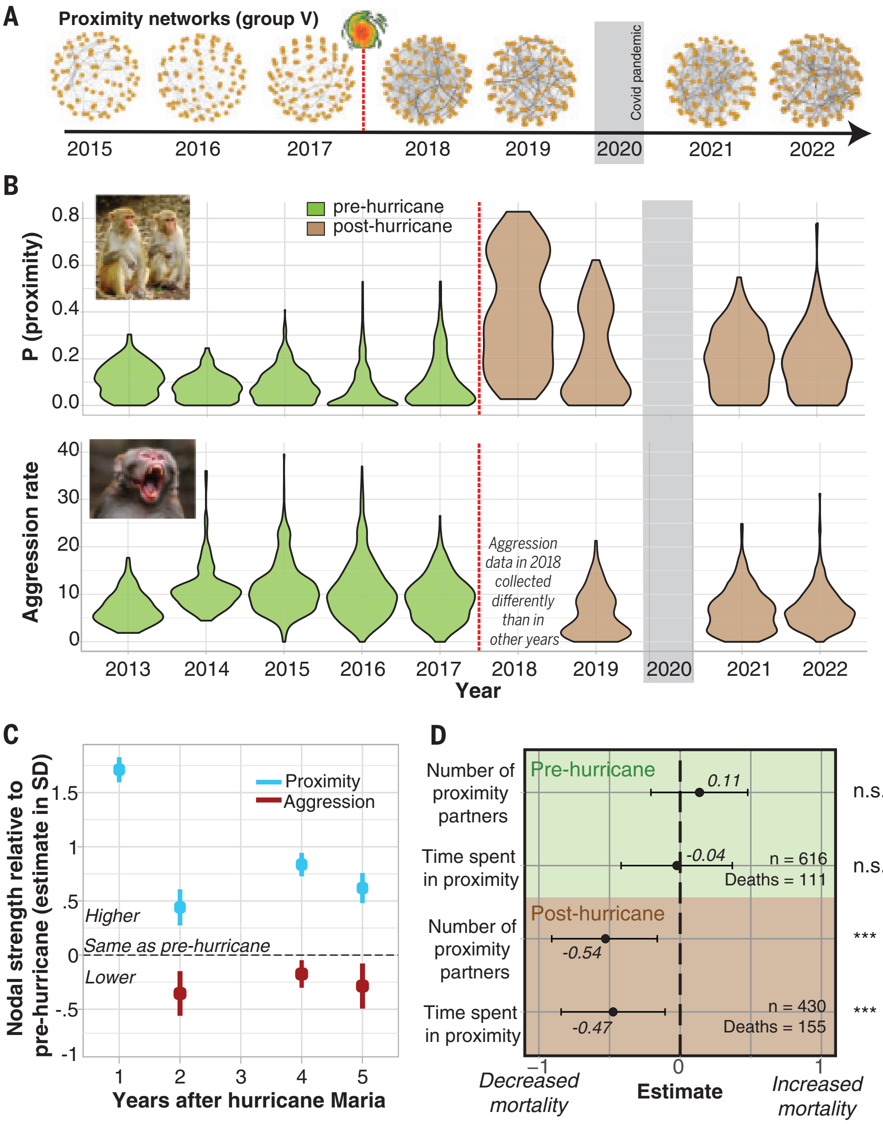


Fig. 2. After Hurricane Maria, Cayo Santiago macaques exhibited persistently higher social tolerance, which predicted their survival. (**A**) Proximity networks for social group V from 2013 to 2022 (excluding 2020 because of the COVID-19 crisis) are shown as example networks. Each node represents an individual, and edges represent the probability for two individuals to be in proximity to one another. The thicker the line, the higher the probability of proximity. The red dotted line marks Hurricane Maria. (**B**) (Top) Probability of being observed in proximity to another monkey from 2013 to 2022, across 790 individuals in seven social groups, 24 group-year combinations. (Bottom) Same as the top panel but for aggression rate (number of aggression events per hours observed). In 2018, aggression data were collected differently and therefore were excluded from this visualization. (**C**) Linear mixed model estimates for the long-term change in standardized proximity and aggression network strength compared with prehurricane levels (for grooming results, see fig. S2 and data S1). Error bars represent the 95% CI, pooled across 100 network iterations (i.e., distribution of possible networks given the observed interaction data). Blue, proximity; red, aggression. (**D**) Estimates from survival models for the impact of degree and strength in the proximity networks on mortality risk. Estimates < 1: lower mortality risk; estimates > 1: higher mortality. Hazard ratio = e^{estimate}. Error bars represent the 95% CI pooled across Bayesian-generated social networks. Green, prehurricane; brown, posthurricane. ***$P < 0.01$; n.s., not significant. [Photo credits: Lauren Brent for monkeys in proximity and Gomen S. for aggressive macaque]

account for repeated measures and between-group and between-year differences.

We found that both degree and strength of proximity predicted survival after the hurricane, indicating that behavioral responses were adaptive (Fig. 2D). Each additional standard deviation in an individual's number of proximity partners (+12.7 partners) resulted in a 42.69% lower mortality risk (model estimate = −0.54, hazard ratio = 0.58, 95% CI = [0.40, 0.86]; data S1), and an additional standard deviation in the propensity to be in proximity to another macaque (i.e., proximity strength) led to 37.3% lower mortality risk (model estimate = −0.48, hazard ratio = 0.62, 95% CI = [0.41, 0.95]; data S1). Focusing on individuals for which we had data before and after the hurricane (N_{ID} = 258), we found that interindividual variation in increased tolerance after the hurricane predicted survival. However, this was only the case for individuals that had low social tolerance before the hurricane (Spearman correlation between plasticity and survival time = 0.84, CI = [0.83, 0.87]; fig. S4). This relationship did not hold for individuals that were already highly tolerant before the hurricane (Spearman's rho = 0.22, CI = [−0.12, 0.56]). In other words, being highly tolerant after the event is the critical predictor of survival, regardless of how this was achieved (through plastic change or by being highly tolerant to begin with).

To test whether the adaptive benefits of proximity were due to the storm and its aftermath, we tested whether proximity also predicted survival before Hurricane Maria. Using the same approach but focusing on prehurricane years (study period: January 2013 to September 2017; N = 617 individuals, 111 deaths), we found no evidence that proximity strength or degree predicted survival (Fig. 2D and data S1). Furthermore, aggression received or given did not predict survival before or after the hurricane (data S1). These findings indicate that, by modifying Cayo Santiago's ecosystem, Hurricane Maria altered the adaptive benefits of tolerating conspecifics in close proximity.

Stronger thermoregulatory pressure posthurricane drove the adaptive benefits of social tolerance

The drastic drop in shade caused by Hurricane Maria elevated thermoregulatory burdens on Cayo Santigo's macaques. Temperatures in non-shaded areas regularly exceed 40°C (fig. S1C), and persistent exposure to these temperatures increases the risk of heat shock and generally degrades health and physical condition (*8*). Anecdotal reports from observers on the ground suggested that macaques used shady areas afforded by the remaining vegetation after Hurricane Maria to lower their body temperature (Fig. 3A). From these observations, we

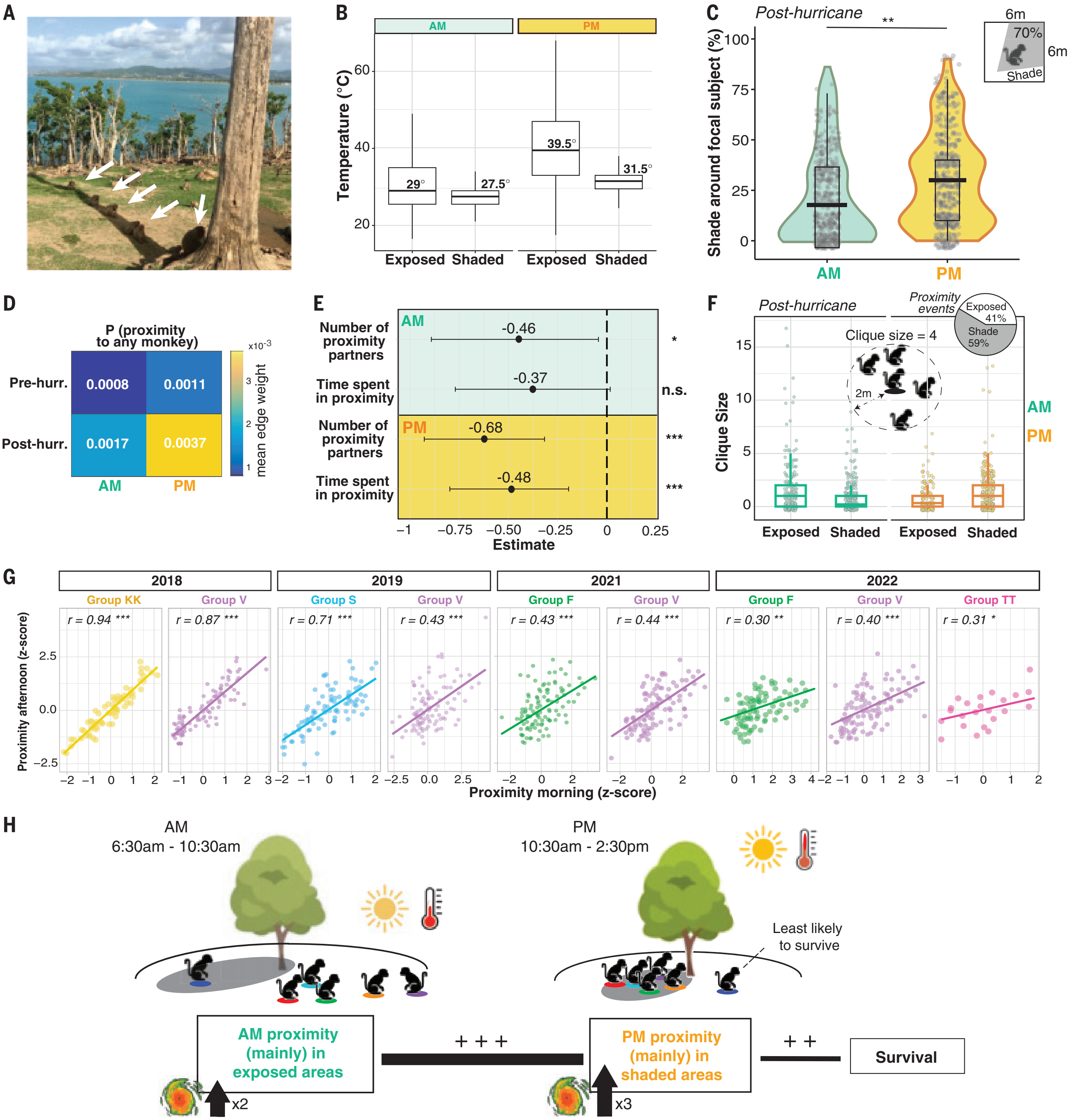


Fig. 3. Thermoregulatory needs drove macaques to become more tolerant after Hurricane Maria. (**A**) Photo taken 2 years after the hurricane, showcasing macaques concentrated in a narrow shady area cast by dead tree. (**B**) Temperature (°C) in exposed versus shaded areas in the morning (left) versus afternoon (right). (**C**) Shade (%) in 6 m^2 around the focal animal split by morning and afternoon (t test, $P < 0.01$). (**D**) Mean edge weight of proximity networks in the morning versus afternoon pre- and posthurricane. (**E**) Estimates from survival models testing the impact of proximity on mortality risk after the hurricane split by time of day (morning, green; afternoon, orange). Bars represent 95% CI. (**F**) Number of individuals in proximity to a focal monkey (i.e., clique size) in exposed versus shaded areas, morning versus afternoon (based on data collected May through August 2019). (**G**) Spearman's correlation between morning and afternoon proximity for all groups and years posthurricane. (**H**) Schematic diagram summarizing our results: Social proximity in the morning doubled after the hurricane compared with before, largely occurs in exposed areas of the island, and is strongly correlated with afternoon proximity. Proximity is more likely in the afternoon than in the morning, mostly occurs in the shade, and predicts survival. Thus, the hurricane led to a generalized change in patterns of social interaction, even in nonthermoregulatory contexts, facilitating access to shade when most needed. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. [Photo credit: Sébastien Tremblay]

hypothesized that individuals' increased proximity to conspecifics after the hurricane and the adaptive benefits of this social response reflected access to scarce shade to deal with the intensified thermoregulatory burden. As cause of death was not part of our longitudinal data collection, we used temperature, shade status, and survival data to test this hypothesis. We found that the temperature difference between shaded and exposed areas was higher in the afternoon (8°C) than during the morning hours (2°C; Fig. 3B). Concurrently, macaques were found in areas with more shade in the afternoon (t test, $P < 0.01$; Fig. 3C) and were more likely to be in proximity to another monkey during the hottest hours of the day (posthurricane a.m. mean proximity edge weight = 0.0017; p.m. = 0.0037; Fig. 3D). Finally, we found that proximity during hotter hours of the day (p.m.) better predicted survival than proximity during cooler daytime hours (a.m.; Fig. 3E and data S1). These findings endorse the hypothesis that increased tolerance after the hurricane reflects an adaptive behavioral response to increased thermoregulatory stress.

Next, we investigated whether proximity between monkeys after the hurricane solely reflected concentration in cool shaded areas or whether it also occurred in other contexts, suggesting a general increase in social tolerance. We collected shade use and proximity data from May to August 2019 on 200 individuals across all social groups. We hypothesized that although social tolerance is most important in the afternoon to access shade, changes in patterns of social interaction will be apparent in other nonthermoregulatory contexts, such as during cool morning hours, reflecting changes in social relationships rather than simply spatial preference for shade during the hottest hours of the day. We found that proximity to conspecifics frequently occurred in exposed areas of the island (41% of observed proximity events), particularly in the morning, when temperatures are coolest (Fig. 3F). In fact, morning proximity between individuals almost doubled after the hurricane compared with before (Fig. 3D). If proximity interactions reflect social relationships, then we would expect a strong positive correlation between morning and afternoon proximity—which we observe across all groups and years posthurricane (Spearman correlations; Fig. 3G). These results suggest that increased proximity after the hurricane was not only a by-product of individuals congregating in narrow shaded spaces during hot hours of the day but also reflects active social decisions in other contexts, which in turn predict access to valuable shade when most needed. In other words, Cayo Santiago macaques changed how they engage with others or the nature of their social relationships, facilitating access to a scarce thermoregulatory resource in response to intensified heat stress posthurricane (Fig. 3H).

Discussion

We found that monkeys were persistently more tolerant of others in their vicinity and less aggressive for up to 5 years after Hurricane Maria. Relationships based on social tolerance became more numerous and predicted individual survival after the storm, especially during the hottest hours of the day. These findings support our hypothesis that the adaptive benefits of social tolerance are linked to accessing a thermoregulating resource—shade—in response to increased heat stress. Monkeys were not simply being passively "squeezed" into now-limited shaded spaces but instead showed a generalized increase in social tolerance, including outside of thermoregulatory contexts, suggesting a fundamental change in how they engaged with others. Notably, social tolerance did not predict survival before the hurricane, demonstrating that hurricane-induced drastic changes in ecological pressures altered the benefits individuals gain from social relationships. These results show that an extreme climatic event and its aftermath altered selective pressures on a social phenotype and identify ecosystem fluctuations as potential evolutionary drivers of sociality in group-living animals.

The hurricane-induced change in selective pressures on Cayo Santiago's macaques and ensuing social flexibility may partially reflect macaques' inability to leave the island—a situation that is not uncommon among animal populations, including humans. Animals living on islands or in fragmented habitats have only limited, if any, opportunities to migrate to new territories when their habitat is degraded (*34–37*), thus forcing rapid adjustments to new environmental conditions (*38*). In humans, marginalized populations of low socioeconomic status also suffer from limited capacity to move away from disaster zones (*39*, *40*).

Social tolerance after the hurricane was associated with a 42% decrease in mortality risk, an effect comparable to the survival impacts of social support in humans (*41*). If fitness-enhancing tolerance is heritable and the environmental conditions favoring this phenotype do not change, ecological degradation may select for more-tolerant individuals, leading to the evolution of an overall more-tolerant macaque society on Cayo Santiago in the future. However, the adaptive benefits of increased tolerance may change dynamically over time as new challenges emerge. For example, persistent increases in proximity or close contact may increase disease transmission (*42*), altering the balance between the costs and benefits of social tolerance.

Our study provides rare evidence of an abrupt change in selection on sociality in the face of a large and persistent ecological disturbance. These findings show the potential of social flexibility to provide resilience to rapid and unpredictable environmental fluctuations in animals (*43–46*) and emphasize a dynamic link between the environment and fitness consequences of social behavior.

REFERENCES AND NOTES

1. K. Balaguru, G. R. Foltz, L. R. Leung, K. A. Emanuel, *Nat. Commun.* **7**, 13670 (2016).
2. K. Emanuel, *Bull. Am. Meteorol. Soc.* **98**, 495–501 (2017).
3. S. R. Weiskopf *et al*., *Sci. Total Environ.* **733**, 137782 (2020).
4. C. M. Donihue *et al*., *Nature* **560**, 88–91 (2018).
5. D. T. Blumstein, L. D. Hayes, N. Pinter-Wollman, *Trends Ecol. Evol.* **38**, 337–345 (2023).
6. D. N. Fisher *et al*., *Biol. Rev. Camb. Philos. Soc.* **96**, 2661–2693 (2021).
7. D. D. Kane, *N. Engl. J. Med.* **379**, e30 (2018).
8. W. G. Whitford, R. E. Sherman, *Comp. Biochem. Physiol. A Comp. Physiol.* **53**, 337–341 (1976).
9. K. L. Ebi *et al*., *Lancet* **398**, 698–708 (2021).
10. Y. G. Kamberov *et al*., *J. Hum. Evol.* **125**, 99–105 (2018).
11. A. Best, J. M. Kamilar, *J. Hum. Evol.* **117**, 33–43 (2018).
12. R. McFarland *et al*., *Am. J. Phys. Anthropol.* **171**, 407–418 (2020).
13. R. McFarland, B. Majolo, *Biol. Lett.* **9**, 20130428 (2013).
14. G. Schino, A. Troisi, *Physiol. Behav.* **47**, 1125–1128 (1990).
15. L. A. Isbell, *Behav. Ecol.* **2**, 143–155 (1991).
16. M. L. Boccia, M. Laudenslager, M. Reite, *Am. J. Primatol.* **16**, 123–130 (1988).
17. R. L. Chancellor, L. A. Isbell, *Anim. Behav.* **75**, 1939–1947 (2008).
18. E. B. Cooper *et al*., *eLife* **11**, e78169 (2022).
19. C. Testard *et al*., *Curr. Biol.* **31**, 2299–2309.e7 (2021).
20. S. E. DeTroy, D. B. M. Haun, E. J. C. van Leeuwen, *Evol. Anthropol.* **31**, 30–44 (2022).
21. R. Ventura, B. Majolo, N. F. Koyama, S. Hardie, G. Schino, *Am. J. Primatol.* **68**, 1138–1149 (2006).
22. C. Dubuc, K. D. Hughes, J. Cascio, L. R. Santos, *Am. J. Phys. Anthropol.* **148**, 73–80 (2012).
23. O. Schülke *et al*., *Behav. Ecol. Sociobiol.* **76**, 50 (2022).
24. R. E. Morrison, W. Eckardt, F. Colchero, V. Vecellio, T. S. Stoinski, *eLife* **10**, e62939 (2021).
25. C. Corradino, *Primates* **31**, 351–362 (1990).
26. L. J. N. Brent, J. Lehmann, G. Ramos-Fernández, *Am. J. Primatol.* **73**, 720–730 (2011).
27. N. A. Thompson, *Behaviour* **156**, 859–908 (2019).
28. R. E. Morrison, W. Eckardt, T. S. Stoinski, L. J. N. Brent, *Proc. Biol. Sci.* **287**, 20201026 (2020).
29. P. Belisle, B. Chapais, *Behaviour* **138**, 487–509 (2001).
30. K. N. Balasubramaniam, C. M. Berman, *Behaviour* **154**, 1145–1176 (2017).
31. J. D. A. Hart, M. N. Weiss, D. W. Franks, L. J. N. Brent, *Methods Ecol. Evol.* **14**, 2411–2420 (2023).
32. S. Sweet, *Int. J. Mass Emerg. Disasters* **16**, 321–331 (1998).
33. Z. Zhang, J. Reinikainen, K. A. Adeleke, M. E. Pieterse, C. G. M. Groothuis-Oudshoorn, *Ann. Transl. Med.* **6**, 121 (2018).
34. N. M. Haddad *et al*., *Sci. Adv.* **1**, e1500052 (2015).
35. N. L. Volpe, W. D. Robinson, S. J. K. Frey, A. S. Hadley, M. G. Betts, *PLOS ONE* **11**, e0167513 (2016).
36. K. R. Crooks *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7635–7640 (2017).
37. A. C. Delciellos, S. E. Ribeiro, M. V. Vieira, *J. Mammal.* **98**, 1129–1136 (2017).
38. W. Jetz, D. R. Rubenstein, *Curr. Biol.* **21**, 72–78 (2011).
39. S. Hallegatte, A. Vogt-Schilb, J. Rozenberg, M. Bangalore, C. Beaudet, *Econ. Disaster. Clim. Chang.* **4**, 223–247 (2020).
40. "Greater Impact: How disasters affect people of low socioeconomic status," Disaster Technical Assistance Center Supplemental Research Bulletin (Substance Abuse and Mental Health Services, 2017).
41. J. Holt-Lunstad, T. B. Smith, J. B. Layton, *PLOS Med.* **7**, e1000316 (2010).
42. C. J. Carlson *et al*., *Nature* **607**, 555–562 (2022).
43. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637–669 (2006).
44. B. B. M. Wong, U. Candolin, *Behav. Ecol.* **26**, 665–673 (2014).

45. E. A. Beever et al., Front. Ecol. Environ. 15, 299–308 (2017).
46. U. Kalbitzer, C. A. Chapman, in Primate Life Histories, Sex Roles, and Adaptability: Essays in Honour of Linda M. Fedigan, U. Kalbitzer, K. M. Jack, Eds. (Springer International Publishing, 2018), pp. 283–310.

ACKNOWLEDGMENTS

We thank the Caribbean Primate Research Center staff for their important roles in data collection and in the maintenance of Cayo Santiago. We also thank G. Albery, E. Siracusa, and D. de Moor for insightful comments that enriched our analyses and discussion of the results. **Funding:** Support for this research was provided by the National Institutes of Health (R01MH118203, U01MH121260, R01MH096875, P40OD012217, R01AG060931, R00AG051764, R56AG071023, R01AG084706, and R24-AG065172), a RAPID award from the National Science Foundation (1800558), a Royal Society Research Grant (RGS/R1/191182), and a European Research Council Consolidator Grant (864461 - FriendOrigins). Cayo Santiago is supported by grant 8-P40 OD012217-25 from the National Center for Research Resources (NCRR) and the Office of Research Infrastructure Programs (ORIP) of the National Institutes of Health. **Author contributions:** C.T. and L.J.N.B. conceptualized the study; J.H. developed methodology specific to this study; C.T. and C.S. analyzed the data with input from L.J.N.B. and J.I.S.-S.; C.T. designed and edited figures with input from J.I.S.-S.; J.E.N.-D.V., D.P., A.A.-I., A.B., and M.M.W. collected the data; N.S.-M., J.P.H., M.J.M., M.L.P., and L.J.N.B. secured funding for this work; C.T. and L.J.N.B. drafted the manuscript, with input from all authors; L.J.N.B. supervised this work. All authors read and approved the manuscript. **Competing interests:** M.L.P. is a scientific advisory board (SAB) member, consultant, and/or cofounder of Blue Horizons International (SAB), NeuroFlow (SAB), Amplio, Cogwear Technologies (cofounder), Burgeon Labs (SAB), Glassview Inc. (SAB), Speak Up Now (SAB), Lazul (SAB), and Business Collaborative for Brain Health (SAB). M.L.P. also receives research funding from AIIR Consulting, Glassview Inc., Mars Inc., Slalom Inc., the Lefkort Family Research Foundation, Sisu Capital, the Philadelphia Flyers, and Benjamin Franklin Technology Partners. All other authors declare that they have no competing interests. **Data and materials availability:** Data and code are available on GitHub at https://github.com/camilletestard/Cayo-Maria-Survival. Data S1, which includes all model outputs from this study, can also be found in this GitHub repository.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk0606
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References (47–61)
MDAR Reproducibility Checklist
Data S1

Submitted 31 July 2023; accepted 15 April 2024
10.1126/science.adk0606

CARBON CYCLING

Bomb radiocarbon evidence for strong global carbon uptake and turnover in terrestrial vegetation

Heather D. Graven[1]*, Hamish Warren[1], Holly K. Gibbs[2], Samar Khatiwala[3], Charles Koven[4], Joanna Lester[1], Ingeborg Levin[5]†, Seth A. Spawn-Lee[6,7], Will Wieder[8]

Vegetation and soils are taking up approximately 30% of anthropogenic carbon dioxide emissions because of small imbalances in large gross carbon exchanges from productivity and turnover that are poorly constrained. We combined a new budget of radiocarbon produced by nuclear bomb testing in the 1960s with model simulations to evaluate carbon cycling in terrestrial vegetation. We found that most state-of-the-art vegetation models used in the Coupled Model Intercomparison Project underestimated the radiocarbon accumulation in vegetation biomass. Our findings, combined with constraints on vegetation carbon stocks and productivity trends, imply that net primary productivity is likely at least 80 petagrams of carbon per year presently, compared with the 43 to 76 petagrams per year predicted by current models. Storage of anthropogenic carbon in terrestrial vegetation is likely more short-lived and vulnerable than previously predicted.

The processes contributing to the net sink of CO_2 in the terrestrial biosphere are not yet well understood and will likely change in the future (*1*), making it difficult to predict future climate change and create effective mitigation and adaptation policies. Future climate predictions require robust representation of the global carbon cycle, which is challenging when basic properties still have large uncertainties. In particular, observational constraints on global net primary productivity (NPP), the rate of creation of new plant tissues and products, and on carbon turnover rates are lacking. Estimates of global NPP rely on statistical or model-based estimates that use site-scale data (*2*); however, it is very difficult to measure all components of NPP (*3*), and there are not many sites with comprehensive measurements, especially in the tropics (*4*). A large range of global NPP of 43 to 76 petagrams of carbon (PgC) per year is currently simulated by models (*5*, *6*), and these models do not generally show a strong trend over the 20th century. This is in contrast to the trend found for gross primary productivity (GPP) (+30%) (*7*), which is typically twice as large as NPP. Here, we provide global-scale constraints on NPP and carbon turnover by analyzing radiocarbon (^{14}C) produced by nuclear bomb testing and models of the terrestrial biosphere and vegetation.

[1]Department of Physics, Imperial College London, London, UK. [2]Nelson Institute for Environmental Studies and the Department of Geography, University of Wisconsin-Madison, Madison, WI, USA. [3]Department of Earth Sciences, University of Oxford, Oxford, UK. [4]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany. [6]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA. [7]The Nature Conservancy, Arlington, VA, USA. [8]Climate & Global Dynamics, National Center for Atmospheric Research, and Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA.
*Corresponding author. Email: h.graven@imperial.ac.uk
†Deceased.

Global bomb radiocarbon budget

Nuclear bomb testing in the 1950s and 1960s produced excess ^{14}C in the atmosphere (Fig. 1A), which was assimilated into the terrestrial biosphere and ocean through photosynthesis and air-sea gas exchange over time. Tracking how ^{14}C accumulated in the terrestrial biosphere after the bomb testing can therefore enable evaluation of the rates of carbon uptake and turnover (*8*). However, the global accumulation of ^{14}C in the biosphere cannot be observed directly; from new leaves to highly aged soil carbon, there is too much heterogeneity in ^{14}C content in the biosphere.

We use a budgeting approach to diagnose the ^{14}C accumulation in the terrestrial biosphere caused by bomb testing to evaluate carbon cycling in terrestrial biosphere models. In this approach, the ^{14}C accumulation in the terrestrial biosphere is calculated using observations in the stratosphere and troposphere and observationally constrained ocean models to close the ^{14}C budget. In contrast to prior work (*9*) that examined the period from 1945 to 2005, we focus here on the period 1963 to 1967, when atmospheric ^{14}C was highly elevated relative to the biosphere but no strong detonations took place (green area in Fig. 1A) (*10*). Therefore, total ^{14}C in the Earth system was roughly constant but exchanged between reservoirs over 1963 to 1967. This allows us to focus on the period when there was good observational coverage of the stratosphere by aircraft and balloon sampling and to avoid uncertainty and assumptions with calculating the total ^{14}C produced by the bombs and estimating the pre-bomb ^{14}C content. Another advantage of focusing on 1963 to 1967 is that we sharpen the constraint on ^{14}C uptake and turnover in vegetation, where the ^{14}C first entered the terrestrial biosphere, before much ^{14}C was transferred to litter and soil pools.

We used stratospheric data originally published in reports of the Health and Safety Laboratories, which were reassessed and recalculated with corrected standard values (*11*–*13*) and used in an atmospheric model to calculate global stratospheric ^{14}C inventories (*12*) (Fig. 1). Tropospheric ^{14}C inventories were calculated from global compilations recently produced for modeling purposes (*14*, *15*). Ocean ^{14}C simulations (*16*–*19*) that match revised ocean ^{14}C inventories (*20*, *21*) from the 1970s (GEOSECS)

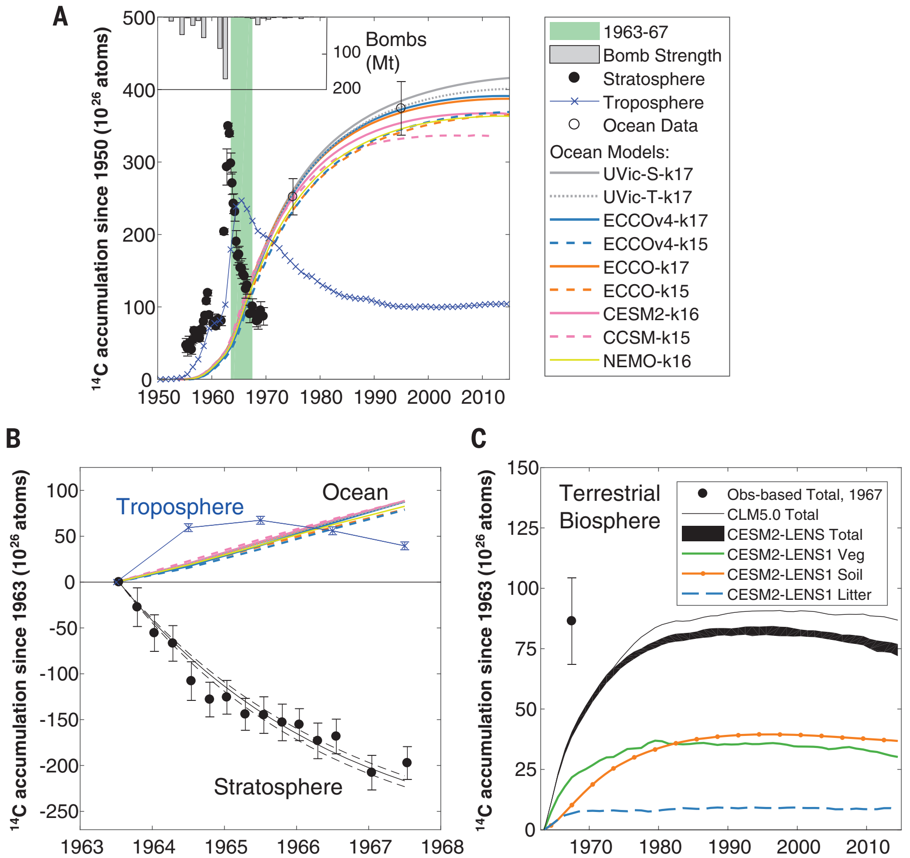


Fig. 1. Budget of excess ^{14}C from nuclear bomb testing. (**A**) Accumulation of ^{14}C in the stratosphere (*12*), troposphere (*14*, *15*), and ocean since 1950 based on observations and simulated accumulation of ^{14}C in ocean models (*16–18*) selected to match observations (*20*, *21*) in the 1970s and 1990s. Inset shows annual nuclear bomb strength in units of megatons of TNT equivalents (*10*). The period 1963 to 1967 with no strong nuclear detonations is highlighted in green. (**B**) ^{14}C accumulation in the stratosphere, troposphere, and ocean since 1963 focusing on the period 1963 to 1967 [green area in (A)]. The black solid line shows an exponential fit to the stratospheric data, and the dashed lines show the 1-σ uncertainty in the χ^2 fit. (**C**) Our new observation-based estimate of ^{14}C accumulation in the terrestrial biosphere in 1967 relative to 1963 (black circle) based on the budgeting approach and simulations of the CLM5.0 model driven with observed climate data (CLM5.0-unc) or as part of CESM2-LENS. The black area shows the range of ^{14}C accumulation in the terrestrial biosphere across nine ensemble members. ^{14}C accumulation in vegetation, soils, and litter (including coarse woody debris) are shown for CESM2-LENS ensemble member 1001.001.

and 1990s (WOCE) were used for ocean ^{14}C inventories.

After the ^{14}C was initially deposited in the stratosphere, the stratosphere lost ~200 × 10^{26} atoms of ^{14}C through mixing of the ^{14}C into the troposphere over 1963 to 1967, which experienced a net gain of about 40 × 10^{26} atoms. The ocean gained about 80 × 10^{26} atoms through air-sea exchange (Fig. 1B). We estimate that the terrestrial biosphere therefore must have accumulated 86 ± 18 ×10^{26} atoms [95% confidence interval (CI)] (*22*) over 1963 to 1967 (Fig. 1C) as the assimilation of ^{14}C outpaced the turnover of ^{14}C back to the air.

Terrestrial biospheric ^{14}C accumulation in the CESM2 model

The terrestrial biospheric ^{14}C accumulation over 1963 to 1967 provides a new constraint on coupled climate-carbon cycle models (also known as Earth system models or ESMs), which are used to inform global climate policy but have particularly uncertain terrestrial carbon cycle components because of the heterogeneity and complexity of land ecosystems. Simulations of the only such land model to simulate ^{14}C explicitly within an ESM, the Community Land Model version 5.0 (CLM5.0) (*23*), accumulate a much lower amount of ^{14}C in the terrestrial biosphere (~40 × 10^{26} atoms) than our observation-based estimate (86 ± 18 ×10^{26} atoms; Fig. 1C). Simulations of CLM5.0 driven with observed climate data (CLM5.0-unc, where “unc” means “uncoupled”) (*24*) and coupled model simulations of the Community Earth System Model 2 (*25*) Large Ensemble Project (CESM2-LENS) (*26*, *27*) following the Coupled Model Intercomparison Project (CMIP) phase 6 historical (concentration-driven) simulation protocol show similar ^{14}C accumulation, and the spread across nine ensemble members is small (Fig. 1C). CLM5.0-unc results are similar to another offline simulation of CLM5.0 that suggested the ^{14}C accumulated in the terrestrial biosphere in the 1960s could be too small (*28*).

In 1963 to 1967, not much bomb ^{14}C had yet entered the soil, as most biospheric bomb ^{14}C was in vegetation (Fig. 1C). In CESM2-LENS, 56% of the ^{14}C accumulated in vegetation, with only 18% in litter and coarse woody debris and 26% in soils over 1963 to 1967. If the ^{14}C accumulation in vegetation in CESM2 were correct, then the ^{14}C accumulation in nonvegetation pools would have to be >3 times larger than simulated in CESM2-LENS and >75% of the total ^{14}C accumulation to match the observation-based estimate. It is unlikely that more than half of the biospheric ^{14}C accumulation over 1963 to 1967 occurred in dead plant material and soils because the peak in global mean tropospheric ^{14}C occurred only in 1964–1965.

We thus conclude that the ^{14}C accumulation in vegetation over 1963 to 1967 in CESM2 is too low (Fig. 1). The underestimate for vegetation could be because the NPP in the model is too low, so not enough ^{14}C enters the vegetation, and/or because carbon is misallocated between short-lived versus long-lived pools, so ^{14}C is turned over too quickly.

Vegetation model emulators and model-data comparisons

CESM2 is the only Earth system model with explicit simulations of ^{14}C available. Therefore, to simulate the ^{14}C accumulation in other models and to explore the sensitivity of the ^{14}C accumulation to NPP and carbon stocks, we needed to construct emulator models. We found that the variables included in CMIP were not sufficient to construct a reliable emulator model for the whole terrestrial biosphere for CESM2, but ^{14}C in vegetation could be modeled reliably (Figs. 2 to 4).

We focus now on analyzing the ^{14}C accumulation only in vegetation in models over 1963 to 1967. We constructed a simple emulator model for woody (long-lived: stem and coarse roots) and nonwoody (short-lived: leaves, fine roots, and other pools) vegetation biomass run on each model grid cell (*22*). We applied the emulator model to CESM2-LENS member 1001.001 and to models from CMIP5 and CMIP6 that reported the necessary variables. We examined global sums for woody and nonwoody pools across all biomes and grid cells (Figs. 2 to 4 and fig. S1), so the global nonwoody vegetation biomass includes the nonwoody vegetation biomass in forests as well as other biomes. We compared these with satellite-based vegetation carbon products (*29–31*) that omit leaf carbon in forests, so we

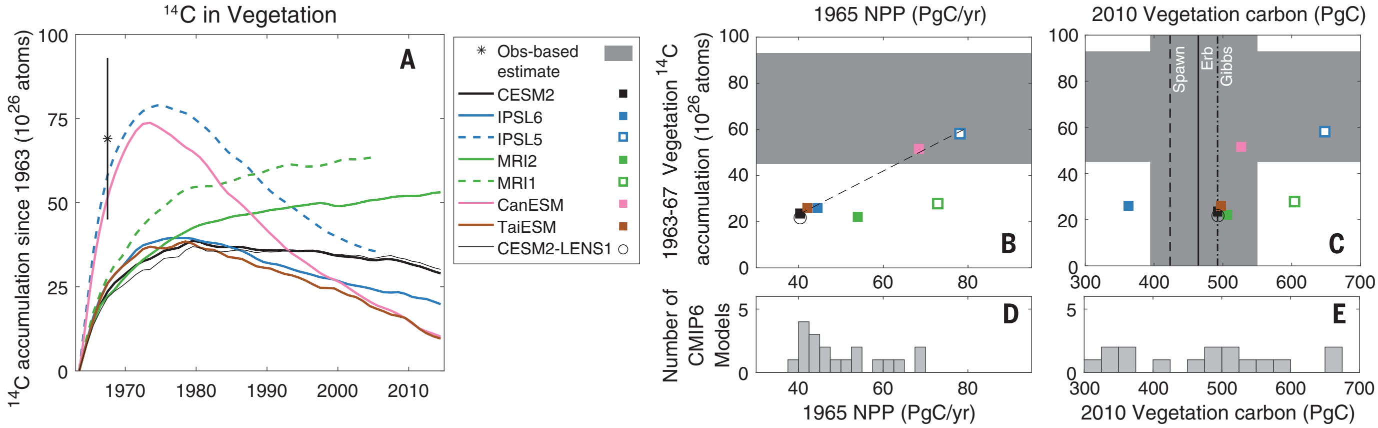


Fig. 2. Model-data comparison for vegetation in the emulator models and in CESM2. (**A**) Simulated accumulation of ^{14}C in vegetation since 1963 compared with the observation-based estimate of ^{14}C accumulation in vegetation over 1963 to 1967. (**B**) Accumulation of ^{14}C in vegetation over 1963 to 1967 versus NPP in 1965 in each emulator model and CESM2, including a regression line for emulator models excluding MRI models. Gray area shows the uncertainty range in the observation-based estimate of ^{14}C accumulation. (**C**) Accumulation of ^{14}C in vegetation over 1963 to 1967 versus carbon stock in vegetation in 2010 (2005 for MRI1 and IPSL5) in each emulator model and CESM2, including observation-based estimates of vegetation carbon stock (*29*–*31*). The gray area reflects the uncertainty from Erb *et al*. (*31*) and uncertainty in ^{14}C accumulation. (**D** and **E**) Histograms of NPP in 1965 and carbon stock in vegetation in 2010 in CMIP6 models, including additional models that could not be included in the vegetation emulator simulations because the available CMIP6 output for these models lacked the necessary variables to run the emulator model (table S2). The explicit simulation of ^{14}C in vegetation in CESM2-LENS member 1001.001 (CESM2-LENS1) is shown in (A) to (C) for comparison with the CESM2 vegetation emulator model.

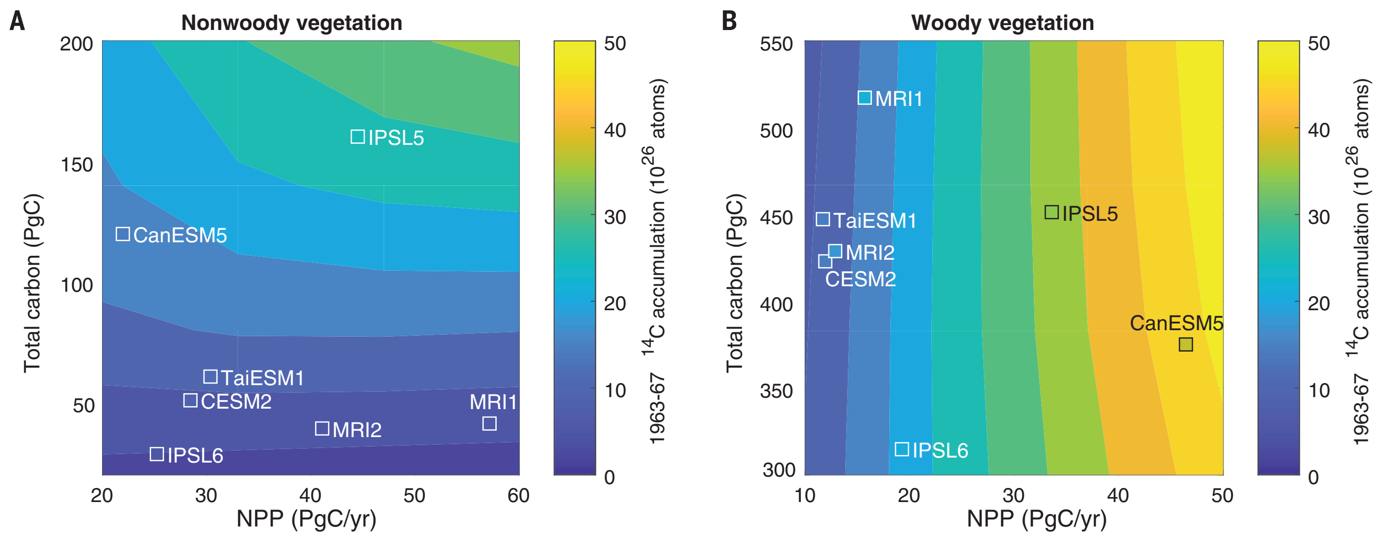


Fig. 3. Sensitivity of ^{14}C accumulation to NPP and total carbon. Accumulation of ^{14}C over 1963 to 1967 in nonwoody (**A**) and woody (**B**) vegetation biomass plotted in color with NPP and total carbon stock in 1965 on the *x* and *y* axes. Contours reflect relationships across 16 simulations of the CESM2 emulator, where NPP and total carbon stock were scaled across the range shown here. Symbols show ^{14}C accumulation in the emulator models using the same color bar.

estimated global total leaf carbon in forests to be 14.3 PgC [based on table S5 in (*32*)] and added this to the observation-based estimates of vegetation carbon stocks.

To evaluate the vegetation ^{14}C simulations, we estimated the true ^{14}C accumulation in vegetation by subtracting the ^{14}C accumulation in litter, coarse woody debris, and soils simulated by CESM2-LENS member 1001.001 from the observation-based total terrestrial biosphere ^{14}C accumulation over 1963 to 1967. We allowed the uncertainty in nonvegetation ^{14}C accumulation to be ±100% (95% CI) (*22*), even though CESM2/CLM5 is in fact likely to overestimate this ^{14}C accumulation because its proportion of fresh carbon in both surface and subsurface soils has been shown to be too high (*33*). Our estimate of vegetation ^{14}C accumulation is $69 \pm 24 \times 10^{26}$ atoms (95% CI) over 1963 to 1967, which allows for a possible range of 43 to 100% of biospheric ^{14}C accumulation in vegetation.

Most of the CMIP5 and CMIP6 vegetation emulator models (five of seven) underestimate the observation-based vegetation ^{14}C accumulation over 1963 to 1967 (Fig. 2). The two models that match the observation-based vegetation ^{14}C accumulation have high NPP of >68 PgC/yr in 1965 (Fig. 2B, fig. S1, and table S1). One of the two models is from CMIP5 (IPSL5), whereas the CMIP6 version of that model (IPSL6) has much lower NPP and underestimates the observation-based vegetation bomb ^{14}C inventory. The other model matching the observation-based vegetation bomb ^{14}C inventory, CanESM5 from CMIP6,

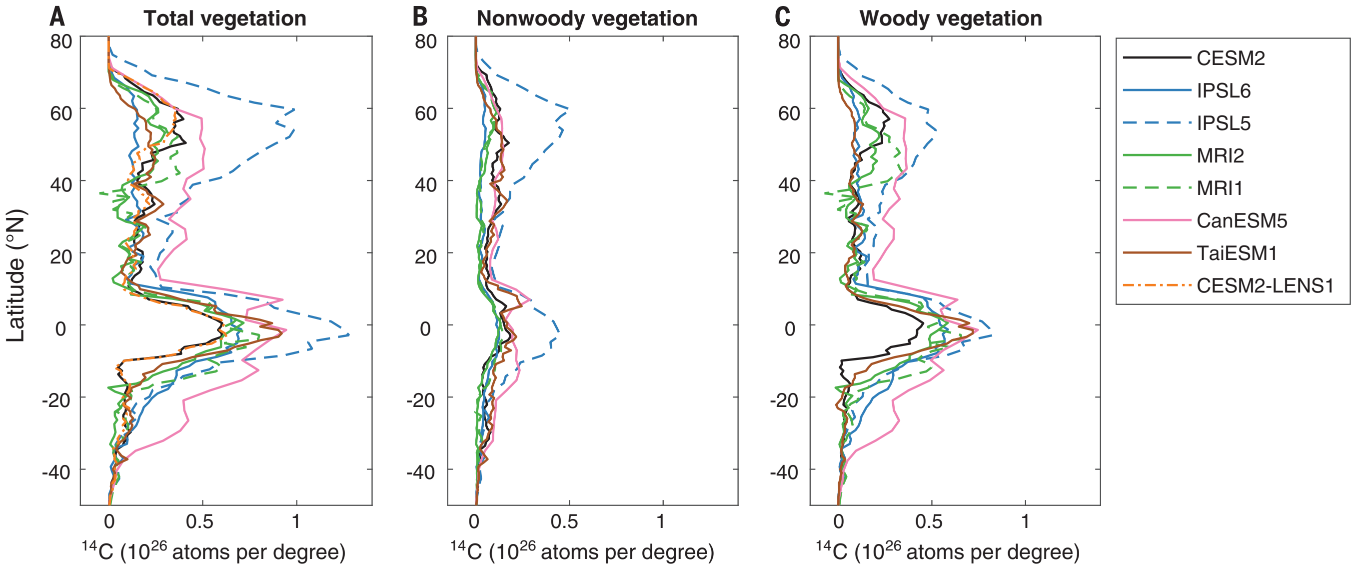


Fig. 4. Spatial distribution of ^{14}C accumulation simulated in vegetation in the emulator models and in CESM2. Accumulation of ^{14}C over 1963 to 1967 per degree latitude in total (**A**), nonwoody (**B**), and woody (**C**) vegetation biomass integrated over all longitudes. The explicit simulation of ^{14}C in CESM2-LENS1 is shown in (A) for comparison with the CESM2 emulator model.

has high NPP and allocates a large fraction of its NPP to wood (68% in 1965), in contrast to other models allocating 22 to 43% of NPP to wood (table S1).

Overall, the ^{14}C accumulation in vegetation over 1963 to 1967 shows a strong relationship with NPP but not with vegetation carbon stock (Fig. 2). This indicates that higher NPP increases ^{14}C accumulation in vegetation over 1963 to 1967, but higher carbon stock (and slower turnover rate) generally does not. Two versions of the MRI model lie below a regression line between ^{14}C accumulation in vegetation and NPP for the other five models (Fig. 2B). The MRI models allocate the highest fraction of NPP to nonwoody vegetation (76 to 78% to nonwoody and 22 to 24% to woody), and their nonwoody annual NPP is similar to their nonwoody carbon stock (table S1), which indicates a very high level of productivity per unit biomass and a fast turnover rate. Therefore, the flux of ^{14}C into nonwoody vegetation in the MRI models is large but is turned over quickly, and the ^{14}C accumulation in nonwoody vegetation is among the lowest (Fig. 3A).

There are differing controls on ^{14}C accumulation over 1963 to 1967 in nonwoody versus woody vegetation biomass in the emulator models (Fig. 3 and figs. S2 and S3). Accumulation of ^{14}C in longer-lived woody vegetation is sensitive to NPP, whereas accumulation of ^{14}C in shorter-lived nonwoody vegetation is more sensitive to the carbon stock. At higher stocks of nonwoody vegetation carbon, ^{14}C accumulation in nonwoody vegetation is also sensitive to NPP. The patterns found for scaling experiments in the CESM2 vegetation emulator (contours in Fig. 3) are similar to the patterns found for the other vegetation model emulators (colored symbols in Fig. 3).

The patterns in Fig. 3 indicate that underestimated ^{14}C accumulation in vegetation over 1963 to 1967 is due to underestimated NPP or underestimated nonwoody vegetation biomass in models. Only IPSL6 underestimates the total vegetation carbon stock estimated with satellite data (Fig. 2C and fig. S4), so increasing nonwoody carbon stock in the models requires that carbon shifts from woody biomass (stems and coarse roots) to nonwoody biomass (leaves, fine roots, and other biomass) by adjustment of their turnover rates. The models tend to underestimate belowground vegetation carbon stocks (*29*, *30*) (fig. S5), so shifting aboveground woody carbon (stems) to belowground nonwoody carbon (fine roots) may be required. Conversely, NPP in woody (or nonwoody) vegetation could be increased in the models without necessarily affecting carbon stocks if modeled turnover rates are simultaneously increased.

The regression between vegetation ^{14}C accumulation and NPP ($R^2 > 0.99$), excluding the MRI models that have very high nonwoody NPP, suggests that NPP in 1965 should have been at least 63 PgC per year (the value of NPP at the intersection of the regression line and ^{14}C accumulation uncertainty range in Fig. 2B). However, only 16% of all CMIP6 models have NPP higher than 63 PgC per year in 1965 (Fig. 2D and table S2). Considering that total carbon assimilation (GPP) increased by ~30% over the 20th century (*7*), if carbon uptake efficiency (NPP/GPP) did not change significantly, then NPP should be at least 80 PgC per year presently, but it is only 43 to 76 PgC per year in current models (*5*).

Implications for the carbon cycle

The simulations of ^{14}C that we analyzed provide evidence that CESM2 and most other CMIP6 models underestimate the magnitude of NPP in the 1960s. The minimum NPP of 63 PgC per year in 1965 and 80 PgC per year recently [applying a 30% increase according to (*7*)] that is implied by our analysis of bomb ^{14}C in vegetation is higher than simulated in most CMIP6 models (*5*) (Fig. 2) but within the higher end of the range of observation-based estimates of GPP (*34*–*37*), assuming ~50% NPP/GPP. The global NPP/GPP ratio might increase slightly in the future (*38*), but we are not aware of any evidence for a historical trend. The average NPP in CMIP6 models actually decreased compared with CMIP5 models (*5*, *39*), which likely degraded the model cohort rather than improved it.

Our results highlight parametric and structural uncertainties in model simulations of leaf-level photosynthesis and stomatal conductance, nutrient limitation, autotrophic respiration, carbon allocation, mortality, and turnover. For example, replacing the widely used assumption of homogeneity in wood carbon turnover rates at a given location (*40*) with vegetation demographic models (*41*) that allow distinct populations of fast-growing versus long-lived trees may improve ^{14}C accumulation, where the former are able to rapidly take up ^{14}C whereas the latter dominate the overall biomass pool (*42*). However, because ^{14}C accumulation over 1963 to 1967 is higher in woody than nonwoody vegetation (Fig. 3 and figs. S1 and S4), it is likely that increasing NPP to woody vegetation in models that underestimate ^{14}C accumulation is required. Satisfying observational constraints on carbon stocks while increasing NPP will require that the rate of carbon turnover in the models also increases.

A range of 41 to 64 PgC per year for NPP was found in a previous study using a ^{14}C budget to diagnose the bomb-produced ^{14}C in the biosphere (*9*) and then using this budget to fit parameters in a simple three-box global biosphere model (*43*). Our evaluation of state-of-the-art global biosphere models suggests that the ^{14}C budget in the 1960s cannot be met with NPP lower than 63 PgC per year in current model formulations (Fig. 2B). This is in fact consistent with (*9*), in which the budget was not closed in the 1960s and instead included a residual "hidden sink" that must be in the terrestrial biosphere.

Radiocarbon data provide powerful and unique insights on carbon cycling and model evaluation, but they have been underused because of the low number of models simulating ^{14}C. In addition to the observation-based global ^{14}C accumulation used here and soil carbon ^{14}C data used previously to evaluate CMIP models (*33*, *44*), other data including ^{14}C in specific soil compounds, in respiration, or in atmospheric CO_2 could be used to evaluate more processes in models that simulate ^{14}C. Analyzing the 1963 to 1967 period allowed us to focus on vegetation, but longer analysis of subsequent decades would enable critical insights on whole-ecosystem cycling, including litter and soil (Fig. 1C). Within vegetation alone, ^{14}C simulations strongly diverge over time (Fig. 2A), and there are large differences between models in spatial distribution of ^{14}C accumulation, NPP, and carbon stock (Fig. 4 and fig. S6). Spatial differences in ^{14}C accumulation between models are at least a factor of two but up to a factor of 10 for nonwoody vegetation in northern temperate and boreal regions. Additional ^{14}C data-model comparisons will enable more constraints on various processes. In addition, because we estimated the 1963 to 1967 ^{14}C accumulation in litter and soils based on the CESM-LENS simulations (with ±100% uncertainty), further analysis of ^{14}C through all biospheric pools would help to refine the constraints on vegetation.

The vegetation emulator model that we used here represents the ^{14}C explicitly simulated in CESM2 well (Figs. 2 to 4), but the emulator could not be evaluated for other models, and emulators for litter and soil pools could not be constructed with the limited variables in the CMIP output. Ensuring an accurate representation of ^{14}C in biospheric models requires that the models explicitly simulate ^{14}C, which only requires one additional tracer to be added in a simple way (*22*). New methods for fast spin-up could be exploited (*45*–*47*). As requested for CMIP6 (*48*), we strongly recommend that modeling groups implement ^{14}C in ESMs and in stand-alone models and report these results to CMIP and related activities to enable model assessment and scientific understanding.

Accurate simulation of vegetation and total biospheric carbon uptake and turnover is critical to understanding historical and future anthropogenic carbon storage in terrestrial ecosystems, both for natural sinks of CO_2 and for "nature-based solutions" that aim to remove atmospheric CO_2 by increasing land ecosystem carbon. Our analysis shows that the uptake of carbon through NPP and the rate of carbon turnover in models must both be increased, which will increase the turnover of anthropogenic carbon in the terrestrial biosphere. Because the uptake and turnover of carbon are the main controls on the anthropogenic CO_2 sink in the terrestrial biosphere, the results of our study suggest that the storage of anthropogenic carbon in the terrestrial biosphere is likely more short-lived and more vulnerable to future changes than previously thought.

REFERENCES AND NOTES

1. P. Friedlingstein *et al*., *Earth Syst. Sci. Data* **15**, 5301–5369 (2023).
2. A. Ito, *Glob. Change Biol.* **17**, 3161–3175 (2011).
3. D. A. Clark *et al*., *Ecol. Appl.* **11**, 356–370 (2001).
4. R. J. Olson, J. M. O. Scurlock, S. D. Prince, D. L. Zheng, K. R. Johnson, "NPP Multi-Biome: Global Primary Data Initiative Produces, R2" (ORNL DAAC, 2013); https://doi.org/10.3334/ORNLDAAC/617.
5. R. M. Varney, S. E. Chadburn, E. J. Burke, P. M. Cox, *Biogeosciences* **19**, 4671–4704 (2022).
6. M. Zhao, F. A. Heinsch, R. R. Nemani, S. W. Running, *Remote Sens. Environ.* **95**, 164–176 (2005).
7. J. E. Campbell *et al*., *Nature* **544**, 84–87 (2017).
8. S. E. Trumbore, *Ecol. Appl.* **10**, 399–411 (2000).
9. T. Naegler, I. Levin, *J. Geophys. Res.* **114**, D17302 (2009).
10. United Nations, "UNSCEAR 2000 report volume 1" (UN Scientific Committee on the Effects of Atomic Radiation, 2000); https://www.unscear.org/unscear/en/publications/2000_1.html.
11. T. Naegler, Simulating bomb radiocarbon: Consequences for the global carbon cycle, PhD Thesis, University of Heidelberg, Heidelberg, Germany (2005).
12. T. Naegler, I. Levin, *J. Geophys. Res.* **111**, D12311 (2006).
13. V. Hesshaimer, I. Levin, *J. Geophys. Res.* **105** (D9), 11641–11658 (2000).
14. H. Graven *et al*., *Geosci. Model Dev.* **10**, 4405–4417 (2017).
15. M. Meinshausen *et al*., *Geosci. Model Dev.* **10**, 2057–2116 (2017).
16. H. D. Graven, N. Gruber, R. Key, S. Khatiwala, X. Giraud, *J. Geophys. Res.* **117**, C10005 (2012).
17. S. Khatiwala, H. Graven, S. Payne, P. Heimbach, *Geophys. Res. Lett.* **45**, 5617–5626 (2018).
18. B. Pookkandy, H. Graven, A. Martin, *Clim. Dyn.* **61**, 3223–3235 (2023).
19. J. G. Lester, Changes in oceanic radiocarbon and CFCs since the 1990s, Zenodo (2023); https://doi.org/10.5281/zenodo.8242144.
20. T. Naegler, *Tellus B Chem. Phys. Meterol.* **61**, 372–384 (2009).
21. S. Peacock, *Global Biogeochem. Cycles* **18**, 2003GB002211 (2004).
22. Materials and methods are available as supplementary materials online.
23. D. M. Lawrence *et al*., *J. Adv. Model. Earth Syst.* **11**, 4245–4287 (2019).
24. CLM5.0 data for: H. D. Graven, H. Warren, H. K. Gibbs, S. Khatiwala, C. Koven, J. Lester, I. Levin, S. A. Spawn-Lee, W. Wieder, Bomb radiocarbon evidence for strong global carbon uptake and turnover in terrestrial vegetation, Climate Data Gateway (2024); https://doi.org/10.5065/d6154fwh.
25. G. Danabasoglu *et al*., *J. Adv. Model. Earth Syst.* **12**, e2019MS001916 (2020).
26. K. B. Rodgers *et al*., *Earth Syst. Dyn.* **12**, 1393–1411 (2021).
27. CESM2-LENS data for: H. D. Graven, H. Warren, H. K. Gibbs, S. Khatiwala, C. Koven, J. Lester, I. Levin, S. A. Spawn-Lee, W. Wieder, Bomb radiocarbon evidence for strong global carbon uptake and turnover in terrestrial vegetation, NCAR Community Earth System Model (2024); https://www.cesm.ucar.edu/community-projects/lens2.
28. T. Frischknecht, A. Ekici, F. Joos, *Global Biogeochem. Cycles* **36**, e2021GB007042 (2022).
29. H. K. Gibbs, A. Ruesch, "New IPCC tier-1 global biomass carbon map for the year 2000" (US Department of Energy, Office of Scientific and Technical Information, 2008); https://www.osti.gov/dataexplorer/biblio/dataset/1463800.
30. S. A. Spawn, C. C. Sullivan, T. J. Lark, H. K. Gibbs, *Sci. Data* **7**, 112 (2020).
31. K.-H. Erb *et al*., *Nature* **553**, 73–76 (2018).
32. P. B. Reich *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13721–13726 (2014).
33. Z. Shi *et al*., *Nat. Geosci.* **13**, 555–559 (2020).
34. C. Beer *et al*., *Science* **329**, 834–838 (2010).
35. L. R. Welp *et al*., *Nature* **477**, 579–582 (2011).
36. G. Badgley, L. D. L. Anderegg, J. A. Berry, C. B. Field, *Glob. Chang. Biol.* **25**, 3731–3740 (2019).
37. M.-C. Liang *et al*., *Sci. Rep.* **13**, 2162 (2023).
38. J. M. Mathias, A. T. Trugman, *Ecol. Lett.* **25**, 498–508 (2022).
39. N. Wei *et al*., *J. Clim.* **35**, 5483–5499 (2022).
40. A. Wolf *et al*., *Global Biogeochem. Cycles* **25**, n/a (2011).
41. R. A. Fisher *et al*., *Glob. Chang. Biol.* **24**, 35–54 (2018).
42. J. A. Lutz *et al*., *Glob. Ecol. Biogeogr.* **27**, 849–864 (2018).
43. T. Naegler, I. Levin, *J. Geophys. Res.* **114**, D17303 (2009).
44. Y. He *et al*., *Science* **353**, 1419–1424 (2016).
45. Y. Luo *et al*., *J. Adv. Model. Earth Syst.* **14**, e2022MS003008 (2022).
46. H. Metzler *et al*., *J. Adv. Model. Earth Syst.* **12**, e2019MS001776 (2020).
47. S. Khatiwala, *J. Adv. Model. Earth Syst.* **15**, e2022MS003447 (2023).
48. C. D. Jones *et al*., *Geosci. Model Dev.* **9**, 2853–2880 (2016).
49. CMIP5 and CMIP6 data for: H. D. Graven, H. Warren, H. K. Gibbs, S. Khatiwala, C. Koven, J. Lester, I. Levin, S. A. Spawn-Lee, W. Wieder, Bomb radiocarbon evidence for strong global carbon uptake and turnover in terrestrial vegetation, Earth System Grid Federation (2024); https://esgf-node.llnl.gov/projects/esgf-llnl/.

ACKNOWLEDGMENTS

We dedicate this paper to the memory of Ingeborg Levin, a pioneering, dedicated, and beloved scientist whose ambitions and efforts to measure and use ^{14}C to understand the carbon cycle laid the foundation for this study. We acknowledge the CESM2 Large Ensemble Community Project and supercomputing resources provided by the IBS Center for Climate Physics in South Korea; the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6 and CMIP5; the climate modeling groups for producing and making available their model output; the Earth System Grid Federation (ESGF) for archiving the data and providing access; and the multiple funding agencies who support CMIP6, CMIP5, and ESGF. **Funding:** This work was supported by the Leverhulme Trust, Philip Leverhulme Prize PLP-2018-038; Royal Society Wolfson Fellowship RSWF\FT\191013; Schmidt Sciences, LLC for LEMONTREE (Land Ecosystem Models based On New Theory, obseRvations and ExperimEnts); and by the Office of Science, Office of Biological and Environmental Research of the US Department of Energy under contract DE-AC02-05CH11231 through the Regional and Global Model Analysis (RUBISCO SFA) and Environmental Systems Science (NGEE-Tropics) Programs **Author contributions:** Conceptualization: H.D.G.; Methodology: H.D.G., H.W., C.K., W.W.; Investigation: H.D.G., H.W., H.K.G., S.K., C.K., J.L., I.L., S.S., W.W.; Visualization: H.D.G., J.L.; Writing – original draft: H.D.G.; Writing – review and editing: H.D.G., H.W., H.K.G., S.K., C.K., J.L., I.L., S.S., W.W. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** CMIP5 and CMIP6 model output data are available at (*49*), CLM5.0 data are available at (*24*), and CESM2-LENS data are available at (*26*). Other data are available from references (*10*), (*12*), (*14*–*19*), and (*30*–*32*). All data plotted in the figures are tabulated in data S1 as an Excel spreadsheet.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl4443
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
Data S1

Submitted 18 October 2023; accepted 9 May 2024
10.1126/science.adl4443

QUANTUM OPTICS

Topological Hong-Ou-Mandel interference

Max Ehrhardt[1], Christoph Dittel[2,3], Matthias Heinrich[1], Alexander Szameit[1]*

The interplay of topology and optics provides a route to pursue robust photonic devices, with the application to photonic quantum computation in its infancy. However, the possibilities of harnessing topological structures to process quantum information with linear optics, through the quantum interference of photons, remain largely uncharted. Here, we present a Hong-Ou-Mandel interference effect of topological origin. We show that this interference of photon pairs—ranging from constructive to destructive—is solely determined by a synthetic magnetic flux, rendering it resilient to errors on a fundamental level. Our implementation establishes a quantized flux that facilitates exclusively destructive quantum interference. Our findings pave the way toward the development of next-generation photonic quantum circuitry and scalable quantum computing protected by virtue of topologically robust quantum gates.

Topological quantum photonics combines two emerging fields of physics: In one of these, topological states in condensed-matter physics with their robust wave transport have inspired efforts to apply this universal concept for wave systems to electromagnetics (*1*), photonics (*2*, *3*), acoustics (*4*), cold atoms (*5*, *6*) and mechanics (*7*). By contrast, quantum photonics uses interference (*8*) and entanglement (*9*, *10*) among many-photon states for quantum computation and quantum information processing (*11*, *12*). As the scalability of the latter is limited by scattering losses and the detrimental impact of manufacturing imperfections, bringing topological features to bear on photonic quantum circuits is a particularly promising approach to establish robustness in photonic quantum computing (*13*). The first demonstrations of topological transport on edge channels of single photons (*14*) were soon followed by proposals to use similar principles to safeguard delicate biphoton properties such as correlations (*15*) and entanglement (*16*). Furthermore, it has been experimentally shown that even one-dimensional topological systems outperform their conventional counterparts (*17*, *18*). It is also recognized that chiral quantum-optic interfaces (*19*) may serve as topological sources of distinguishable (*20*) and indistinguishable photon pairs (*21*) with inherent robustness in spectral correlations between the photon pairs. Beyond the synthesis and transport of such states, Hong-Ou-Mandel (HOM) interference, two-photon interference in a beam splitter or directional coupler leading to the suppression of coincidences among the output of the photonic device (*8*, *22*, *23*), is a fundamental building block for linear optical quantum computing.

Here, we introduce topological HOM interference equipped with intrinsic topological protection against imperfections. Specifically, we apply Aharonov-Bohm (AB) interference of a single particle (*24*) to photon pairs, such that their interference is solely dependent on the magnetic flux and the associated Berry phase (*25*, *26*). The hallmark two-photon suppression occurs at a flux of π and has to be implemented by means of artificial gauge fields, because photons as uncharged particles are immune to external fields. We tailor the birefringence of waveguides in a directional coupler to leverage the additional dynamics in the polarization degree of freedom to implement artificial gauge fields with a quantized magnetic flux of π. Note that in momentum space (*27*), this relates to a Chern number (*28*) as topological invariant. Consequently, coincidences of indistinguishable photons are fully suppressed regardless of the length of the directional coupler. This is in stark contrast to conventional designs wherein the visibility of quantum interference periodically oscillates upon propagation (*29*, *30*). Further we demonstrate experimentally that our topological architecture protects the HOM interference against perturbations such as impurities of the input states or deviations in coupling strengths.

Photonic system—theory and experimental platform

Our system is a square plaquette (in the following symbolized as $\boxdot$, with Hamiltonian $\hat{\mathcal{H}}_{\boxdot}$) of four sites with equal couplings $C_{j,k}$ between the pairs of sites $(j,k) \in \mathcal{M} = \{(1,2),(1,3),(2,4),(3,4)\}$ that is permeated by a flux Φ by virtue of an artificial magnetic field (Fig. 1A). The propagation of a single photon along the z direction, created on site j with respective creation operator $\hat{a}_j^\dagger$, is then described by the Heisenberg equation of motion

$$i\frac{\partial}{\partial z}\hat{a}_j^\dagger = \left[\hat{a}_j^\dagger, \hat{\mathcal{H}}_{\boxdot}\right],$$
$$\hat{\mathcal{H}}_{\boxdot} = C\sum_{(j,k)\in\mathcal{M}} e^{i\varphi_{j,k}}\hat{a}_j^\dagger\hat{a}_k + h.c. \qquad (1)$$

where h.c. denotes the Hermitian conjugate and $\varphi_{j,k} = \arg C_{j,k}$ are the arguments of the complex-valued coupling strengths $|C_{j,k}| = C$. In total, Peierls' substitution (*31*) gives rise to the flux $\Phi = \oint_{\boxdot} d\varphi$ enclosed by the plaquette. The resulting single-photon propagation is closely related to the evolution of an electron in the AB effect (Fig. 1B): Single photons with the initial state $\hat{a}_1^\dagger|\mathbf{0}\rangle = |1000\rangle$ propagate clockwise (blue path) and counterclockwise (orange path) around the flux-threaded plaquette and interfere at site $j = 4$ (gray sphere) with flux-induced phase differences that are accumulated along the different paths, where $|\mathbf{0}\rangle$ denotes the multimode vacuum state. Thus, the probability of finding the photon on site $j = 4$ scales with $\propto \cos^2\frac{\Phi}{2}$, in particular establishing constructive (destructive) interference at $\Phi = 0$ ($\Phi = \pi$), respectively, as has been recently demonstrated with classical laser light in an integrated photonic setting (*32*, *33*). We apply this effect to a two-photon evolution on the plaquette. Figure 1C illustrates the contributing single-photon trajectories (blue and orange arrows) that convert the initial two-photon state $\hat{a}_1^\dagger\hat{a}_4^\dagger|\mathbf{0}\rangle = |1001\rangle$ into $\hat{a}_2^\dagger\hat{a}_3^\dagger|\mathbf{0}\rangle = |0110\rangle$. The underlying principle behind the considered two-photon interference is contained in the transition from distinguishable to indistinguishable photons: The twofold Hamiltonian of two distinguishable photons a and b, $\hat{\mathcal{H}}_2 = \hat{\mathcal{H}}_{\boxdot}^a \oplus \hat{\mathcal{H}}_{\boxdot}^b$ on the square plaquette can be illustrated by the mesh shown in Fig. 1D. Notably, this resembles a torus (Fig. 1E) where the toroidal and poloidal coordinates encode the positions of photons a and b on the square plaquette. Crucially, each photon is exclusively affected by the specific flux that it is subjected to. In the transition from distinguishable to indistinguishable photons depicted in Fig. 1F, two-photon states, which are identical under the photon's permutation, are removed by folding the mesh along the line of states in which the photons coincide. The resulting two-photon Hamiltonian (Fig. 1G) instantiates a Möbius strip (*34*) that encloses the same flux Φ (*35*) as the square plaquette does for single photons [details provided in the supplementary text (*36*)]. Furthermore, the single-photon AB effect depicted in Fig. 1B is repeated in the two-photon trajectories that transfer to clockwise (blue) and counterclockwise (orange) trajectories around the Möbius strip (Fig. 1G). Hence, the interference of the two paths dictates a scaling of the probability of finding the photons in the state $|0110\rangle$ with $\propto \cos^2\frac{\Phi}{2}$, and, in particular, allows for this state to be suppressed by means of destructive quantum interference when $\Phi = \pi$, whereas distinguishable photons can still be found in this configuration.

Next, we turn to the implementation of a quantized $\Phi = \pi$ flux on a square plaquette as basis for error-resilient HOM interference. To this end, we use a birefringent directional coupler (BDC) composed of two waveguides with, in principle, independently customizable principal axis orientations as well as strengths of

[1]University of Rostock, Institute of Physics, Albert-Einstein-Str. 23, 18059 Rostock, Germany. [2]Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Hermann-Herder-Str. 3, 79104 Freiburg, Germany. [3]EUCOR Centre for Quantum Science and Quantum Computing, Albert-Ludwigs-Universität Freiburg, Hermann-Herder-Str. 3, 79104 Freiburg, Germany.
*Corresponding author. Email: alexander.szameit@uni-rostock.de

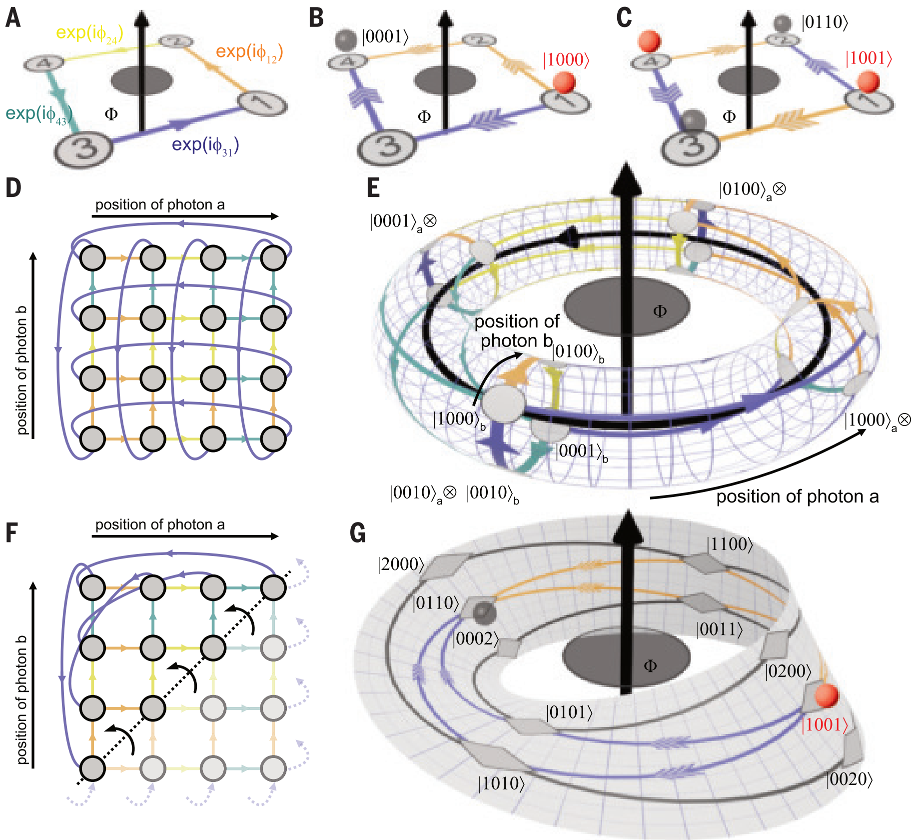


Fig. 1. The two-particle Aharonov-Bohm effect on a square plaquette. (**A**) Schematic illustration of a square plaquette as a Hamiltonian with complex hopping rates. According to Peierls' substitution, the integral of the coupling rate's arguments on the closed loop gives rise to the magnetic flux Φ shown in the center. (**B**) A single particle in the initial state $\hat{a}_1^\dagger|\mathbf{0}\rangle = |1000\rangle$ (red sphere) at site 1 of a square plaquette encloses a magnetic flux Φ with its clockwise and counterclockwise trajectories (marked blue and orange via site 3 and 2, respectively) on the way to its final state $\hat{a}_4^\dagger|\mathbf{0}\rangle = |0001\rangle$ (gray sphere). (**C**) Two photons launched in the initial state $\hat{a}_1^\dagger\hat{a}_4^\dagger|\mathbf{0}\rangle = |1001\rangle$ at the sites 1 and 4 (red spheres) can take either the blue or orange marked trajectories to the final state $\hat{a}_2^\dagger\hat{a}_3^\dagger|\mathbf{0}\rangle = |0110\rangle$, where the photons are found on sites 2 and 3, respectively. (**D**) The Hamiltonian for distinguishable biphotons as a twofold single-photon Hamiltonian $\hat{\mathcal{H}}_2 = \hat{\mathcal{H}}_\square^a \oplus \hat{\mathcal{H}}_\square^b$ resembles (**E**), a torus with the position of a photon a and b as a toroidal and poloidal coordinate, respectively. The color scheme of the connections corresponds to that in (A). (**F**) The mesh of the torus is converted to the Hamiltonian for indistinguishable photons. Hence, sites, that are identical under photon exchange are removed by folding the twofold Hamiltonian along the line of its orbifold points (points of coinciding single photons). (**G**) Reconnecting the mesh in the same manner as the torus, the connectivity pattern instantiates a Möbius strip threaded with a magnetic flux Φ through the center, where each site (diamond) and each connection represents one two-particle configuration and connection of an individual particle on the square plaquette, respectively. The marked paths in (C) now represent clockwise (blue) and counterclockwise (orange) paths from initial to final two-particle states on a connectivity tree shaped as a Mobius strip threaded with a magnetic flux Φ through the center.

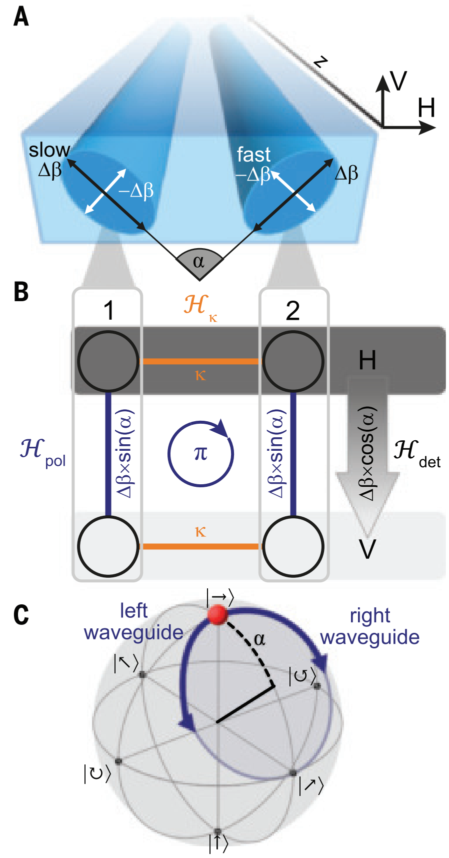


Fig. 2. Single-photon dynamics in a birefringent directional coupler. (**A**) Two identical, birefringent waveguides with slow axes rotated clockwise and counterclockwise by $\pm\alpha/2$ with respect to the vertical reference. Therefore, the slow axes intersect at an angle α. (**B**) The resulting single-photon dynamics in this waveguide system are driven by the evanescent coupling ($\hat{\mathcal{H}}_\kappa$, orange) with the rate κ between the two waveguides (1, 2), the polarization coupling ($\hat{\mathcal{H}}_{\mathrm{pol}}$, blue) with rate $\Delta\beta \sin\alpha$, and the detuning ($\hat{\mathcal{H}}_{\mathrm{det}}$) between horizontal (H) and vertical (V) polarizations with the rate $\Delta\beta \cos\alpha$. (**C**) Depending on which waveguide it occupies, the polarization of an initially horizontally polarized photon launched in the BDC performs a left- or right-handed rotation on the Poincare sphere around an axis that is located in the S_1-S_2 plane. The rotation axis and the S_1 axis enclose the angle α.

birefringence. For the sake of simplicity, we consider waveguides with identical birefringence $\Delta\beta = (\beta_f - \beta_s)/2$ (Fig. 2A). To establish single-photon dynamics in the polarization degree of freedom (*37*), a horizontal and vertical (H and V) reference frame is chosen such that the slow axes are oriented at angles of $\pm\alpha/2$ with respect to the V axis. Consequently, the single-photon Hamiltonian $\hat{\mathcal{H}}_{\mathrm{BDC}} = \hat{\mathcal{H}}_\kappa + \hat{\mathcal{H}}_{\mathrm{pol}} + \hat{\mathcal{H}}_{\mathrm{det}}$ comprises evanescent coupling $\hat{\mathcal{H}}_\kappa$ with a coupling rate κ (similar to a conventional directional coupler) and polarization coupling $\hat{\mathcal{H}}_{\mathrm{pol}} \propto \pm\Delta\beta \sin\alpha$ (sign changes among the waveguides) as well as polarization detuning $\hat{\mathcal{H}}_{\mathrm{det}} \propto \Delta\beta \cos\alpha$ in each waveguide [see supplementary text for details (*36*)]. The joint dynamics (Fig. 2B) mediated by $\hat{\mathcal{H}}_\kappa + \hat{\mathcal{H}}_{\mathrm{pol}}$ give rise to a connectivity pattern in a rectangularly shaped plaquette, whereas the polarization detuning $\hat{\mathcal{H}}_{\mathrm{det}}$ can be interpreted as a scalar electric potential of strength $\Delta\beta \cos\alpha$, i.e., an artificial electric field that vanishes for perpendicular slow axes. Crucially, $\hat{\mathcal{H}}_{\mathrm{pol}}$ entails real-valued coupling rates with opposite sign. Hence, the π difference in the arguments in the polarization coupling establishes the $\Phi = \pi$-flux when combined with the evanescent coupling rates with $\arg\kappa = 0$ of $\hat{\mathcal{H}}_\kappa$. The manifestation of the $\Phi = \pi$-flux in the BDC is illustrated in Fig. 2C: During its propagation, a horizontally polarized photon rotates around one axis in the S_1-S_2 plane (orientation given by α) on the Poincaré sphere. Whether the rotation is left- or right-handed depends on which waveguide the photon occupies. The opposite handedness does not change with α, unless the trajectories shrink into a single point for

$\alpha = 0°$ $(180°)$. Thus, the artificial magnetic flux over a full roundtrip is quantized as

$$\Phi = \begin{cases} 0 & \alpha = 0°, \\ \pi & \text{otherwise} \end{cases}$$

Note that in the case $\alpha = 0°$ (both principal axes parallel to the input state), in which no polarization oscillation occurs in the H/V reference frame, arbitrarily rotated reference frames provide rectangularly shaped plaquettes without flux (*37*) as an exception from the otherwise always present flux of $\Phi = \pi$.

Propagation-invariant coincidence suppression

We fabricated BDCs (Fig. 3A) by means of femtosecond laser direct writing in fused silica (*38*, *39*). Each waveguide exhibits birefringence (*40*) with principal axes aligned to the horizontal and vertical edges of the sample. Thus, the horizontally and vertically polarized incident photons maintain their polarization as the waveguides converge from an initial separation of 127 μm (effectively zero coupling) to the eventual separation of 29.5 μm within the interaction region that stretches until the end of the sample [see materials and methods for details on the implementation (*36*)]. Simultaneously, the principal axes of the left and right waveguide are rotated in opposite directions by about $\pm(45.3\pm0.7)°$ by inscribing heavily detuned defect lines with slower inscription speed in close proximity to the original waveguides. Owing to the strong detuning, coupling of photons to these defects is negligible, allowing their stress fields to be utilized to modify the birefringent properties of the waveguides as a function of their relative position (*41*). The perpendicular principal axis alignment among the two waveguides ($\alpha = 90°$) yields a vanishing detuning $\mathcal{H}_{\text{det}} \propto \cos\alpha$ and subjects the photons solely to the artificial magnetic flux $\Phi = \pi$. Moreover, issues of polarization-dependent coupling rates (*42*) vanish because coupling exclusively occurs among fast and slow axes of different waveguides [for details, see the supplementary text (*36*)]. To realize a square plaquette, we match the evanescent coupling strength $\kappa = (0.043 \pm 0.002)$ mm^{-1} to the birefringence $\Delta\beta = (0.0398 \pm 0.0008)$ mm^{-1} [see materials and methods for details (*36*)]. After the propagation through the waveguide circuit, coincidences between horizontally and vertically polarized photons in the respective right and left waveguides are detected. Figure 3B displays an exemplary HOM dip (*8*) in the BDC at $z = 27$ mm that was recorded by varying the arrival time of one photon with respect to the other one: Indistinguishable photons, which arrive with $\tau = 0$ time delay at the BDC (marked with the red line in Fig. 3B), show suppressed coincidence C_{ind} compared to coincidence rates of distinguishable photons C_{dis}, which arrive at the BDC with nonzero delays, e.g., $|\tau| = 1$ ps (marked with a black line as threshold for distinguishable photons). Figure 3C shows the coincidence rates measured in 12 BDCs with propagation lengths ranging from 6 to 39 mm. The coincidence rates of indistinguishable photons are found to be suppressed independently of the propagation distance, even though the coincidence rates of distinguishable photons oscillate. As a result, we were able to consistently observe a near-perfect HOM visibility $v = \frac{C_{\text{ind}}}{C_{\text{dis}}} - 1$ of -0.969 ± 0.008 (Fig. 3D), confirming the predicted propagation-invariant suppression of coincidences. Moreover, the indistinguishable photons that are suppressed in the output $|0110\rangle$ are found to be bunched in the state $|0200\rangle$ or $|0020\rangle$, as illustrated by the peaks ($v \approx +1$) in the measured HOM traces (Fig. 3, E and F) for these output states.

Topological robustness

We then analyzed the error resilience of the HOM interference in the BDC by comparing the two-photon suppression to the HOM interference in a conventional directional coupler under the influence of similar perturbations. We take into account potential errors in input state preparation as well as structural imperfections affecting the ratio between evanescent coupling and birefringence $\kappa/\Delta\beta$. An imperfect input state may be prepared in either the polarization degree of freedom or among the waveguides. Without loss of generality, we restrict the imperfections to the two photons' initial polarization states $\vec{\psi}_{\text{pol},1} = (1,1,0,0)^T$ and $\vec{\psi}_{\text{pol},2} = (1,-1,0,0)^T$ [represented as normalized Stokes components $(S_0 = 1, S_1, S_2, S_3)^T$] that correspond to the photons' initial horizontal ($S_1 = 1$) and vertical ($S_1 = -1$) polarization, respectively. Thus, perturbations of strength ΔS_1 yield $S_1 = \pm(1 - \Delta S_1)$, and the S_2 and S_3 components acquire random, nonzero values (Fig. 4A) [see materials and methods for details on the experiment (*36*)]. Therefore, we measure

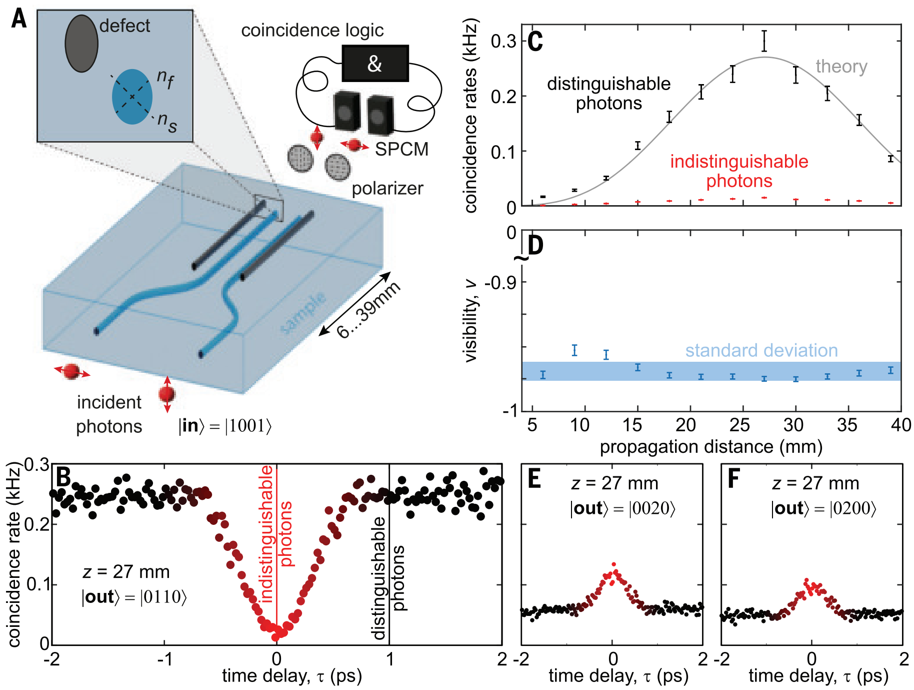


Fig. 3. Propagation-invariant suppression. (**A**) A horizontally and a vertically polarized photon are launched in the two waveguides of a BDC and subsequently detected in coincidence between the two output waveguides with single-photon counting modules (SPCMs) if the polarization of each photon flipped with respect to the input (the filtering is realized by polarizers). Different realizations of the BDC with lengths of 6 to 39 mm are embedded in a 50-mm-long sample. In their active region, the BDCs comprise two coupled waveguides and, additionally, heavily detuned defect traces serving as stress rods to set the waveguides' slow axes perpendicular to one another. (**B**) The predicted HOM dip that is the two-photon coincidence rate as a function of photons' delay in arrival time τ, is measured in a BDC at a propagation distance $z = 27$ mm. Indistinguishable photons are available in the center of the dip (shaded red), whereas the photons become indistinguishable outside of the dip, for instance, at a time delay $\tau = 1$ ps (black line). (**C**) The coincidence rates of distinguishable (black) and indistinguishable photons (red) as a function of the propagation distance give rise to (**D**), propagation-invariant suppression indicated by quantum interference visibilities (HOM visibility) $v = C_{\text{ind}}/C_{\text{dis}} - 1$ ($v = -1$ for suppression with $C_{\text{dis(ind)}}$ coincidence rates of (in)distinguishable photons) of $\bar{v} = -0.969$ and standard deviation $\sigma = 0.008$ (blue shaded area). (**E** and **F**) HOM traces for the output states $|0200\rangle$ and $|0020\rangle$. The coincidences are recorded by splitting the corresponding output after the polarizer with a balanced beam splitter and recording the coincidences among the beam splitters' outputs.

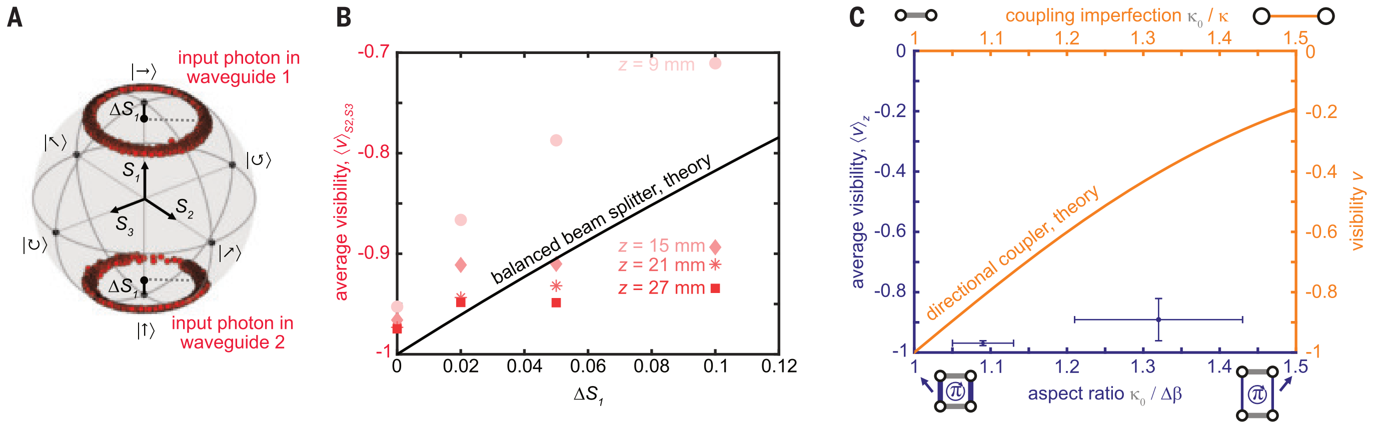


Fig. 4. Error resilience of two-photon suppression. (**A**) A representative measurement of perturbed polarization input state on both photons by ΔS_1 in the $S_1 = \pm(1 - \Delta S_1)$ Stokes component shows a randomization in the S_2 and S_3 components. Each red dot represents the measured polarization state within a 20-ms time interval. (**B**) Launching these perturbed input states in the birefringent directional coupler of propagation distance $z = 9$, 15, 21, and 27 mm yields average HOM visibilities $\langle v\rangle_{S2,S3}$ [assumed to be measured in independent combinations for both photons in (A)]. A balanced beam splitter (black) that is subjected to the same perturbations serves as a theoretical benchmark. (**C**) The measured visibility $\langle v_z\rangle$ (averaged over all available propagation distances z, blue) changes slightly as a function of the aspect ratio $\kappa_0/\Delta\beta$ of the plaquette. Here, a conventional directional coupler with coupling rate κ, which governs suppression when $\kappa = \kappa_0$, serves as a theoretical benchmark for the change in HOM visibility v.

the HOM visibility without knowledge about the actual initial polarization state, which gives rise to averaged visibilities $\langle v\rangle_{S2,S3}$ over all possible $S_2 - S_3$ combinations for sufficiently long measurement times. Figure 4B compares the measured $\langle v\rangle_{S2,S3}$ for perturbations $\Delta S_1 \leq 0.1$ and BDCs at four propagation distances to a balanced beam splitter that is subjected to similar perturbations (details provided in the supplementary text). BDCs with very short propagation distances (i.e., $z = 9$ mm in our model system) are less capable of providing highly visible quantum interference when dealing with perturbed input states than balanced beam splitters, because the parasitic parts of the input state lead directly to coincidences and imperfect suppression due to the limited correction possibilities in short BDCs. In turn, removing the parasitic parts of the input state (the photons are allowed to initially occupy sites 2 and 3) to the unobserved sites 1 and 4 of the square plaquette by means of single-photon AB dynamics [details provided in the supplementary text (*36*)] reestablishes perfect suppression for specific propagation distances. Consequently, BDCs with longer propagation distances (i.e., $z = 15$, 21, and 27 mm) achieve quantum interference of in general higher visibility than the balanced beam splitter. The optimal performance of the BDC occurs at $z = 27$ mm (at the maximum of suppressed photon pairs in Fig. 3C), for which the average visibility becomes almost independent on the perturbation strength.

To test the latter issue, a directional coupler with $L = \frac{\pi}{4\kappa_0}$ serves as a reference. With precisely matched coupling rate ($\kappa = \kappa_0$), such a component exhibits perfect suppression, whereas any deviations from this design point cause a rapid deterioration of the visibility of the HOM dip, with v exceeding the threshold of $v = -0.5$ for relative deviations of κ_0/κ on the order of 1.3. By contrast, even when we intentionally modify a BDC by adjusting the birefringence to $\Delta\beta = (0.0329 \pm 0.012)\ \mathrm{mm}^{-1}$, resulting in a heavily unbalanced configuration with $\frac{\kappa}{\Delta\beta} = (1.32 \pm 0.11)$, the mean visibility over the different propagation distances is only marginally affected [$v_z = (-0.89 \pm 0.07)$, Fig. 4C], again highlighting the error resilience of topological q3uantum interference. Likewise, the fact that v remains independent of the propagation distance z is grounded in the topological character of the system of the two-photon interference [details are provided in the supplementary text (*36*)].

Concluding remarks

Our findings pave the way for harnessing topological features in the next generation of quantum photonic architectures with superior stability and resilience against imperfections. We anticipate that photonic devices using photonic degrees of freedom (*43*) to introduce topological protection in many-particle dynamics will find use cases in a broad range of new quantum technological applications alongside non-Abelian geometric phases (*44*), nonconservative systems (*45*, *46*), and many-photon quantum states (*47*) with interactions (*48*, *49*).

REFERENCES AND NOTES

1. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljacić, *Nature* **461**, 772–775 (2009).
2. M. C. Rechtsman *et al.*, *Nature* **496**, 196–200 (2013).
3. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, *Nat. Photonics* **7**, 1001–1005 (2013).
4. Z. Yang *et al.*, *Phys. Rev. Lett.* **114**, 114301 (2015).
5. M. Aidelsburger *et al.*, *Nat. Phys.* **11**, 162–166 (2015).
6. G. Jotzu *et al.*, *Nature* **515**, 237–240 (2014).
7. R. Süsstrunk, S. D. Huber, *Science* **349**, 47–50 (2015).
8. C. K. Hong, Z. Y. Ou, L. Mandel, *Phys. Rev. Lett.* **59**, 2044–2046 (1987).
9. S. J. Freedman, J. F. Clauser, *Phys. Rev. Lett.* **28**, 938–941 (1972).
10. A. Aspect, J. Dalibard, G. Roger, *Phys. Rev. Lett.* **49**, 1804–1807 (1982).
11. L. S. Madsen *et al.*, *Nature* **606**, 75–81 (2022).
12. J. M. Arrazola *et al.*, *Nature* **591**, 54–60 (2021).
13. T. Rudolph, *APL Photonics* **2**, 030901 (2017).
14. T. Kitagawa *et al.*, *Nat. Commun.* **3**, 882 (2012).
15. S. Mittal, V. V. Orre, M. Hafezi, *Opt. Express* **24**, 15631–15641 (2016).
16. M. C. Rechtsman *et al.*, *Optica* **3**, 925–930 (2016).
17. A. Blanco-Redondo, B. Bell, D. Oren, B. J. Eggleton, M. Segev, *Science* **362**, 568–571 (2018).
18. M. Wang *et al.*, *Nanophotonics* **8**, 1327–1335 (2019).
19. S. Barik *et al.*, *Science* **359**, 666–668 (2018).
20. S. Mittal, E. A. Goldschmidt, M. Hafezi, *Nature* **561**, 502–506 (2018).
21. S. Mittal, V. V. Orre, E. A. Goldschmidt, M. Hafezi, *Nat. Photonics* **15**, 542–548 (2021).
22. C. Dittel *et al.*, *Phys. Rev. Lett.* **120**, 240404 (2018).
23. M. C. Tichy, M. Tiersch, F. de Melo, F. Mintert, A. Buchleitner, *Phys. Rev. Lett.* **104**, 220405 (2010).
24. Y. Aharonov, D. Bohm, *Phys. Rev.* **115**, 485–491 (1959).
25. J. Brendel, W. Dultz, W. Martienssen, *Phys. Rev. A* **52**, 2551–2556 (1995).
26. P. Mehta, J. Samuel, S. Sinha, *Phys. Rev. A* **82**, 034102 (2010).
27. L. Lu, J. D. Joannopoulos, M. Soljačić, *Nat. Photonics* **8**, 821–829 (2014).
28. D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, *Phys. Rev. Lett.* **49**, 405–408 (1982).
29. J. Janszky, C. Sibilia, M. Bertolotti, Y. Yushin, *J. Mod. Opt.* **35**, 1757–1765 (1988).
30. W. K. Lai, V. Buek, P. L. Knight, *Phys. Rev. A* **43**, 6323–6336 (1991).
31. K. Fang, Z. Yu, S. Fan, *Nat. Photonics* **6**, 782–787 (2012).
32. S. Mukherjee, M. Di Liberto, P. Öhberg, R. R. Thomson, N. Goldman, *Phys. Rev. Lett.* **121**, 075502 (2018).
33. M. Kremer *et al.*, *Nat. Commun.* **11**, 907 (2020).
34. J. M. Leinaas, J. Myrheim, *Nuovo Cimento B* **37**, 1–23 (1977).
35. D. N. Klyshko, *Phys. Lett. A* **140**, 19–24 (1989).
36. See supplementary materials.
37. M. Ehrhardt *et al.*, *Sci. Adv.* **7**, eabc5266 (2021).
38. K. M. Davis, K. Miura, N. Sugimoto, K. Hirao, *Opt. Lett.* **21**, 1729–1731 (1996).
39. A. Szameit, S. Nolte, *J. Phys. At. Mol. Opt. Phys.* **43**, 163001 (2010).
40. P. Yang, G. R. Burns, J. Guo, T. S. Luk, G. A. Vawter, *J. Appl. Phys.* **95**, 5280–5283 (2004).

41. R. Heilmann, M. Gräfe, S. Nolte, A. Szameit, *Sci. Rep.* **4**, 4118 (2014).
42. S. Rojas-Rojas *et al.*, *Phys. Rev. A* **90**, 063823 (2014).
43. M. Ehrhardt, S. Weidemann, L. J. Maczewsky, M. Heinrich, A. Szameit, *Laser Photonics Rev.* **17**, 2200518 (2023).
44. V. Neef *et al.*, *Nat. Phys.* **19**, 30–34 (2023).
45. M. Ehrhardt, M. Heinrich, A. Szameit, *Nat. Photonics* **16**, 191–195 (2022).
46. K. Wang, A. Dutt, C. C. Wojcik, S. Fan, *Nature* **598**, 59–64 (2021).
47. C. Dittel, G. Dufour, G. Weihs, A. Buchleitner, *Phys. Rev. X* **11**, 031041 (2021).
48. M. Jürgensen, S. Mukherjee, M. C. Rechtsman, *Nature* **596**, 63–67 (2021).
49. M. Jürgensen, S. Mukherjee, C. Jörg, M. C. Rechtsman, *Nat. Phys.* **19**, 420–426 (2023).
50. M. Ehrhardt, A. Szameit, Dataset for Manuscript "Topological Hong-Ou-Mandel interference," Rostock University Publication Server (2024); https://doi.org/10.18453/rosdok_id00004569.

ACKNOWLEDGMENTS

We thank C. Otto for preparing the high-quality fused silica samples used for the inscription of all photonic structures used in this work. We thank R. Keil and T. A. W. Wolterink for fruitful discussions. **Funding:** A.S. acknowledges funding from the Deutsche Forschungsgemeinschaft (grants SZ 276/9-2, SZ 276/19-1, SZ 276/20-1, SZ 276/21-1, SZ 276/27-1, and GRK 2676/1-2023, "Imaging of Quantum Systems," project no. 437567992). A.S. also acknowledges funding from the Krupp von Bohlen and Halbach Foundation as well as from the FET Open Grant EPIQUS (grant no. 899368) within the framework of the European H2020 program for Excellent Science. A.S. and M.H. acknowledge funding from the Deutsche Forschungsgemeinschaft via SFB 1477, "Light–Matter Interactions at Interfaces" (project no. 441234705). C.D. acknowledges the Georg H. Endress Foundation for support. **Author contributions:** M.E. devised the conceptual idea, designed and fabricated the sample, and carried out the experiments. M.E. and C.D. developed the theoretical framework. M.H. contributed analysis tools. A.S. supervised the project. All authors discussed the results and cowrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All experimental data and any related experimental background information not mentioned in the text can be found at the Rostock University Publication Server repository (*50*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ado8192
Materials and Methods
Supplementary Text
Figs. S1 to S4
References (*51*, *52*)

Submitted 24 February 2024; accepted 3 May 2024
10.1126/science.ado8192

FLUID DYNAMICS

Selective directional liquid transport on shoot surfaces of *Crassula muscosa*

Ling Yang[1]†, Wei Li[1]†, Jiaoyuan Lian[2], Hengjia Zhu[1], Qiyu Deng[1], Yiyuan Zhang[1], Jiaqian Li[1,3]*, Xiaobo Yin[1,4]*, Liqiu Wang[2,5]*

Directional liquid transport has been widely observed in various species including cacti, spiders, lizards, the pitcher plant *Nepenthes alata*, and *Araucaria* leaves. However, in all these examples the liquid transport for a specific liquid is completely restricted in a fixed direction. We demonstrate that *Crassula muscosa* shoot surfaces have the ability to transport a specific liquid unidirectionally in either direction. This is accomplished through the presence of asymmetric reentrant leaves with varying reentrant angles, which yields the variation in liquid meniscus heterogeneity. These findings enable engineered biomimetic structures capable of selective directional liquid transport, with functions such as intelligent flow direction switching, liquid distribution, and mixing.

Directional liquid transport plays a crucial role in numerous applications including microfluidics (*1*, *2*), water harvesting (*3*, *4*), oil-water separation (*5*, *6*), solar desalination (*7*, *8*), and heat transfer (*9*, *10*). This property has been observed in various species such as cacti, spiders, lizards, the pitcher plant *Nepenthes alata*, and *Araucaria* leaves. However, surface liquid transport in all these species proceeds in the same manner, with a specific liquid moving in a fixed direction. For example, cacti transport harvested fog from their spine tips to their roots (*11*); spider silk transports captured fog or dew from periodic spindle-knots to joints (*12*); lizards transport water from the air into their snout through interconnected capillary channels (*13*); *Nepenthes alata* delivers liquids or nectar directionally from the inner to outer edge of its peristome using multiscale structures (*14*); and *Araucaria* leaves transport a specific liquid in a fixed direction by utilizing a capillary ratchet effect (*15*). Are there alternative modes of liquid transport in nature? The answer could further inspiration for the designs of directional liquid transport.

We show a different liquid transport mode found on shoot surfaces of the succulent plant *Crassula muscosa*, indigenous to South Africa and Namibia. In contrast to reported species, we observe that liquids deposited on the surfaces of *C. muscosa* shoots exhibit a spontaneous and unidirectional motion in opposite directions on different horizontal shoots. This selective directional liquid transport is attributed to the distinctive asymmetric reentrant structure of the *C. muscosa* leaf.

[1]Department of Mechanical Engineering, The University of Hong Kong, Hong Kong SAR 999077, P. R. China. [2]Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR 999077, P. R. China. [3]School of Energy and Power Engineering, Shandong University, Jinan 250061, P. R. China. [4]Department of Physics, The University of Hong Kong, Hong Kong SAR 999077, P. R. China. [5]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR 999077, P. R. China.
***Corresponding author. Email: liqiu.wang@polyu.edu.hk (L.W.); xbyin@hku.hk (X.Y.); jiaqianli@sdu.edu.cn (Jiaq.L.)**
†These authors contributed equally to this work.

Results

Crassula muscosa

Crassula muscosa is a succulent plant with centimeter-long shoots, characterized by thin stems decussately packed with asymmetric reentrant leaves (Fig. 1A and fig. S1). These leaves exhibit an upper-tilting reentrant angle (ω_1) toward the stem tip and a lower-tilting reentrant angle (ω_2) toward the root (Fig. 1A). The values of these two reentrant angles, ranging from ~20° to ~90° for ω_1 and ~20° to ~70° for ω_2, vary across different shoots (figs. S2 and S3). The leaves and stems are protected by epidermal cells covered with films of epicuticular waxes (figs. S4 to S6), which have similar chemical compositions (figs. S7 to S12). These cells also vary in size and shape (figs. S13 and S14), resulting in distinctly different wettability on various parts of the shoots (fig. S15). *C. muscosa* shoots can be easily wetted by water fog (*16*) and collect water between the leaves and stems (fig. S16). This allows the plant to obtain water through direct foliar uptake (*17*, *18*) or root absorption of water that drips from the shoots to the soil (*19*) (fig. S16), enabling its survival in arid and foggy regions of South Africa and Namibia (*20*).

Selective directional liquid transport

To investigate the spreading behavior of liquids on *C. muscosa* surfaces we infused ethanol—which wets the leaves better than water—on two horizontally placed shoots at a flow rate of 2.8 μL/s. The average upper-tilting reentrant angle, ω_1, are 32° ± 8.5° and 55° ± 5.2°, respectively, on shoot I (Fig. 1B) and shoot II (Fig. 1C). The average lower-tilting reentrant angle, ω_2, are 37° ± 6.6° and 34°± 4.8°, respectively, on shoot I and shoot II. Interestingly, the deposited ethanol exhibited distinct transport characteristics on the two shoots. On shoot I, ethanol propagated unidirectionally toward the tip whereas it was effectively blocked in the opposite direction (Fig. 1B and movie S1A). Conversely, on shoot II ethanol preferentially spread toward the root and remained pinned in all other directions (Fig. 1C and movie S1B). Figure 1D shows the evolution of liquid transport length (l) between the flow front and the base of the leaf where the liquid was deposited. We selected ethanol to better illustrate this selective directional transport phenomenon; this is also observed when water is used on shoots with surfaces wetted with a thin water layer (fig. S17). It is worth noting

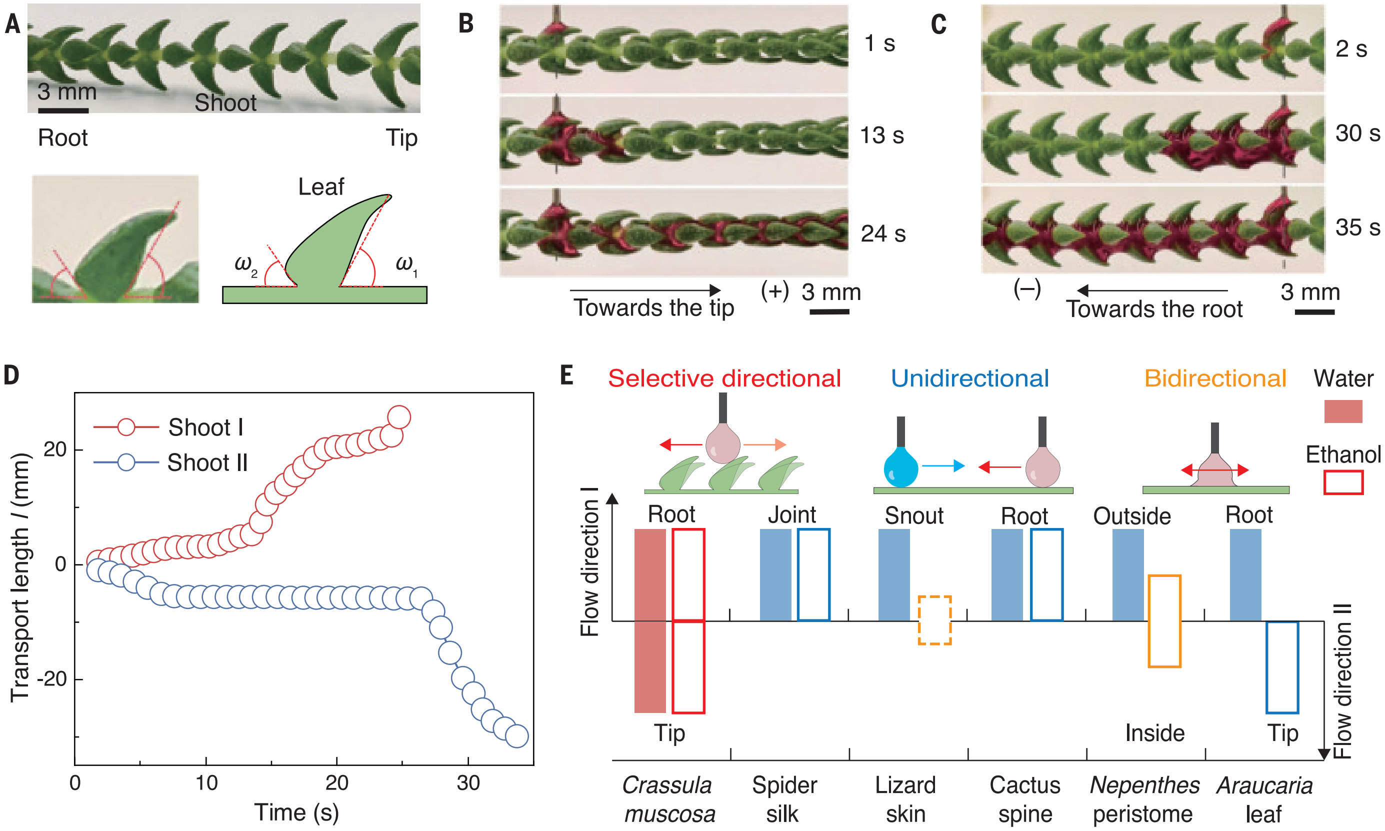


Fig. 1. Selective directional liquid transport on shoot surfaces of *C. muscosa*. (**A**) Optical images of one *C. muscosa* shoot and structural characteristics of its leaves. Each leaf has an upper-tilting reentrant angle ω_1 and a lower-tilting reentrant angle ω_2. (**B**) Optical images of shoot I transporting ethanol (dyed red) unidirectionally toward its tip (positive direction) while pinning in the reverse direction. (**C**) Optical images of shoot II transporting ethanol unidirectionally toward its root (negative direction) while pinning in the tip direction. (**D**) Plot of liquid transport length of ethanol versus time for selective directional liquid transport on shoots I and II. (**E**) Comparison of different transport modes of water and ethanol in different species. Water and ethanol are indicated by solid and hollow bars, respectively. Selective directional, unidirectional, and bidirectional liquid transport are indicated by red, blue, and orange colors, respectively. The dashed orange hollow bar represents ethanol is transported in all directions on lizard skin.

that this selective directional liquid transport phenomenon is not observed in previous literature (*11–15*), in which specific liquids such as water or ethanol are transported in fixed directions (Fig. 1E and fig. S18).

It is the asymmetry of reentrant leaves that distorts the liquid meniscus in two directions, thereby influencing the direction of liquid transport. To confirm this we designed two *C. muscosa*-inspired arrays (CMIAs) using three-dimensional (3D) printing techniques. These arrayed fin structures, labeled CMIA I and CMIA II, mimic the reentrant leaves with different tilting reentrant angles (Fig. 2, A and B). We use the term "fins" for these artificially made structures. CMIA I has reentrant angles ω_1 = 70° and ω_2 = 40° (Fig. 2D) and CMIA II has reentrant angles ω_1 = 51° and ω_2 = 57° (Fig. 2E). Both CMIAs share the same upper height *H* (the height from the upper vertex to the base), lower height *h* (the height from the lower vertex to the base), and bottom-to-bottom spacing *d* (Fig. 2, A and B). Ethanol has an intrinsic contact angle of 11.5° and an advancing contact angle of 21.5° on both CMIAs, reflecting a similar wetting behavior to that of ethanol on the shoot surfaces of *C. muscosa* (figs. S19 and S20). For simplicity, we defined the upper-tilting direction as the positive direction (+) and the lower-tilting direction as the negative direction (–) (Fig. 2A).

As expected, selective directional liquid transport occurs on the bioinspired CMIAs (Fig. 2, C to E). On CMIA I ethanol is transported unidirectionally toward the negative direction while pinning in the positive direction (Fig. 2D, fig. S21, and movie S2). Conversely, on CMIA II ethanol propagates unidirectionally toward the positive direction while being stably thwarted in the reverse direction (Fig. 2E, fig. S22, and movie S3). This selective liquid flow is demonstrated by the evolution of the normalized liquid transport length l/d (Fig. 2C). The normalized lengths l/d periodically increase in their preferential directions (negative direction on CMIA I and positive direction on CMIA II). Moreover, the l/d ratio increases significantly faster during crossing periods (for example, 2.67 s to 2.75 s for CMIA I and 3.13s to 3.17s for CMIA II) than during spreading periods.

Upon depositing the liquid on the first fin of the two artificial CMIAs, the liquid meniscus gradually forms and subsequently propagates heterogeneously in both the positive and negative directions as a result of the reentrant angles in both directions. This heterogeneity is further enhanced when the liquid meniscus is partially pinned by the lower fin edge in the negative direction (1.06 s in Fig. 2D and 0.16 s in Fig. 2E). When the front edge of this heterogeneous liquid meniscus reaches the adjacent fin, the liquid quickly crosses that fin as a result of the prominent capillary rise at the corners (*14*, *21*, *22*). In all other directions, the liquid is thwarted. The velocity of liquid transport across the adjacent fin becomes nearly 20 times faster than the averaged velocity propagating across the spacing between two fins (fig. S23). Subsequently, a new liquid meniscus forms and spreads toward the next adjacent fin in the same direction. In this way, the liquid is

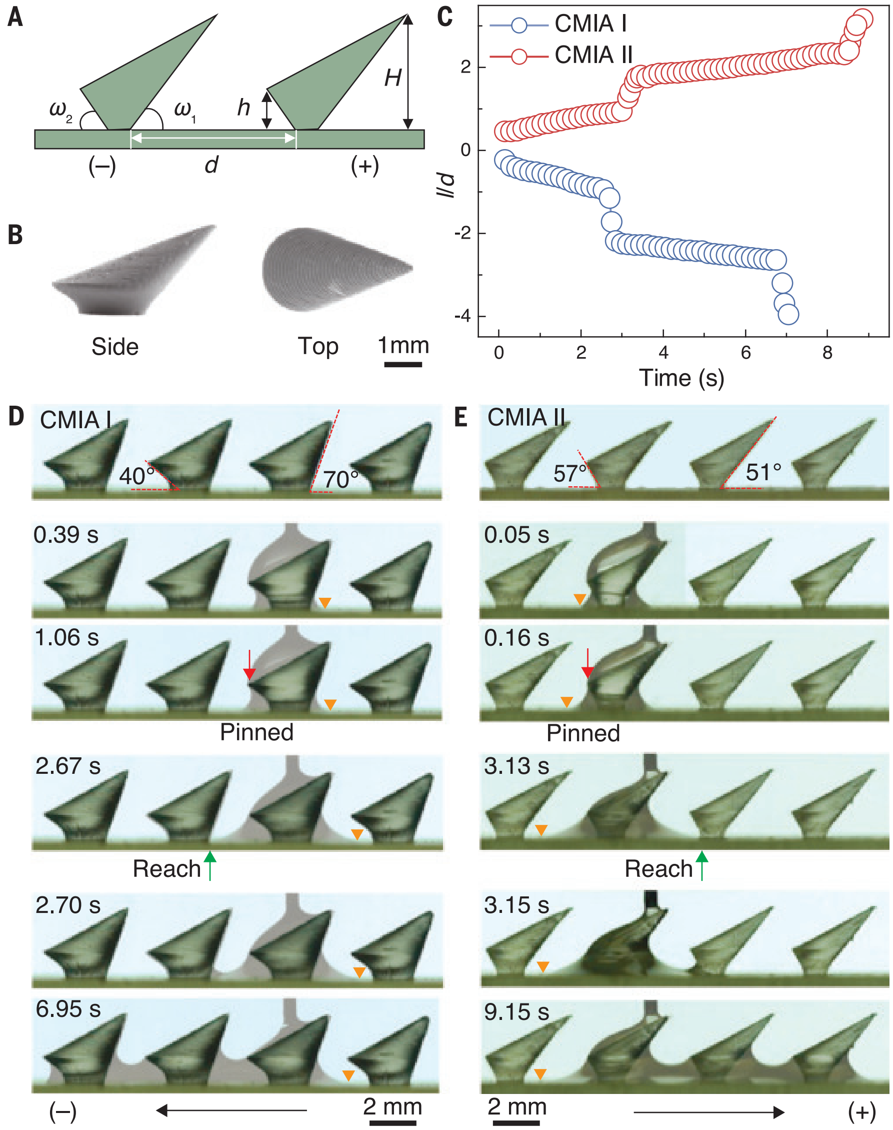


Fig. 2. Selective directional liquid transport on artificial CMIAs. (**A**) Schematic of the CMIA. Structural characteristics consists of the upper height (*H*), lower height (*h*), bottom to bottom spacing (*d*), upper-tilting reentrant angle (ω_1) and lower-tilting reentrant angle (ω_2). (**B**) Scanning electron microscope images of a 3D printed fin on the CMIA with side view and top view. (**C**) Normalized lengths (*l*/*d*) versus time on CMIA I with ω_1 = 70°, ω_2 = 40° (blue symbols) and CMIA II with ω_1 = 51°, ω_2 = 57° (red symbols). (**D** and **E**) Timelapse images of selective directional liquid transport on two artificial CMIAs. Ethanol is transported in the negative direction and pinned toward the positive side (indicated by orange triangles) on CMIA I, while it spreads in the positive direction and is thwarted in the reverse side (indicated by orange triangles) on CMIA II. This selective directional liquid transport phenomenon manifests through the heterogeneous liquid meniscus and is separated by two sites in which the liquid meniscus is pinned by the lower fin edge (red arrows) and the front edge of the liquid meniscus reaches the adjacent fin (green arrows).

periodically transported in its preferential direction (*23*) whereas the pinning sites facilitate spontaneous and continuous liquid transport.

Heterogeneous meniscus model

A heterogeneous meniscus model is developed to illustrate the selective directional liquid transport (see supplementary text for details). The liquid menisci in the positive and negative directions are noted as m+ and m− and their spreading distances are defined as s_+ and s_-, respectively (Fig. 3A). Assuming a low liquid flow rate and equal Laplace pressures forcing at m+ and m− of a reentrant fin, it can be concluded that m+ and m− have the same radius *R* (*24*). Consequently, the spreading distances of the liquid meniscus m+ and m− are derived as $s_i = R\cos(\omega_i/2+\theta)/\sin(\omega_i/2)$, where i represents + or − and θ represents the advancing contact angle of the liquid on the CMIAs (Fig. 3B). When the meniscus m− is partially pinned by the lower fin edge, the spreading distance s_- reaches a critical value $s_{\text{critical}} = h/\sin\omega_2$. As the meniscus propagates, the spreading distance s_- becomes $s_- = h\cot\omega_2 + h\cot\delta$, where $h/\sin\delta = 2R\sin(\delta-\theta)$ and δ is the intersection angle between the chord of the meniscus m− and the base (Fig. 3C).

To quantify selective directional liquid transport, a theoretical criterion $r_s = s_+/s_-$ is introduced, defined as the ratio of the spreading distance of liquid meniscus in the positive direction (s_+) to that in the negative direction (s_-),

$$r_s = \begin{cases} \dfrac{\cos(\omega_1/2+\theta)\sin(\omega_2/2)}{\cos(\omega_2/2+\theta)\sin(\omega_1/2)}, \text{when } s_- < s_{\text{critical}} \\ \dfrac{\cos(\omega_1/2+\theta)}{h\sin(\omega_1/2)(\cot\omega_2+\cot\delta)}R, \text{when } s_- \geq s_{\text{critical}} \end{cases}$$

The liquid meniscus radius (*R*) is limited to less than the capillary length $l_c = \sqrt{\gamma/\rho g}$, where γ is the liquid surface tension, ρ is the liquid density, and *g* is the gravitational constant (see supplementary text). If the liquid meniscus reaches the adjacent fin before m− is pinned by the lower fin edge, r_s is solely determined by reentrant angles and is independent of *R*. Otherwise, r_s increases approximately linearly with *R*, with the slope primarily influenced by the reentrant angles (Fig. 3D). Here, *R* is specified when the liquid meniscus reaches the adjacent fin, implying that the larger value between s_+ and s_- equals the spacing *d*. r_s, therefore, can be controlled by both reentrant angles and the spacing *d* (fig. S24). The sign of (r_s – 1) determines the direction in which the liquid meniscus contacts the adjacent fin first and, consequently, the direction of liquid transport. For r_s > 1, the liquid transport on the CMIA occurs in the positive direction whereas for r_s < 1 the liquid transport occurs in the negative direction. The predicted behavior is

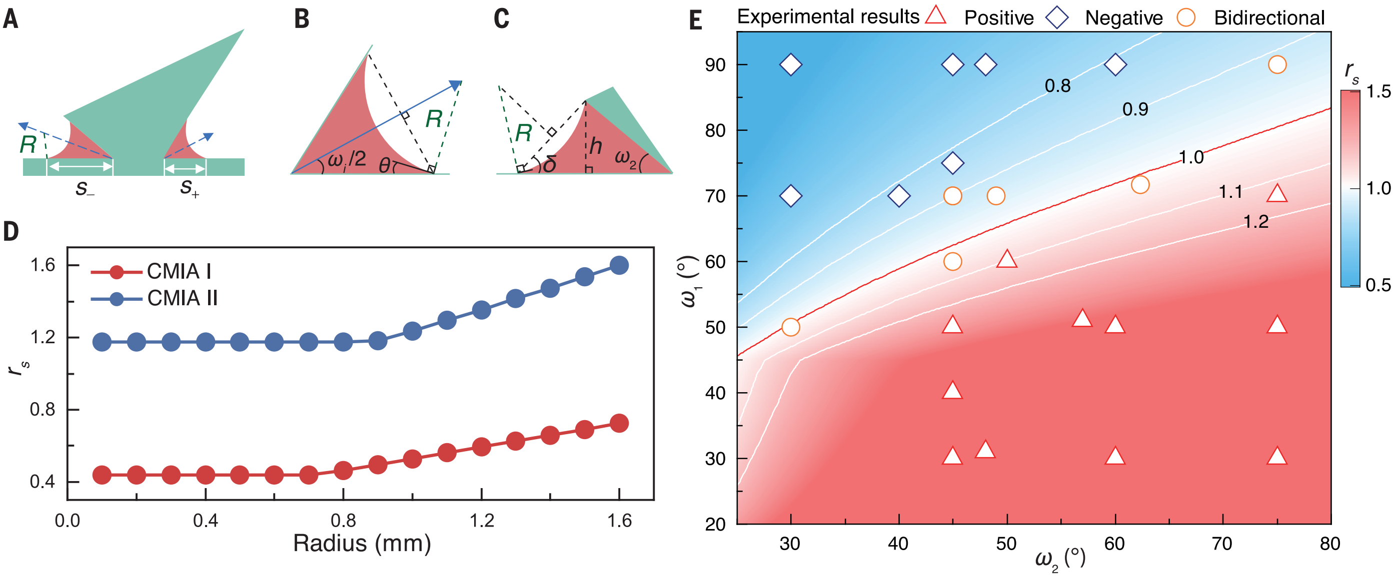
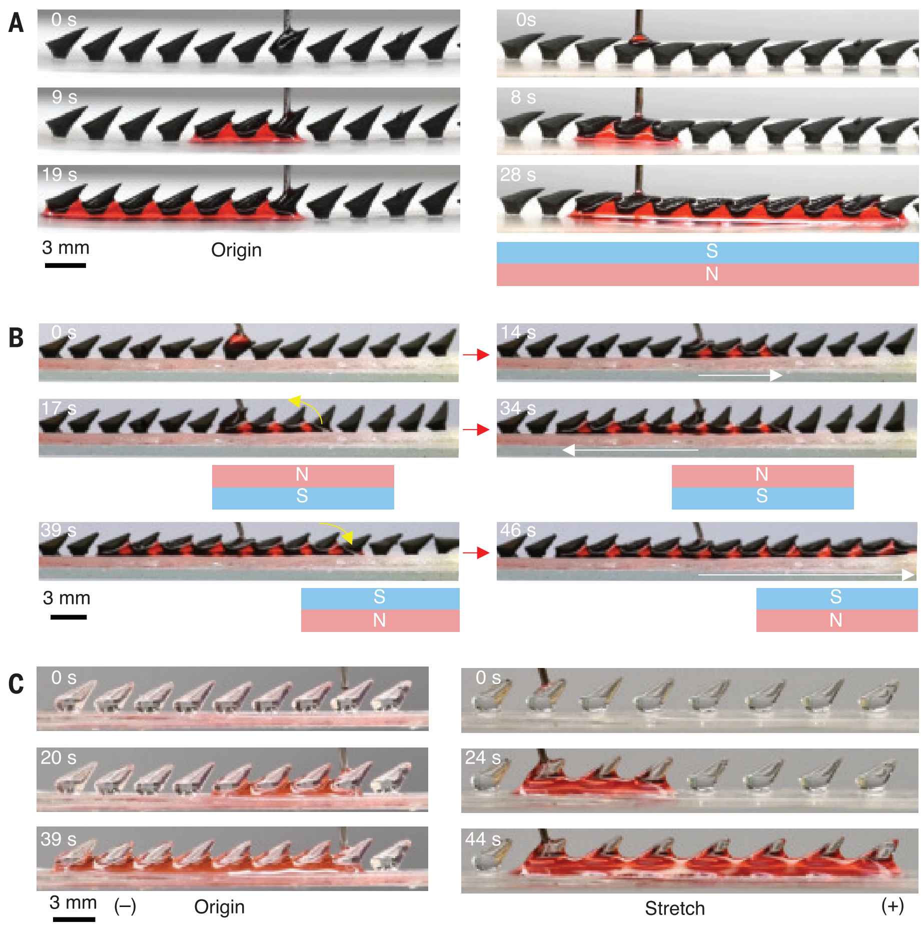


Fig. 3. Theoretical model and experimental results. (**A** to **C**) Schematic of heterogeneous meniscus model. The s_+ and s_- represent the spreading distances of the liquid menisci in the positive and negative directions, respectively. The R represents the radius of the liquid meniscus. The θ and δ represent the advancing contact angle of the liquid on the CMIA and the intersection angle between the chord of the liquid meniscus in the – direction (m–) and the base, respectively. As the m– reaches the lower fin edge, m– is partially pinned, and the profiles of heterogeneous meniscus change from B to C. (**D**) Calculated criterion $r_s = s_+/s_-$ versus R. (**E**) Contour map of the r_s and experimental results on artificial CMIAs with the spacing d of ~3.2 mm. The triangles, diamonds, and circles show experimental results of positive, negative, and bidirectional liquid transport, respectively. Bidirectional liquid transport occurs when $0.9 \le r_s \le 1.1$ as the liquid meniscus gradually propagates and eventually unpins in pinning directions as a result of fabrication defects.

Fig. 4. Smart control of directional liquid transport with external stimuli on magnetic CMIA and elastic CMIA. (**A**) Remote control of directional liquid transport by magnetic stimuli. Ethanol (dyed red) is transported in the negative direction on the CMIA without applying a magnet (left), and in the positive direction (right) resulted from decreasing ω_1 and increasing ω_2 by magnetic stimuli. The black shapes are fins made of Ecoflex doped with magnetic NdFeB particles. (**B**) Real-time and in situ control of directional liquid transport by magnets. The transporting route of liquid is adjusted by strategically operating the magnet to tune reentrant angles of fins. The yellow arrows illustrate the deformation of fins, white arrows indicate the liquid transport direction in time, and red arrows indicate images of the next time. (**C**) Control of directional liquid transport by mechanical stimuli. On the elastic CMIA, ethanol is transported in the negative direction (left) whereas it is converted to the positive direction by the mechanical stretching-induced spacing rise (right).



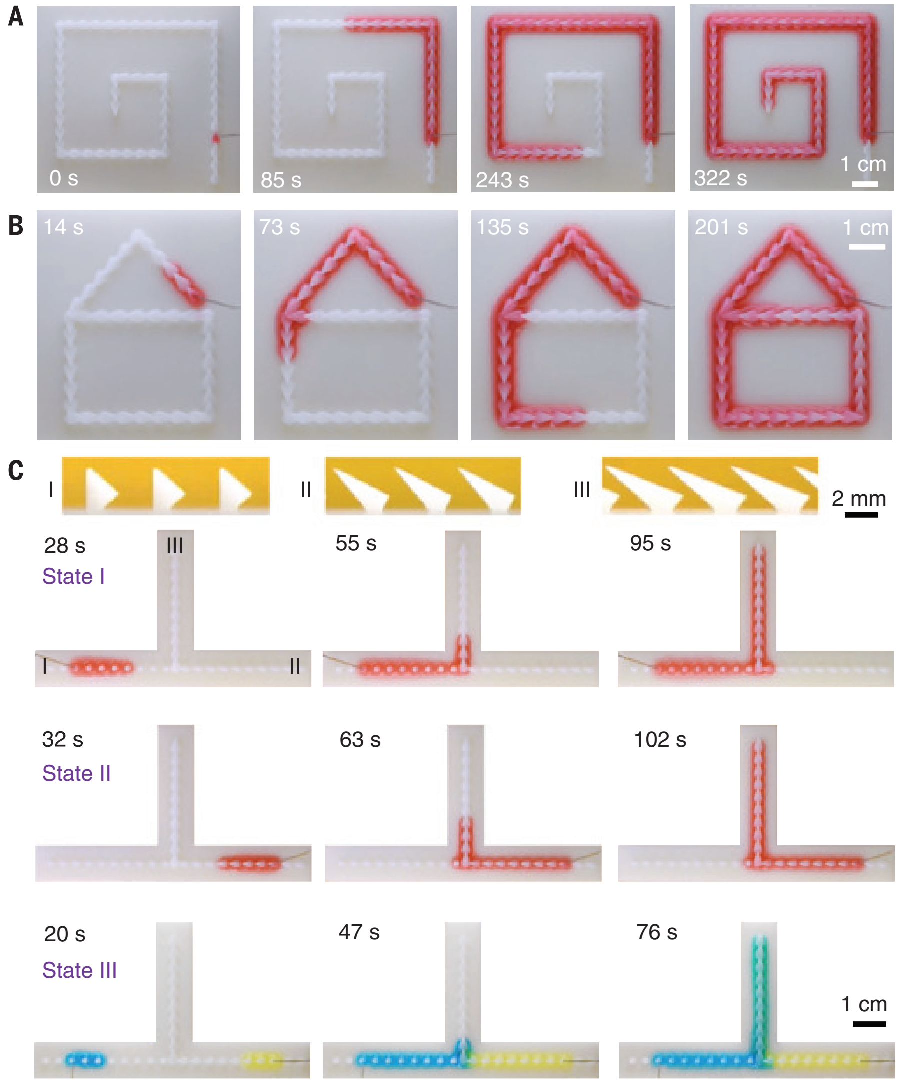


Fig. 5. Applications of directional liquid transport on rigid CMIA. (**A** and **B**) Ethanol is transported continuously and directionally on CMIA along complex paths: (A) a zigzag pattern and (B) a cartoon house pattern. (**C**) CMIA serves as a functional three-way valve. The first row of images shows CMIA structures of inlet I, inlet II, and outlet III. In states I and II, the liquid (ethanol dyed red) deposited on either inlet I or inlet II can be transported to outlet III, respectively. In state III, the liquid (ethanol dyed blue and yellow, respectively) is deposited on both inlets I and II and can flow along designed paths, then be mixed to a green color and flow out from outlet III. The liquid depositing flow rate is 2.2 μL/s.

confirmed by our observations, showing that CMIA I transports the liquid unidirectionally in the negative direction with $r_s < 1$ whereas CMIA II transports the liquid in the positive direction with $r_s > 1$. Therefore, reentrant angles of asymmetric reentrant fins generate heterogeneous meniscus, which directly determines the selective directional liquid transport.

We further investigated how asymmetric reentrant angles regulate liquid transport directions (Fig. 3E and fig. S25). Figure 3E shows the contour map of the calculated criterion r_s and the experimentally observed liquid transport directions on CMIAs with the same spacing but various different asymmetric reentrant angles ω_1 and ω_2. CMIAs with large ω_2 and small ω_1 enable the positive liquid transport (Fig. 3E, triangles, $r_s > 1.1$), while CMIAs with large ω_1 and small ω_2 facilitate the negative liquid transport (Fig. 3E, diamonds, $r_s < 0.9$). Within a narrow region with r_s ~1, bidirectional liquid transport happens as the liquid meniscus gradually propagates and eventually unpins in pinning directions (Fig. 3E, circles, $0.9 \le r_s \le 1.1$). Therefore, by adjusting the reentrant angles the direction of liquid transport can be effectively controlled.

The selective directional liquid transport of ethanol on surfaces of *C. muscosa* shoots can be explained by enhancing the aforementioned model with different advancing contact angles on different shoot organs (figs. S26 and S27). Using this improved model, the calculated criterion r_s for ethanol is found to be 1.19 and 0.55 on shoot I and shoot II, respectively. The internode size of approximately 1.5 mm is smaller than the capillary length. The predicted flow directions determined by the criterion r_s are consistent with the experimentally observed flow behaviors.

Direction-reconfigurable liquid transport

Taking advantage of the selective flow property of *C. muscosa*, smart and reconfigurable liquid transport is achieved through the use of external stimuli. For instance, magnetic neodymium-iron-boron (NdFeB) fillers (fig. S28) (*25*) are introduced to change the tilting reentrant angles and manipulate the tetherless liquid transport on magnetic CMIAs (movie S4). In the absence of a magnetic field, the CMIA transports liquid in the negative direction (Fig. 4A, left) but switches to the positive direction when a magnetic field is applied (Fig. 4A, right). Real-time and in situ control of the magnetic CMIA allows for reprogrammable, pixelated liquid transport (Fig. 4B and movie S5). Diligent manipulation of the magnetic field at specific fin sites enables selective liquid transport directions on demand. Our technique offers continuous bulk liquid transport along dynamically reconfigurable pathways, distinguishing it from approaches involving the addition of magnetic nanoparticles into liquid or using magnetic robot (*26*), tubular (*27*) or microplate arrays (*28*). Additionally, changing the spacing *d* can also switch liquid transport direction, allowing for reconfigurable liquid transport through other techniques like mechanical stimuli (Fig. 4C, fig. S29, and movie S6). Compared to acoustic (*29*), photopyroelectric (*30*) or photothermal (*31*) microfluidics technologies, our magnetic and mechanical reprogrammed liquid transport eliminates the heating induced by ultrasonics and photo, which is particularly advantageous in biomedical and chemical applications requiring precise control of liquid temperature for various reactions.

Inspiration

By incorporating fin arrangements on CMIAs, functional patterned liquid transport is achieved. The ability to transport liquid along desired paths is crucial in various fields such as microfluidics, chemical synthesis, and biomedical diagnostics. Artificial CMIAs enable spontaneous liquid transport along designed trajectories, including continuous and long-distance transport of ethanol along zigzag and house-like patterns (Fig. 5, A and B, and movie S7). Other liquid operations such as mixing and distributing are also achievable. A "T-shape" CMIA can function similarly to a three-way valve, distributing or mixing liquids from different inlets (Fig. 5C and movie S8). Furthermore, this liquid transport method is applicable to a wide range of liquids with surface tensions less than 30 mN m^{-1}, including ethanol, silicon oil, acetone, hexadecane, and more (fig. S30). It is also effective for high surface tension liquids, such as water on hydrophilically treated CMIAs (fig. S31).

Concluding Remarks

In summary, our work unveils the liquid transport mechanism on the shoot surfaces of *C. muscosa*, where asymmetric reentrant leaves generate heterogeneous liquid meniscus profiles, resulting in selective unidirectional motion of the liquid. This discovery has inspired the development of innovative structures with functional liquid transport capabilities, including reconfigurable liquid transport, intelligent modulation of transport directions, and spontaneous and long-distance directional transport. As a result, real-time and in situ programmable integrated microfluidics can be achieved, facilitating diverse microfluidic applications in biomedical and chemical reactions and analyses.

REFERENCES AND NOTES

1. Y. Gao *et al.*, *Sci. Adv.* **7**, eabg9614 (2021).
2. Q. Sun *et al.*, *Nat. Mater.* **18**, 936–941 (2019).
3. C. Li, C. Yu, S. Zhou, Z. Dong, L. Jiang, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23436–23442 (2020).
4. A. R. Parker, C. R. Lawrence, *Nature* **414**, 33–34 (2001).
5. C. Li, L. Wu, C. Yu, Z. Dong, L. Jiang, *Angew. Chem. Int. Ed.* **56**, 13623–13628 (2017).
6. Y. Li *et al.*, *Adv. Funct. Mater.* **32**, 2201035 (2022).
7. L. Wu *et al.*, *Nat. Commun.* **11**, 521 (2020).
8. Y. Shi, O. Ilic, H. A. Atwater, J. R. Greer, *Nat. Commun.* **12**, 2797 (2021).
9. K.-C. Park *et al.*, *Nature* **531**, 78–82 (2016).
10. N. A. Dudukovic *et al.*, *Nature* **595**, 58–65 (2021).
11. J. Ju *et al.*, *Nat. Commun.* **3**, 1247 (2012).
12. Y. Zheng *et al.*, *Nature* **463**, 640–643 (2010).
13. P. Comanns *et al.*, *J. R. Soc. Interface* **12**, 20150415 (2015).
14. H. Chen *et al.*, *Nature* **532**, 85–89 (2016).
15. S. Feng *et al.*, *Science* **373**, 1344–1348 (2021).
16. T. Mouterde *et al.*, *Nat. Mater.* **16**, 658–663 (2017).
17. H. R. Tolken, "Taxonomic studies on the genus *Crassula* L." thesis, University of Cape Town (1974).
18. C. E. Martin, D. J. von Willert, *Plant Biol. (Stuttg.)* **2**, 229–242 (2008).
19. E. B. Limm, K. A. Simonin, A. G. Bothman, T. E. Dawson, *Oecologia* **161**, 449–459 (2009).
20. R. Spirig, R. Vogt, C. Feigenwinter, *Atmos. Res.* **262**, 105765 (2021).
21. F. Higuera, A. Medina, A. Linan, *Phys. Fluids* **20**, 102102 (2008).
22. A. Ponomarenko, D. Quéré, C. Clanet, *J. Fluid Mech.* **666**, 146–154 (2011).
23. K.-H. Chu, R. Xiao, E. N. Wang, *Nat. Mater.* **9**, 413–417 (2010).
24. P.-G. Gennes, F. Brochard-Wyart, D. Quéré, *Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves* (Springer, 2004).
25. Y. Kim, H. Yuk, R. Zhao, S. A. Chester, X. Zhao, *Nature* **558**, 274–279 (2018).
26. A. Li *et al.*, *Sci. Adv.* **6**, eaay5808 (2020).
27. W. Lei *et al.*, *Sci. Adv.* **4**, eaau8767 (2018).
28. K. Shao *et al.*, *Adv. Funct. Mater.* **32**, 2205831 (2022).
29. Z. Yuan *et al.*, *Sci. Adv.* **9**, eadg2352 (2023).
30. W. Li, X. Tang, L. Wang, *Sci. Adv.* **6**, eabc1693 (2020).
31. G. L. Liu, J. Kim, Y. Lu, L. P. Lee, *Nat. Mater.* **5**, 27–32 (2006).

ACKNOWLEDGMENTS

Funding: L.W. acknowledges the financial support from the Research Grants Council of Hong Kong (GRF 17213823, 17205421, and 17204420) and the Hong Kong Polytechnic University (SHS Chair Professor: P0045687). X.Y. acknowledges support from the New Cornerstone Science Foundation through the XPLORER Prize and the Hong Kong Jockey Club. **Author contributions:** L.Y. and L.W. conceived the project. L.W. and X.Y. supervised the research. L.Y., W.L., and L.W. developed the theoretical model. L.Y. and W.L. performed the experiments. W.L. conducted the XPS analysis. H.Z. and Q.D. assisted in fabricating elastic samples. Jiaq.L., Jiao.L., and Q.D. assisted in the experimental apparatus. L.Y., Jiaq.L., W.L., and L.W. analyzed the data. L.Y. and L.W. wrote the original draft. L.Y., Jiaq.L., W.L., X.Y. and L.W. revised the manuscript. All authors commented on the paper. **Competing interests:** Authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk4180
Materials and Methods
Supplementary Text
Figs. S1 to S31
Movies S1 to S8

Submitted 28 August 2023; resubmitted 19 March 2024
Accepted 13 May 2024
10.1126/science.adk4180

PHOTOSYNTHESIS

Cryo–electron microscopy reveals hydrogen positions and water networks in photosystem II

Rana Hussein[1]†*, André Graça[2,3]†, Jack Forsman[2], A. Orkun Aydin[3], Michael Hall[2], Julia Gaetcke[1], Petko Chernev[3], Petra Wendler[4], Holger Dobbek[1], Johannes Messinger[3,5]*, Athina Zouni[1]*, Wolfgang P. Schröder[2,5]*

Photosystem II starts the photosynthetic electron transport chain that converts solar energy into chemical energy and thus sustains life on Earth. It catalyzes two chemical reactions: water oxidation to molecular oxygen and plastoquinone reduction. Coupling of electron and proton transfer is crucial for efficiency; however, the molecular basis of these processes remains speculative owing to uncertain water binding sites and the lack of experimentally determined hydrogen positions. We thus collected high-resolution cryo–electron microscopy data of fully hydrated photosystem II from the thermophilic cyanobacterium *Thermosynechococcus vestitus* to a final resolution of 1.71 angstroms. The structure reveals several previously undetected partially occupied water binding sites and more than half of the hydrogen and proton positions. This clarifies the pathways of substrate water binding and plastoquinone B protonation.

By oxidizing water into molecular oxygen, photosystem II (PSII) liberates electrons and protons for CO_2 fixation in the Calvin-Benson-Bassham cycle (*1*). PSII is thus at the root of two gas conversion reactions that have shaped Earth's bio-, atmo-, hydro-, and geosphere (*2*, *3*). The efficiency of PSII in splitting water with a catalyst made of earth-abundant elements makes it the blueprint for emerging technologies for the sustainable production of fuels and chemicals (*4*–*6*).

The active PSII complex is a dimer (dPSII) of about 700-kDa molecular weight of which each monomer comprises 20 proteins and nearly 100 cofactors (Fig. 1A). Water oxidation is driven by light-induced charge separations in the PSII reaction center [P_{680} and pheophytin ($Pheo_{D1}$)] that is connected to the Mn_4CaO_5 cluster by means of a redox-active tyrosine (Y_Z) (Fig. 1C). The electrons extracted from water are transferred from $Pheo_{D1}$ to the two plastoquinone (Q_A and Q_B) molecules at the acceptor side of PSII (Fig. 1B). Because of the high quantum efficiency of PSII, both reaction cycles can be driven stepwise by applying saturating microsecond flashes (Fig. 1D). The four-electron, four-proton water oxidation (Kok) cycle starts from the dark-stable S_1 state, and the successively higher indices generally reflect Mn oxidation events (except possibly $S_3 \rightarrow S_4$) (*1*).

Biophysical experiments have revealed a tight coupling between electron and proton transfer in PSII (*7*–*9*). Altering this coupling by dehydration or mutation of critical amino acids reduces the efficiency of PSII and may even inhibit it (*10*). For example, mutation of D1–D^{61} (Fig. 1C) to alanine slows drastically the rates of the $S_2 \rightarrow S_3$ transition and of O_2 formation during the $S_3 \rightarrow S_4 \rightarrow S_0$ transition (*11*). Similarly, removal of the bicarbonate (BCT) ligand of the nonheme Fe^{2+} between Q_A and Q_B (Fig. 1B) or its replacement by formate strongly slows electron transfer between the two quinones.

[1]Humboldt-Universität zu Berlin, Department of Biology, D 10099 Berlin, Germany. [2]Department of Chemistry, Umeå University, SE 90187 Umeå, Sweden. [3]Molecular Biomimetics, Department of Chemistry- Ångström Laboratory, Uppsala University, SE 75120 Uppsala, Sweden. [4]Institute of Biochemistry and Biology, Department of Biochemistry, University of Potsdam, Karl-Liebknecht Strasse 24-25, D 14476, Potsdam-Golm, Germany. [5]Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Sweden.
***Corresponding author. Email: husseinr@hu-berlin.de (R.H.); johannes.messinger@kemi.uu.se (J.M.); athina.zouni@hu-berlin.de (A.Z.); wolfgang.schroder@umu.se (W.P.S.)**
†These authors contributed equally to this work.

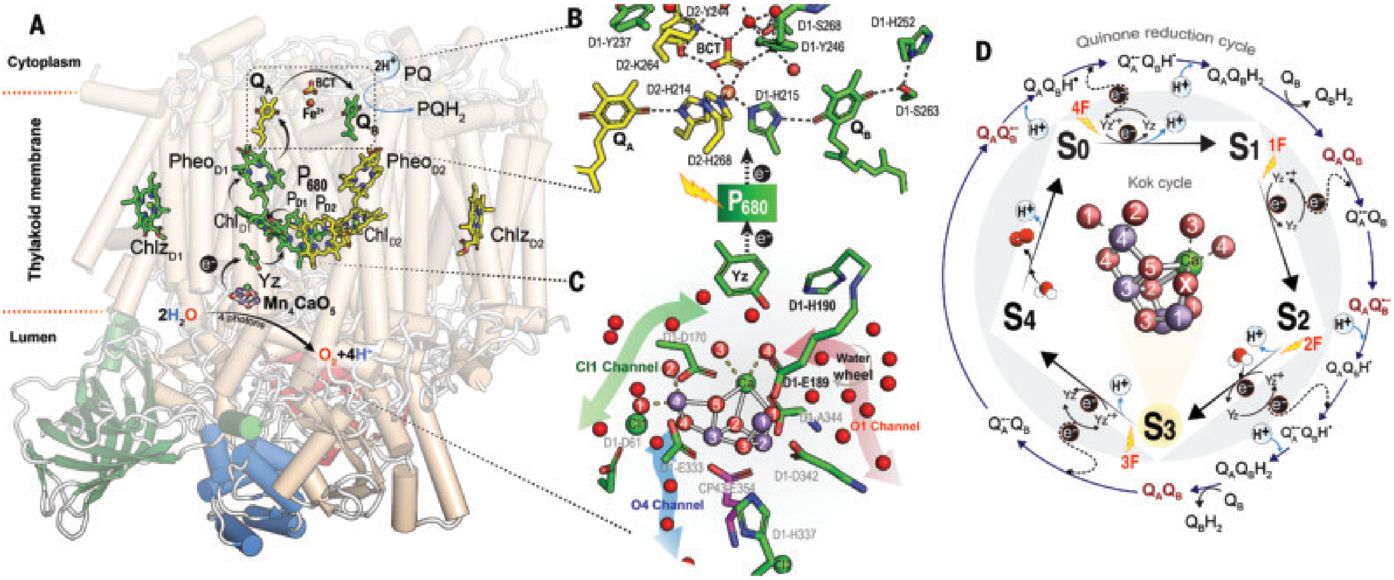


Fig. 1. PSII and its catalytic reactions. (**A**) Structure of monomeric PSII with cofactors for light-induced charge separation between P_{680} (chlorophylls: Chl_{D1}, P_{D1}, P_{D2}, and Chl_{D2}) and $Pheo_{D1}$. Charge separation is stabilized by electron transfer to plastoquinones Q_A and Q_B as well from the Mn_4CaO_5 cluster to P_{680}^{+} through Y_Z (*1*). (**B**) Detailed view of the acceptor site harboring Q_A and Q_B. (**C**) The Mn_4CaO_5 cluster and its three channels: O1 (red), O4 (blue), and Cl1 (green). (**D**) Kok cycle for water oxidation and its connection to quinone reduction. The S_0 through S_4 states represent successively higher oxidation states of the Mn_4CaO_5 cluster reached from the dark-stable S_1 state by one to four flashes (1F to 4F). The Mn_4O_5Ca-OX structure of the S_3 state (PDB: 7RF8) is displayed in the center. Residues and cofactors are colored according to the subunits they belong to: green (D1), yellow (D2), and magenta (CP43). The membrane-embedded helices are shown in light beige, and extrinsic protein subunits are shown in green (PsbO), blue (PsbU), and red (PsbV). Spheres represent Mn ions (purple), water or oxygen (red), Ca^{2+} or Cl^- (green), and Fe^{2+} (brown). Arrows represent electron (black) and proton (blue) transfer. Single-letter abbreviations for the amino acid residues referenced throughout this paper are as follows: A, Ala; C, Cys; E, Glu; R, Arg; N, Asn; Q, Gln; P, Pro; S, Ser; T, Thr; Y, Tyr; G, Gly; V, Val; H, His; D, Asp; K, Lys; F, Phe; I, Ile.

The structure of PSII is known at high resolution based on x-ray crystallography of single crystals at synchrotrons and serial crystallography at x-ray free electron lasers (XFELs) with present record resolutions of 1.85 Å (synchrotron) and 1.89 Å (XFEL) (*12*, *13*). These structures reveal that both the Mn_4CaO_5 cluster and Q_B are situated at buried sites and that proton exchange with the bulk is facilitated by water-filled channels (Fig. 1C). Using XFELs allows obtaining structures of transient species during water insertion, proton release, and O_2 formation (*14–16*), indicating that substrate water insertion into the Mn_4CaO_5 cluster occurs through the "water wheel" in the O1 channel (Fig. 1C). This leads to the formation of an additional oxo-bridge between Ca and Mn1 in the S_3 state (X, Fig. 1D), which supports O–O bond formation between O5 and Ox (also known as O6) (*17–20*).

Recently, this emerging picture has been fundamentally challenged by two studies: (i) On the basis of a reanalysis of XFEL data, the presence of Ox in the S_3 state was disputed (*21*), (ii) and a theoretical study (*22*) concluded that the pretreatment procedures used in crystallography for enhancing the resolution remove not only water molecules from interfaces between PSII dimers but also critical water molecules from within PSII. Thus, those authors of latter study proposed that none of the present crystal structures depict the physiological hydration state of PSII, making it potentially impossible to draw meaningful conclusions. Additionally, ambiguities remain because the present x-ray structures do not resolve hydrogen and proton positions, which need to be known for understanding proton-coupled electron transfer reactions.

Over recent years, cryo–electron microscopy single-particle analysis (cryo-EM SPA) has made enormous progress by reaching resolutions comparable to those of x-ray crystallography, with the advantage that noncrystalline samples are used. This not only simplifies data collection but also ensures full hydration of the samples and avoids structural and functional effects due to crystal contacts (*12*). Notably, it has been demonstrated that the Coulomb potential maps obtained by cryo-EM are more sensitive to H atoms than electron densities visualized by x-ray diffraction. For apoferritin, H atoms were clearly resolved at resolutions of 1.19 to 1.25 Å (*23–26*).

In this work, we performed cryo-EM SPA on dPSII core complexes (dPSIIcc) isolated from the thermophilic cyanobacterium *Thermosynechococcus vestitus*, similar to preparations previously used for XFEL studies (*13*). We obtained a 1.71-Å resolution structure, which is higher than resolutions of prior PSII complex structures regardless of method or species and at par with the highest resolution yet obtained by cryo-EM for a membrane protein complex (*24*). By using omitted difference maps, we extracted >50% of the protons for the reaction center proteins, demonstrating the ability of cryo-EM SPA for resolving H positions and thus increasing our understanding of how substrate water is inserted into the Mn_4CaO_5 cluster and Q_B is reduced and protonated.

Results

High-resolution cryo-EM structure of PSII

Structural data were collected of dPSIIcc from *T. vestitus* BP-1 by cryo-EM SPA. Right before vitrification, the samples were exposed to one or two flashes (1F or 2F)to enrich the S_2 or S_3 and Q_B– or Q_B states, respectively (see sections 1.1 and 1.2 in materials and methods). Initial data analysis revealed that the Mn_4CaO_5 cluster was reduced by the electron beam to nonphysiological states below S_0. Thus, the data were combined to improve the overall quality. The combined dataset had a resolution of 1.71 Å (see sections 1.3 and 1.4 in materials and methods) (Fig. 2A, figs. S1 to S3, and table S1).

The refined structure obtained from the combined dataset reliably detected the Mn_4CaO_5 cluster and other cofactors involved in charge separation and water oxidation (Fig. 2, B to F). For the Mn_4CaO_5 cluster, clear densities were

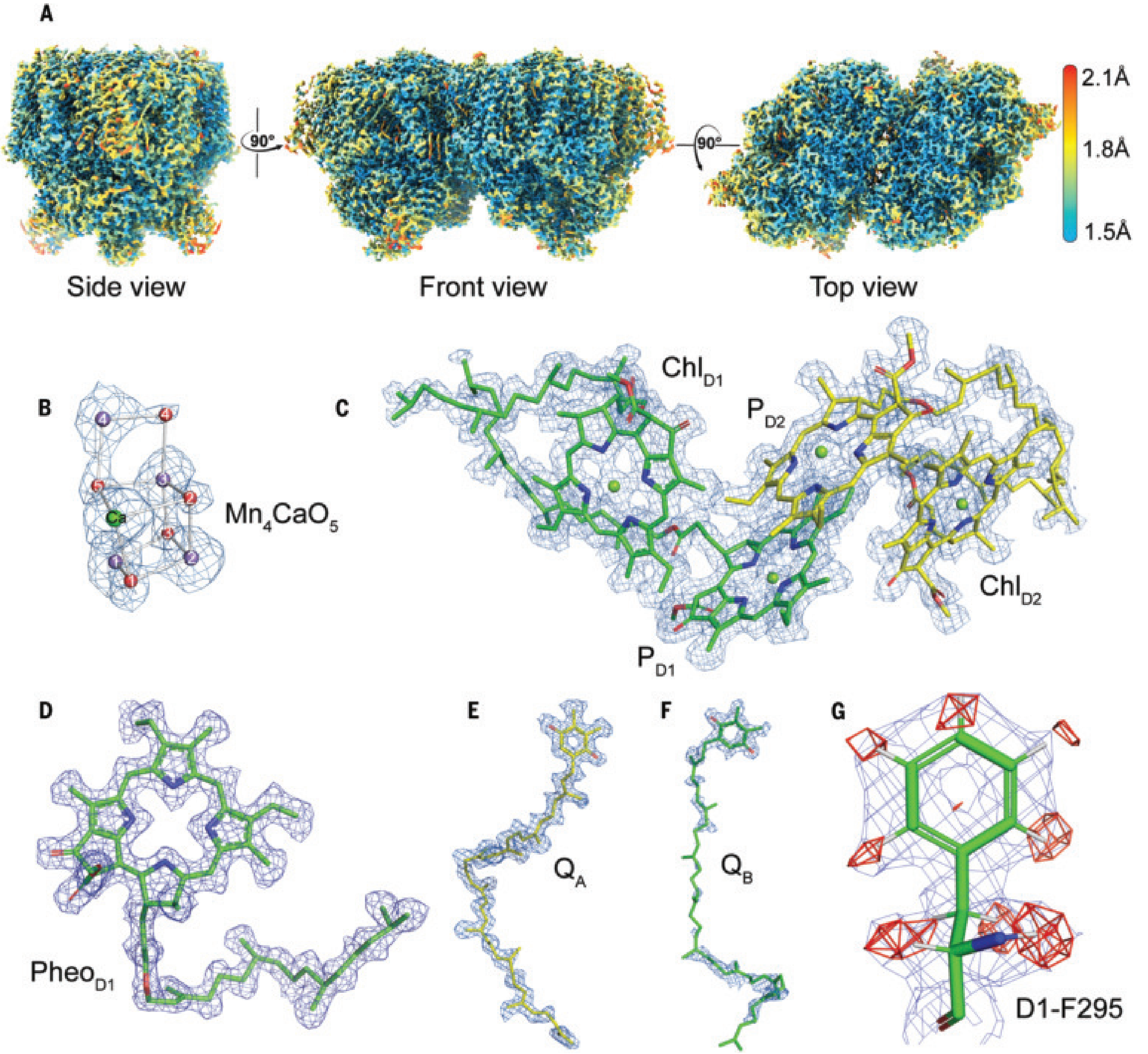


Fig. 2. Cryo-EM structure of dimeric PSII from *T. vestitus* (PDB9EVX). (**A**) Local resolution map shown at different views of the dPSIIcc. The map is colored according to the local resolution values at the FSC threshold of 0.143. (**B** to **F**) PSII redox-active cofactors with their densities: (B) Mn_4CaO_5 cluster, (C) P_{680} (Chl_{D1}, P_{D1}, P_{D2}, and Chl_{D2}), (D) $Pheo_{D1}$, (E) Q_A, and (F) Q_B. All densities are shown as blue mesh and contoured at 5 RMSD, except for the heavy metals, which are depicted at 8 RMSD. (**G**) The hydrogen atoms detected in the omitted difference map in one of the amino acid residues of the D1 subunit. The map (blue mesh) and the omit map (red mesh) were contoured at 4 RMSD and 8 σ, respectively.

observed for the four Mn ions and the Ca ion, which allows precise modeling of their positions based on the peak maxima. Moreover, all five oxygens connecting the metal ions were directly revealed, removing the need for difference map analysis (Fig. 2B). The structure found is in agreement with previous reports at lower resolution (*12*, *18*, *27*, *28*), but the metal-metal, metal-oxo, and metal-ligand distances are elongated by up to 0.2, 0.4, and 0.2 Å, respectively (figs. S4 and S5 and table S2). This aligns with previous observations that high-valent metal centers are rapidly reduced by electron beam exposure during cryo-EM data collection (*29*, *30*). Despite exposing half of the samples to two flashes, no density for Ox was detected (Fig. 2B). We ascribe this to disorder in bridge positions caused by the reduction of the Mn ions. Nevertheless, the radiation damage in our structure appears to be relatively moderate. Clear densities are evident for all ligands of the Mn_4CaO_5 cluster, including D1–A^{344} with only one conformation (fig. S6A). Furthermore, only 50% of the disulfide bond between PsbO–C^{19} and PsbO–C^{44} is disrupted (fig. S6B), unlike in earlier reports. (*29*, *30*).

For the acceptor side, the data show a complete density for Q_A, whereas for Q_B, full density was only resolved at a lower root mean square deviation (RMSD) level of the density values of the map, especially for the isoprenoid tail (Fig. 2, E and F, and fig. S7). This is most probably due to the generally higher mobility of Q_B and the flash-induced difference in the redox state of Q_B between the two data sets but may also reflect a lower Q_B occupancy.

Our high-resolution cryo-EM map of dPSIIcc demonstrates potential for elucidating the positions of protons and hydrogen atoms within

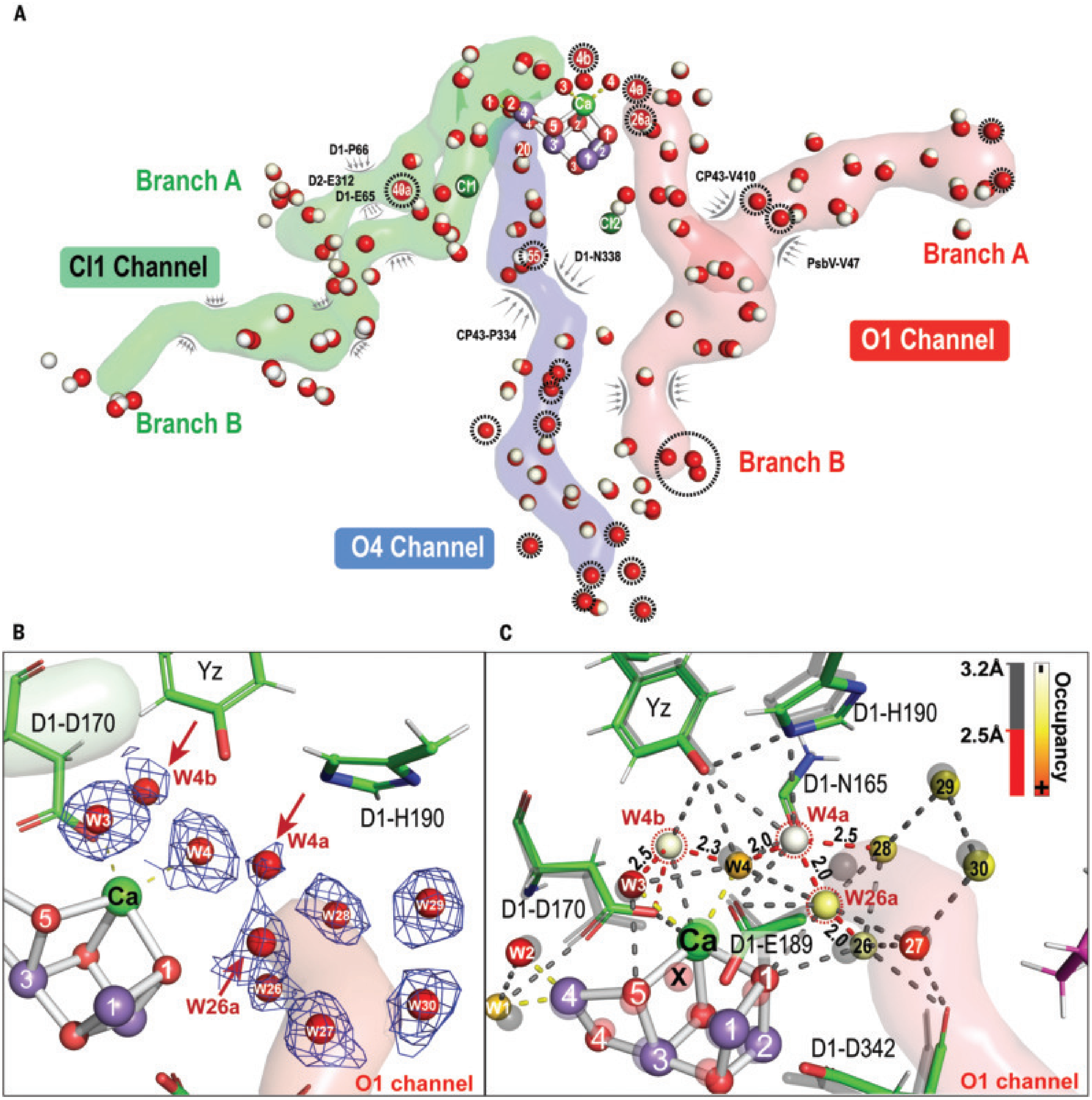


Fig. 3. Water molecules in water or proton channels enabling water oxidation by the Mn_4CaO_5 cluster. (**A**) The O1, O4, and Cl1 channels are depicted in red, blue, and green, respectively. Water molecules within 5 Å of these channels are shown as red (present cryo-EM structure; PDB: 9EVX) and white spheres (1.89-Å XFEL structure; PDB: 7RF1). Waters present only in our cryo-EM structure are marked by black dashed circles. Channel bottlenecks are indicated by gray brackets and arrows. (**B** and **C**) Waters found in the O1 channel near the Mn_4CaO_5 cluster. Red arrows and dashed circles highlight newly detected waters. The blue mesh density maps of these waters are contoured in (B) at 3.0 RMSD, and the colors of corresponding spheres in (C) reflect their occupancy (range, 0.2 to 1.0; for details, see materials and methods). Interatomic distances are depicted as dashed, color-coded lines. The occupancy and distance scales are shown in the upper right. The 2F XFEL model (PDB ID: 7RF8) is displayed in transparent colors.

PSII (Fig. 2G) owing to the higher scattering potential of electrons for lighter elements compared with x-rays, which markedly improves the contrast for hydrogen atoms (*31*).

Water networks enabling water splitting

We confirmed in our EM map the positions of water molecules in the network extending through the three channels connecting the Mn_4CaO_5 cluster to the lumen (Fig. 1C), which were described previously by room temperature XFEL experiments (*13*, *15*, *19*, *32*) (Fig. 3A). Our maps revealed sites featuring reduced

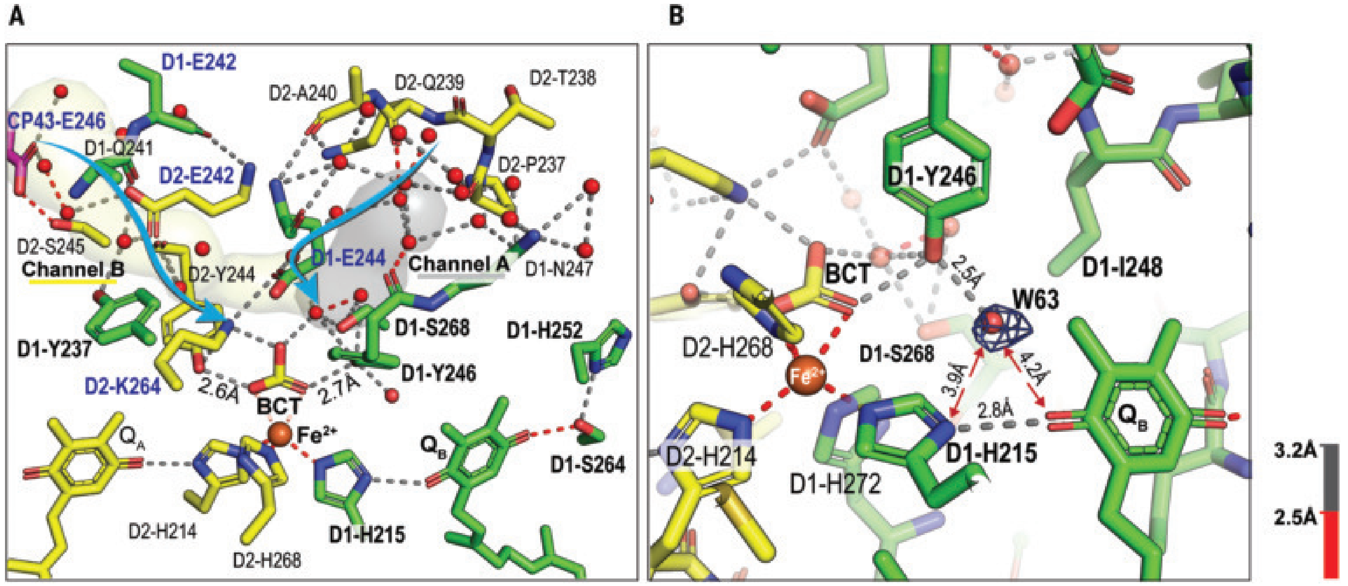


Fig. 4. Water or proton channels supporting Q_B reduction and protonation. (**A**) Water channels displayed in gray (channel A) and in pale yellow (channel B) were generated by CAVER (*34*) and connect the BCT to the cytosol. Cyan arrows depict suggested protonation pathways of BCT ligated at the nonheme iron. Charged residues are indicated in bold dark blue. (**B**) Q_B binding pocket showing the extra density for water W63. The density map of this water is shown as a blue mesh and contoured at 3.0 RMSD. Interatomic distances are color coded (see bottom right). The residues belonging to the D1 and D2 subunits are colored green and yellow, respectively. All water molecules are shown as red spheres.

density due to high mobility and/or partial occupancy, which require high resolution and cryogenic conditions for detection. Three of these additional water sites (W26a, W4a, and W4b) were found close to the origin of the O1 channel (Fig. 3, A to C; for water numbering, see fig. S8 and table S3). Analysis of the hydrogen bond network in this area (Fig. 3C) showed that W26a is hydrogen bonded to the O1 oxo-bridge of the Mn_4CaO_5 cluster and at a very close distance (2.0 Å) to W26, suggesting that it might represent a second position of W26. W4a and W4b are connected to the Mn_4CaO_5 cluster through H bonding to the W3 and/or W4 water ligands of Ca^{2+}. We propose that these newly identified sites represent intermediates along the water insertion route into the Mn_4CaO_5 cluster during the $S_2 \rightarrow S_3$ transition, i.e., from the water wheel to the Ox binding site located between Ca and Mn1 (*13*) (Figs. 1 and 3C). A higher positional flexibility of W3 and W4 compared with that of W1 and W2 is in line with recent reports (*18*, *30*, *33*), but we cannot fully exclude structural effects of Mn_4CaO_5 cluster reduction.

Going out further in the O1 channel toward the lumen, we detected two additional water binding sites near a bottleneck at the entrance to the "A branch" that is formed by CP43–V^{410} and PsbV–K^{47} (Fig. 3A). The remaining five additional water sites of the O1 channel are at the exits of the A and B branches, i.e., in areas with very high mobility at room temperature.

In the Cl1 channel, all water positions are consistent with those detected by XFEL (*13*) (fig. S9). One additional water molecule, W40a, was detected at partial occupancy (fig. S9B) near the proton gate residues D2–E^{312}, D1–E^{65}, and D1–R^{334} (fig. S9B). This water was detectable previously only at the 250-μs time point after the second flash collected by XFEL (*13*). The proximity to W40 and the occupancy pattern suggest that W40a represents an alternative position of W40.

For the O4 channel, 10 additional water molecules were found, but none are close to the Mn_4CaO_5 cluster (Fig. 3A). The closest, W55, is around 15 Å away, located just before the first bottleneck region (Fig. 3A and fig. S10, A and B). W55 was detected previously in x-ray diffraction data collected under cryogenic conditions (fig. S10C) (*12*, *28*). It's absence in other structures indicates high mobility at room temperature, which may affect the H network in this area (fig. S10B). Notably, despite cryogenic conditions and the high resolution achieved in this study, no water molecules were detectable in the subsequent bottleneck region (Fig. 3A). Therefore, our data are not in line with recent molecular dynamics (MD) calculations suggesting that this break is due to artifacts as a result of dehydration of PSII crystals (*22*). Instead, they provide strong support for previous proposals that this bottleneck breaks the H-bonding network of the O4 channel. The remaining nine additional water molecules were all found past this region, most near the exit of the channel (Fig. 3A). At the start of the O4 channel, W20 was visible in our data, albeit with a notably lower peak height compared with that of adjacent water molecules (fig. S11). W20 was disordered in previous 1F (S_2) and 2F (S_3) XFEL data (*13*, *18*, *19*), suggesting a potential contribution of S_1 or S_0 states to our current structure. However, the observed stabilization of this water molecule could also be a result of the cryogenic conditions under which the data were collected or the reduction of the Mn_4CaO_5 cluster.

H-bonding network facilitating plastoquinone reduction

The water-proton network at the acceptor side was investigated by using CAVER (section 1.5, materials and methods) (*34*). Two main channels, labeled A and B, were detected that connect the nonheme iron (NHI) by its ligated BCT to the cytosol side of the thylakoid membrane (Fig. 4A). Channel A is comparatively short, wide, and open to the bulk surface through the D1–N^{247}, D2–Q^{239}, and D2–P^{237} residues (fig. S12). It harbors several water molecules that form an H-bond network connecting the BCT and D1–S^{268} to the bulk (Fig. 4A). Notably, the center path of this wide channel appears largely empty, likely because the waters present there are highly mobile. Channel A was previously identified based on quantum mechanics/molecular mechanics (QM/MM) (*35*) and MD (*22*) calculations and was proposed to provide a path by which BCT or CO_2 enters to or exits from the NHI, respectively (*35*).

Channel B is narrow (fig. S12) and features two bottlenecks that restrict water movements. The first is formed by the BCT, D2–T^{243}, D1–E^{244}, and D2–Y^{244}, whereas the second is located just before the channel opens to the bulk and is formed by D2–E[242], D1–G[240], and D2–V[247]. Channel B comprises several charged amino acid residues, and H-bond network analysis (Fig. 4A) suggests that channel B provides a better protonation path for the BCT than channel A. D1–E[244], which is a residue found at the branching point of both channels, is proposed to be involved in the protonation of BCT, especially at lower pH (*35*, *36*). Recent MD calculations (*22*) report two additional water channels; however, we found no evidence for them, possibly owing to their proposed dynamic nature (fig. S13).

In the Q_B pocket, our structure identifies the density for a water molecule, W63, that is situated within a 2.5-Å distance from D1–Y[246] (Fig. 4B). Previously, W63 was experimentally observed only for the ΔPsbM mutant of PSII (*37*) and predicted based on QM/MM calculations (*38*), whereas another theoretical study suggested the presence of two water molecules in this region (*22*). Thus, our finding provides strong experimental support for the proposal that W63 contributes to the reprotonation of D1–H[215] through BCT and D1–Y[246] following the formation of Q_BH_2 (see 2F in Fig. 5E) (*22*, *35*, *38*, *39*).

Hydrogens and protons

Investigation of the percentage of detectable hydrogen atoms was performed by calculating difference maps with Servalcat (section 1.6, materials and methods) (*40*). By using the D1 subunit as a representative example, a value of >50% was obtained.

Accurate hydrogen positioning refines structural models by reducing side chain rotamer errors, notably for asparagine and glutamine (*41*). For example, D1–N[338], situated at the bottleneck of the O4 channel, has been recently proposed to function as a drawbridge for water molecules in this region based on QM/MM

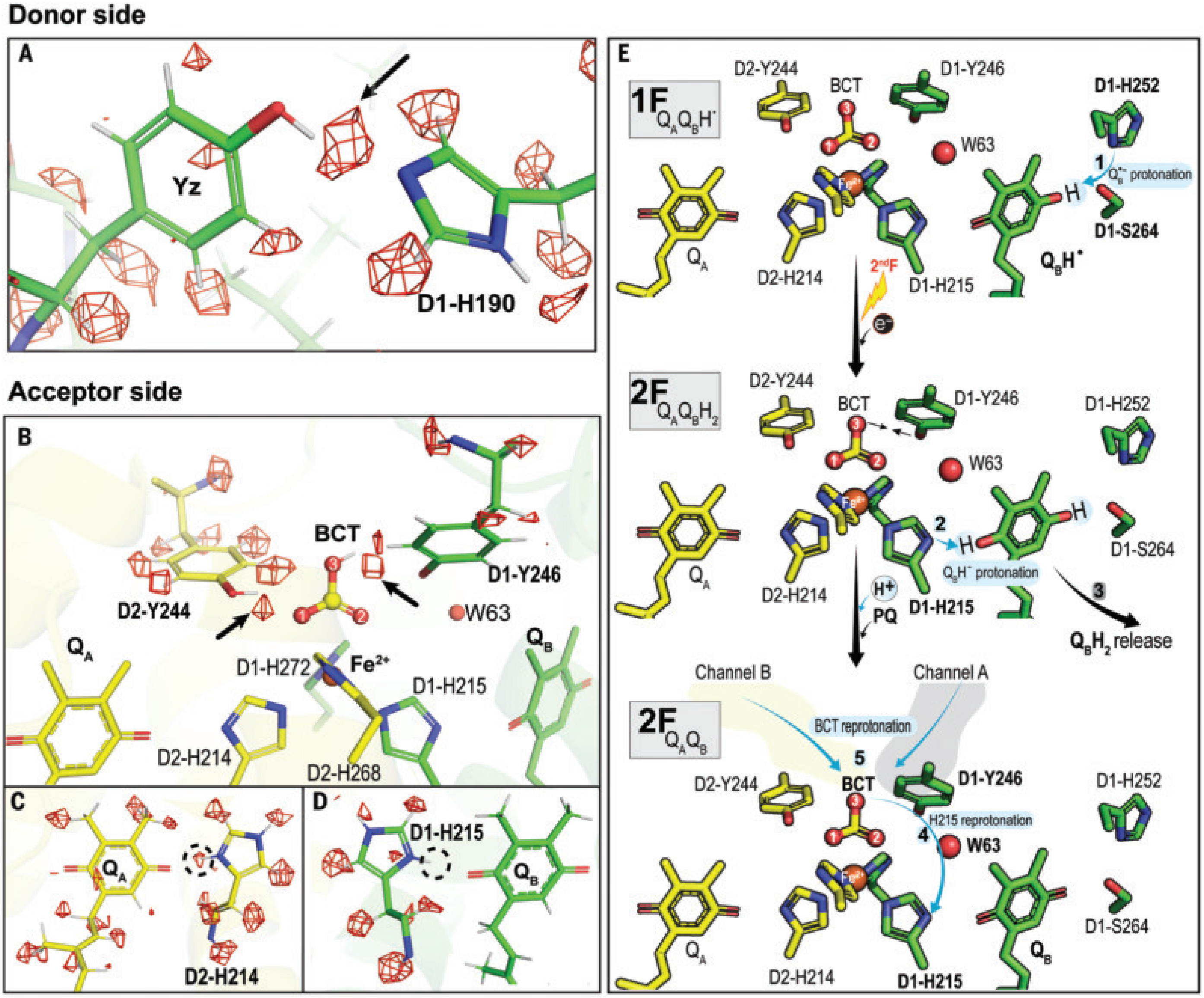


Fig. 5. Hydrogens or protons detected at the donor and acceptor sides. (**A** to **D**) Hydrogen omitted-difference maps of key residues at the donor (A) and acceptor [(B) to (D)] sides. The omit map of the hydrogens and protons are shown as a red mesh and contoured at 8 σ. The black arrows and dashed circles highlight the detected or missed protons at specific residues. (**E**) Schematic summary of the sequence of events at the Q_B site after the first and second flashes given to dark-adapted PSII. The scheme starts with the 1F (Q_B–) state reached after the first flash (see also Fig. 1) and depicts the protonation of Q_B– by D1–H[252] (step 1). After the second flash (2F state), Q_BH^- is formed and protonated from D1–H[215] (step 2), which is followed by the release of Q_BH_2 (step 3), rebinding of Q_B and reprotonation of D1–H[215] from BCT (step 4), and the reprotonation of BCT by protons from the cytosol through channel A (gray) and/or B (yellow) (step 5). Blue arrows indicate the proposed proton transfer path.

calculations (*33*). In several previous structures (*14*, *18*), the side chain of D1–N^{338} has been reported in a flipped position to that of the current structure (fig. S14).

The omitted difference map for protons also identifies a density peak for the shared proton between Y_Z and D1–H^{190} (Fig. 5A). The peak maximum of this density is closer to the N(ε) of H^{190} than to the phenolic oxygen of Y_Z (1.24 and 1.57 Å, respectively), whereas the total distance of 2.75 Å confirms the presence of a hydrogen bond, albeit a weaker one than expected (≤2.6 Å) (*13*, *14*). Although the present result clearly demonstrates the potential of cryo-EM to address such questions, we note that for a full analysis, further studies are required. Firstly, in cryo-EM, the peak maxima for hydrogen or proton positions are affected by the resolution and B factor of the dataset and thus typically deviate from their expected locations (table S4) (*24*, *25, 42*). Secondly, this proton position may be affected by the reduction of the nearby Mn_4CaO_5 cluster.

On the acceptor side, the hydrogen-omitted difference map reveals a density for a proton shared between the phenolic oxygen of D2–Y^{244} and O1 of BCT, whereas no density was detected for the proton of the hydroxyl group of D1–Y^{246} (Fig. 5B). This result corroborates the earlier conclusion that only one of the tyrosine residues, either D2–Y^{244} or D1–Y^{246}, can form a hydrogen bond with the BCT (*43*). Furthermore, a proton density was also detected near O3 of BCT and in plane with D1–Y^{246}, which has its peak maximum within 1.5 Å from O3 of BCT and 2.3 Å from the phenolic group of D1–Y^{246}. Notably, the distance between O3 of BCT and D1–Y^{246} is considerably shorter compared with previously reported distances in 0F (Q_B-rich) and 1F (Q_B–-rich) structures (*13*, *14*) (fig. S15).

Our analysis showed that all hydrogen atoms and protons of D2–H^{214}, situated close to Q_A, were detected (Fig. 5C). By contrast, for D1–H^{215}, which is positioned near Q_B, the proton closest to the proximal oxygen of Q_B was not observed (Fig. 5D). These findings may indicate that D1–H^{215} is deprotonated and/or that the proton moved to the proximal oxygen of Q_B (Fig. 5E). However, the higher B factor of the Q_B region relative to the Q_A region (figs. S16 and S17), which reflects the differences in the protein environment of the two quinones that enable their disparate functions, results in several undetected hydrogen positions at Q_B and its surrounding residues, such as D1–S^{264} and D1–H^{252}.

Overall, the structural data are consistent with the suggestion (*38*) that Q_BH^- is protonated by D1–H^{215} and that D1–H^{215} is reprotonated through D1–Y^{246}, W63, and BCT (Fig. 5E).

Conclusions

The advancements of cryo-EM have allowed a transformative change in the field of structural biology, enabling the capture of atomic resolution three-dimensional structural data from highly purified and homogeneous protein complexes flash-frozen from their dispersions in aqueous buffer (*23*). Although concerns presently remain regarding radiation damage to high-valent metal centers, these should not distract from the exceptional insights provided by this technique.

The present 1.71 Å–resolution cryo-EM structure of dimeric PSII from *T. vestitus* unveils partially occupied water-binding sites near the water-splitting Mn_4CaO_5 cluster, strongly supporting water delivery to the Mn_4CaO_5 cluster through the O1 channel. Similarly, it provides experimental detection of W63 in intact PSII. By confirming this crucial structural feature of a theoretical calculation (*22*, *35*, *38*), our data settles the protonation pathways of Q_BH^- and for the subsequent reprotonation of the proton donor (Fig. 5E).

A notable feature of cryo-EM is its enhanced scattering potential for light elements, such as hydrogen, which offers distinct advantages over x-ray crystallography at similar resolution. Our analysis shows >50% of the hydrogen or proton positions within PSII, which greatly extends present structural models. Knowledge of the protonation states of amino acids, cofactors, and/or substrates near the catalytic sites of PSII is essential for revealing how their complex interactions enable efficient proton-coupled electron transfer and water oxidation. The various degrees of proton or hydrogen detection in different regions of the protein are suggestive of local protein dynamics that may play a role in catalysis, exemplified by the marked differences in the Q_A and Q_B binding sites.

REFERENCES AND NOTES

1. D. Shevela, J. F. Kern, G. Govindjee, J. Messinger, *Photosynth. Res.* **156**, 279–307 (2023).
2. R. A. Berner, *Geochim. Cosmochim. Acta* **70**, 5653–5664 (2006).
3. E. Nisbet, C. M. R. Fowler, *Chin. Sci. Bull.* **56**, 4–13 (2011).
4. M. Grätzel, *Acc. Chem. Res.* **14**, 376–384 (1981).
5. W. Lubitz, E. J. Reijerse, J. Messinger, *Energy Environ. Sci.* **1**, 15–31 (2008).
6. T. Faunce *et al*., *Energy Environ. Sci.* **6**, 1074 (2013).
7. A. Klauss, M. Haumann, H. Dau, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16035–16040 (2012).
8. F. Allgöwer, A. P. Gamiz-Hernandez, A. W. Rutherford, V. R. I. Kaila, *J. Am. Chem. Soc.* **144**, 7171–7180 (2022).
9. P. Greife *et al*., *Nature* **617**, 623–628 (2023).
10. T. Noguchi, M. Sugiura, *Biochemistry* **41**, 2322–2330 (2002).
11. M. Hundelt, A.-M. A. Hays, R. J. Debus, W. Junge, *Biochemistry* **37**, 14450–14456 (1998).
12. A. Tanaka, Y. Fukushima, N. Kamiya, *J. Am. Chem. Soc.* **139**, 1718–1721 (2017).
13. R. Hussein *et al*., *Nat. Commun.* **12**, 6531 (2021).
14. M. Ibrahim *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12624–12635 (2020).
15. A. Bhowmick *et al*., *Nature* **617**, 629–636 (2023).
16. H. Li *et al*., *Nature* **626**, 670–677 (2024).
17. P. E. M. Siegbahn, *Acc. Chem. Res.* **42**, 1871–1880 (2009).
18. J. Kern *et al*., *Nature* **563**, 421–425 (2018).
19. M. Suga *et al*., *Science* **366**, 334–338 (2019).
20. C. de Lichtenberg *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2319374121 (2024).
21. J. Wang, W. H. Armstrong, V. S. Batista, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023982118 (2021).
22. A. Sirohiwal, D. A. Pantazis, *J. Am. Chem. Soc.* **144**, 22035–22050 (2022).
23. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, *Nature* **587**, 157–161 (2020).
24. T. Nakane *et al*., *Nature* **587**, 152–156 (2020).
25. S. Maki-Yonekura, K. Kawakami, K. Takaba, T. Hamaguchi, K. Yonekura, *Commun. Chem.* **6**, 98 (2023).
26. T. Bick, P. M. Dominiak, P. Wendler, *BBA Adv.* **5**, 100113 (2024).
27. Y. Umena, K. Kawakami, J. R. Shen, N. Kamiya, *Nature* **473**, 55–60 (2011).
28. M. Suga *et al*., *Nature* **517**, 99–103 (2015).
29. K. Kato *et al*., *Commun. Biol.* **4**, 382 (2021).
30. C. J. Gisriel *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2116765118 (2022).
31. K. Yonekura *et al*., *IUCrJ* **5**, 348–353 (2018).
32. M. Suga *et al*., *Nature* **543**, 131–135 (2017).
33. M. D. Doyle *et al*., *J. Am. Chem. Soc.* **145**, 14621–14635 (2023).
34. E. Chovancova *et al*., *PLOS Comput. Biol.* **8**, e1002708 (2012).
35. Y. Sugo, H. Ishikita, *Front. Plant Sci.* **13**, 934736 (2022).
36. H. Ishikita, E. W. Knapp, *J. Am. Chem. Soc.* **127**, 14714–14720 (2005).
37. S. Uto *et al*., *Faraday Discuss.* **198**, 107–120 (2017).
38. K. Saito, A. W. Rutherford, H. Ishikita, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 954–959 (2013).
39. F. Müh, A. Zouni, *Photosynth. Res.* **116**, 295–314 (2013).
40. K. Yamashita, C. M. Palmer, T. Burnley, G. N. Murshudov, *Acta Crystallogr. D Struct. Biol.* **77**, 1282–1291 (2021).
41. I. K. McDonald, J. M. Thornton, *Protein Eng.* **8**, 217–224 (1995).
42. M. T. B. Clabbers, M. W. Martynowycz, J. Hattne, T. Gonen, *J. Struct. Biol. X* **6**, 100078 (2022).
43. R. Takahashi, A. Boussac, M. Sugiura, T. Noguchi, *Biochemistry* **48**, 8994–9001 (2009).

ACKNOWLEDGMENTS

We acknowledge M. Ibrahim, A. Bhowmick, J. F. Kern, J. Yano, and V. K. Yachandra for helpful discussions. **Funding:** The data were collected at Umeå Core Facility for Electron Microscopy, which is funded by the Knut and Alice Wallenberg, Family Erling Persson, and Kempe Foundations; SciLifeLab; Stockholm University; and Umeå University. This work was supported by Vetenskapsrådet grant 2020-03809 (J.M.); Germany's Excellence Strategy grant, coordinated by T.U. Berlin (project EXC 2008/1-390540038) (A.Z., H.D., and P.W.); German Research Foundation (DFG) grant through the Collaborative Research Center SFB1078 (Humboldt Universität zu Berlin, grant no.TP A5) (R.H., J.G., H.D., and A.Z.); Carl Tryggers Foundation (CTS 19.324) (W.P.S.); Kempe (JCK-2030 2021-2023) (W.P.S. and J.M.); and Lawski (date 230525) (A.G. and J.F.). **Author contributions:** J.M., A.Z., and W.P.S. designed the experiments, coordinated the research, and provided funding; J.G. prepared the dPSIIcc under the guidance of A.Z.; A.O.A. and P.C. developed the flash set up and protocol in the group of J. M.; A.G. and J.F. prepared the grids with the support of W.P.S.; A.G. designed the data collection parameters and, together with M.H., collected the data; R.H. and A.G. performed the data analysis with the support of P.W. and H.D.; and R.H. and J.M. wrote the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The cryo-EM map and structure model have been deposited to the Electron Microscopy Databank and Protein Data Bank with accession numbers EMD-50019 and PDB 9EVX.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn6541
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 to S4
References (*44–51*)
MDAR Reproducibility Checklist

Submitted 18 January 2024; accepted 16 May 2024
10.1126/science.adn6541

FREQUENCY COMBS

Observation of topological frequency combs

Christopher J. Flower[1]†, Mahmoud Jalali Mehrabad[1]†, Lida Xu[1]†, Gregory Moille[1], Daniel G. Suarez-Forero[1], Oğulcan Örsel[2], Gaurav Bahl[2], Yanne Chembo[3], Kartik Srinivasan[1], Sunil Mittal[4]*, Mohammad Hafezi[1]*

On-chip generation of optical frequency combs using nonlinear ring resonators has enabled numerous applications of combs that were otherwise limited to mode-locked lasers. Nevertheless, on-chip frequency combs have relied predominantly on single-ring resonators. In this study, we experimentally demonstrate the generation of a novel class of frequency combs, the topological frequency combs, in a two-dimensional lattice of hundreds of ring resonators that hosts fabrication-robust topological edge states with linear dispersion. By pumping these edge states, we demonstrate the generation of a nested frequency comb that shows oscillation of multiple edge state resonances across ≈40 longitudinal modes and is spatially confined at the lattice edge. Our results provide an opportunity to explore the interplay between topological physics and nonlinear frequency comb generation in a commercially available nanophotonic platform.

Nonlinear effects, in particular the Kerr effect, in ring resonators provide a compact route to the generation of optical frequency combs in integrated photonic chips (*1–4*). These combs have led to a plethora of applications including spectroscopy (*5*, *6*), precision timekeeping, on-chip signal synthesis, ranging and detection (*7*, *8*), and optical neural networks. Although Kerr combs have been demonstrated in a wide variety of integrated material platforms, device design has been predominantly limited to single-ring resonators. Coupled resonator systems have only very recently been investigated as a means to engineer the dispersion and, subsequently, the comb spectrum (*9–13*). Beyond dispersion engineering, nonlinear coupled resonator systems can exhibit coherent solutions that are not possible using single-ring resonators (*13–16*).

Concurrently, topological photonics has emerged as a powerful paradigm for the design of photonic devices with novel functionalities (*17–22*). More specifically, topological systems exhibit chiral or helical edge states that are confined to the boundary of the system and are exceptionally robust against imperfections common to integrated photonic devices (*23–25*). Examples include robust optical delay lines (*26*), chiral quantum optics interfaces (*27–29*), slow-light engineering (*30*), waveguides, tapers, and reconfigurable routers (*31–33*). Whereas early efforts in topological photonics focused on linear devices, more recent demonstrations have included nonlinear effects, extending the scope of possible applications to include lasers (*34–37*), parametric amplifiers (*38*, *39*), and quantum light sources (*40–43*). Additionally, it was theoretically shown that nonlinear effects in large two-dimensional (2D) topological arrays of ring resonators can lead to the generation of coherent nested temporal solitons exhibiting an order of magnitude higher efficiency compared with single-ring combs (*44*).

We experimentally demonstrate the generation of the first topological frequency comb in a 2D lattice of >100 ring resonators, fabricated using a commercially available integrated silicon nitride (SiN) nanophotonic platform. As we pump within a topological edge band, we observe the generation of a frequency comb confined within the edge bands across ≈40 longitudinal modes. Using an ultra-high-resolution spectrum analyzer, we reveal the distinctive nested structure of the comb, wherein each comb tooth is further split into a set of finer teeth. Furthermore, we directly image a set of comb teeth and verify that their spatial profile is indeed confined to the edge of the lattice. As such, each comb tooth constitutes a topological edge state that is robust against 90° bends in the lattice and demonstrates the preservation of topology in a highly nonlinear system. This novel modulation instability comb is the first example of a new family of frequency combs and paves the way for the development of coherent topological frequency combs and nested temporal solitons (*44*).

[1]Joint Quantum Institute, National Institute of Standards and Technology, University of Maryland, College Park, MD, USA. [2]Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD, USA. [4]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA.
***Corresponding author. Email: s.mittal@northeastern.edu (S.M.); hafezi@umd.edu (M.H.)**
†These authors contributed equally to this work.

Design

Our topological system consists of an array of coupled ring resonators that simulates the anomalous quantum Hall (AQH) model for photons (*44–46*), as shown in Fig. 1. The 180 "site-ring" resonators form a square lattice, where nearest and next-nearest sites are coupled together by means of an interspersed lattice of 81 detuned "link-ring" resonators. The link rings are detuned by engineering a path-length difference with respect to the site rings, ensuring that close to site-ring resonances the intensity present in the link rings will be negligible. As a result, the link rings act as waveguides and introduce a direction-dependent hopping phase of $\pm\pi/4$ for nearest-neighbor couplings and 0 for next-nearest neighbor couplings. We note that our system implements a copy of the AQH model at each of the longitudinal mode resonances ($\omega_{0,\mu}$) of the ring resonators. Therefore, the linear dynamics of the system are described by a multiband tight-binding Hamiltonian $\left(\hat{H}_{\mathrm{L}}\right)$

$$\begin{aligned}\hat{H}_{\mathrm{L}} = & \sum_{m,\mu}\omega_{0,\mu}\hat{a}^{\dagger}_{m,\mu}\hat{a}_{m,\mu} \\ & -J\sum_{\langle m,n\rangle,\mu}\hat{a}^{\dagger}_{m,\mu}\hat{a}_{n,\mu}e^{-i\phi_{m,n}} \\ & -J\sum_{\langle\langle m,n\rangle\rangle,\mu}\hat{a}^{\dagger}_{m,\mu}\hat{a}_{n,\mu}\end{aligned} \quad (1)$$

Here, $\hat{a}^{\dagger}_{m,\mu}$ is the photon creation operator at a site ring m and longitudinal mode μ. The coupling strength between site rings for both nearest and next-nearest neighbors is given by J. The resonance frequency of the site-ring resonators for a longitudinal mode with index μ is denoted $\omega_{0,\mu} = \omega_0 + D_1\mu + \frac{D_2}{2}\mu^2$, where D_1 is the free spectral range (FSR), D_2 is the second-order dispersion, and ω_0 is the resonance frequency of the pumped longitudinal mode, $\mu = 0$. This Hamiltonian leads to the existence of an edge band, spectrally located between two bulk bands, near each of the longitudinal mode resonances, as shown in Fig. 1D. Furthermore, simulated transmission shows each edge band hosting another set of resonances. These resonances are associated with the longitudinal modes of the super-ring resonator formed by the edge states, giving rise to a nested mode structure. For the associated band structure, refer to fig. S1.

We also emphasize that the system is time-reversal invariant, and the topological edge states are helical in nature. More specifically, the clockwise (CW) and counterclockwise (CCW) circulation of light in the site rings (also referred to as the pseudospin) leads to edge states that are circulating around the lattice boundary in the CCW and CW directions, respectively. By choosing the port of excitation, we can selectively excite a given edge state (Fig. 1).

In the presence of a strong pump, the intrinsic Kerr nonlinearity of SiN leads to four-wave mixing and, subsequently, the generation of optical frequency combs in the lattice. This nonlinear interaction is described by the following Hamiltonian

$$\hat{H}_{\mathrm{NL}} = -\beta\sum_{m,\mu}\hat{a}^{\dagger}_{m,\mu_1}\hat{a}^{\dagger}_{m,\mu_2}\hat{a}_{m,\mu_3}\hat{a}_{m,\mu_4}\delta_{\mu_1+\mu_2,\mu_3+\mu_4} \quad (2)$$

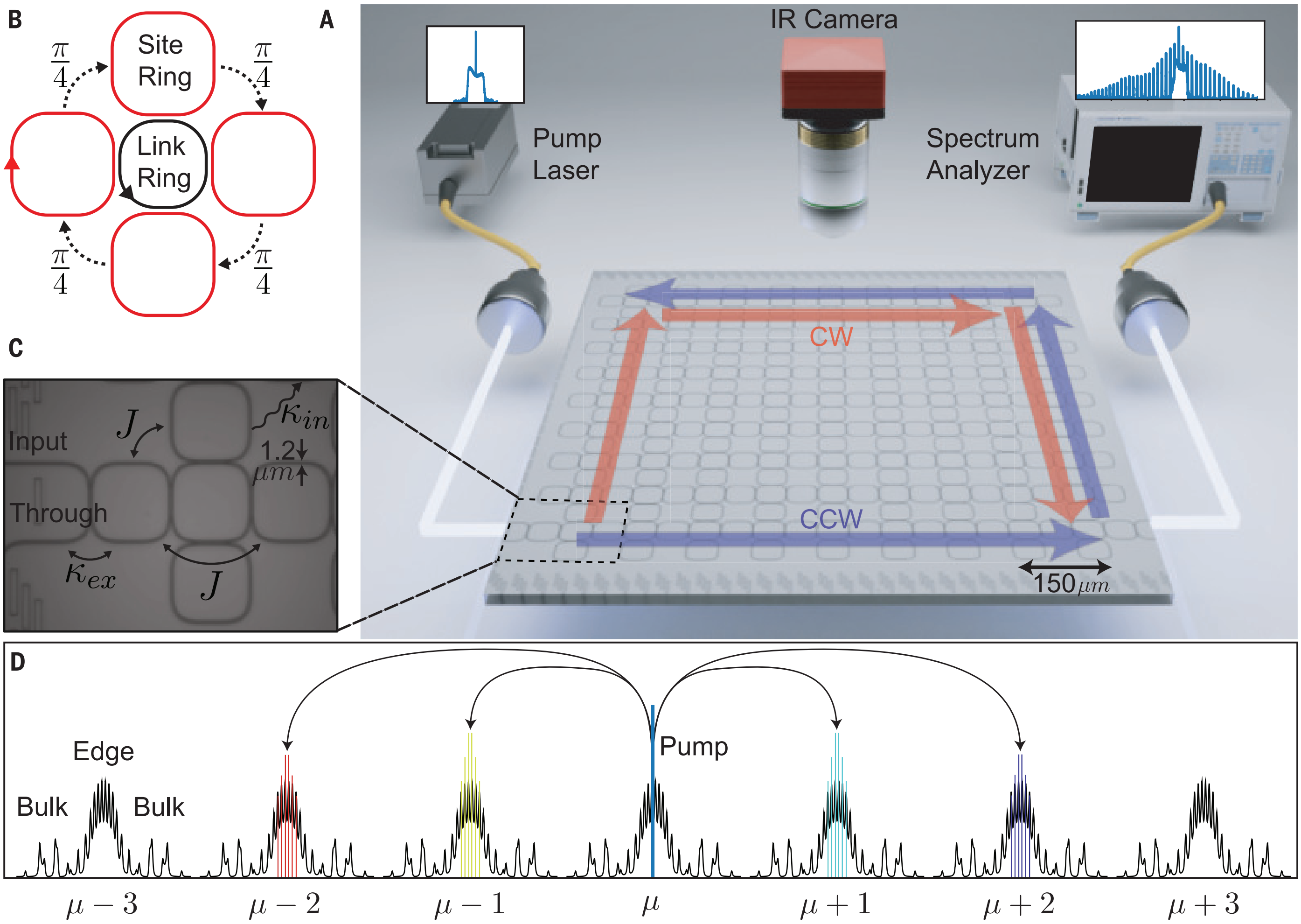


Fig. 1. Generation of the topological frequency comb. (**A**) Schematic of the pump, the spectral measurement setup, and an optical image of the device. A tunable pump is coupled into the lattice at the input port and circulates around the edge of the 2D AQH SiN lattice. The generated topological frequency comb spectra are collected from the drop port and analyzed with a spectrum analyzer. The paths followed by CW and CCW edge modes are highlighted in red and blue, respectively. (**B**) Photons acquire a nonzero phase $\pi/4$ when they hop to an adjacent site ring (red) via a link ring (black). (**C**) Close-up high-resolution optical image of the fabricated AQH lattice. Parameters and the input and through ports are marked. (**D**) Simulated linear transmission of the device with the four-wave mixing process schematically depicted.

where β is the interaction strength between photons [see the supplementary materials (SM) for definition]. Finally, we note that our lattice is a driven-dissipative system. The dissipation includes both the intrinsic decay rate κ_{in} in each individual site and the extrinsic decay rate κ_{ex} introduced by the coupling between input or output rings and probe waveguides, as shown Fig. 1. The nonlinear dynamics of this coupled resonator system is described by a modified Lugiato-Lefever formalism that predicts the formation of nested optical frequency combs (*44*). In particular, pumping the lattice in one of the edge bands leads to efficient light generation only in the edge bands centered around other resonances $\omega_{0,\mu}$. This is because the spatial confinement of the edge modes ensures excellent spatial overlap between different edge modes while minimizing overlap between edge and bulk modes. Furthermore, the linear dispersion of the edge modes ensures that the anomalous dispersion from the waveguides is the dominant contribution, as is typical with single-ring combs (*40*, *44*).

Device fabrication

This model is experimentally realized in a thick SiN platform patterned through deep-ultraviolet lithography in a commercial foundry (*47*). The device itself consists of a 2D array of 261 coupled photonic ring resonators with two coupled input-output waveguides. A high-resolution optical image in Fig. 1 shows the topological photonic lattice used in this work. The waveguides are embedded in silicon dioxide, and their dimensions are chosen to be 1200 nm wide by 800 nm thick in order to operate in the anomalous dispersion regime. Simulated mode profiles and dispersion can be found in fig. S2 (*48*). Each ring is a racetrack design, composed of 12-μm straight coupling regions and 90° Euler bend regions with a 20 μm effective radius, giving rise to an FSR of ≈0.75 THz. We specifically use Euler bends as opposed to round bends to reduce the mode mixing that occurs at the straight-bent interfaces within each ring (*49*). Although this mode mixing and its impact in perturbing dispersion has also been of concern for single-racetrack combs (*50*), its impact can be physically distinct in our coupled resonator lattice. In particular, spurious hopping phases can be generated through mode conversion during hopping between adjacent rings (*51*). We also note that previous implementations of such topological devices have avoided this problem by operating in the single-mode regime (*46*). The constraints

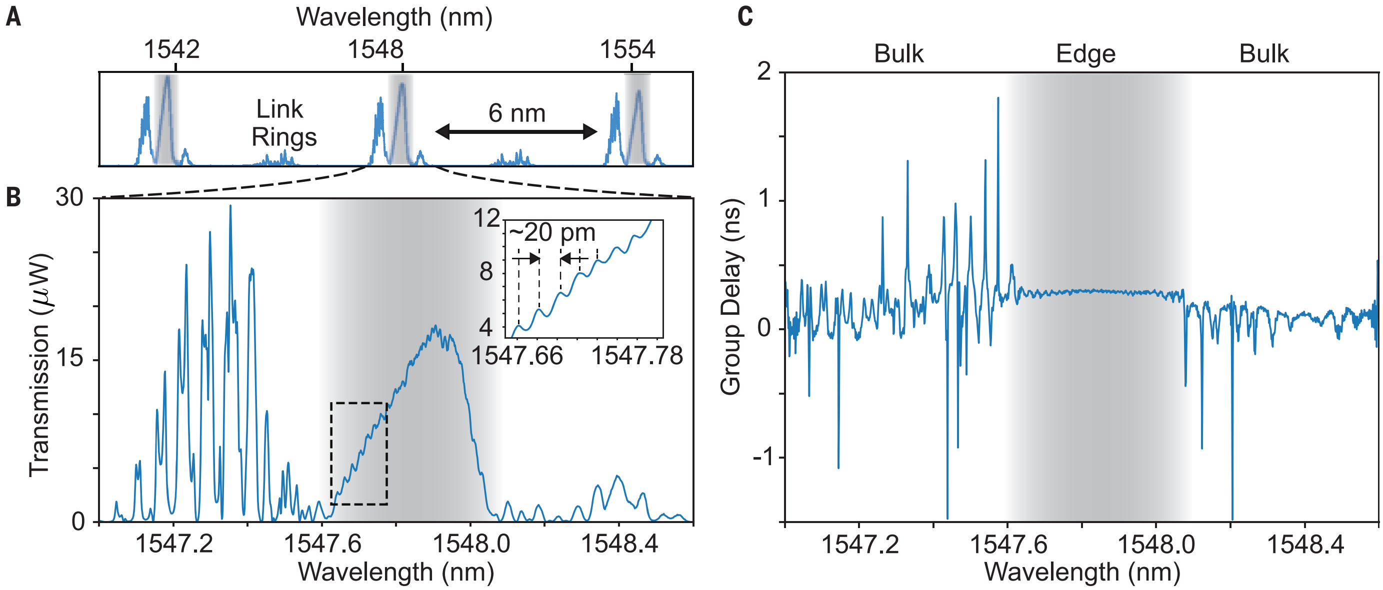


Fig. 2. Experimental characterization of the topological lattice. (**A**) Measured drop transmission spectrum of the topological lattice showing bulk and edge bands for three longitudinal modes and detuned link-ring resonances. (**B**) Zoomed drop spectrum of the topological lattice on one set of edge and bulk bands. (**C**) The group delay spectrum showing a flat edge band.

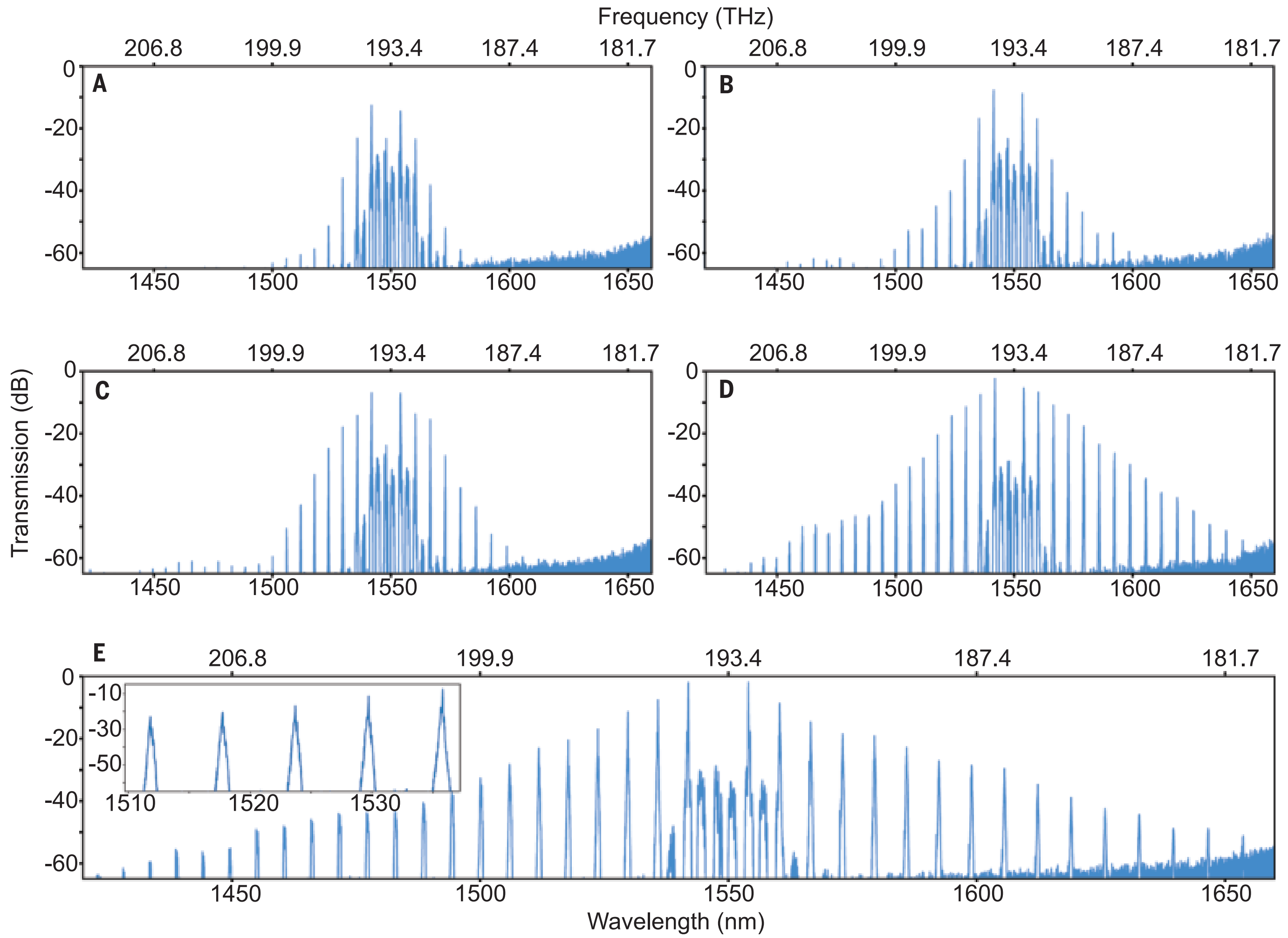


Fig. 3. Formation of the topological frequency comb. (**A** to **E**) Comb spectra measured with a pump laser wavelength of 1547.97 nm and peak powers of approximately 70, 78, 86, 92, and 100 W, respectively. The inset in (E) shows a zoomed spectrum of five comb teeth.

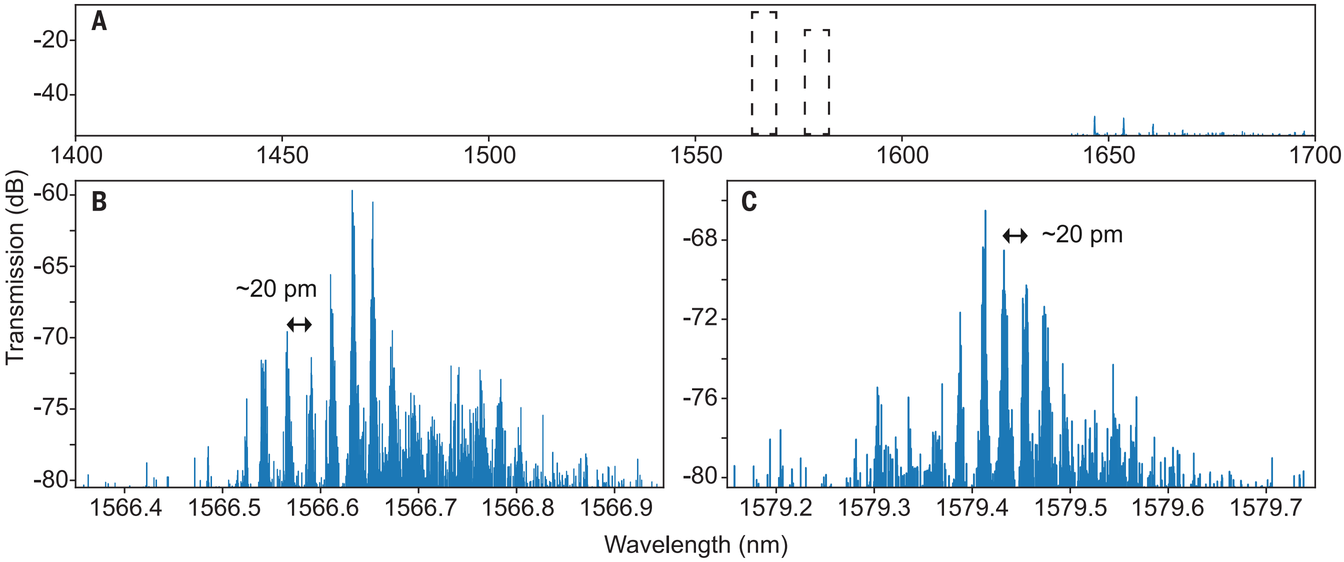


Fig. 4. High-resolution spectra of individual comb teeth. (**A**) Low-resolution, broadband comb spectrum with a pump laser wavelength of 1547.97 and on-chip peak power of ~85 W. Individual comb teeth selected for high-resolution analysis are indicated with dashed lines. (**B** and **C**) High-resolution spectra of individual comb teeth showing a nested substructure around 1566.6 and 1579.4 nm, respectively.

placed on waveguide geometry to access anomalous dispersion necessitate the use of wider waveguides that support higher-order modes.

The coupling gaps between the resonators, as well as those between the input-output waveguides and the resonators, are 300 nm, corresponding to an approximate value of $2\pi \times 25$ GHz for the coupling strength, J. The extrinsic and intrinsic couplings (κ_{ex} and κ_{in}) are estimated to be $2\pi \times 30$ and $2\pi \times 2$ GHz, respectively (fig. S3). For details on these calculations, see the SM.

Linear measurements

We begin by characterizing the transmission spectrum of the device in the linear (low-power) regime, where Kerr phase shifts are negligible. Figure 2, A and B, shows the measured drop port transmission spectrum of the device over three longitudinal modes of the site rings, as well as a single higher-resolution spectrum across one transmission band. The edge bands are shaded in gray, and the spacing of the individual edge modes is highlighted in a zoomed inset. While the topological edge states are robust against disorder, the bulk states are prone to reduced transmission. We also note that while the individual edge state resonances in our experiment are not well resolved given that the lattice is strongly coupled to input-output waveguides, the edge mode splitting can be approximated as 20 pm. This is in agreement with the estimated edge bandwidth of the device divided by the number of individual edge modes.

To highlight the linear dispersion of the edge states, we measure the group delay through the lattice (Fig. 2C) using an optical vector analyzer. As expected, the linear dispersion of the edge states leads to a flat group delay response in the edge band, indicating that the dispersion of the super-modes is small, and therefore, the device dispersion is dominated by the single-ring dispersion. The group delay through the bulk states, which do not have a well-defined momentum, shows prominent variations throughout the bulk band.

Nonlinear measurements

To observe the formation of topological frequency combs in the ring resonator array, we pump the array using a 5-ns pulsed laser with a repetition rate of 250 kHz and on-chip peak powers up to ≈100 W. We specifically choose a long pulse laser with a low duty cycle so that we can achieve a high peak power while keeping the average power low enough to avoid serious thermal effects. Furthermore, the 5-ns pulse duration is longer than any relevant timescale of the lattice dynamics, including the round-trip time in the super-ring resonator (≈400 ps). In other words, the longer pulse duration facilitates a selective quasi-continuous wave excitation of the edge band. We note that operating in this regime is necessary given the particular challenges of fabricating a resonant structure of this scale. See the SM for additional details of these challenges, the nonlinear measurement setup (fig. S4), and the pump laser spectrum (figs. S5 and S6).

We pump the system at the edge band and show the emergence of the topological frequency comb as a function of increasing pump power. In particular, the drop port spectra displayed in Fig. 3 were taken with a pump wavelength of 1547.97 nm and on-chip peak pump powers of 70, 78, 86, 92, and 100 W. We estimate the threshold peak pump power to be ≈70 W, but we also note that the threshold power changes with the pump wavelength. The full comb bandwidth at the highest pump power is ~250 nm wide with about 65 dB contrast from the most prominent sidebands to the noise floor of the measurement. The inset of Fig. 3E shows a zoomed region of the spectrum around 1524 nm, spanning five longitudinal modes. The observed FSR of the comb is ~6 nm, in agreement with the single-ring FSR. For the broadest comb, the contrast between the height of the pump laser and the most prominent sideband (shown in fig. S7) is ~2.9 dB.

To show the nested structure of the topological frequency comb within each comb tooth, we measure the comb output at the through port using an ultra-high-resolution (0.04 pm) heterodyne-based optical spectrum analyzer. A reference low-resolution spectrum is shown in Fig. 4A. We select two individual comb teeth, as indicated, for high-resolution analysis. Within each of these comb teeth, we observe the oscillation of another set of well-resolved modes that correspond to the individual edge modes (Fig. 4, B and C). The spacing between the oscillating edge modes is about 20 pm, which corresponds to the FSR of the super-ring formed by the edge states and agrees with linear measurements. The linewidths of the individual edge modes vary in the range of 3 to 5 pm. For comparison to bulk and single-racetrack comb spectra that lack this nested structure, see fig. S8.

To show that the topological frequency comb inherits the topological properties of the linear system and is indeed confined to the boundary of the lattice, we perform direct imaging of the generated comb. Although the system is designed to confine light in-plane, there is a certain

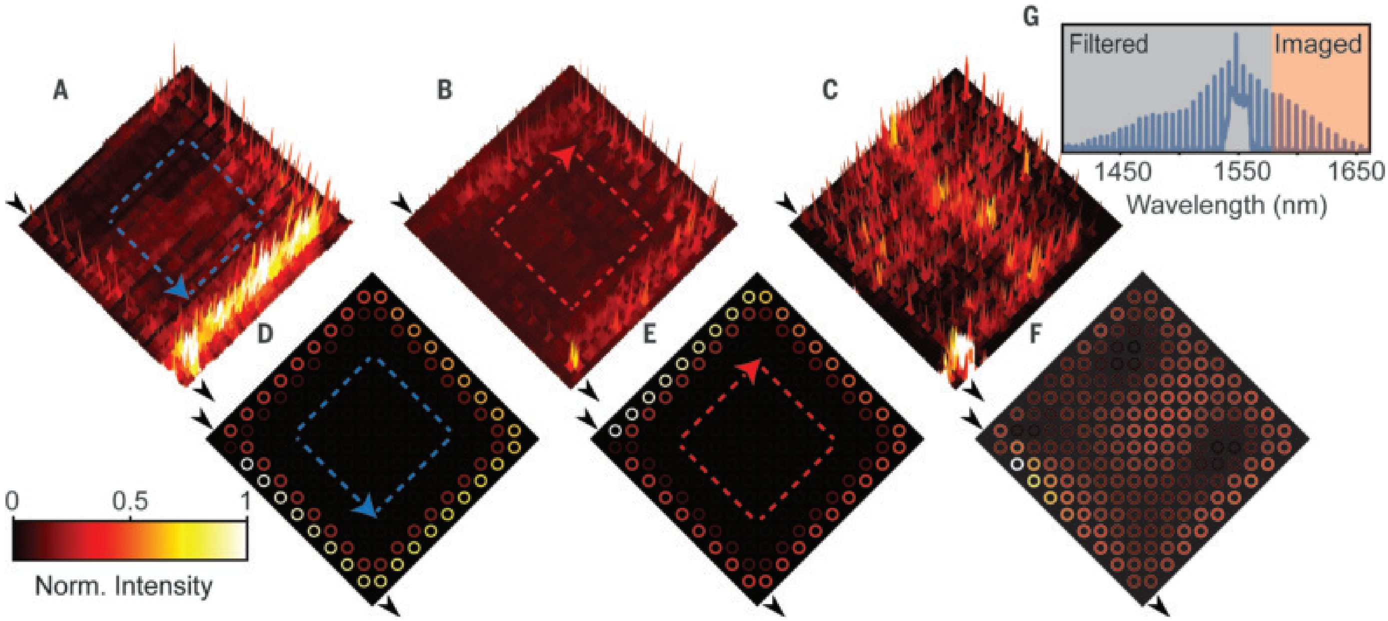


Fig. 5. Spatial imaging of the topological frequency comb. (**A** to **C**) Measured spatial imaging of the topological frequency comb for CCW, CW, and bulk modes, respectively. (**D** to **F**) Simulated spatial profile of the CCW, CW, and bulk modes in the linear regime, respectively. (**G**) Integration bandwidth used for the top-down imaging.

amount of out-of-plane scattering caused by fabrication imperfections and disorder. The light scattered as a result of surface roughness is collected from above with a 10× objective lens and imaged on an infrared (IR) camera. In addition, we use a 1580-nm long-pass filter to remove the pump and only collect part of the generated comb light.

Figure 5 shows the measured spatial intensity profile of three types of generated frequency combs. First, we observe that the generated comb light is confined to the edge of the lattice and that light travels from the input to the output port in the CCW direction. Note that the lattice is not critically coupled to the bus waveguides, therefore the light continues to circulate around the lattice after reaching the first output port in its path. Moreover, the propagation is robust, and no noticeable scattering into the bulk is observed from the two sharp 90° corners. These characteristics show that the comb teeth are indeed generated within the topological edge band and that the topology is preserved even in the presence of strong nonlinearity.

Next, by pumping the system in the other pseudospin, we generate the comb in the CW edge state. We observe similar confinement of the topological frequency comb, but here the light travels in the opposite direction around the lattice, as expected.

In sharp contrast to the CW and CCW edge band excitation, when we excite the lattice in the bulk band, the spatial intensity distribution of generated frequencies exhibits no confinement and occupies the bulk of the lattice. These images represent a novel look into the spatial profile of frequency comb formation, enabled by the distinctive geometry and scale of the topological lattice. For the spectra (fig. S9) and details on the generation of each of these types of frequency combs, see the SM. Additionally, Fig. 5 shows simulated spatial distributions of CCW, CW, and bulk modes in the linear regime for comparison, as well as a schematic illustrating the filtered and imaged regions of the spectrum. We note that in these linear simulation results, we observe a uniform decay in intensity due to propagation loss from the input to the output port. In contrast, our experimentally observed intensity profiles do not show a uniform decay, likely because of competition between the linear loss and the nonlinear parametric gain. For a comparison with nonlinear simulation results, see fig. S10.

Outlook

Here we have demonstrated the first topological frequency comb using an array of >100 coupled resonators. Our results entail the first realization of a new class of frequency combs that also includes coherent dissipative solutions, such as nested solitons and phase-locked Turing rolls, that are not accessible using single resonators (*44*). The distinctive nested spectral structure of these combs, characterized by two disparate frequency scales, could lead to a host of new applications. For instance, this nested structure could be useful in certain spectroscopic measurements where there are multiple regions of interest that each require a high-resolution analysis but are separated by a large frequency gap. Moreover, in this work, we have used a commercially available SiN platform in order to operate in the telecom wavelength regime. However, our device design can be easily translated to other frequency domains and photonic material platforms that can exhibit much higher nonlinearities, such as aluminum gallium arsenide (*52*, *53*) and lithium niobate (*54*).

On a more fundamental level, our results provide a new platform to study the interplay of topology and optical nonlinearities (*21*, *55*, *56*), as well as intriguing topological physics specific to bosons. For example, although optical nonlinearities have been used to demonstrate topological phase transitions and restructured bulk-edge correspondence (*57*, *58*), in this work we observe that the system retains the topological behavior of its linear counterpart even in the presence of such strong nonlinear effects. These results could enable novel applications where topological physics is used to engineer the underlying band structure (or dispersion) of a linear system and optical nonlinearities provide additional functionalities.

REFERENCES AND NOTES

1. T. J. Kippenberg, R. Holzwarth, S. A. Diddams, *Science* **332**, 555–559 (2011).
2. A. Pasquazi *et al.*, *Phys. Rep.* **729**, 1–81 (2018).
3. A. L. Gaeta, M. Lipson, T. J. Kippenberg, *Nat. Photonics* **13**, 158–169 (2019).
4. S. A. Diddams, K. Vahala, T. Udem, *Science* **369**, eaay3676 (2020).
5. M.-G. Suh, Q.-F. Yang, K. Y. Yang, X. Yi, K. J. Vahala, *Science* **354**, 600–603 (2016).
6. D. A. Long *et al.*, *Nat. Photonics* **18**, 127–131 (2024).
7. J. Riemensberger *et al.*, *Nature* **581**, 164–170 (2020).
8. R. Chen *et al.*, *Nat. Photonics* **17**, 306–314 (2023).
9. S. A. Miller *et al.*, *Opt. Express* **23**, 21527–21540 (2015).

10. S. Kim et al., Nat. Commun. 8, 372 (2017).
11. H. Bao et al., Nat. Photonics 13, 384–389 (2019).
12. X. Xue, X. Zheng, B. Zhou, Nat. Photonics 13, 616–622 (2019).
13. Z. Yuan et al., Nat. Photonics 17, 977–983 (2023).
14. J. Vasco, V. Savona, Phys. Rev. Appl. 12, 064065 (2019).
15. A. Tikan et al., Nat. Phys. 17, 604–610 (2021).
16. A. Tusnin, A. Tikan, K. Komagata, T. J. Kippenberg, Commun. Phys. 6, 317 (2023).
17. L. Lu, J. D. Joannopoulos, M. Soljačić, Nat. Photonics 8, 821–829 (2014).
18. T. Ozawa et al., Rev. Mod. Phys. 91, 015006 (2019).
19. H. Price et al., JPhys Photonics 4, 032501 (2022).
20. X. Zhang, F. Zangeneh-Nejad, Z.-G. Chen, M.-H. Lu, J. Christensen, Nature 618, 687–697 (2023).
21. D. Smirnova, D. Leykam, Y. Chong, Y. Kivshar, Appl. Phys. Rev. 7, 021306 (2020).
22. M. Jalali Mehrabad, S. Mittal, M. Hafezi, Phys. Rev. A 108, 040101 (2023).
23. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljacić, Nature 461, 772–775 (2009).
24. M. C. Rechtsman et al., Nature 496, 196–200 (2013).
25. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, Nat. Photonics 7, 1001–1005 (2013).
26. S. Mittal et al., Phys. Rev. Lett. 113, 087403 (2014).
27. S. Barik et al., Science 359, 666–668 (2018).
28. M. Jalali Mehrabad et al., Optica 10, 415 (2023).
29. S. Guddala et al., Science 374, 225–227 (2021).
30. J. Guglielmon, M. C. Rechtsman, Phys. Rev. Lett. 122, 153904 (2019).
31. M. I. Shalaev, W. Walasik, A. Tsukernik, Y. Xu, N. M. Litchinitser, Nat. Nanotechnol. 14, 31–34 (2019).
32. C. J. Flower et al., ACS Photonics 10, 3502–3507 (2023).
33. H. Zhao et al., Science 365, 1163–1166 (2019).
34. P. St-Jean et al., Nat. Photonics 11, 651–656 (2017).
35. B. Bahari et al., Science 358, 636–640 (2017).
36. M. A. Bandres et al., Science 359, eaar4005 (2018).
37. L. Yang, G. Li, X. Gao, L. Lu, Nat. Photonics 16, 279–283 (2022).
38. V. Peano, M. Houde, F. Marquardt, A. A. Clerk, Phys. Rev. X 6, 041026 (2016).
39. B.-U. Sohn et al., Nat. Commun. 13, 7218 (2022).
40. S. Mittal, E. A. Goldschmidt, M. Hafezi, Nature 561, 502–506 (2018).
41. A. Blanco-Redondo, B. Bell, D. Oren, B. J. Eggleton, M. Segev, Science 362, 568–571 (2018).
42. S. Mittal, V. V. Orre, E. A. Goldschmidt, M. Hafezi, Nat. Photonics 15, 542–548 (2021).
43. T. Dai et al., Nat. Photonics 16, 248–257 (2022).
44. S. Mittal, G. Moille, K. Srinivasan, Y. K. Chembo, M. Hafezi, Nat. Phys. 17, 1169–1176 (2021).
45. D. Leykam, S. Mittal, M. Hafezi, Y. D. Chong, Phys. Rev. Lett. 121, 023901 (2018).
46. S. Mittal, V. V. Orre, D. Leykam, Y. D. Chong, M. Hafezi, Phys. Rev. Lett. 123, 043201 (2019).
47. A. Rahim et al., IEEE J. Sel. Top. Quantum Electron. 25, 1–18 (2019).
48. K. Luke, Y. Okawachi, M. R. E. Lamont, A. L. Gaeta, M. Lipson, Opt. Lett. 40, 4823–4826 (2015).
49. T. Fujisawa, S. Makino, T. Sato, K. Saitoh, Opt. Express 25, 9150–9159 (2017).
50. X. Ji et al., Commun. Phys. 5, 84 (2022).
51. L. D. Tzuang, K. Fang, P. Nussenzveig, S. Fan, M. Lipson, Nat. Photonics 8, 701–705 (2014).
52. L. Chang et al., Nat. Commun. 11, 1331 (2020).
53. M. Pu, L. Ottaviano, E. Semenova, K. Yvind, Optica 3, 823 (2016).
54. M. Zhang et al., Nature 568, 373–377 (2019).
55. M. Jürgensen, S. Mukherjee, M. C. Rechtsman, Nature 596, 63–67 (2021).
56. N. Mostaan, F. Grusdt, N. Goldman, Nat. Commun. 13, 5997 (2022).
57. L. J. Maczewsky et al., Science 370, 701–704 (2020).
58. M. S. Kirsch et al., Nat. Phys. 17, 995–1000 (2021).
59. C. J. Flower et al., Observation of Topological Frequency Combs [Dataset], Dryad (2024); https://doi.org/10.5061/dryad.xwdbrv1mw.

ACKNOWLEDGMENTS

The authors acknowledge fruitful discussions with E. Goldschmidt, A. Dutt, R. Vasanth, D. Session, and E. Mechtel. **Funding:** This work was supported by Airforce Office of Scientific Research grant FA9550-22-1-0339, Office of Naval Research grant N00014-20-1-2325, Army Research Lab grant W911NF1920181, National Science Foundation grant DMR-2019444, and the Minta Martin and Simons Foundations. **Author contributions:** C.J.F. designed the devices and constructed the experimental setup. C.J.F. performed the measurements, with assistance from L.X. and M.J.M. C.J.F. and L.X. performed the simulations. C.J.F., L.X., and M.J.M. analyzed the data. All work was supervised by G.M., K.S., S.M., and M.H. All authors discussed the results and contributed to the manuscript. **Competing interests:** S.M. and M.H. are inventors on a patent (US Patent 11599006) dated 7 March 2023 that covers the generation of nested frequency combs in a topological source. The authors declare that they have no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the figures of the paper and/or the supplementary materials. The data can also be accessed at Dryad (59).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ado0053
Supplementary Text
Figs. S1 to S10
Reference (60)

Submitted 11 January 2024; accepted 9 May 2024
10.1126/science.ado0053

NEUROSCIENCE

Top-down brain circuits for operant bradycardia

Airi Yoshimoto[1], Shota Morikawa[2], Eriko Kato[1], Haruki Takeuchi[2], Yuji Ikegaya[1,3,4]*

Heart rate (HR) can be voluntarily regulated when individuals receive real-time feedback. In a rat model of HR biofeedback, the neocortex and medial forebrain bundle were stimulated as feedback and reward, respectively. The rats reduced their HR within 30 minutes, achieving a reduction of approximately 50% after 5 days of 3-hour feedback. The reduced HR persisted for at least 10 days after training while the rats exhibited anxiolytic behavior and an elevation in blood erythrocyte count. This bradycardia was prevented by inactivating anterior cingulate cortical (ACC) neurons projecting to the ventromedial thalamic nucleus (VMT). Theta-rhythm stimulation of the ACC-to-VMT pathway replicated the bradycardia. VMT neurons projected to the dorsomedial hypothalamus (DMH) and DMH neurons projected to the nucleus ambiguus, which innervates parasympathetic neurons in the heart.

Physiological parameters such as heart rate (HR), blood pressure, and body temperature, which are predominantly controlled by the autonomic nervous system, can be intentionally modulated through specialized training that provides real-time feedback to the individual (*1*). Despite its potential for a wide range of clinical applications, the neural basis underlying brain-to-organ control during biofeedback remains poorly understood. Inspired by a previous investigation (*2*), we developed an experimental model of HR feedback using freely behaving rats to elucidate the neural mechanisms that govern HR feedback training and investigated the underlying neurophysiological activity.

Operant HR control through HR feedback to the rat brain

Our biofeedback training paradigm was designed to reduce HR in rats (Fig. 1A and movie S1). Electrocardiograms (ECGs) were monitored from the pectoralis major muscle (*3*) and the mean HR during a 1-s window was calculated every 100 ms from the RR intervals (the time between two consecutive R waves). The target HR to be achieved by the rats was set to be 15% lower than the mean baseline HR for 1 min before HR feedback training. Rats were informed every 2 s of the difference between the current HR and the target HR by the number of stimulation pulses delivered to the primary somatosensory (S1) cortex in the right or left hemisphere. Specifically, the higher the current HR was above the baseline HR, the more pulses were delivered to the right barrel cortex, whereas the lower the HR was below the baseline, the more pulses were delivered to the left barrel cortex (up to five pulses per stimulation). If rats reduced their HR below the target HR, the medial forebrain bundle (MFB) was stimulated as a neural reward (*4*). When rats received a total of 10 rewards, target HRs were updated to be 15% lower than the mean HR for the preceding 1 min so that rats continuously needed to reduce HR. This training was performed for 3 hours each day and was repeated for 5 consecutive days (Fig. 1B). The baseline HR before training on day 1 was 428 ± 21 beats per minute (bpm) (mean ± SD of 17 rats) and all rats tested were able to start lowering their HR within the first 30 min of training on day 1 (Fig. 1C). The HR dropped to 231 ± 14 bpm at the end of training on day 5, representing a 46.0 ± 4.8% decrease from the initial baseline on day 1 (Fig. 1D; note that detailed statistics are summarized in table S1). The MFB stimulation alone induced no apparent change in HR (fig. S1, A and B), and the barrel cortex stimulation did not alter whisker movement (fig. S1, C and D).

[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo 113-0033, Japan. [2]Graduate School of Science, The University of Tokyo, Tokyo 113-0032, Japan. [3]Institute for AI and Beyond, The University of Tokyo, Tokyo 113-0033, Japan. [4]Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita City, Osaka 565-0871, Japan.
*Corresponding author. Email: yuji@ikegaya.jp

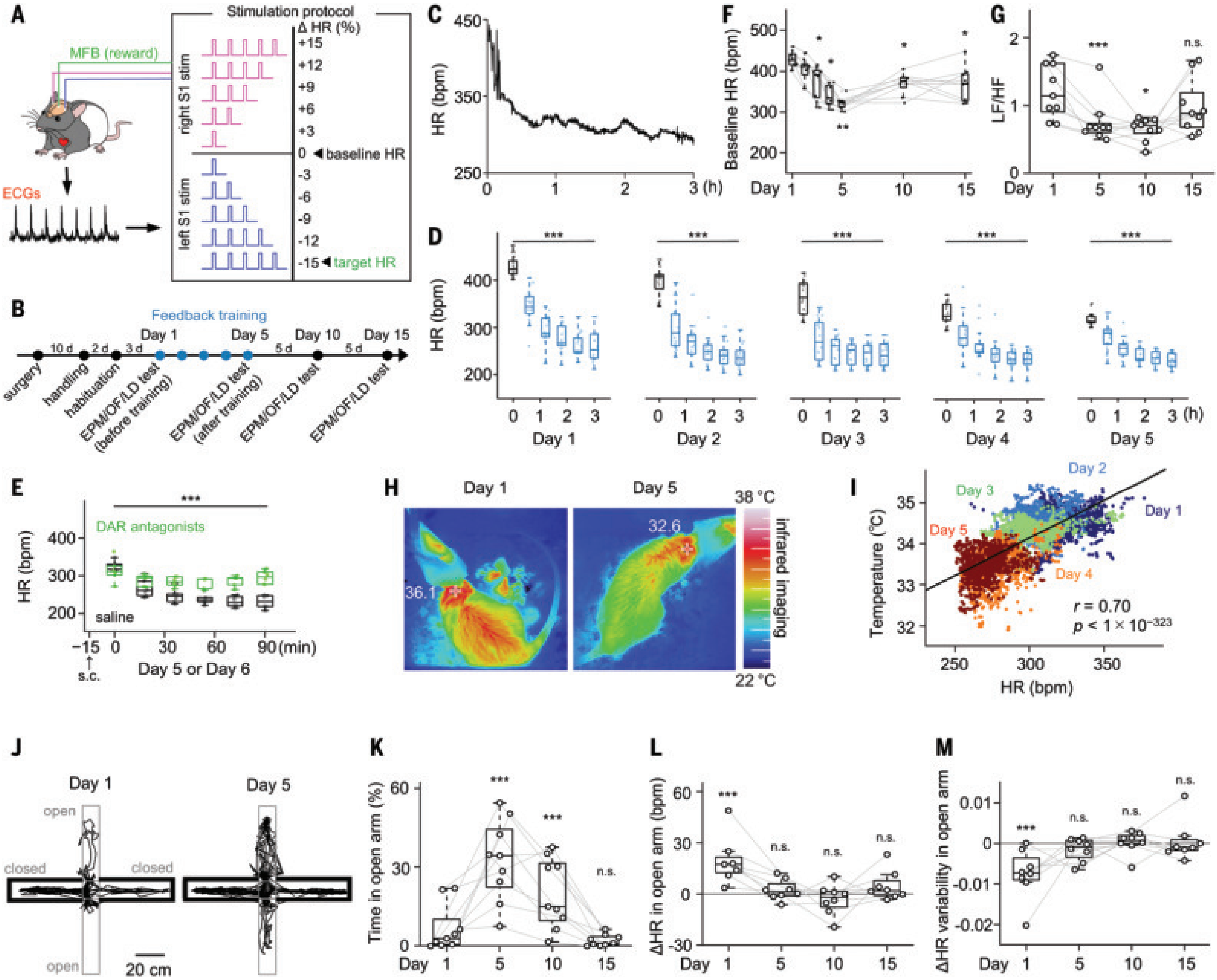


Fig. 1. Closed-loop HR feedback to the rat brain enables operant HR control. (**A** and **B**) Experimental design (A) and schedule (B) for HR feedback training and behavioral tasks. HR feedback training was conducted for 5 days while the elevated plus maze (EPM), open field (OF), and light-dark box (LD) tests were performed before training on day 1, after training on day 5, and on days 10 and 15. (**C**) Time course of the HR during feedback training in a rat on day 1. (**D**) Time course of the HR during feedback training for 5 days collected from 17 rats. Each dot indicates data from a single rat. Baseline HR before the start of training on each day is plotted in black and HR during training is plotted in blue. ***$P < 0.0001$, Jonckheere-Terpstra test for each day. Detailed statistics are summarized in table S1. (**E**) HR of rats during training with subcutaneous administration of saline or dopamine receptor antagonists (0.03 mg/kg SCH-23390 and 0.03 mg/kg eticlopride, 1 ml/kg, s.c.). $F_{1,70} = 24.01$, ***$P = 5.97 \times 10^{-6}$, $n = 6$ rats, one-way analysis of variance (ANOVA) with Bonferroni correction. (**F**) The pretraining baseline HR decreased gradually during the training period of days 1 to 5 and bradycardia persisted in home cages for at least 10 days after the training period (days 5 to 15). Each dot indicates a single rat. *$P < 0.05$, **$P < 0.001$ versus day 1, $n = 9$ rats, multiple correspondence bootstrap test with Bonferroni correction. (**G**) Same as (F) but for LF/HF. *$P < 0.05$, ***$P < 0.0001$ versus day 1. (**H**) Infrared thermal images of body surface temperature at the beginning of training on day 1 and at the end of training on day 5. (**I**) Positive correlation between HR and body temperature during HR feedback training. The black line represents the regression line with least squares regression for 7 rats. $F_{1,5403} = 71.85$, $r = 0.70$, $P < 1 \times 10^{-323}$, t test of the correlation coefficient. (**J**) Trajectories of a representative rat exploring in an elevated plus maze on days 1 and 5. (**K**) Total time spent in the open arms. ***$P < 0.0001$ versus day 1, $n = 9$ rats, multiple correspondence bootstrap test with Bonferroni correction. (**L** and **M**) Same as (K) but for the changes in HR (L) and HR variability (M) in the open arms. ***$P < 0.0001$ versus 0, $n = 8$ rats, bootstrap test.

HR feedback to the primary visual cortex also induced a reduction in HR (fig. S2, A and D; $n = 6$ rats). However, rats were less efficient at reducing their HR when they received only the reward of MFB stimulation without HR feedback to the brain (fig. S2, B and D; $n = 6$ rats). In addition, rats were unable to lower their HR when the barrel cortex and MFB were stimulated at randomized times, even when the total numbers of stimuli matched those in HR feedback experiments (here referred to as the random S1/MFB group; fig. S2, C and D; $n = 5$ rats). Thus, appropriate HR feedback stimulation to the brain was required for top-down control of bradycardia. Rats were treated subcutaneously with saline or a cocktail of dopamine receptor antagonists (SCH-23390 and eticlopride) in a randomized order on either day 5 or 6 of training. The treatment attenuated the effect of training (Fig. 1E, $n = 6$ rats).

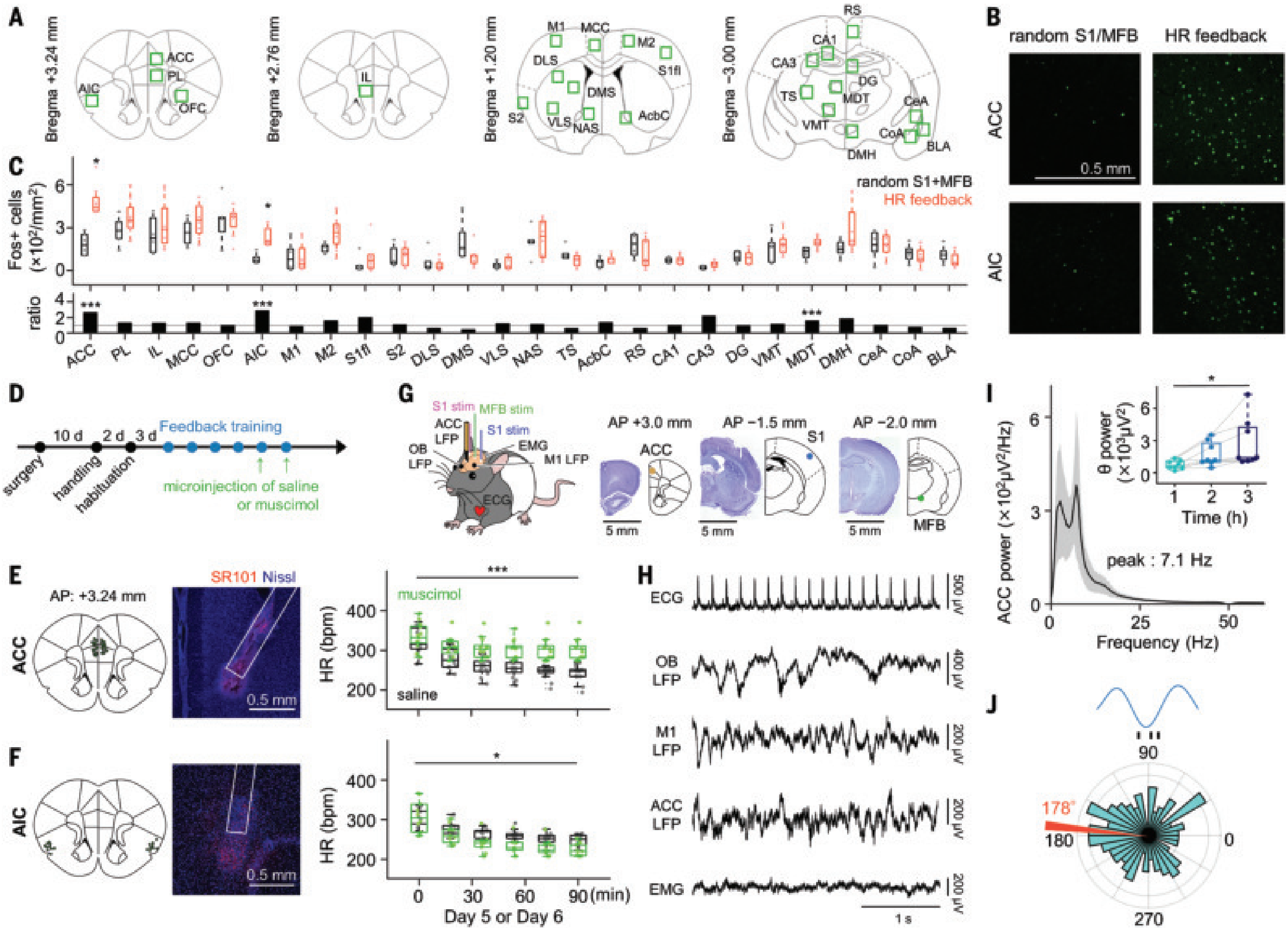


Fig. 2. HR feedback-induced bradycardia requires ACC neuronal activity. (**A**) Regions in which data are plotted in (C) are indicated in the schematic coronal sections. See table S2 for abbreviations. (**B**) Confocal images of c-Fos immunostaining in the ACC and the AIC after HR feedback training on day 5. In the control random S1/MFB group, the same numbers of stimuli were randomly applied to the barrel cortex and MFB, independent of HR. (**C**) The densities of c-Fos-positive cells in the HR feedback and random S1/MFB stimulation (top) and their ratio (bottom) are plotted for each brain region. $^{*}P < 0.05$, multiple correspondence bootstrap test; $^{***}P < 0.0001$, bootstrap test with Bonferroni correction, $n = 6$ (random S1/MFB) and 8 (HR feedback) regions of interest. (**D**) Experimental schedule for neuronal inactivation. Each rat serially received 0.05 μg muscimol or saline in a random order on either day 5 or 6. (**E**) (Left) Brain map and photograph of the identified locations of the cannula tips in the ACC. SR-101 was coinjected with drugs and its fluorescence was confirmed post hoc. (Right) HR of rats during training that were treated with muscimol (green) or saline (black) in the ACC. $F_{1,154} = 62.92$, $^{***}P = 4.10 \times 10^{-13}$, $n = 13$ rats, one-way ANOVA with Bonferroni correction. (**F**) Same as (E) but for the AIC. $F_{1,118} = 9.55$, $^{*}P = 2.49 \times 10^{-3}$, $n = 10$ rats. (**G**) Schematic diagram for LFP recordings from the ACC, M1, and olfactory bulb (OB), together with ECG and electromyogram (EMG) data. Nissl-stained section images on the right and the corresponding brain maps represent post hoc confirmation of the locations of the electrode tracks in the ACC, S1, and MFB. (**H**) Traces of ECG, LFP, and EMG data. (**I**) Fourier power spectrum of ACC LFP during operant training. The data are the mean (line) ± standard deviation (SD) (gray area) of 5 days in 8 rats and exhibited the largest peak at 7.1 Hz. The mean theta (7 to 8 Hz) power increased during training. $^{*}P = 2.14 \times 10^{-3}$, $JT = 2.86$, $n = 8$ rats, Jonckheere-Terpstra test. (**J**) Polar histogram of spikes from putative excitatory ACC neurons along the phases of the LFP theta cycle (7 to 8 Hz) during training. The spike distribution was skewed toward the mean of 176° (arrowhead). (Top) spike times along theta cycles. $P = 0.038$, $Z = 3.27$, $n = 17$ cells from 9 rats, Rayleigh test for uniformity.

Under isoflurane anesthesia, rats did not exhibit HR feedback-induced bradycardia (fig. S3A, $n = 8$ rats). Rats very occasionally fell asleep during training, after which the reduced HR reverted toward baseline levels (fig. S3B, $n = 4$ rats). The effect of sleep was not quantified due to the inability to experimentally regulate the timing, duration, and depth of spontaneous sleep events. However, the observation suggests that HR feedback training induced more pronounced bradycardia than that which naturally occurs during sleep. Biofeedback-induced bradycardia did not coincide with a significant decrease in respiratory rate (fig. S4).

Long-lasting systemic effects of operant bradycardia

The baseline HR before each day of training gradually decreased over the 5-day training period (Fig. 1D black box plots, $n = 17$ rats), indicating that bradycardia occurs chronically as well as acutely. The gradual increase in HR variability over 5 days of training also suggests chronic parasympathetic dominance (fig. S2E). We monitored the HR of rats in their home cages for 24 hours after training on day 5, during which the HR remained lower than the initial baseline HR before training on day 1, even in the absence of MFB stimuli (fig. S5, A and B, n = 6 rats). The prolonged bradycardia was still observed on days 10 and 15 (Fig. 1F, n = 9 rats) and was accompanied by a prolonged decrease in blood pressure (fig. S5C). The power ratio of the low-frequency (LF) (0.2 to 0.8 Hz) to high-frequency (HF) (0.8 to 3.5 Hz) band in the HR variability spectrogram (LF/HF), an index of autonomic activity (*5*), was significantly lower than the pretraining level on day 1 until day 10 (Fig. 1G). Chronic bradycardia cannot be explained by cardiac dysfunction as echocardiographic and histopathologic inspections revealed no apparent abnormalities in cardiac motion or myocardial cells (fig. S5, D and E).

Using infrared thermography to monitor changes in body temperature we observed a decrease in body temperature during the later phase of HR feedback training, and the minimum body temperature reached a mean of 29.7 ± 1.8°C on day 5 (Fig. 1, H and I, n = 7 rats). Moreover, blood oxygen saturation measurements showed that training on day 1 significantly decreased SpO_2 (fig. S5, F and G, n = 8 rats). However, training on day 4, despite even lower HR, did not reduce SpO_2, suggesting that prolonged bradycardia induced some compensatory improvement in cardiovascular function. As one compensatory mechanism, we observed an increase in the red blood cell count (fig. S5H, n = 8 rats). Video observation in the open field revealed that despite chronic bradycardia and hypothermia rats did not reduce their locomotor activity (fig. S6, A to C).

HR variability has been associated with anxiety (*6*, *7*). Rats underwent the elevated plus maze test before and after 5 days of feedback training to examine whether prolonged bradycardia affects anxiety-related behavior (Fig. 1B). Rats spent more time in the open arms, which are thought to trigger anxiety, on days 5 and 10 (but not day 15) compared with the pretraining levels on day 1 (Fig. 1, J and K, n = 9 rats). Prior to training on day 1, rats exhibited an increase in HR and a decrease in HR variability upon entering the open arms, indicative of increased stress (*8*), which were not observed after 5 days of training (Fig. 1, L and M). The increase in time spent in the open arms was not observed in the random S1/MFB group (fig. S6, F to H). In the open field test, rats spent more time in the central area away from the walls on day 5 (fig. S6D), and the increased HR in the central area were not observed after 5 days of training (fig. S6E). In the light-dark box test, where time spent in the light chamber is an indicator of low-stress states, rats spent more time in the light chamber after 5 days of feedback training (fig. S6I, n = 6 rats).

ACC activation during operant bradycardia

To identify brain regions involved in operant bradycardia, we searched for cells activated during HR feedback in whole-brain sections that were immunohistochemically labeled for the immediate-early gene c-Fos (Fig. 2, A and B, and fig. S7). Data from the HR feedback group were compared to those from the random S1/MFB group. In the neocortex of rats with HR feedback, two regions, namely, the ACC and the anterior insular cortex (AIC),

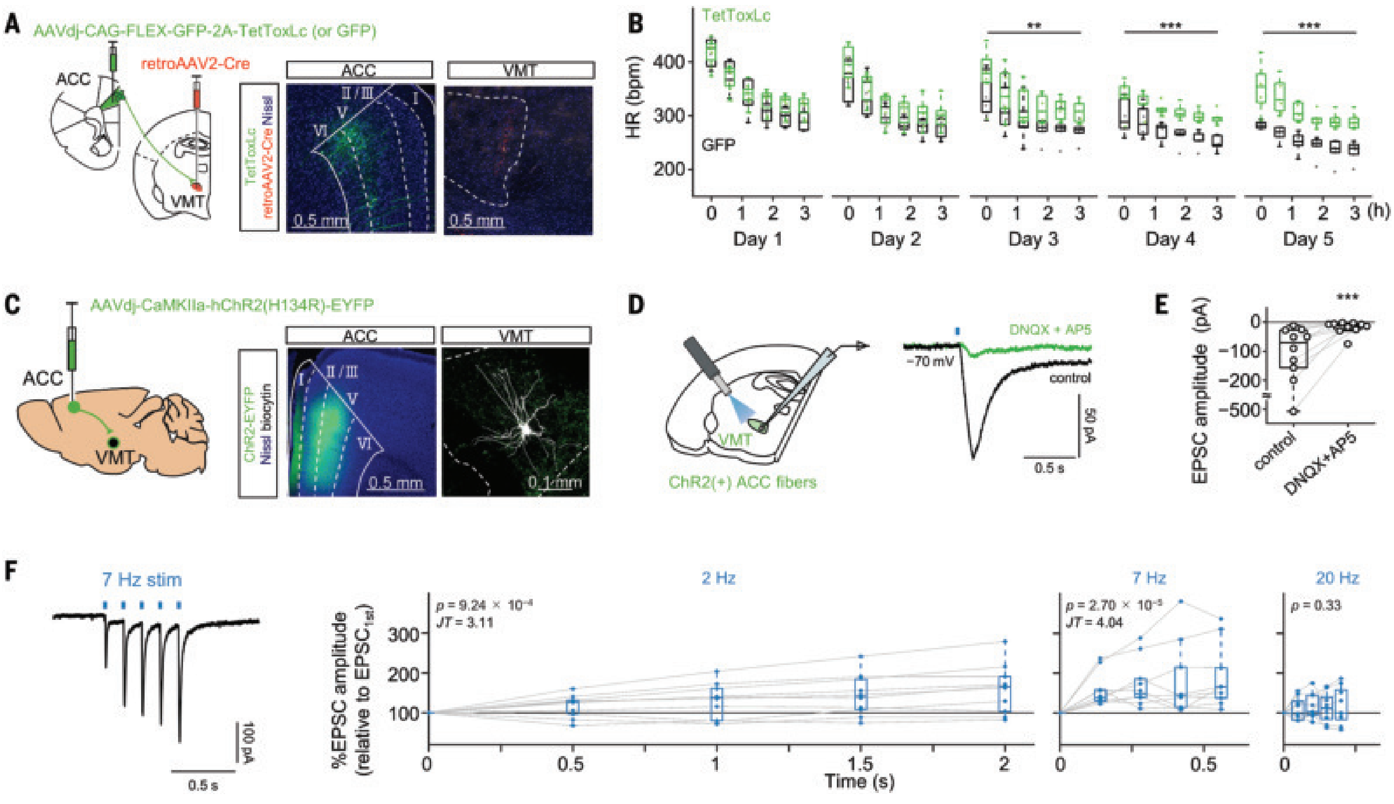


Fig. 3. HR feedback-induced bradycardia requires neurotransmission from the ACC to the VMT. (**A**) Left diagram shows viral injection of AAVdj-CAG-FLEX-GFP-2A-TetToxLc or AAVdj-CAG-FLEX-GFP into the ACC and retroAAV2-pmSyn1-EBFP-Cre into the VMT. The confocal images on the right show the injection sites in the ACC and the VMT in a rat. (**B**) Expression of TetToxLc in ACC neurons projecting to the VMT reduced the effect of HR feedback training. ***P* < 0.001, ****P* < 0.0001, n = 7 and 8 rats for TetToxLc and GFP, respectively, one-way ANOVA with Bonferroni correction. (**C**) The left schematic shows viral expression of ChR2 in the ACC. Right confocal images show the injection sites in the ACC and a patch-clamped VMT neuron. (**D**) Responses of a VMT neuron to blue light illumination to the VMT in a thalamocortical slice in the presence and absence of 10 μM DNQX and 20 μM AP5. (**E**) EPSC amplitudes are pooled from 10 neurons from 3 rats. ****P* < 0.0001, paired bootstrap test. (**F**) (Left) EPSCs recorded in response to five blue light pulses applied at 7 Hz to the VMT. Right plots indicate the relative changes in EPSC amplitudes to five pulse stimulation at different frequencies. n = 10 neurons in 3 rats for 2 Hz and 7 Hz stimulation, n = 9 neurons from 3 rats for 20 Hz stimulation, Jonckheere-Terpstra test.

exhibited a significant increase in the number of c-Fos-positive cells (Fig. 2C, n = 6 and 8 regions of interest for each brain area in random S1/MFB and HR feedback groups, respectively). To investigate whether these regions are involved in operant bradycardia, we suppressed their neuronal activity by local injection of muscimol into the bilateral ACC or AIC. We treated each rat with muscimol or saline on either day 5 or 6 of training in a randomized order (Fig. 2D). Feedback-induced bradycardia was impaired by muscimol injection into the ACC (Fig. 2E, n = 13 rats) but not by injection into the AIC (Fig. 2F, n = 10 rats). Because water immersion–induced bradycardia was not inhibited by injection of muscimol into the ACC (fig. S8, A to C), the ACC plays a role in operant bradycardia but not in physiological reflex bradycardia.

Local field potentials (LFPs) were recorded from the ACC (Fig. 2, G and H). The power spectrum exhibited theta-band oscillations, peaking at 7.1 Hz, during biofeedback training (Fig. 2I, n = 8 rats). The power of theta oscillations (7 to 8 Hz) increased with training time (Fig. 2I inset). Spikes emitted by ACC neurons were significantly phase-locked to ACC theta oscillations, with a mean phase of 178° (Fig. 2J). Theta oscillations were not evident in LFPs of the primary motor cortex or olfactory bulb recorded simultaneously with the ACC (fig. S9, A and B).

Anterograde tracing by injection of AAVdj-hSyn-GFP into the ACC revealed that GFP-positive axon fibers were abundant in the ventromedial thalamic nucleus (VMT) and mediodorsal thalamic nucleus (MDT) (fig. S10). Retrograde tracing from the VMT or MDT to the ACC using red RetroBeads revealed that RetroBeads signal and c-Fos immunoreactivity coexisted in ACC neurons in rats after 3 hours of HR feedback training (fig. S11A). Labeling VMT-projecting ACC neurons by dual infection of AAVdj-EF1a-Flex-Synaptophysin-mCherry into the ACC and retroAAV2-pmSyn1-EBFP-Cre into the VMT revealed that mCherry-positive neurons existed in layers 5 and 6 in the ACC and that mCherry-positive neurites reached the MDT, indicating that VMT-projecting layer 5/6 ACC neurons also send their axon collaterals to the MDT (fig. S11B) (*9*).

Theta oscillations of ACC-VMT glutamatergic transmission

We prevented neuronal transmission from the ACC to the VMT or the MDT using tetanus toxin light chain (TetToxLc) by bilateral injection of AAVdj-CAG-FLEX-GFP-2A-TetToxLc or AAVdj-CAG-FLEX-GFP (control) into the ACC and retroAAV2-pmSyn1-EBFP-Cre injection into the VMT or the MDT (Fig. 3A and fig. S11C). Expression of TetToxLc in VMT-projecting or MDT-projecting ACC neurons attenuated the effects of HR feedback training, especially on days 4 and 5 (VMT: Fig. 3B, n = 7 and 8 rats for TetToxLc and GFP, respectively, MDT: fig. S11C, n = 7 and 6 rats).

We examined the gene expression patterns of VMT-projecting ACC neurons using RNA

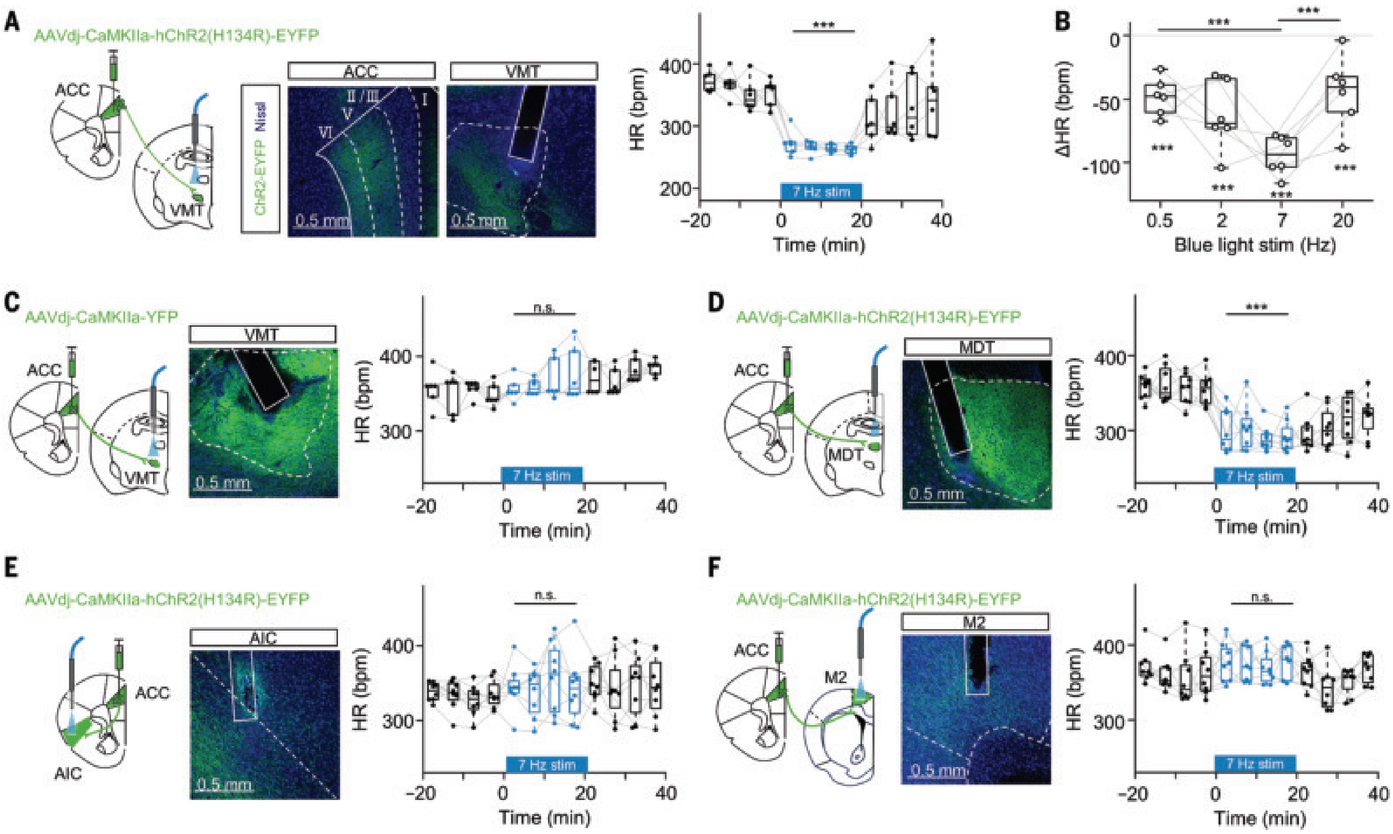


Fig. 4. ACC-VMT neural pathway mediates operant bradycardia. (**A**) Left schematic shows the injection of AAV-CaMKIIa-ChR2-EYFP or AAV-CaMKIIa-EYFP into the ACC. Confocal images show the injection sites in the ACC and optical fiber track in the VMT. The right plot indicates HR in response to blue-light stimulation at 7 Hz. $F_{1,46}$ = 323.01, $^{***}P = 1.98 \times 10^{-22}$, n = 6 rats, one-way ANOVA versus the -20-to-0 min baseline. (**B**) Effects of various frequencies of blue-light stimulation to the VMT on the mean HR during the period from 0 to 20 min. $^{***}P < 0.0001$ versus 0, n = 6 rats, bootstrap test. $^{***}P < 0.0001$ versus 7 Hz, n = 6 rats, multiple correspondence bootstrap test with Bonferroni correction. (**C**) Injection of AAV-CaMKIIa-YFP into the ACC and optical fiber implantation in the VMT. HR did not change with 7-Hz blue-light stimulation of the ACC-VMT pathway that did not express ChR2. $F_{1,46}$ = 3.42, $^{n.s.}P$ = 0.071, n = 6 rats, one-way ANOVA. (**D** to **F**) Injection of AAV-CaMKIIa-ChR2-EYFP into the ACC and optical fiber implantation into the (D) MDT, (E) AIC, or (F) M2. HR decreased with 7-Hz blue-light stimulation of the ACC-to-MDT pathway but not the ACC-to-AIC or ACC-to-secondary motor cortex (M2) pathway. n = 8 experiments of 5 rats each. $^{***}P < 0.0001$, one-way ANOVA.

sequencing with the translating ribosome affinity purification method (fig. S12A) (10). Several genes were enriched in the immunoprecipitated RNA in VMT-projecting ACC neurons compared to total lysates (control) (fig. S12B). Clustering analyses based on gene ontology terms (fig. S12C) and individual genes (fig. S12D) revealed that VMT-projecting ACC neurons represent a distinct class in ACC cells.

We expressed the genetically encoded calcium indicator GCaMP7f in VMT-projecting ACC neurons by injecting AAVdj-hSyn-FLEX-jGCaMP7f into the ACC and retroAAV2-pmSyn1-EBFP-Cre into the VMT and performed fiber photometry recordings from ACC layer 5 (fig. S13A). As in LFPs, theta oscillations occurred in calcium activity during HR feedback

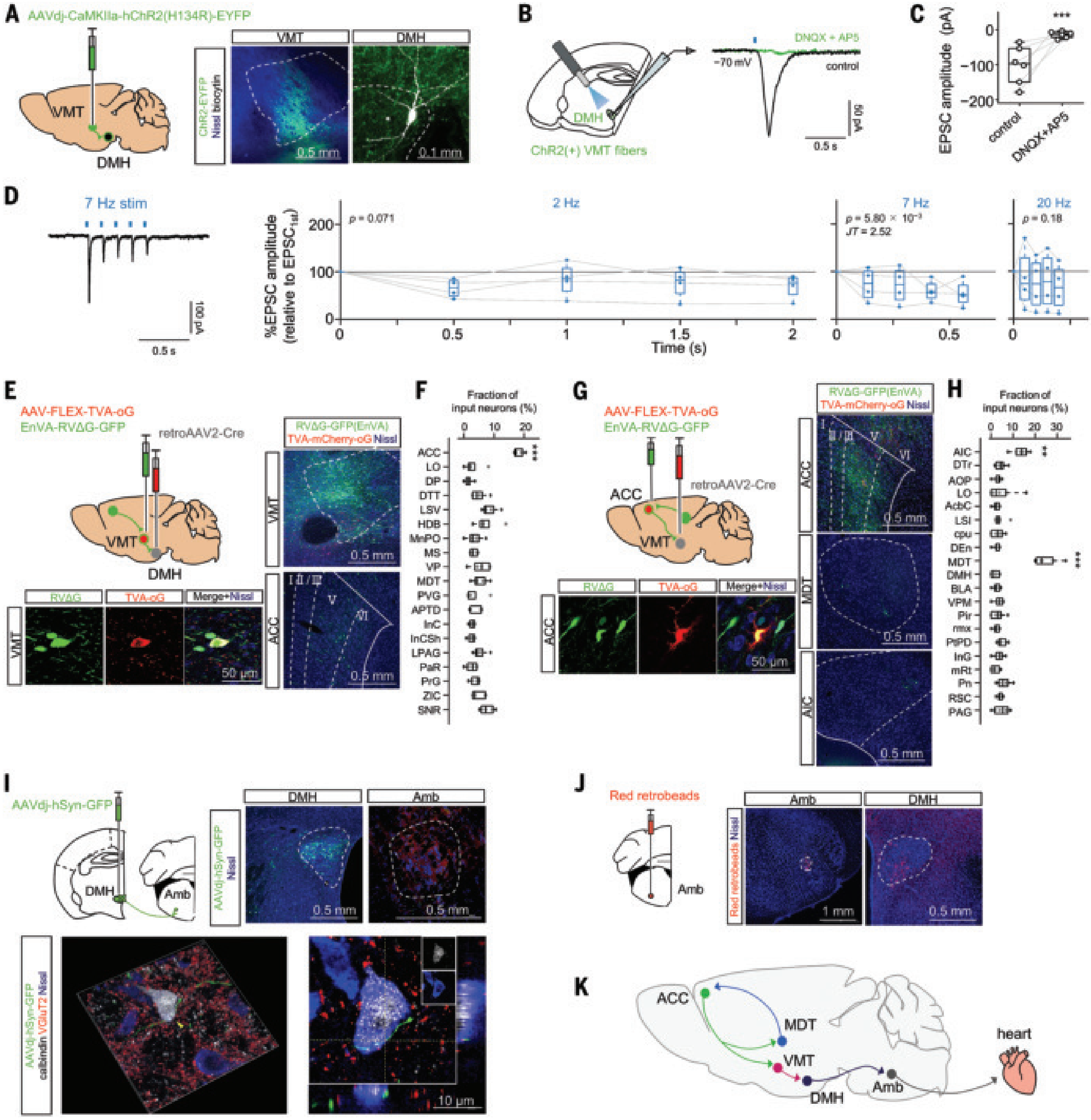


Fig. 5. ACC→VMT→DMH→Amb pathway constitutes a top-down route underlying operant bradycardia. (**A**) The left schematic shows viral expression of ChR2 in the VMT. The right confocal images show the injection sites in the VMT and a patch-clamped DMH neuron. (**B**) Responses of a DMH neuron to blue light illumination to the DMH in a thalamocortical slice in the presence and absence of 10 μM DNQX and 20 μM AP5. (**C**) EPSC amplitudes are pooled from 6 neurons from 3 rats. ***$P < 0.0001$, paired bootstrap test. (**D**) Same as Fig. 3F but for light stimulation of VMT-DMH fibers. $n = 4$ neurons from 2 rats, Jonckheere-Terpstra test. (**E**) The top left schematic shows the injection sites for projection-specific monosynaptic rabies virus. Confocal images show the injection sites in the VMT (top right: wide-field image; left bottom: high-magnification image) and input neurons (bottom right, green) in the ACC. (**F**) Ratios of input cells to DMH-projecting VMT neurons. ***$P < 0.0001$, $n = 5$ rats, one-way ANOVA. See table S2 for abbreviations. (**G**) Same as (E) but for input neurons projecting to ACC neurons that project to the VMT. Confocal images show the injection sites in the ACC and input neurons in the MDT and the AIC. (**H**) Proportions of input cells to VMT-projecting ACC neurons. **$P < 0.001$, ***$P < 0.0001$, $n = 6$ rats, one-way ANOVA. (**I**) The left top illustration depicts the injection of AAVdj-hSyn-GFP into the DMH to detect axons of DMH neurons to the Amb. The right top confocal images show the injection site in the DMH and axons of DMH neurons in the Amb. Sections were immunostained with anti-calbindin and anti-VGluT2 and counterstained with fluorescent Nissl. (**J**) The left illustration depicts the injection of red RetroBeads into the Amb. Confocal images on the right show the injection site in the Amb and DMH neurons projecting to the Amb. (**K**) Putative neural pathway of top-down HR control.

training, and their intensity increased as training progressed (fig. S13, B and C). However, theta oscillations were only weakly observed when GCaMP7f was nonspecifically expressed in ACC neurons (fig. S13, D to G).

To electrophysiologically confirm the monosynaptic projection from the ACC to the VMT, we expressed channelrhodopsin 2 (ChR2) in ACC neurons by injecting AAVdj-CaMKIIa-hChR2(H134R)-EYFP into the ACC (Fig. 3C) and patch-clamped VMT neurons under voltage-clamp configuration in acute thalamocortical slices (Fig. 3D). Upon blue light illumination of ACC axon terminals, VMT neurons exhibited excitatory postsynaptic currents (EPSCs) that were blocked by a cocktail of 10 μM DNQX and 20 μM AP5 (Fig. 3E, n = 10 neurons from 3 rats). This glutamatergic transmission exhibited short-term synaptic facilitation for repetitive activation, that is, the amplitudes of EPSCs gradually increased when blue light pulses were repeatedly applied at short intervals. The facilitation was greatest when the frequency of repetitive stimulation was set to a theta rhythm of 7 Hz (Fig. 3F, n = 9 to 10 neurons from 3 rats).

When VMT-projecting ACC neurons were photostimulated in vivo at 7 Hz, rats exhibited a decrease in HR (Fig. 4A, n = 6 rats). Unlike biofeedback-induced bradycardia, optogenetically induced bradycardia was not impaired by the administration of dopamine receptor antagonists (fig. S14A, n = 6 rats). We also applied blue light stimulation at other frequencies, 0.5, 2, and 20 Hz (fig. S14, B to D, n = 6 rats). All stimulus frequencies reduced HR but the 7-Hz stimulus had the greatest bradycardic effect (Fig. 4B). The bradycardic effect was not observed in YFP-expressing control rats (Fig. 4C, n = 6 rats). Bradycardia was also induced by 7-Hz stimulation of MDT-projecting ACC neurons (Fig. 4D, n = 8 experiments from 5 rats) but not by stimulation of AIC-projecting ACC neurons (Fig. 4E, n = 8 experiments from 5 rats) or to the secondary motor cortex (Fig. 4F, n = 8 experiments from 4 rats).

Multisynaptic pathway from neocortex to heart

Previous anatomical studies have demonstrated that the VMT is synaptically connected to the dorsomedial hypothalamus (DMH) (*11*), which is involved in HR regulation (*12*, *13*). We confirmed this synaptic connection by observing glutamatergic EPSCs evoked in DMH neurons by optogenetic activation of ChR2-expressing VMT axon terminals in acute brain slices (Fig. 5, A to C, n = 4 neurons). VMT-DMH transmission showed a short-lasting depression at 7 Hz repetitive stimulation (Fig. 5D). To search for upstream neurons that project to DMH-projecting VMT neurons, we used a transsynaptic tracing method to monosynaptically label inputs to a specific population of projection neurons (*14*). Specifically, 3 weeks after injection of AAVdj-CMV-FLEX-TVA-mCherry-2A-oG into the VMT and retroAAV2-pmSyn1-EBFP-Cre into the DMH, we injected EnvA-RVΔG-GFP into the VMT (Fig. 5E). GFP-labeled presynaptic neurons were distributed in multiple brain regions, with the highest abundance observed in the ACC (Fig. 5F, n = 5 rats). No significant presynaptic neurons were detected in the MDT, indicating that the ACC, but not the MDT, innervates DMH-projecting VMT neurons. Using the same tracing method, we searched for presynaptic neurons that project to VMT-projecting ACC neurons (Fig. 5G) and found that the MDT and AIC contained a significant number of GFP-positive neurons (Fig. 5H, n = 6 rats).

DMH neurons are reported to project to the nucleau ambiguus (Amb) (*15*), a brainstem structure that sends long projections to the heart through the vagus nerve (*16*, *17*). We also confirmed that the DMH had axonal innervation to the Amb by anterograde labeling from DMH neurons (Fig. 5I) and by retrograde labeling from the Amb (Fig. 5J). DMH axons traveled near calbindin-positive Amb neurons (Fig. 5I), which are known to regulate cardiac function (*16*).

Discussion

In the present study we established a rodent HR feedback system and demonstrated that rats can regulate their HR through biofeedback-based reinforcement learning. The reduction in HR was mediated by central dopamine signaling but was not primarily driven by respiratory modulation or did not occur under anesthesia. Therefore, the operant bradycardia was likely produced by a volitional, top-down mechanism.

We detected ACC theta-rhythm activity in biofeedback-induced bradycardia. Because the bradycardic reflex induced by physical stimuli, such as water immersion, hypoxia, and ultrasound, occurs even in anesthetized or decerebrated animals (*18*) and does not require the ACC, the ACC may serve as a top-down center for HR regulation. This idea may be relevant to the increase in ACC theta oscillations during human meditation (*19*). Fiber photometric recordings from the ACC showed that theta oscillations were not absent during training but were very weak, whereas VMT-projecting ACC neurons exhibited strong theta oscillations. This observation suggests that a subset of ACC cells (i.e., VMT-projecting ACC neurons) are recruited for theta oscillations. The ACC may consist of multiple parallel circuits that can simultaneously process different levels of information, consistent with the fact that the ACC is involved in diverse functions, including pain, cognition, attention, and emotion.

The ACC has been implicated in the regulation of the autonomic nervous system (*20*), and electrical stimulation of the ACC induces autonomic responses (*21*). Indeed, ACC neurons project to primary autonomic structures such as the hypothalamus and brainstem (*22*, *23*). However, the detailed neural pathway from the ACC to the heart remains to be elucidated. We employed a combination of monosynaptic and disynaptic tracing techniques and revealed the entire neural route from the ACC to the heart. We also found that both VMT-projecting and MDT-projecting ACC neurons were activated during training. We speculate that the ACC and MDT form a reciprocal circuit that functions as a generator of theta oscillations (*24*). Then, the activity of this circuit is transmitted from the ACC to the VMT and subsequently relayed to the DMH, the Amb, and the heart (Fig. 5K). Although we do not exclude other possible cardiac control pathways, this study opens new avenues for research on the cardio-cerebral interplay.

REFERENCES AND NOTES

1. P. M. Lehrer, R. Gevirtz, *Front. Psychol.* **5**, 756 (2014).
2. N. E. Miller, L. DiCara, *J. Comp. Physiol. Psychol.* **63**, 12–19 (1967).
3. A. Yoshimoto, K. Yamashiro, Y. Ikegaya, N. Matsumoto, *Biol. Pharm. Bull.* **44**, 789–797 (2021).
4. A. Yoshimoto, Y. Shibata, M. Kudara, Y. Ikegaya, N. Matsumoto, *eNeuro* **9**, ENEURO.0521-21.2022 (2022).
5. M. Kuwahara *et al*., *J. Electrocardiol.* **27**, 333–337 (1994).
6. H. D. Critchley, N. A. Harrison, *Neuron* **77**, 624–638 (2013).
7. B. Hsueh *et al*., *Nature* **615**, 292–299 (2023).
8. A. A. Walf, C. A. Frye, *Nat. Protoc.* **2**, 322–328 (2007).
9. D. P. Collins, P. G. Anastasiades, J. J. Marlin, A. G. Carter, *Neuron* **98**, 366–379.e4 (2018).
10. M. Heiman, R. Kulicke, R. J. Fenster, P. Greengard, N. Heintz, *Nat. Protoc.* **9**, 1282–1291 (2014).
11. E. G. Jones, *The Thalamus* (Cambridge Univ. Press, ed. 2, 2007).
12. T. F. S. Brasil *et al*., *Front. Pharmacol.* **10**, 1547 (2020).
13. J. A. DiMicco, S. Sarkar, M. V. Zaretskaia, D. V. Zaretsky, *Auton. Neurosci.* **126-127**, 106–119 (2006).
14. L. A. Schwarz *et al*., *Nature* **524**, 88–92 (2015).
15. Y. Kono *et al*., *Sci. Rep.* **10**, 1 (2020).
16. A. Veerakumar, A. R. Yung, Y. Liu, M. A. Krasnow, *Nature* **606**, 739–746 (2022).
17. A. Standish, L. W. Enquist, J. A. Escardo, J. S. Schwaber, *J. Neurosci.* **15**, 1998–2012 (1995).
18. W. M. Panneton, Q. Gan, R. Juric, The rat: A laboratory model for studies of the diving response, **108**, 811–820 (2010).
19. L. I. Aftanas, S. A. Golocheikine, *Neurosci. Lett.* **310**, 57–60 (2001).
20. K. Yoshihara *et al*., *Neuroimage* **134**, 522–531 (2016).
21. J. L. Pool, J. Ransohoff, *J. Neurophysiol.* **12**, 385–392 (1949).
22. X. An, R. Bandler, D. Ongür, J. L. Price, *J. Comp. Neurol.* **401**, 455–479 (1998).
23. D. Ongür, J. L. Price, *Cereb. Cortex* **10**, 206–219 (2000).
24. S. Parnaudeau *et al*., *Neuron* **77**, 1151–1162 (2013).

ACKNOWLEDGMENTS

We thank A. Ishigaki for her technical support. We are grateful to Nagai Memorial Research Scholarship from the Pharmaceutical Society of Japan for the scholarship to A.Y. We used generative AI to proofread for grammatical errors and typos in the text. **Funding:** This work was funded by the following: JST ERATO (JPMJER1801) (to Y.I.); AMED CREST (22 g1510002h0002) (to Y.I.); JSPS Grants-in-Aid for Scientific Research (22K21353) (to Y.I.); and The Institute for AI and Beyond of the University of Tokyo (to Y.I.). **Author contributions:** Conceptualization: A.Y. and Y.I. Methodology: A.Y., S.M., and Y.I. Investigation: A.Y., E.K., and Y.I. Visualization: A.Y. and Y.I. Funding acquisition: Y.I. Project administration: A.Y. and Y.I. Supervision: Y.I. Writing – original draft: A.Y. and Y.I. Writing – review and editing: A.Y., S.M., and Y.I. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All primary data that support the findings of this study and the code used for data processing and

analysis are available from the following URL: https://drive.google.com/drive/folders/1VzZmL4RSekHOzbZaAGFayzXadZMPE_g3.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl3353
Figs. S1 to S14
Tables S1 and S2
Materials and Methods
References (*25–39*)
MDAR Reproducibility Checklist
Movie S1

Submitted 12 October 2023; resubmitted 13 February 2024
Accepted 13 May 2024
10.1126/science.adl3353

SPINTRONICS

Spin torque–driven electron paramagnetic resonance of a single spin in a pentacene molecule

Stepan Kovarik†‡, Richard Schlitz†§, Aishwarya Vishwakarma, Dominic Ruckert, Pietro Gambardella*, Sebastian Stepanow*

Control over quantum systems is typically achieved by time-dependent electric or magnetic fields. Alternatively, electronic spins can be controlled by spin-polarized currents. Here, we demonstrate coherent driving of a single spin by a radiofrequency spin-polarized current injected from the tip of a scanning tunneling microscope into an organic molecule. With the excitation of electron paramagnetic resonance, we established dynamic control of single spins by spin torque using a local electric current. In addition, our work highlights the dissipative action of the spin-transfer torque, in contrast to the nondissipative action of the magnetic field, which allows for the manipulation of individual spins based on controlled decoherence.

The initialization, control, and readout of the electron spin underlie the processing and storing of information in both classical and quantum systems. In spintronic devices such as magnetic tunnel junctions, spin-polarized currents are routinely used to achieve electrical control of the magnetization through spin torques and readout of the magnetic state through the tunneling magnetoresistance (TMR) (*1*, *2*). Spin torques rely on the transfer of angular momentum between a spin current and a magnetic layer, mediated by the exchange interaction and scattering between conduction and localized electrons. This control has enabled the development of magnetic random-access memories (*3*, *4*) and spin torque nano-oscillators (*5–7*), which serve as tunable microwave sources (*8*), rectifiers (*9*), and neuromorphic emulators (*10*). Despite these important achievements, the quantum mechanical nature of spin torques is still poorly understood. With limited exceptions (*11–15*), most theoretical and experimental realizations of spin torques consider the magnetic moments subjected to the torque as classical vectors obeying Landau-Lifshitz-Gilbert dynamics, which is appropriate for mesoscopic devices but not for single spins. Conversely, methods that rely on radiofrequency (RF) magnetic fields (*16*), electric fields (*17*, *18*), and optical probes (*19*) for achieving control of single spins are well established and have widespread applications in quantum information technology, quantum sensing, and high-resolution spectroscopy (*20*, *21*). The potential of spin-polarized currents in this context has been much less explored.

Experiments on individual spins (*22*, *23*) using scanning tunneling microscopy (STM) with magnetic tips (*24*) have revealed that the incoherent spin flips induced by a spin-polarized DC current result in a change of the spin expectation value $\langle \boldsymbol{S} \rangle$ along the tip magnetization direction (*25*). This phenomenon allows for switching the orientation of atomic-scale magnets by a local electric current (*23*, *26*). Additionally, theoretical models predict the capability of spin currents to initialize individual spins into a superposition state with arbitrary components $\langle S_{x,y,z} \rangle$ and to induce electron paramagnetic resonance (EPR) (*27–29*). Harnessing spin currents to drive the initialization and precession of single spins would open new routes for the manipulation of quantum states that are complementary to methods reliant on RF electric and magnetic fields.

Here, we show that the injection of an RF spin-polarized current into the singly occupied and unoccupied molecular orbitals (SOMO and SUMO, respectively) of a pentacene molecule leads to the nonequilibrium initialization of the molecular spin and resonant excitation of EPR by spin-transfer torque. Our experiments discriminate the dissipative (damping-like) and the energy-conserving (field-like) components of the spin torque. We refer to the first as the spin-transfer torque (STT) and include the second into the response of the spin to the effective magnetic field. By comparing field- and current-driven EPR, we illustrate differences in the manipulation of single spins due to the different action of the STT compared with a magnetic field. Finally, using STT-driven EPR in an STM, we demonstrate the possibility of mapping the spin distribution of extended orbitals in organic molecules and achieving strong EPR signals independently of temperature because of the strong initialization of the spin states.

Department of Materials, ETH Zurich, CH-8093 Zürich, Switzerland.
*Corresponding author. Email: pietro.gambardella@mat.ethz.ch (P.G.); sebastian.stepanow@mat.ethz.ch (S.S.)
†These authors contributed equally to this work.
‡Present address: nanotech@surfaces Laboratory, Empa–Swiss Federal Laboratories for Materials Science and Technology, CH-8600 Dübendorf, Switzerland.
§Present address: Department of Physics, University of Konstanz, 78457 Konstanz, Germany.

Pentacene as a single-spin model system

To study the influence of spin torques at the quantum level, we focused on pentacene ($C_{22}H_{14}$) adsorbed on a thin MgO layer deposited on an Ag(001) crystal as a model quantum dot system that hosts a single spin in a molecular orbital weakly coupled to a metallic reservoir. We used STM to probe the topography and local density of states of pentacene through the differential tunneling conductance dI/dU and to locally inject a spin-polarized current. The magnetic moment of pentacene on MgO originates from one electron transferred from the Ag substrate into its lowest unoccupied molecular orbital (LUMO), resulting in the formation of SOMO and SUMO states (*30*). Accordingly, the dI/dU spectrum peaks at the energy of the SOMO and SUMO below and above the Fermi level, respectively (Fig. 1A). Both orbitals present the nodal structure of the pristine LUMO (insets of Fig. 1A), which is evidence that the decoupling provided by the MgO layer is sufficient to preserve the electronic structure of pentacene and that the charging electron is fully localized on the molecule, in agreement with previous studies (*30–32*).

To measure the effect of a spin-polarized current on the pentacene spin, we applied an RF voltage U_{RF} to the tunnel junction and used a magnetic tip to detect the modulation of the TMR associated with the DC and RF response of the spin (*33–35*). This technique, called EPR-STM, has recently achieved unprecedented resolution to detect and manipulate spins at surfaces (*36–42*), including the coherent manipulation of one or more atomic spins (*43*, *44*) (fig. S1). Previous investigations using EPR-STM have focused on spins localized at transition metal atoms with strong electronic hybridization with the substrate, which hinders the observation of orbital-specific EPR. In these systems, the EPR is excited by a time-dependent magnetic field B_1 originating from the adiabatic oscillation of the probed atoms in the inhomogeneous magnetic field of the STM tip, which has both exchange and

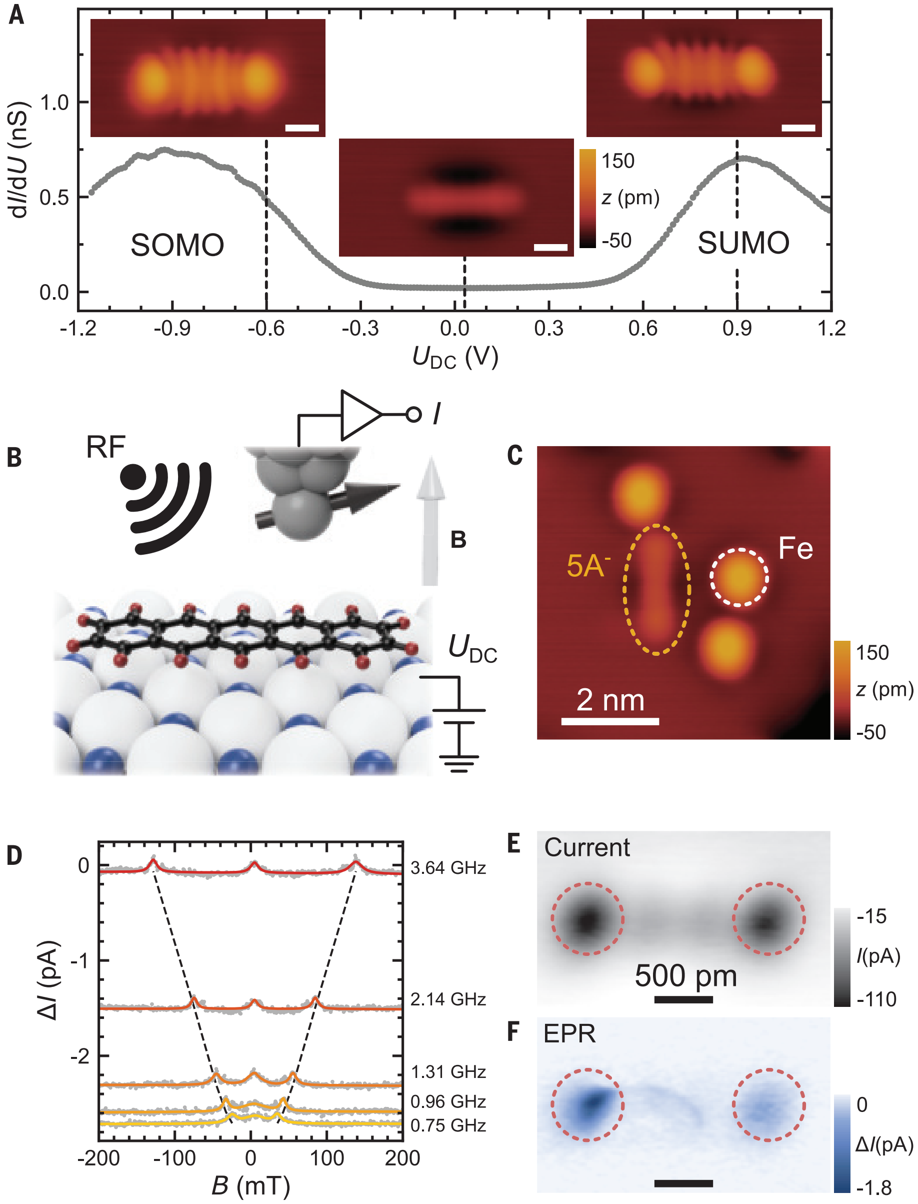


Fig. 1. Scanning tunneling spectroscopy and EPR of individual pentacene molecules. (**A**) Differential conductance dI/dU taken at constant height (feedback opened at I = 500 pA and U_{DC} = –1.2 V). Because of the unpaired electron in the SOMO, pentacene on MgO is an organic radical with spin S = 1/2. Due to the vanishing density of states of the molecule near the Fermi level (56), we did not observe the usual spin-flip excitation steps near the Fermi energy in the tunneling spectra. The topography maps in the insets were taken at I = 30 pA and U_{DC} as indicated by the dashed lines. Scale bars, 500 pm. (**B**) Sketch of the EPR-STM junction. The magnetic tip is shown on top of a pentacene molecule on MgO/Ag(100). The DC bias voltage U_{DC} was applied to the sample, and an RF voltage U_{RF} was capacitively coupled to the tunneling junction through an antenna. (**C**) Constant current topography of a charged pentacene molecule surrounded by three Fe atoms on two monolayers of MgO (I = 30 pA and U_{DC} = –350 mV). (**D**) EPR spectra as a function of B acquired at different frequencies (I = 50 pA, U_{DC} = –520 mV, and U_{RF} ≈ 90 mV). The dashed lines show the shift of the resonance field B_0 with frequency. (**E**) Current map taken at constant height (feedback opened at I = 150 pA and U_{DC} = –450 mV on the molecular lobe). (**F**) Simultaneous map of the EPR signal acquired with frequency modulation at B = 1.28 T and f = 36.22 GHz.

dipolar components. This oscillation is a piezoelectric response to the RF electric field associated with U_{RF} (*33*, *38*, *45*). In the following, we show that a second EPR excitation mechanism exists in addition to B_1, which is consistent with the STT driving a single spin and largely dominating B_1 upon direct electron tunneling in the SOMO and SUMO.

EPR of a single spin in a delocalized molecular orbital

A schematic view of the EPR-STM junction is shown in Fig. 1B. We prepared a spin-polarized STM tip by picking up several Fe atoms codeposited with pentacene on two monolayers of MgO (Fig. 1C). The tip magnetization has both out-of-plane (z) and in-plane (x) components because of the magnetic anisotropy of the Fe cluster assembled on the tip. We used lock-in modulation of U_{RF} superimposed to the DC bias U_{DC} to measure the change ΔI of the tunneling current I due to the excitation of EPR as a function of the external magnetic field B applied along z (*35*). The EPR spectra measured at a temperature of 4.5 K (Fig. 1D) consist of three peaks, of which the outer two shift linearly with the frequency f of U_{RF}, in agreement with the change of Zeeman energy expected for a spin S = 1/2 (fig. S2). The spatial extent of the spin density was revealed by mapping the EPR signal at constant f and B (*46*, *47*). The maps of I and ΔI in Fig. 1, E and F, respectively, show resonant slices of the spin density, which is distributed over the entire molecule and corresponds to the electronic density of the SOMO [for a discussion of the EPR maps compared with the STM topography, see (*48*) and figs. S3 to S9].

The EPR spectra confirmed the paramagnetic character of pentacene. Moreover, they revealed two features that are not expected for EPR driven by a time-dependent magnetic field B_1. The first is the large amplitude of the EPR lines, which is independent of B (fig. S2) despite the relatively high temperature of our experiment ($k_B T$ ≈ 0.4 meV) compared with the Zeeman energy splitting ($E_Z = 2\mu_B B$ ≈ 3 – 150 μeV) (figs. S10 to S12). For B_1-driven EPR, in the absence of spin initialization, the signal amplitude should be proportional to the difference in thermal population of the $\pm S_Z$ spin states, which cannot be resolved for a single spin with $E_Z \sim f \approx$ 1 GHz at 4.5 K. The second feature is the peak near B = 0, which has not been reported in previous EPR-STM studies and is also observed in the absence of RF excitation (*48*) (fig. S13). This feature is consistent

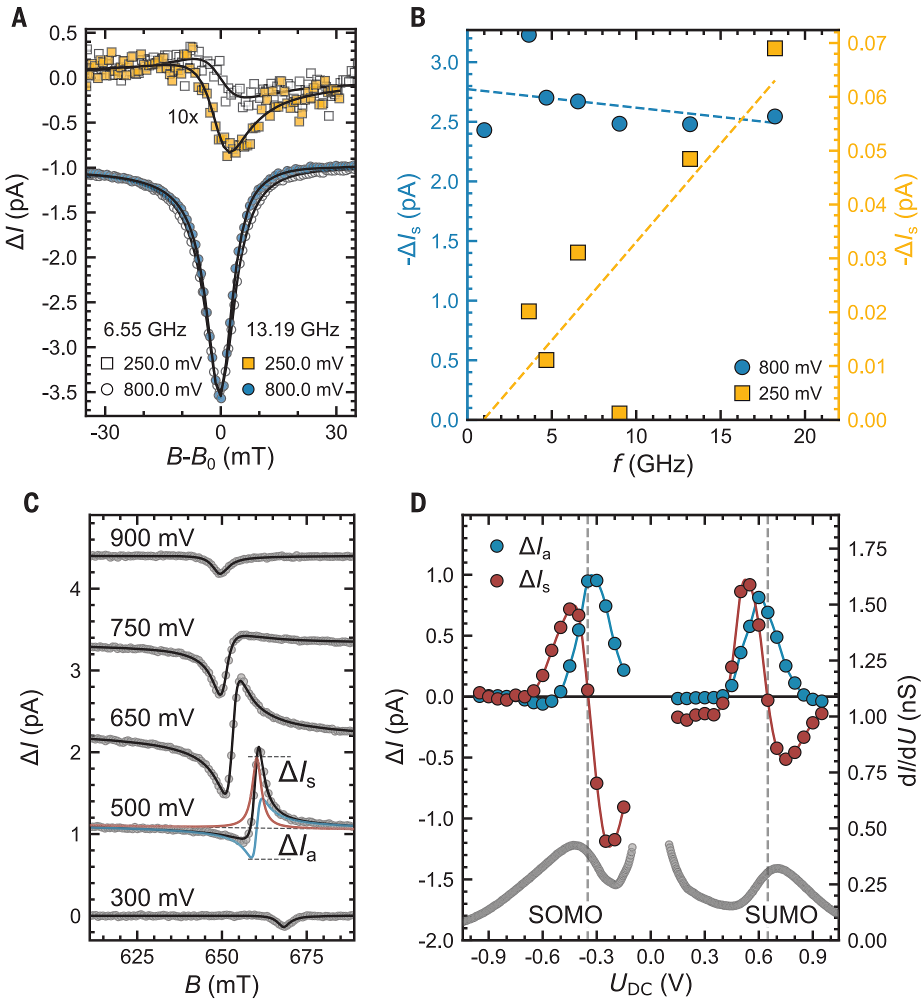


Fig. 2. Spin-transfer torque initialization and driving of a molecular spin. (**A**) EPR spectra as a function of B acquired at U_{DC} = 250 mV (squares) and 800 mV (dots) at f = 6.55 GHz (open symbols) and 13.19 GHz (filled symbols) with I = 100 pA and U_{RF} ≈ 130 mV. (**B**) Symmetric peak amplitude ΔI_s as a function of f (figs. S2 and S18). The dashed lines are linear fits. (**C**) EPR spectra as a function of B acquired at different values of U_{DC} (I = 50 pA, U_{RF} ≈ 90 mV, and f = 18.22 GHz). The dashed lines correspond to the symmetric and antisymmetric components of the EPR line shape with their respective amplitudes, ΔI_s and ΔI_a. (**D**) ΔI_s and ΔI_a as a function of U_{DC} (figs. S17 and S19). The gray points represent the differential conductance of the molecule. The vertical dashed lines indicate the sign inversion of ΔI_s as STT driving overcomes B_1 driving.

with the change of conductance induced by a Hanle-like precession of the transverse $\langle S_x \rangle$ component injected in the SOMO in the magnetic field experienced by the molecule, as expected for a quantum dot attached to magnetic leads (*11*, *12*, *27*, *49*, *50*). The dependence of the zero field conductance on current (fig. S13), the external field, and tip magnetization is rather complex and will be considered in future work (*48*).

To further explore the origin of the EPR signal, we studied its dependence on the DC bias, which determines whether tunneling proceeds through the electronic states of the substrate or the SOMO and SUMO of pentacene. EPR spectra recorded in the SOMO-SUMO gap (U_{DC} = 250 mV) and near the SUMO (U_{DC} = 800 mV) exhibit a markedly different behavior as a function of the excitation frequency (Fig. 2A). The amplitude of the EPR signal at U_{DC} = 250 mV (yellow squares in Fig. 2B) increases linearly with $f \sim B$, consistent with the polarization of a spin $S = 1/2$, proportional to $\tanh(E_Z/2k_BT) \approx E_Z/2k_BT$ (*42*, *51*). At U_{DC} = 800 mV instead (blue dots), the spin-polarized current through the SUMO leads to a frequency-independent EPR intensity that is much larger than the Boltzmann limit, corresponding to the spin polarization expected at millikelvin temperature. We observed the same behavior when tunneling through the SOMO. Thus, we conclude that resonant tunneling into the occupied

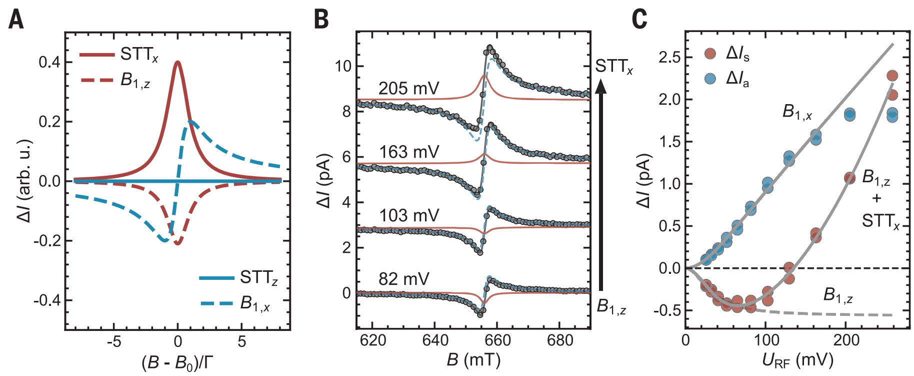


Fig. 3. Power dependence of the EPR signal. (**A**) Calculated line shape of the different contributions to the EPR signal for STT and B_1 driving (*48*). (**B**) EPR spectra for different values of U_{RF} acquired with I = 50 pA and U_{DC} = −350 mV. The solid lines show the symmetric (red) and antisymmetric (dashed blue) component of the fit according to Eq. 1 (black). (**C**) ΔI_s and ΔI_a as a function of U_{RF}. The lines are fits to the data for the different driving mechanisms [see the materials and methods (*48*)]. The deviation of ΔI_a from the expected linear trend at large U_{RF} is ascribed to the bias-dependent junction conductance around U_{DC} = −350 mV (Fig. 2D and fig. S20).

and unoccupied molecular orbitals efficiently polarized the molecule, and that this initialization by the spin-polarized current dominated over all other relaxation channels (*48*) (fig. S14).

The strong initialization of the spin, which was further confirmed by the dependence of the tip standoff distance on bias and magnetic field (*48*) (figs. S11 and S12), can be assigned to a rotation of the spin polarization vector toward the tip magnetization (*27*, *52*). If the latter has a finite x component, then this rotation corresponds to a spin-transfer torque, leading to a superposition state $|S_x\rangle = (|S_z\rangle + |{-S_z}\rangle)/\sqrt{2}$. Thus, the spin is not in an eigenstate and will precess in the magnetic field. In the following, we show that synchronizing the STT with the Larmor precession of the spin in the magnetic field leads to EPR.

Experimental signatures of spin torque–driven EPR

We analyzed the EPR line shape and its bias dependence to reveal the tell-tale signatures of spin torque driving at the single-spin level. We proceeded by analogy with mesoscopic devices, where the ferromagnetic resonance driven by STT is distinguished from that driven by an oscillating magnetic field B_1 by the symmetry of the resonance line shape (*6*, *7*, *9*, *53*). In general, the EPR-STM signal can be decomposed into the sum of symmetric (ΔI_s) and antisymmetric (ΔI_a) Lorentzian functions

$$\Delta I = \Delta I_s \frac{1}{1+\varepsilon^2} + \Delta I_a \frac{\varepsilon}{1+\varepsilon^2} + \Delta I_{off} \qquad (1)$$

where $\varepsilon = \frac{B - B_0}{\Gamma/2}$, B_0 and Γ are the resonance field and width of the EPR line, respectively, and ΔI_{off} is an offset current due to nonmagnetic rectification effects. For the B_1 driving established for single atoms (*34*, *38*, *45*), ΔI_s arises from the time-averaged change of the projection of the atom's spins on the tip magnetization, whereas ΔI_a arises from the homodyne mixing of U_{RF} and time-dependent TMR caused by the rotating component of the spin (*34*, *38*). The EPR spectra in Fig. 2C and the bias dependence of ΔI_s and ΔI_a in Fig. 2D show significant deviations from previous studies of single atoms in which EPR was induced by B_1 (*34*, *38*, *51*).

At low bias, we found that ΔI_s has the same sign for both positive and negative U_{DC}, consistent with B_1-driven EPR and the simultaneous inversion of the TMR and tunneling current with DC bias in pentacene (*48*). As $|U_{DC}|$ approached the energy of the SOMO and SUMO, however, both ΔI_s and ΔI_a increased strongly due to the more efficient initialization and the amplification of the TMR with the growing fraction of current flowing through the molecular orbitals. Upon further increasing $|U_{DC}|$ beyond the onset of the SOMO and

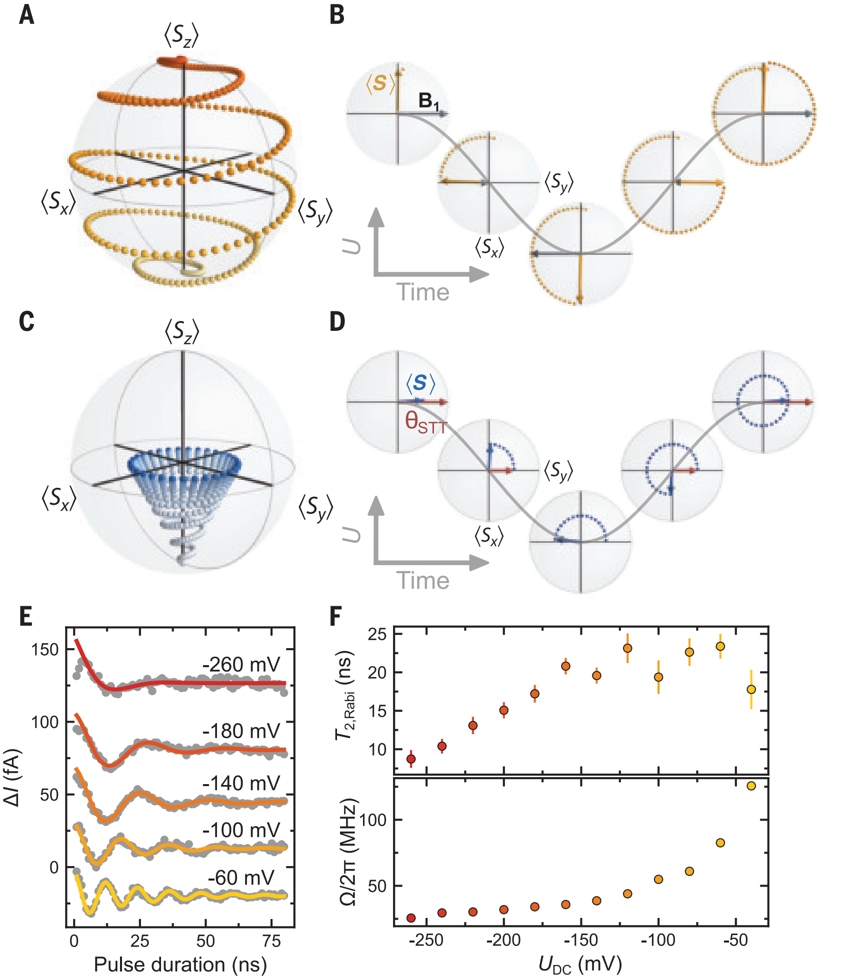


Fig. 4. Evolution of $\langle S_z \rangle$ and $\langle S_{x,y} \rangle$ on resonance. (**A**) Evolution of the spin expectation value for B_1-driven EPR during the initial half of the Rabi cycle. (**B**) The five circular diagrams depict the time evolution of $\langle S_{x,y} \rangle$ (yellow) and B_1 (gray) for B_1-driven EPR. The diagrams are superposed to the time evolution of the RF voltage. (**C**) Evolution of the spin expectation value for STT-driven EPR. (**D**) The five circular diagrams depict the time evolution of $\langle S_{x,y} \rangle$ (blue) and the RF STT (Θ_{STT}, red) superposed to $U_{RF}(t)$. (**E**) Rabi oscillations measured at I = 13 pA and U_{RF} = 240 mV for different values of U_{DC} (see fig. S21 for the complete dataset). The solid lines are fits to the data according to the equation in the caption of fig. S21. (**F**) $T_{2,Rabi}$ time and Rabi rate Ω as a function of U_{DC} extracted from the fits shown in (E). Error bars represent the SDs of the fitted values. The error of $\Omega/2\pi$ is within the symbol size.

SUMO resonances, the conductance increased and the tip retracted to maintain a constant current (*48*). This process in turn led to a much less efficient B_1 driving, as shown by the reduction of ΔI_a already at U_{DC} lower than the SOMO and SUMO peaks in the conductance spectrum (Fig. 2D). Conversely, ΔI_s changed sign on the flanks of the SOMO and SUMO. This behavior was unexpected, because changing the sign of ΔI_s should also lead to an inversion of ΔI_a for B_1-driven EPR [see the materials and methods (*48*)]. Moreover, the complex dependence of $\Delta I_{s,a}$ on bias cannot be ascribed to changes of the rectified current near the SOMO and SUMO (*48*). The two sign changes of ΔI_s at U_{DC} = −350 mV and 650 mV thus indicate the emergence of an EPR-driving mechanism that only generates a symmetric signal upon tunneling through the molecular orbitals and has an opposite sign relative to ΔI_s in the SOMO-SUMO gap.

We associated the symmetric line shape observed when a spin-polarized current tunnels into the SOMO and SUMO to STT-driven EPR. This assignment agrees with theoretical models of coherent quantum transport in an STM junction under an RF bias (*27*–*29*), as well as with the basic properties of angular momentum transfer through a spin-transfer mechanism. For a single spin in a tunnel junction, STT driving appears through the periodic variation of the initialization rate caused by the modulation of the tunneling probability and the coherent evolution of $\langle \boldsymbol{S} \rangle$ in a magnetic field

(*28*, *29*). If the STT modulation rate matches the Larmor frequency $\omega_L = \gamma B$ (where γ is the electronic gyromagnetic ratio), the precession of the in-plane expectation value $\langle S_{x,y}\rangle = \langle S_x\rangle(\hat{x}\cos\omega_L t + \hat{y}\sin\omega_L t)$ about B is phase-locked to U_{RF}, leading to EPR. The symmetric EPR line shape observed in our experiments reflects the in-phase oscillation of $\langle S_x(t)\rangle$ averaged over many subsequent tunneling events. Conversely, for B_1 driving, $\langle S_x(t)\rangle$ is 90° out of phase with respect to U_{RF}, giving rise to the antisymmetric line shape ΔI_a in the homodyne channel.

The phase locking between the precession of the molecular spin and the tunneling current can be described by using either a modified Bloch equation model that includes the STT driving as a time-dependent relaxation rate of $\langle \boldsymbol{S}\rangle$ toward the direction of the tunneling spin (*34*, *38*) or a Lindblad master equation approach that describes the evolution of the reduced density matrix of the molecule (*48*) (fig. S15). Figure 3A summarizes the different contributions to the line shape of the EPR signal from STT driving (STT_z and STT_x) and B_1 driving ($B_{1,z}$ and $B_{1,x}$) obtained from the modified Bloch equation model. Note that the contributions STT_z and $B_{1,z}$ result from the DC change of the tunneling current proportional to $\langle S_z\rangle$, whereas STT_x and $B_{1,x}$ result from the homodyne detection of $\langle S_x(t)\rangle$. Both $B_{1,z}$ and STT_x yield a symmetric line shape, but of opposite sign, supporting the interpretation of the EPR spectra in Fig. 2C.

An additional signature of STT driving was found in the power dependence of the EPR signal. Whereas for B_1 driving, ΔI_s saturates (*38*, *51*) and ΔI_a increases linearly with U_{RF} [see the materials and methods (*48*)], the homodyned STT signal STT_x scales as U_{RF}^2 because the modulation of the tunneling probability enters twice, once for the STT rate and once for the homodyne readout by the TMR (*9*, *28*). To demonstrate the different power scaling of the two contributions to ΔI_s, we plotted the evolution of the EPR signal with U_{RF} (Fig. 3, B and C). From low to high U_{RF}, we observed a negative ΔI_s at low U_{RF}, as expected for B_1 driving at negative DC bias, followed by a nonlinear increase and inversion of sign of ΔI_s, as expected for STT driving at high U_{RF}. Conversely, ΔI_a increased almost linearly with U_{RF}, consistent with B_1 driving. Our observations thus indicate that STT is the dominant mechanism that drives EPR of singly occupied molecular orbitals.

Correspondence of spin torques in the quantum and classical descriptions

In the context of spintronics, the STT driving that emerges from tunneling through the SOMO and SUMO corresponds to the damping-like component of the spin torque, which is dissipative in nature. This torque dominates in the sequential tunneling regime where the rate of spin-polarized electrons tunneling in or out of the molecule is large. The Pauli spin blockade causes a spin filter (spin pumping) effect when tunneling from (to) the molecule into (from) the magnetic tip, conditioned by the evolution of $\langle \boldsymbol{S}(t)\rangle$ in the magnetic field. This process leads to a strong transverse spin accumulation and a consequent rotation of $\langle \boldsymbol{S}\rangle$ antiparallel (parallel) to the tip polarization when averaging over many tunneling events. The B_1 driving associated with the exchange field of the tip, conversely, corresponds to the field-like spin torque. This torque dominates in the Coulomb-blockaded regime between SOMO and SUMO, in which the dwell time of a single electron in the molecule is large and the exchange field is determined by virtual electron hopping to the magnetic tip (*11*). To account for the precession of $\langle \boldsymbol{S}(t)\rangle$ in phase with the driving RF voltage, it is crucial to consider the coherences of the molecular spin in the singly occupied charge state in both cases (*28*, *29*).

Pulsed EPR measurements showed that the longitudinal spin relaxation time of pentacene is $T_1 = 140 \pm 10$ ns at a current of 10 pA (*48*) (fig. S16). Therefore, the spin relaxation rate was slower than the tunneling rate in our experiments. Because of the time-averaged nature of $\langle \boldsymbol{S}(t)\rangle$, considering electrons that tunnel at a rate smaller or larger than the Larmor frequency does not change the conclusions on the STT-driven EPR (*48*).

Temporal evolution of $\langle S\rangle$ due to magnetic field and STT

We finally turn to the evolution of $\langle \boldsymbol{S}(t)\rangle$ on the Bloch sphere for a system driven by B_1 and compare it with the action of the STT. To that end, we modeled the time evolution of $\langle \boldsymbol{S}\rangle$ by solving the modified Bloch equations. Equivalent results can be obtained from the solution of the Lindblad model (*48*). For B_1 driving, we found the conventional Rabi behavior, whereby $\langle S_z\rangle$ oscillates between $+S_z$ and $-S_z$ at a rate proportional to B_1 (Fig. 4A), and $\langle S_x\rangle$ oscillates with a 90° phase shift relative to U_{RF} (Fig. 4B). By contrast, a time-dependent STT yields no Rabi oscillations, showcasing the intrinsic difference of the two driving mechanisms. For STT driving, $\langle S_z\rangle$ decays at a rate proportional to the current (Fig. 4C), and $\langle S_x\rangle$ oscillates in phase with U_{RF} (Fig. 4D).

To reveal the differences in the two driving regimes, we performed a Rabi experiment by applying RF pulses of varying length to the STM junction [for further details, see (*43*) and (*48*)]. We observed Rabi oscillations in the B_1-dominated regime at low $|U_{DC}|$. However, upon further increasing $|U_{DC}|$ beyond 260 mV, the Rabi frequency dropped until no oscillations could be observed in the STT-dominated regime (Fig. 4E). The Rabi rate Ω reduced with increasing $|U_{DC}|$, which is consistent with the reduction of B_1 caused by the retraction of the tip (*48*). Furthermore, the relaxation time $T_{2,Rabi}$ dropped quickly when U_{DC} approached the SOMO (Fig. 4F), as expected from the increased tunneling current flowing through the molecule. The reduction of Ω and $T_{2,Rabi}$ entails a strong reduction of the B_1-driven EPR intensity at high bias, which contrasts with the drastic increase and sign change of the EPR signal reported in Fig. 2. Therefore, the Rabi measurements further support the dominant role of STT driving when tunneling into the molecular orbitals and its distinct phenomenology relative to B_1 driving.

Conclusions and outlook

Our data provide evidence of coherent transfer of angular momentum to a single spin by a spin-polarized current. This transfer occurs through interaction with the tunneling electrons that involve entanglement with the molecular spin, followed by relaxation of the many-body wave function to a single spin state. STT thus allows for the initialization of spins into superposition states with arbitrary spin polarization, overcoming the spatial and thermal population constraints associated with electromagnetic field excitations. Spin rotations induced by the dissipative STT are inherently nonunitary, which limits the coherent control of the spin's phase. However, the field-like spin torque contribution to B_1 provides a voltage-dependent means to coherently manipulate the spin's state. Future work may focus on implementing iterative schemes to achieve fault-tolerant single-spin initialization and rotation operations (*13*) and on the investigation of the nondissipative field-like spin torque component. Spin-polarized currents may be used to control the spin of paramagnetic centers in the gate oxide of field effect transistors and magnetic tunnel junctions within quantum devices (*54*) and entangle separated magnetic qubits (*55*). These capabilities provide complementary control of quantum spin excitations relative to RF electromagnetic fields.

REFERENCES AND NOTES

1. D. C. Ralph, M. D. Stiles, *J. Magn. Magn. Mater.* **320**, 1190–1216 (2008).
2. A. Manchon *et al*., *Rev. Mod. Phys.* **91**, 035004 (2019).
3. J. Z. Sun, *J. Magn. Magn. Mater.* **559**, 169479 (2022).
4. V. Krizakova, M. Perumkunnil, S. Couet, P. Gambardella, K. Garello, *J. Magn. Magn. Mater.* **562**, 169692 (2022).
5. M. Tsoi *et al*., *Nature* **406**, 46–48 (2000).
6. J. C. Sankey *et al*., *Nat. Phys.* **4**, 67–71 (2008).
7. L. Liu, T. Moriyama, D. C. Ralph, R. A. Buhrman, *Phys. Rev. Lett.* **106**, 036601 (2011).
8. T. Chen *et al*., *Proc. IEEE* **104**, 1919–1945 (2016).
9. A. A. Tulapurkar *et al*., *Nature* **438**, 339–342 (2005).
10. J. Torrejon *et al*., *Nature* **547**, 428–431 (2017).
11. M. Braun, J. König, J. Martinek, *Phys. Rev. B Condens. Matter Mater. Phys.* **70**, 195345 (2004).
12. M. Braun, J. König, J. Martinek, *Europhys. Lett.* **72**, 294–300 (2005).
13. B. Sutton, S. Datta, *Sci. Rep.* **5**, 17912 (2015).
14. A. Zholud, R. Freeman, R. Cao, A. Srivastava, S. Urazhdin, *Phys. Rev. Lett.* **119**, 257201 (2017).

15. M. D. Petrović, P. Mondal, A. E. Feiguin, P. Plecháč, B. K. Nikolić, Spintronics Meets Density Matrix Renormalization Group, *Phys. Rev. X* **11**, 021062 (2021).
16. J. J. Pla *et al*., *Nature* **489**, 541–545 (2012).
17. K. C. Nowack, F. H. L. Koppens, Y. V. Nazarov, L. M. K. Vandersypen, *Science* **318**, 1430–1433 (2007).
18. A. Laucht *et al*., *Sci. Adv.* **1**, e1500022 (2015).
19. D. D. Awschalom, R. Hanson, J. Wrachtrup, B. B. Zhou, *Nat. Photonics* **12**, 516–527 (2018).
20. A. J. Heinrich *et al*., *Nat. Nanotechnol.* **16**, 1318–1329 (2021).
21. E. Moreno-Pineda, W. Wernsdorfer, *Nat. Rev. Phys.* **3**, 645–659 (2021).
22. S. Loth *et al*., *Nat. Phys.* **6**, 340–344 (2010).
23. A. A. Khajetoorians *et al*., *Science* **339**, 55–59 (2013).
24. R. Wiesendanger, *Rev. Mod. Phys.* **81**, 1495–1550 (2009).
25. F. Delgado, J. J. Palacios, J. Fernández-Rossier, *Phys. Rev. Lett.* **104**, 026601 (2010).
26. S. Loth, S. Baumann, C. P. Lutz, D. M. Eigler, A. J. Heinrich, *Science* **335**, 196–199 (2012).
27. A. M. Shakirov, Y. E. Shchadilova, A. N. Rubtsov, P. Ribeiro, *Phys. Rev. B* **94**, 224425 (2016).
28. A. M. Shakirov, A. N. Rubtsov, P. Ribeiro, *Phys. Rev. B* **99**, 054434 (2019).
29. J. Reina-Gálvez, C. Wolf, N. Lorente, *Phys. Rev. B* **107**, 235404 (2023).
30. M. Hollerer *et al*., *ACS Nano* **11**, 6252–6260 (2017).
31. J. Repp, G. Meyer, S. Paavilainen, F. E. Olsson, M. Persson, *Science* **312**, 1196–1199 (2006).
32. J. Repp, G. Meyer, S. M. Stojković, A. Gourdon, C. Joachim, *Phys. Rev. Lett.* **94**, 026803 (2005).
33. S. Baumann *et al*., *Science* **350**, 417–420 (2015).
34. Y. Bae *et al*., *Sci. Adv.* **4**, eaau4159 (2018).
35. T. S. Seifert *et al*., *Phys. Rev. Res.* **2**, 013032 (2020).
36. T. Choi *et al*., *Nat. Nanotechnol.* **12**, 420–424 (2017).
37. P. Willke *et al*., *Science* **362**, 336–339 (2018).
38. T. S. Seifert *et al*., *Sci. Adv.* **6**, eabc5511 (2020).
39. M. Steinbrecher *et al*., *Phys. Rev. B* **103**, 155405 (2021).
40. T. S. Seifert, S. Kovarik, P. Gambardella, S. Stepanow, *Phys. Rev. Res.* **3**, 043185 (2021).
41. L. M. Veldman *et al*., *Science* **372**, 964–968 (2021).
42. X. Zhang *et al*., *Nat. Chem.* **14**, 59–65 (2022).
43. K. Yang *et al*., *Science* **366**, 509–512 (2019).
44. P. Willke *et al*., *ACS Nano* **15**, 17959–17965 (2021).
45. J. L. Lado, A. Ferrón, J. Fernández-Rossier, *Phys. Rev. B* **96**, 205420 (2017).
46. P. Willke, K. Yang, Y. Bae, A. J. Heinrich, C. P. Lutz, *Nat. Phys.* **15**, 1005–1010 (2019).
47. S. Kovarik *et al*., *Nano Lett.* **22**, 4176–4181 (2022).
48. See the supplementary materials.
49. A. D. Crisan *et al*., *Nat. Commun.* **7**, 10451 (2016).
50. P. Busz, D. Tomaszewski, J. Barnaś, J. Martinek, *J. Magn. Magn. Mater.* **588**, 171465 (2023).
51. P. Willke *et al*., *Sci. Adv.* **4**, eaaq1543 (2018).
52. P. Mondal, U. Bajpai, M. D. Petrović, P. Plecháč, B. K. Nikolić, *Phys. Rev. B* **99**, 094431 (2019).
53. A. A. Kovalev, G. E. W. Bauer, A. Brataas, *Phys. Rev. B Condens. Matter Mater. Phys.* **75**, 014430 (2007).
54. M. Lamblin *et al*., Encoding information onto the charge and spin state of a paramagnetic atom using MgO tunnelling spintronics. arXiv:2308.16592 [Preprint] (2023).
55. A. Suresh *et al*., *Phys. Rev. A* **109**, 022414 (2024).
56. S. Loth, C. P. Lutz, A. J. Heinrich, *New J. Phys.* **12**, 125021 (2010).
57. Data for: S. Kovarik, R. Schlitz, A. Vishwakarma, D. Ruckert, P. Gambardella, S. Stepanow, Spin torque–driven electron paramagnetic resonance of a single spin in a pentacene molecule, ETH Zurich (2024); https://doi.org/10.3929/ethz-b-000669567.

ACKNOWLEDGMENTS

We thank B. K. Nikolić and J. Reina-Gálvez for helpful discussions. **Funding:** This work was supported by the Swiss National Science Foundation (SNSF project CRSII5_205987) and by ETH (research grant ETH-45 20-2). **Author contributions:** Conceptualization: S.K., R.S., P.G., S.S.; Funding acquisition: P.G., S.S.; Investigation: S.K., R.S., A.V., D.R.; Methodology: S.K., R.S., P.G., S.S.; Project administration: P.G., S.S.; Software: S.K., R.S., D.R.; Supervision: P.G., S.S.; Visualization: S.K., R.S.; Data curation: S.K., R.S.; Writing – original draft: S.K., R.S., P.G., S.S.; Writing – review & editing: S.K., R.S., A.V., D.R., P.G., S.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All of the data needed to evaluate the conclusions in this work are available in the main text or the supplementary materials and have also been deposited to the ETH Zurich database (*57*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh4753
Materials and Methods
Supplementary Text
Figs. S1 to S21
References (*58*, *59*)

Submitted 6 March 2023; resubmitted 9 February 2024
Accepted 10 May 2024
10.1126/science.adh4753

ELECTROCHEMISTRY

Water-hydroxide trapping in cobalt tungstate for proton exchange membrane water electrolysis

Ranit Ram[1]†, Lu Xia[1]†, Hind Benzidi[2]†, Anku Guha[1]†, Viktoria Golovanova[1], Alba Garzón Manjón[3], David Llorens Rauret[3], Pol Sanz Berman[2], Marinos Dimitropoulos[1], Bernat Mundet[3], Ernest Pastor[4,5], Veronica Celorrio[6], Camilo A. Mesa[7], Aparna M. Das[1], Adrián Pinilla-Sánchez[1], Sixto Giménez[7], Jordi Arbiol[3,8], Núria López[2], F. Pelayo García de Arquer[1]*

The oxygen evolution reaction is the bottleneck to energy-efficient water-based electrolysis for the production of hydrogen and other solar fuels. In proton exchange membrane water electrolysis (PEMWE), precious metals have generally been necessary for the stable catalysis of this reaction. In this work, we report that delamination of cobalt tungstate enables high activity and durability through the stabilization of oxide and water-hydroxide networks of the lattice defects in acid. The resulting catalysts achieve lower overpotentials, a current density of 1.8 amperes per square centimeter at 2 volts, and stable operation up to 1 ampere per square centimeter in a PEMWE system at industrial conditions (80°C) at 1.77 volts; a threefold improvement in activity; and stable operation at 1 ampere per square centimeter over the course of 600 hours.

The increasing global energy demand, combined with the urgent need to abate climate change, has accelerated the development of sustainable and clean energy technologies as alternatives to fossil fuels. Water electrolysis (WE) to synthesize hydrogen (H_2) has emerged as a promising strategy to produce clean energy vectors from water and low-carbon electricity, offering a path to decarbonize global industries, such as energy, transport, manufacturing, and agriculture, among others (*1*–*3*).

Among the different WE technologies, the proton exchange membrane water electrolysis (PEMWE), in which cathode and anode electrodes are intimately connected through a proton conductive membrane, exhibits advantages compared with diaphragm- and anion transport–based alternatives in terms of productivity (high–current density operation), energy efficiency, stability, and levelized cost of H_2 (*4*–*6*). In this context, the efficient and sustainable large-scale production of H_2 through WE still faces important challenges. These are associated with the sluggish kinetics of the oxygen evolution reaction (OER) and the reliance on scarce, critical raw materials, such as iridium (Ir)—so far, the prevalent anode catalyst material because of its stability but one of the least-abundant metals on Earth (*7*–*9*).

Alternative approaches using ruthenium (Ru) have shown promising activity but suffer from a strong metal dissolution in acidic media intrinsic to lattice OER mechanisms (*10*–*12*). There is, thus, an urgent need to develop efficient and stable Ir- and Ru-free anodes for PEMWE (*13*–*17*).

Transition metal oxides are interesting catalyst candidates for the OER because their multiple oxidation states could promote activity in a wide pH range (*18*, *19*). First-row transition metal oxides have shown promising activity in the acidic OER (*20*–*25*). Among them, Co-, Ni-, and Mn-based anodes have received special attention considering their

[1]ICFO – Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain. [2]ICIQ-CERCA – Institute of Chemical Research of Catalonia, The Barcelona Institute of Science and Technology, 43007 Tarragona, Spain. [3]Catalan Institute of Nanoscience and Nanotechnology (ICN2), CSIC and BIST, Campus UAB, Bellaterra, 08193 Barcelona, Catalonia, Spain. [4]CNRS, Université de Rennes, IPR (Institut de Physique de Rennes) - UMR 6251, Rennes, France. [5]CNRS, Université de Rennes, DYNACOM (Dynamical Control of Materials Laboratory) - IRL2015, The University of Tokyo, 7-3-1 Hongo, Tokyo 113-0033, Japan. [6]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. [7]Institute of Advanced Materials (INAM), Universitat Jaume I, 12006 Castelló, Spain. [8]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain.
***Corresponding author. Email: pelayo.garciadearquer@icfo.eu**
†These authors contributed equally to this work.

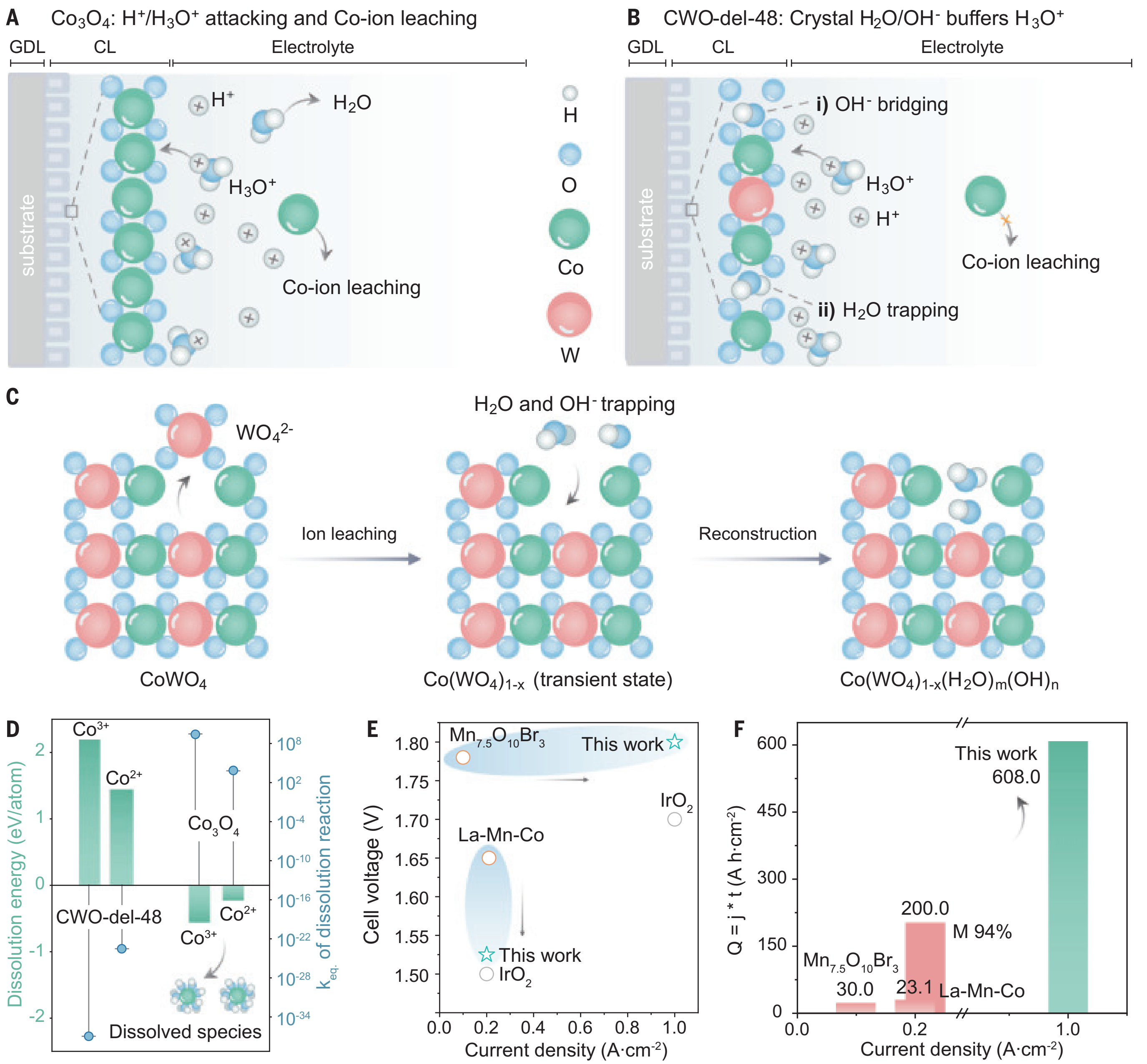


Fig. 1. Water-hydroxide trapping enables active and stable PEMWE. (**A**) Schematic illustration of the dissolution of Co ions in Co_3O_4 through Co–O bond cleaving, followed by their hydration in acidic medium. GDL, gas diffusion layer; CL, catalyst layer. (**B**) A depiction of delaminated cobalt catalysts (CWO-del-48), illustrating the effect of water trapping and hydroxide bridging within the crystal lattice, enabling stability in acid. (**C**) The process and crystallographic representation of CWO delamination into CWO-del-48 through base treatment. (**D**) CWO-del-48 catalyst shows unfavorable Co ion dissolution compared with Co_3O_4. The calculated equilibrium constants of the dissolution reaction for CWO-del-48, $k_{eq} << 1$, suggest nonspontaneity of the process. The structure of dissolved Co^{2+} and Co^{3+} ions are shown in the inset. (**E** and **F**) This equilibrium enables substantial advances in performance cell voltage versus current density (E) and durability (F) compared with state-of-the-art Ir- and Ru-free anodes in PEMWE.

relative abundance and activity prospects (*26*, *27*). Based on theoretical calculations, the OER activity of Co-based oxides should be comparable to that of Ru- and Ir-based oxides (*28*).

However, such prospective activity is challenged by the limited stability of CoO_x in acid because of higher Co ion dissolution (Fig. 1A), even at open circuit potential (*29*, *30*). Several strategies have been proposed to stabilize cobalt oxides in highly concentrated proton environments. Fundamental progress in this direction has mostly been pursued in low-current H cells controlling the valence and ratio of active species (e.g., Co^{2+}/Co^{3+}) through doping with higher-valence metals (e.g., Cr and Mn) (*22*, *24*, *25*, *31*); controlling catalyst reconstruction (*32*); doping with hydrophobic carbon (*33*); and addressing the substrate-catalyst interaction (*29*).

Unfortunately, translating these findings into active and stable PEMWE based on more abundant alternatives to Ir and Ru remains an open challenge (*34*). In Mn systems, phase control (γ-MnO_2) led to stability improvements

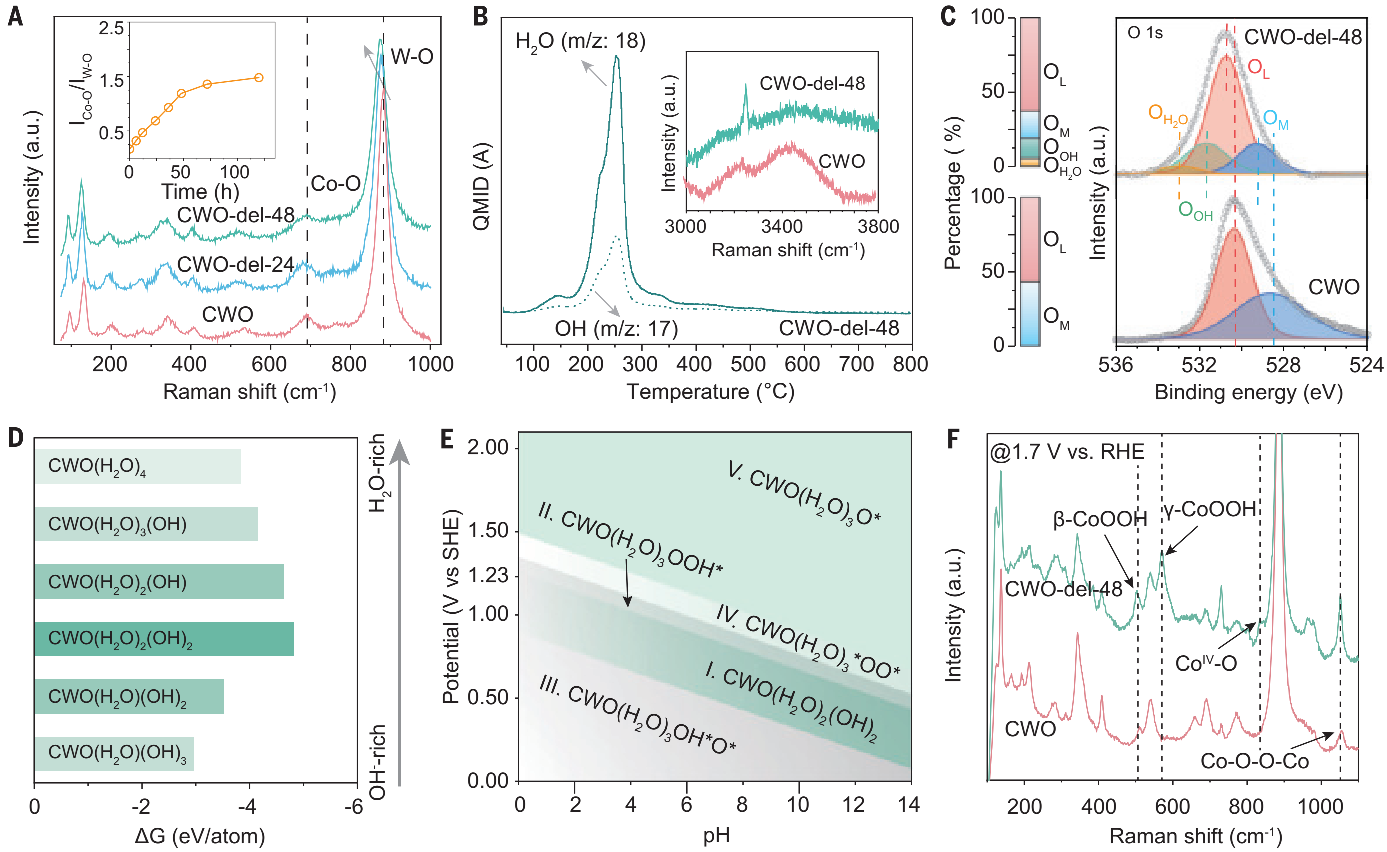


Fig. 2. Water trapping and hydroxide bridging. (**A**) Ex situ Raman spectra of as-synthesized CWO and CWO-del (24 hours and 48 hours) showing a regular red shift associated with the W–O vibrational peaks. (Inset) Intensity ratio of the Co–O and W–O peaks with delamination time. a.u., arbitrary units. (**B**) TGA-MS shows the presence of H_2O and hydroxide in CWO-del-48. (Inset) Ex situ Raman spectra showing presence of different types of water. The sharp H–OH stretching peak of CWO-del-48 indicates presence of strongly H-bonded water. QMID, quasi-multiple ion detection. (**C**) O 1s XPS spectra of CWO and CWO-del-48. O 1s peak in CWO deconvolutes into metal oxygen (O_M) and lattice oxygen (O_L). O 1s peak in CWO-del-48 deconvolutes into four peaks: O_M, O_L, hydroxide (O_{OH}), and water (O_{H2O}). (**D**) Visual representation of the free energy changes involved in the delamination process of CWO to $CWO(H_2O)_m(OH)_n$ under alkaline solution, as a result of water trapping and hydroxide bridging with the removal of tungstate (movie S1). DFT simulations predict that $CWO(H_2O)_2(OH)_2$ is thermodynamically the most favorable composition for CWO-del-48. (**E**) Potential and pH dependence of the intermediates in OER for the CWO-del-48. SHE, standard hydrogen electrode. (**F**) In situ Raman spectra of CWO and CWO-del-48 catalysts at 1.7 V versus RHE suggests the involvement of Co (III), Co (IV), and Co–peroxide as the active OER species.

(12 hours at 100 mA cm^{-2}), which gradually decrease as the γ-MnO_2 phase becomes permanganate (MnO_4^-) (*26*). Mn-oxybromide species resulted in 300 hours of stability at 100 mA cm^{-2} (0.41–mV $hour^{-1}$ degradation) (*23*). In Co-based electrodes, La and Mn doping has enabled a ~0.6–A cm^{-2} current density at 2 V and 110 hours of stability at 210 mA cm^{-2} (~1.65 V at 80°C) (*25*). These experimental observations showcase the challenges in achieving Ir- and Ru-free PEMWE anodes that break the activity-stability trade-off through conventional doping schemes.

Recent works have highlighted the potential of controlling the other half of the electrochemical interface (i.e., water structure and adsorbed oxide species) to improve the OER. The participation of oxygen from adsorbed water in the OER, as opposed to lattice-mediated, would decrease metal dissolution and increase stability (*25*). In general, the interaction of adsorbed bridging oxygen (O_{bri}) species with water could open previously unrealized reaction pathways (*35*). Theoretical studies have shown the critical role of the H bonding network of interfacial water and its impact on proton-electron transfer steps (*36*, *37*). Decreasing the degree of H bonding of interfacial water (n–HB H_2O) has been predicted to reduce the activation energy needed to dissociate water (*38*). The role of interfacial water and the H bonding network, already studied for the hydrogen evolution reaction (*39*), is largely overlooked for acidic OER. This prompted us to jointly address the water and oxide structure, a so-far underexplored path, seeking to improve activity and stability in non-Ir PEMWE anodes.

In this work, we demonstrate control over the OER by modulating the interfacial water structure and intermediate species in a delaminated CoW oxide lattice (Fig. 1, B to D), resulting in active and stable PEMWE. We achieve this by implementing a delamination strategy whereby high-valence sacrificial elements, such as W, when incorporated in a $CoWO_4$ (CWO) crystal structure, could be selectively eliminated in a subsequent water-hydroxide-WO_4^{2-} anion exchange process (Fig. 1C). This results in structural delamination and the subsequent trapping and stabilization of water and hydroxide species in a Co oxide defect network. Such water-hydroxide shielding renders the Co ion dissolution thermodynamically unfavorable for the delaminated CWO in contrast to Co_3O_4 (Fig. 1D and table S1), which shows a marked decrease in Co ion dissolution in acid (fig. S1). The delaminated (CWO-del) catalysts achieve notable performance in a PEMWE, with a current density of 1.8 A cm^{-2} at 2 V—up to a threefold improvement compared with the previous best performance for non-Ir and/or non-Ru (*25*, *40*)—and stable operation for 608 hours at the current density of 1 A cm^{-2} (Fig. 1, E and F).

Anion exchange delamination controls water structure in acid

To incorporate and stabilize OH^- and H_2O into the lattice of MM'_xO_y (where M is Mn, Co, Ni, or Cu), we devised an anion exchange strategy whereby lattice oxyanions (e.g., $M'_xO_y^{z-}$, where M' is S, Mo, or W) would be delaminated and exchanged by OH^- and H_2O species as follows

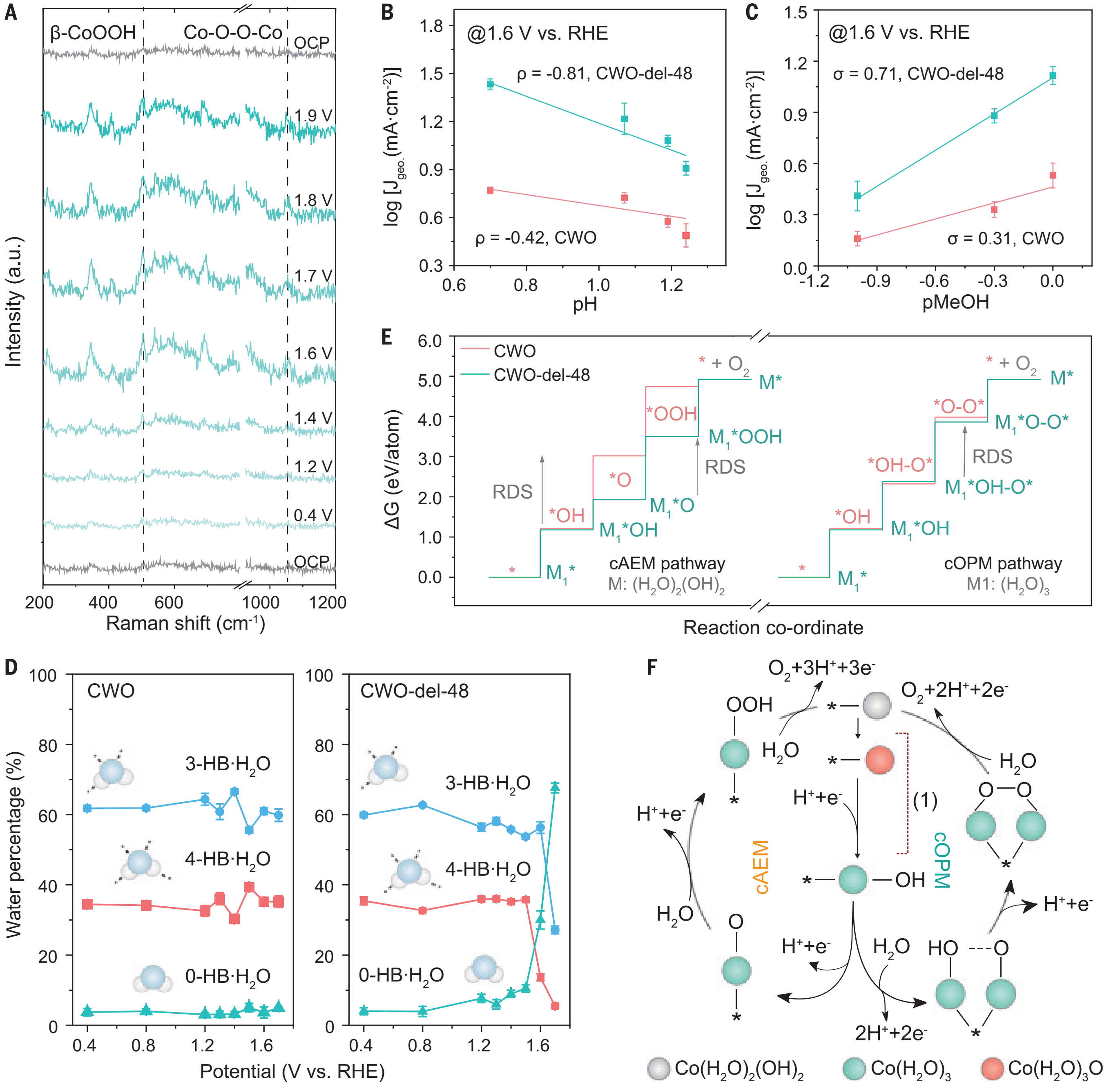


Fig. 3. OER mechanism—operando and DFT studies. (**A**) Operando Raman spectroscopy in CWO-del-48 (on carbon paper, from OCP to 1.9 V versus RHE in 0.5 M H_2SO_4) reveals a correlation between OER activity and the intensity of β–CoOOH and Co–O–O–Co peaks. (**B** and **C**) pH-dependent studies (1.6 V versus RHE as a function of pH) during OER (B) and MOR (C) elucidate the role of surface-trapped water fragments. Error bars correspond to the standard deviation and average of three independent measurements. (**D**) Percentage of different types of interfacial water structures with applied potential for CWO (left) and CWO-del-48 (right). In CWO-del, the percentage of 0–HB·H_2O water increases with applied potential while 4–HB·H_2O structure decreases. These remain almost unchanged for CWO in the applied potential window. (**E**) Free energy profiles of CWO and CWO-del-48 in OER pathways. The involved species and/or intermediates are shown in the corresponding steps. The dynamic involvement of H_2O and OH^- enables favorable confined AEM (cAEM) and confined OPM (cOPM) reaction pathways in CWO-del-48. (**F**) OER catalytic cycle schematically showing the cAEM and cOPM pathways. The first step is the chemical conversion of $Co(H_2O)_2(OH)_2$* to $Co(H_2O)_3O$*. Both the cAEM and cOPM mechanisms share $Co(H_2O)_3$–OH* as a common intermediate.

$$\mathrm{MM'}_x\mathrm{O}_y + m\mathrm{H_2O} + n\mathrm{OH^-} \rightarrow \mathrm{M(M'}_x\mathrm{O}_y)_{1-q}(\mathrm{H_2O})_m(\mathrm{OH})_n + q\mathrm{M'}_x\mathrm{O}_y{}^{z-} \quad (1)$$

The design principles require that such oxyanions have adequate binding energies with OH^- and water species—conditions that promote their sacrificial leaching (*41*)—and that the host lattice could accommodate OH^- and H_2O species to saturate the resulting oxyanion vacancies (supplementary text, section 4, and tables S4 to S6). Our initial theoretical calculations predicted that WO_4^{2-} ions, compared with other anions, such as molybdate or sulfate, provide more favorable defect energy to meet these criteria: The energy associated with defect formation upon removing WO_4^{2-} from the $CoWO_4$ is the lowest energy among SO_4^{2-} and MoO_4^{2-} removal from $CoSO_4$ and $CoMoO_4$, respectively (supplementary text, section 4, and fig. S2).

We thus synthesized CWO using a hydrothermal reaction (supplementary text, section 1, and fig. S3). The sharp peaks in x-ray diffraction (XRD) patterns from different batches indicate a highly crystalline structure matching a monoclinic $CoWO_4$ phase and the high reproducibility of the synthesis procedure (figs. S4 and S5).

To perform the $WO_4^{2-} \rightarrow OH^-$ and H_2O anion exchange, we explored a base treatment dispersing the resulting CWO material in an 0.1 M potassium hydroxide (KOH) aqueous solution for different time periods (supplementary text, section 3, and figs. S6 and S7). We studied the effects of cation (Li^+ to Cs^+), solvent [H_2O, dimethyl sulfoxide (DMSO), or *N*-methylpyrrolidone (NMP)], and pH (both experimental and theoretical) in the process (supplementary text, section 3, and figs. S8 to S13). This study revealed the critical role of K^+ to balance the delamination and the need for H_2O and hydroxide to enable the anion exchange. CWO-del samples retained structural stability after 72 hours of immersion in 0.5 M H_2SO_4 as opposed to Co controls (fig. S14).

Powder XRD patterns show a regular shift in the most intense $\bar{1}11$ peak (table S7), indicating the generation of defects or strain in the crystal structure, whereas the bulk monoclinic crystal phase remains intact. These are consistent with optoelectronic, magnetic, and x-ray absorption spectroscopy studies that show the increased formation of vacancies and defects (figs. S15 to S19).

Transmission electron microscopy (TEM) and scanning transmission electron microscopy (STEM) images show a shape transition from a cube-like CWO (figs. S20 and S21) to a delaminated, flake-like shape after KOH treatment (figs. S22 and S23). The size of the particles after delamination remains comparable to CWO (fig. S24).

High-resolution TEM (HRTEM) images reveal the missing regular (010) crystal plane in CWO-del-48 (figs. S22 and S25), indicative of defects arising from WO_4^{2-} leaching. High-annular angle dark-field STEM (HAADF-STEM) images, and respective energy-dispersive x-ray spectroscopy (EDX) mapping of Co and W, show the uniform distribution of Co and W throughout the pristine and delaminated nanocrystals (figs. S26 and S27). The electron energy-loss spectroscopy (EELS) spectra indicate a reduction in the Co/W atomic ratio after delamination (fig. S28). Additionally, atomic force microscopy (AFM) reveals missing planes and corresponding defects in CWO-del-48 (figs. S29 to S31), in agreement with HRTEM and STEM results.

To assess the atomic arrangement of Co and W sites, we performed integrated differential phase contrast (iDPC) STEM (fig. S32). CWO samples exhibit a regular atomic arrangement of Co, W, and O atoms, consistent with a monoclinic phase. CWO-del-48, on the other hand, reveals a substantial number of vacancies due to WO_4^{2-} leaching (fig. S33).

To get more insights into the dynamics of the oxyanion exchange, we performed ex situ Raman spectroscopy measurements at different delamination times. These reveal a bathochromic shift (~5 to 10 cm^{-1}) for both Co–O (~690 cm^{-1}) and W–O (~886 cm^{-1}) peaks with increasing delamination time (Fig. 2A). The intensity ratio of Co–O to W–O peaks increases up to 48 hours and then saturates (Fig. 2A, inset; fig. S34; and table S8). This suggests a kinetic limitation of W leaching from the CWO matrix. Inductively coupled plasma optical emission spectrometry (ICP-OES) measurements confirm increasing W leaching over time but negligible change in Co concentration (fig. S35 and table S9). The optimal Co/W ratio and OH^- and H_2O trapping considering initial electrochemical studies is achieved for 48-hour delaminated samples (see details in later sections). Density functional theory (DFT) simulations show that (010) is the most thermodynamically favorable crystal facet for the delamination process, which is exothermic only at high pH (figs. S36 to S38, tables S10 and S11, and movie S1).

Water trapping and hydroxide bridging

To assess the presence of water and hydroxyl groups in the delaminated samples, we carried out a series of characterizations. Thermogravimetric analysis coupled with mass spectroscopy (TGA-MS) revealed an ~16% weight loss for CWO-del-48 over a 166° to 396°C range due to water and/or OH^- ions, which is negligible for CWO and Co_3O_4 (Fig. 2B and fig. S39). Four distinct peaks in this range suggest the presence of different water coordination environments (*42*).

To assess these differently coordinated water populations, we initially performed Fourier transform infrared (FTIR) measurements. These reveal HO–H stretching and H–O–H bending vibrational modes with increasing intensity upon delamination, suggesting a higher density of trapped and bridged water and hydroxide groups within the crystals (fig. S40). Further, the narrow single H–OH stretching peak present in the ex situ Raman spectrum of CWO-del-48 indicates the presence of highly H-bonded water (Fig. 2B, inset, and fig. S41) (*43*).

Additionally, we replaced H_2O with D_2O during the delamination process and subjected the samples to mild annealing. This resulted in distinct spectral features corresponding to a D–OD stretching mode, confirming the trapping of D_2O or H_2O in the delaminated materials (fig. S42) (*44*).

We studied the resulting oxygen modes using x-ray photo electron spectroscopy (XPS) (Fig. 2C). The O 1s peak of nondelaminated samples deconvolutes into two peaks at 528.55 and 530.33 eV, which correspond to metal oxygen (O_M) and lattice oxygen (O_L) (*25*). CWO-del-48 spectra, on the other hand, require deconvolution into two additional peaks corresponding to hydroxide (O_{OH}, at 531.6 eV) and water species (O_{H_2O}, at 533.1 eV) (*35*). The O_M and O_L peaks blue shift in this case, which is consistent with the leaching of W in the form of WO_4^{2-} and associated atomic vacancies (O_M) (*45*); this is also supported by ultraviolet-visible (UV-Vis), electron paramagnetic resonance (EPR), and x-ray absorption near-edge structure (XANES) spectra (figs. S15 to S19). The amplitude of O_{OH} and O_{H_2O} signals further suggests the presence of hydroxide and bonded water in the delaminated compound.

To gain insights into the role of delamination and water-hydroxide trapping in rendering catalyst stability, we carried out DFT studies considering water trapping and hydroxide bridging from WO_4^{2-} sites transformation (movie S1 and fig. S12). Because the size of water-trapped structures can also vary, various combinations of structures with different numbers of water molecules and OH^- were calculated. Among the different combinations of CWO-water-hydroxide, $Co(WO_4)_{1-x}(H_2O)_2(OH)_2$ is the thermodynamically most-favorable delaminated species, with a –4.5 eV of Gibbs free energy change for the transformation (Fig. 2D and tables S12 and S14). The calculated Pourbaix diagram (Fig. 2E) confirms the stability of this phase over other structures at pH 0.

The blue shift observed in XANES and distinct extended x-ray absorption fine structure (EXAFS) spectra implies a higher oxidation state of Co and a modified coordination environment in CWO-del-48 compared with CWO (figs. S18 and S19). The corresponding Co–O and Co–Co bond distance of the CWO-del-48 structure are shorter than those in CWO (Fourier-transformed EXAFS, fig. S19; crystal structure, fig. S43), which could facilitate higher-valence Co species, such as CoOOH and Co–peroxide, in OER (*20*, *31*, *46*).

In situ Raman spectroscopy in 0.5 M H_2SO_4 electrolyte reveals three oxide peaks at 1.7 V [versus reversible hydrogen electrode (RHE)] (supplementary text, section 6, and tables S2 and S3), some of which are not visible in ex situ Raman spectra (Fig. 2F). These correspond to layered β–CoOOH (~502 cm^{-1}), γ–CoOOH (~571 cm^{-1}), higher-oxidation Co^{IV}–O (~840 cm^{-1}), and Co–peroxide (~1080 cm^{-1}) sites (*46–48*), in agreement with EXAFS, XPS, and DFT findings (Fig. 2C and figs. S19 and S44).

Structural influence on the OER mechanism

To gain more insights into the peroxide species and the nature of the active sites ensuing from OH^- and H_2O trapping, we performed additional operando Raman spectroscopy studies before and after OER onset potential (Fig. 3A and fig. S43). Both β–CoOOH and Co–peroxide peak intensities steadily increase from open circuit potential (OCP) to 1.9 V versus RHE and vanish as the potential is cycled back to OCP from 1.9 V versus RHE (Fig. 3A). This suggests that both β–CoOOH and Co–peroxide

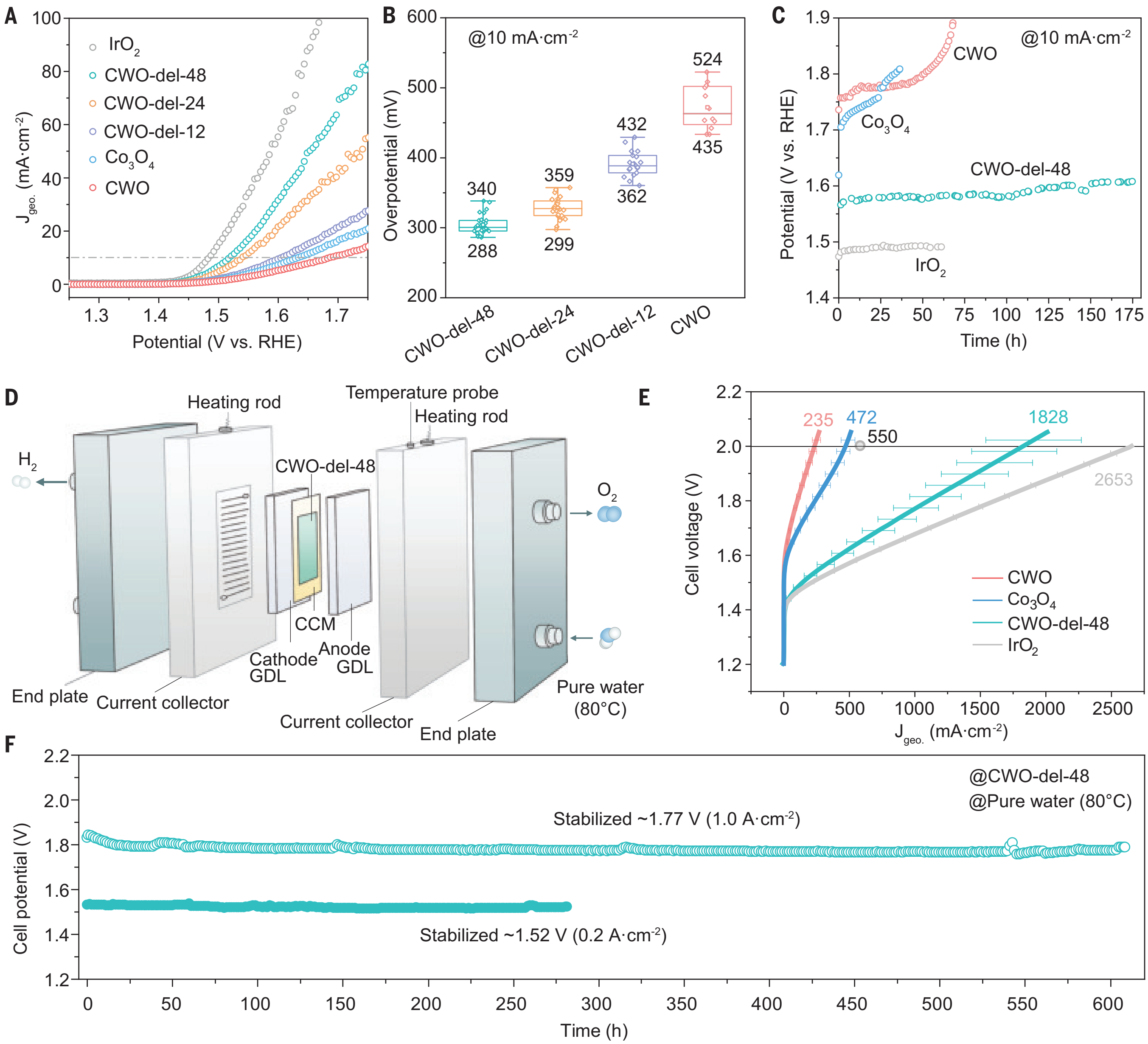


Fig. 4. Electrochemical performance and stable PEMWE at the ampere level. (**A**) Linear-sweep voltammetry of Co-based catalysts at 5 mV s^{-1} in 0.5 M H_2SO_4 electrolyte along with commercial IrO_2 and Co_3O_4 (no *iR* correction). (**B**) Overpotential statistics at 10–mA cm^{-2} current density. (**C**) Chronopotentiometry stability tests in 0.5 M H_2SO_4 electrolyte at constant 10–mA cm^{-2} current density in H-cell setup using a Nafion 117 membrane. (**D**) Illustration of PEMWE along with catalyst-coated membrane (CCM), cathode, and anode GDL. (**E**) The PEMWE-based polarization curves of CWO-del-48, CWO, commercial Co_3O_4, and IrO_2 (without *iR* correction) used as anodic material separately. The PEMWE-based polarization curve is compared with the best reported, La-Mn codoped porous cobalt spinel fibers catalyst (*25*). Commercial 60 wt % Pt/C was used as cathodic material, and Nafion 117 is the membrane. (**F**) The chronopotentiometry stability test of CWO-del-48 at 0.2 A cm^{-2} and at 1.0 A cm^{-2} of current density in PEMWE at 80°C temperature for 278 hours and 608 hours, respectively, after overnight conditioning at 1.7 V (see supplementary text for details). The milli-Q water filling process during operation leads to small voltage transients due to temperature gradients and interface re-equilibration.

are active sites for the OER. To investigate the role of the surface oxides and water-hydroxide trapping in the OER activity, we carried out a suite of pH-dependent electrochemical studies and operando interfacial water structure evaluation using Raman spectroscopy.

Delaminated samples display a very strong pH dependence during the OER, with a reaction order (ρ) of −0.84, nearly 2.5 times that for CWO (Fig. 3B). This can be explained by the presence of trapped water and a higher OH^- coverage, in agreement with both XPS (Fig. 2D) and FTIR studies (fig. S40). The calculated cross-sectional crystal structure (fig. S43) and XPS analysis (fig. S45) suggest a CoOOH-rich arrangement, where water is bonded with Co atoms through the oxygen. This is consistent with Raman spectra and methanol oxidation reaction (MOR) experiments, showing a dominance of MOR over water dissociation in CWO-del-48 across different pH values, in line with the higher amount of surface oxides (Fig. 3C) (*49*).

Next, we studied the role of interfacial H–OH using operando Raman spectroscopy (figs. S46 and S47), deconvoluting three different water structures depending on the number of H bonds: 4-H–bonded water (~3200 cm^{-1}), 3-H–bonded water (~3400 cm^{-1}), and 0-H–bonded water (~3600 cm^{-1}). The activation energy required for water dissociation is predicted to decrease with decreasing degree of H bonding (*39*). As the potential increases from 0.4 to 1.7 V versus RHE, the relative presence of 4–HB H_2O in CWO-del-48 decreases from 35 to 5%, whereas 0–HB H_2O increases from 4 to 68% (Fig. 3D). By contrast, these remain unchanged for CWO throughout the applied potential window. This observation suggests the involvement of interfacial water in the rate-determining step (RDS) for the CWO-del-48 catalyst. Moreover, the substantial change in the Stark slope explains the higher sensitivity of interfacial water structure after OER onset potential (fig. S48).

On the basis of these experimental observations, we performed DFT calculations to assess the energy landscape of the OER mechanisms using the computational hydrogen electrode (CHE) formalism (Fig. 3E). We denote the paths where the confined water fragments can also participate as confined paths. The oxide path mechanism (OPM) has a common RDS (*OH–O* to *O–O*) for both CWO and for the confined water fragments in CWO-del-48 without involving interfacial water. The RDS for CWO-del-48 is 0.09 eV smaller for the OPM pathway than for the adsorbate evolution mechanism (AEM), which makes it thermodynamically more plausible. In the AEM pathway (supplementary text, section 4), the RDS is however different: OH*-to-O* for CWO versus the confined system cAEM and O*-to-OOH* for CWO-del-48 (tables S14 to S17). The higher pH dependence and change of the interfacial water structure with increasing applied potential for CWO-del-48, suggests that CWO-del-48 benefits from the kinetically favorable AEM mechanism (Fig. 3F and fig. S17).

To gain insights into the improved acid resistance under electrolytic conditions, we revisited the surface Pourbaix diagram of CWO-del (Fig. 2E). This diagram is divided into five distinct regions: region I, $CWO(H_2O)_2(OH)_2$; region II, $CWO(H_2O)_3OOH^*$; region III, $CWO(H_2O)_3OH^*O^*$; region IV, $CWO(H_2O)_3{}^*OO^*$; and region V, $CWO(H_2O)_3O^*$ (details in fig. S49). In conditions of low pH, the H^+ ions from the acidic environment interact with H_2O–OH cluster in $CWO(H_2O)_2(OH)_2$ (region I), which transforms into $CWO(H_2O)_3OH^*O^*$ (region III). When the potential lies between 1.20 and 1.37 V versus RHE, interfacial water molecules (region I) interact with another surface oxygen to form a hydroxyl group: $O^* + H_2O \rightarrow OOH^* + H^+ + e^-$, the cation on the surface undergoes oxidation and is stabilized by OOH*. Simultaneously, the oxygen on the surface binds with H* to create a stable structure of $CWO(H_2O)_3{}^*OO^*$ (region IV), and the two O* species combine directly to form O_2, leading to the release of oxygen without the formation of *OOH intermediate. As the potential rises to ≥1.46 V (versus RHE), the surface is shielded by O*, leading to the formation of a stable $CWO(H_2O)_3O^*$ structure (region V).

Electrochemical performance

We compared the polarization curves of CWO-del catalysts with commercial Co_3O_4 and IrO_2 (Fig. 4A) (supplementary text, section 6). A minimum 288-mV overpotential at 10–mA cm^{-2} current density was obtained for CWO-del-48 [no *iR* (*i*, current; *R*, resistance) correction], as opposed to 392 and 259 mV for commercial Co_3O_4 and IrO_2, respectively (see figs. S50 to S52 and table S18 for details on particle size, loading, and double-layer capacitance). Statistical analysis (Fig. 4B) and extended measurements confirm the reproducibility of these trends (figs. S53 and S54). The Faradaic efficiency for O_2 generation was 96.6 ± 5.2% at 10–mA cm^{-2} current density (figs. S55 to S60).

Tafel analysis reveals a slope of 85 mV $decade^{-1}$ for CWO-del-48 versus 63 mV $decade^{-1}$ for IrO_2 and 227 mV $decade^{-1}$ for CWO. This showcases the improved OER kinetics of CWO-del-48 over reference samples, approaching that of IrO_2 (fig. S61).

CWO-del-48 exhibits the highest stability (>175 hours) over commercial Co_3O_4 and CWO (Fig. 4C) in a H-cell at 10 mA cm^{-2}. Structural, compositional, and electrochemical postanalysis, suggest that the crystallinity, microstructure, and size remain comparable after electrolysis (figs. S24 and S62 to S67 and table S19).

We then assessed Co leaching and the potential role of dissolved species. Inductively coupled plasma mass spectroscopy (ICP-MS) reveals that Co leaching remains unchanged after 50 hours of OER in a H-cell at 10 mA cm^{-2} with an ~2.76 parts per million (ppm) concentration of dissolved Co ions in anolyte after 100 hours of chronopotentiometry (fig. S68). A fraction of Co ions (0.3 ppm) leaches out during OCP, before the OER stability test. The concentration of dissolved Co ions in the catholyte remained negligible [<30 parts per billion (ppb); fig. S69], and no traces were found in the membrane (fig. S70 and table S20) or graphite rod (fig. S71). The resulting stability number (S number) (*50*, *51*) after 100 hours of stability test at 10 mA cm^{-2} is 14771 ± 768 (fig. S72). Additional experiments confirm the electrochemical stability at different mass loadings (0.075 to 2 mg cm^{-2}) (figs. S73 and S74).

Encouraged by this, we implemented the CWO-del-48 catalyst in a PEMWE system (Fig. 4D and fig. S75) and studied the cell performance under industrial operational settings, including 80°C temperature and a high current density of 0.2 to 1 A cm^{-2} (see supplementary text, section 6, for details). The polarization curve of CWO-del-48–based cells reaches a nominal current density of 1.8 A cm^{-2} at 2 V (Fig. 4E)—an improvement in rate over the previous-best Ir- and Ru-free anodes of up to 3× for a comparable membrane (*25*) and of 1.8× for advanced, thinner membranes (*40*).

During electrolysis at a fixed current density of 0.2 A cm^{-2}, the voltage range (1.53 to 1.56 V) is ~130 mV lower than prior-best Co-based PEMWE catalysis (Fig. 4F) and matches that of Ir black (1.50 V) at one-quarter the loading. This showcases the potential of CWO-del-48 catalysts compared with the well-established iridium oxide, along with its lower cost and potentially higher availability of Co and W. This performance (~1.52 V at 0.2 A cm^{-2}) is retained for at least >278 hours of continuous operation (limited by pump failure).

ICP-OES analysis of the anolyte reveals a Co concentration of ~1.7 ppm after 20 hours of electrolysis (fig. S76). We also performed single-pass electrolyte flow experiments to study the potential role of ion accumulation in the electrolyte reservoir. The cell voltage remains stable (within 1.53 to 1.54 V) for at least 32 hours at 0.2 A cm^{-2} at similar PEMWE conditions but with continuous, fresh electrolyte flushing (fig. S77).

We further challenged the stability of the CWO-del-48 catalyst at 1–A cm^{-2} operation (Fig. 4F)—a so far elusive benchmark for Ir- and Ru-free catalysts. Electrodes show comparable composition after an initial 20 hours of study (figs. S78 to S81 and table S21). The PEMWE cell exhibited stable performance with a stabilized cell voltage of ~1.77 V during 608 hours of durability study (see table S22 for prior benchmarks). The observed slight voltage fluctuations might arise as a result of a combination of temperature gradients during electrolyte replenishment, modifying electrolysis rates and gas dissolution, and catalyst dissolution (ion and particulate detachment) events.

The calculated S number at these conditions is ~31% of that achieved at 10 mA cm^{-2} in the H-cell (supplementary text). This showcases the impact of high rate and temperature operation stress, which accelerate catalyst degradation.

Conclusion and future scope

The reported strategy still faces several challenges and opportunities. CWO-del samples exhibit high polarization voltage under high current density, which should be improved by further optimization of the integration of key components to enhance electrolysis performance and subsequently improve stability at higher currents that approach state-of-the-art Ir systems.

In the future, achieving industrial performance benchmarks, such as energy-efficient and stable operation in the 2- to 3-A cm^{-2} range (a milestone of IrO_2 in PEMWE) will require tailored engineering of catalysts electrodes and membranes. In that sense, obtaining further insights into the properties of these interfaces and into the roles of water trapping and hydroxide bridging in the OER at increasing current densities will be crucial. Operando spectroscopies and modeling at these settings are fundamental enabling tools that can help in that direction. Quantifying the catalyst dissolution dynamics through inline ICP-MS studies, under relevant PEMWE target conditions, is crucial to further understand the local dissolution-redeposition equilibria. Combined with morphology and elemental analysis, this could help elucidate the dissolution pathways through accelerated stress tests (*52*).

Progressing toward fully Earth-abundant metal PEMWE catalysts—including the cathodic side—and recyclable systems demands further innovation, spanning materials, electrode structures, membrane-electrode interfaces, and process control. Whereas this study focuses on abundant Co-based OER catalysts, further advances in alternative materials (e.g., Mn and Ni), which consider geopolitical barriers and environmental aspects related to metal extraction and purification, remain urgently needed.

REFERENCES AND NOTES

1. S. Chu, A. Majumdar, *Nature* **488**, 294–303 (2012).
2. I. Staffell *et al*., *Energy Environ. Sci.* **12**, 463–491 (2019).
3. Z. W. Seh *et al*., *Science* **355**, eaad4998 (2017).
4. M. Carmo, D. L. Fritz, J. Mergel, D. Stolten, *Int. J. Hydrogen Energy* **38**, 4901–4934 (2013).
5. C. Spöri, J. T. H. Kwan, A. Bonakdarpour, D. P. Wilkinson, P. Strasser, *Angew. Chem. Int. Ed.* **56**, 5994–6021 (2017).
6. L. An *et al*., *Adv. Mater.* **33**, e2006328 (2021).
7. P. C. K. Vesborg, T. F. Jaramillo, *RSC Adv.* **2**, 7933–7947 (2012).
8. L. Yang *et al*., *Nat. Commun.* **9**, 5236 (2018).
9. Y. Pi, Q. Shao, P. Wang, J. Guo, X. Huang, *Adv. Funct. Mater.* **27**, 1700886 (2017).
10. S. Cherevko *et al*., *ChemCatChem* **6**, 2219–2223 (2014).
11. E. A. Paoli *et al*., *Chem. Sci.* **6**, 190–196 (2015).
12. A. Grimaud *et al*., *Nat. Chem.* **9**, 457–465 (2017).
13. Z.-Y. Wu *et al*., *Nat. Mater.* **22**, 100–108 (2023).
14. L. C. Seitz *et al*., *Science* **353**, 1011–1014 (2016).
15. S. Hao *et al*., *Nat. Nanotechnol.* **16**, 1371–1377 (2021).
16. Y. Wang *et al*., *ACS Appl. Mater. Interfaces* **12**, 37006–37012 (2020).
17. H. Liu *et al*., *Joule* **7**, 558–573 (2023).
18. M. W. Kanan, D. G. Nocera, *Science* **321**, 1072–1075 (2008).
19. J. B. Gerken *et al*., *J. Am. Chem. Soc.* **133**, 14431–14442 (2011).
20. Z.-F. Huang *et al*., *Nat. Energy* **4**, 329–338 (2019).
21. B. Zhang *et al*., *Nat. Catal.* **3**, 985–992 (2020).
22. B. Zhang *et al*., *Science* **352**, 333–337 (2016).
23. S. Pan *et al*., *Nat. Commun.* **13**, 2294 (2022).
24. J. Huang *et al*., *Nat. Commun.* **12**, 3036 (2021).
25. L. Chong *et al*., *Science* **380**, 609–616 (2023).
26. A. Li *et al*., *Angew. Chem. Int. Ed.* **58**, 5054–5058 (2019).
27. I. A. Moreno-Hernandez *et al*., *Energy Environ. Sci.* **10**, 2103–2108 (2017).
28. I. C. Man *et al*., *ChemCatChem* **3**, 1159–1165 (2011).
29. J. S. Mondschein *et al*., *Chem. Mater.* **29**, 950–957 (2017).
30. M. Etzi Coller Pascuzzi, M. van Velzen, J. P. Hofmann, E. J. M. Hensen, *ChemCatChem* **13**, 459–467 (2021).
31. N. Wang *et al*., *J. Am. Chem. Soc.* **145**, 7829–7836 (2023).
32. M. Chatti *et al*., *Nat. Catal.* **2**, 457–465 (2019).
33. P.-Y. Cheng *et al*., *J. Colloid Interface Sci.* **623**, 327–336 (2022).
34. B. Rodríguez-García *et al*., *Sustain. Energy Fuels* **2**, 589–597 (2018).
35. Y. Wen *et al*., *Nat. Commun.* **13**, 4871 (2022).
36. J. Rossmeisl, Z.-W. Qu, H. Zhu, G.-J. Kroes, J. K. Nørskov, *J. Electroanal. Chem.* **607**, 83–89 (2007).
37. Y. Ping, R. J. Nielsen, W. A. I. I. I. Goddard III, *J. Am. Chem. Soc.* **139**, 149–155 (2017).
38. L. F. Shen *et al*., *Angew. Chem. Int. Ed.* **59**, 22397–22402 (2020).
39. Y.-H. Wang *et al*., *Nature* **600**, 81–85 (2021).
40. S. Kong *et al*., *Nat. Catal.* **7**, 252–261 (2024).
41. L. Gao *et al*., *Chem. Mater.* **32**, 1703–1747 (2020).
42. S. E. Kelch *et al*., *J. Phys. Chem. C* **125**, 26552–26565 (2021).
43. K. Oka *et al*., *Sci. Rep.* **9**, 223 (2019).
44. Q. Hu, H. Zhao, S. Ouyang, *Phys. Chem. Chem. Phys.* **19**, 21540–21547 (2017).
45. C.-Q. Li *et al*., *J. Mater. Chem. A* **7**, 17974–17980 (2019).
46. A. Moysiadou, S. Lee, C.-S. Hsu, H. M. Chen, X. Hu, *J. Am. Chem. Soc.* **142**, 11901–11914 (2020).
47. C. Jing *et al*., *ACS Catal.* **12**, 10276–10284 (2022).
48. M. Zhang, M. de Respinis, H. Frei, *Nat. Chem.* **6**, 362–367 (2014).
49. H. Liu *et al*., *Sci. Rep.* **7**, 11421 (2017).
50. S. Geiger *et al*., *Nat. Catal.* **1**, 508–515 (2018).
51. C. Wei *et al*., *ACS Catal.* **13**, 14058–14069 (2023).
52. M. Milosevic *et al*., *ACS Energy Lett.* **8**, 2682–2688 (2023).
53. H. Benzidi, Co-del_OER, *ioChem-BD* (2024); https://doi.org/10.19061/iochem-bd-1-320.
54. R. Ram, F. P. García de Arquer, STEM and XRD files of Water and Hydroxide Trapping in Cobalt Tungstate for Proton Exchange Membrane Water Electrolysis, Data set, Zenodo (2024); https://doi.org/10.5281/zenodo.11161891.

ACKNOWLEDGMENTS

Funding: ICFO receives funding from the Fundació Cellex, Fundació Mir-Puig, the La Caixa Foundation (100010434, European Union Horizon 2020 Marie Skłodowska-Curie grant agreement 847648), and the 2022 Leonardo Grant for Researchers in Physics from the BBVA Foundation. The ICIQ receives funding from the Spanish Ministry of Science and Innovation (ref. no. PID2021-122516OB-I00), and the Barcelona Supercomputing Center (BSC-RES) is acknowledged for providing computational resources. ICN2 receives funding from Generalitat de Catalunya through 2021SGR00457, IU16-014206 (METCAM-FIB); the Advanced Materials program supported by the MCIN with European Union NextGenerationEU (PRTR-C17.I1) and Generalitat de Catalunya funds; NANOGEN (PID2020-116093RB-C43) and RED2022-134508-T (CAT&SCALE), funded by MCIN/AEI/10.13039/501100011033/; and "ERDF A way of making Europe," by the European Union. A.G.M. acknowledges grant RYC2021 - 033479. The authors acknowledge the use of instrumentation and advice provided by the Joint Electron Microscopy Center at ALBA (JEMCA) and the Diamond Light Source for access to the B18 and I18 beamlines. ICFO (CEX2019-000910-S), ICIQ (CEX2019-000925-S), and ICN2 (CEX2021-001214-S) receive Severo Ochoa Center of Excellence (MCIN/AEI/10.13039/501100011033) funding and funding from CERCA and Generalitat de Catalunya. S.G. acknowledges PID2020-116093RB-C41 funded by MCIN/AEI/10.13039/501100011033. C.A.M. thanks the Generalitat Valenciana (APOSTD/2021/251) and the MinCiencias Colombia through the Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación "Francisco José de Caldas," call 848-2019. E.P. acknowledges support from the CNRS and the French Agence Nationale de la Recherche (ANR), under grant ANR-22-CPJ2-0053-01. This work was funded/cofunded by the European Union (ERC, PhotoDefect, 101076203). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. **Author contributions:** F.P.G.d.A. proposed and supervised the project. R.R. designed and conducted the experiments, data analysis of electrochemical experiments, crystallography, microscopy, and spectroscopic experiments. L.X. led the PEMWE implementation, data analysis, and illustrations. H.B., P.S.B., and N.L. carried out DFT calculations. A.G. and A.P.-S. optimized the in situ Raman spectroscopy setup. A.G. performed in situ Raman spectroscopy experiments and electrochemical data analysis. V.G. performed SEM. A.G.M., B.M., D.L.R., and J.A. conducted and analyzed STEM experiments. V.C., C.A.M., E.P., and S.G. carried out XAS experiments and data analysis. M.D. carried out AFM experiments and data analysis. A.M.D. performed XRD. R.R., L.X., A.G., and F.P.G.d.A. composed the manuscript. All authors discussed the results and edited and commented on the manuscript. **Competing interests:** R.R., L.X., A.G., and F.P.G.d.A. are inventors on patent application EP147926 submitted by ICFO that covers the OER catalyst reported here. The authors declare no other competing interests. **Data and materials availability:** A dataset collection of computational results is available in an ioChem-BD repository (*53*). STEM and XRD data files are available on Zenodo (*54*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk9849
Supplementary Text
Figs. S1 to S81
Tables S1 to S22
References (*55*–*67*)
Movie S1

Submitted 21 September 2023; resubmitted 15 January 2024
Accepted 13 May 2024
10.1126/science.adk9849

By **Carlo G. Quintanilla**

Science and fiction

As I stepped out of the punishing Arizona heat and into the cool air-conditioned halls of the convention center, I saw a sea of costumed attendees. Some wore elaborate steampunk attire; others portrayed their favorite Marvel or *Star Wars* characters. "Why did I agree to this?" I thought as I made my way to the room where I'd be giving a talk about the science behind the classic *Dune* books by Frank Herbert. I had been struggling for years to find new ways to communicate science to a broad audience. My hope was to inspire admiration and awe of what science can teach us about the world through the imaginative lens of science fiction. At first I felt like an imposter at this pop culture convention. Then I saw the audience's excitement.

Going into my Ph.D. program, I knew I loved to communicate science to the public as much as I liked doing science. But research soon began to consume most of my time. I often worked well past nightfall, and my weekends were largely absorbed by preparing experiments, gathering data, or catching up on the seemingly endless pile of scientific papers on my desk.

When I spoke about my research, it was mostly directed at people within my Ph.D. bubble: colleagues, collaborators, thesis committee members. The few times I presented to broader audiences, I struggled to explain how my research might apply in the real world. The more absorbed I became in practicing science, the more I felt I was losing touch with the bigger picture of why it matters.

"It suddenly occurred to me that this was my path back to science communication."

The problem came to a head 5 years into my Ph.D. when I took a trip to Maine. I was in desperate need of a getaway, so I jumped at an opportunity to spend a weekend with some family and friends at a cabin. Many guests were voracious readers and excited to share stories about the latest books they'd consumed. I could vividly recall every detail of the papers piling up back in the lab, yet I couldn't remember the last book I had read for fun. And so I found myself mostly talking about my research. As I struggled to convey its broader relevance, I noticed that the enthusiasm in my voice waned with each new conversation. "Is my Ph.D. killing my love of science?" I thought. I knew I had been living in a scientific bubble, but that bubble seemed to grow smaller and smaller as the weekend progressed.

When I returned home, I resolved to dedicate more of my time to doing things I love outside the lab, especially reading for pleasure. My eldest brother had recommended a science fiction book, so I started with that. It was like nothing I had ever read before. I was blown away by the scientifically plausible technologies and scenarios the author had crafted within a wildly imaginative story line. It was refreshing to think about science in a new way.

I kept reading science fiction in the years after I finished my Ph.D. and moved on to a science policy fellowship. But I still hadn't gotten back to communicating science to the public. Then, one night I found myself in a conversation with my new policy colleagues about science fiction concepts and their ethical implications for society. One invited me to speak at the pop culture convention, which he had participated in for years.

When the day came to present, I felt apprehensive until I was behind the podium and saw the bright-eyed, eager attendees. After I finished, I was inundated with questions and comments that spilled over beyond the allotted time. It suddenly occurred to me that this was my path back to science communication. Since then, I've given more than 20 talks at those kinds of conventions, exploring concepts like genetic engineering, brain-machine interfaces, and mind-controlling microbes.

Given the various demands and pressures on early-career researchers, it is easy to slide into being trapped in a scientific bubble. I'm glad science fiction gave me a way to break out of it. It not only provided something fun and fulfilling to do in my spare time. It also helped me professionally, fueling my ability to communicate the wonders of science to people I would have never spoken to otherwise. ■

Carlo G. Quintanilla is a health science policy analyst at the National Institute of Child Health and Human Development. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER


Leaders in scientific innovation.
Pioneers in medical discovery.
Scientists and clinicians in immunology at the Icahn School of Medicine at Mount Sinai are transforming patient care. Find out how in "Frontiers of Medical Research: Immunology," a special supplement included with this issue of *Science*.
Listen to Dr. Miriam Merad, inaugural Dean for Translational Research and Therapeutic Innovation, and Dr. Brian Brown, Director of the Icahn Genomics Institute, on the June 13 Science magazine podcast **www.science.org/podcasts**
WE FIND A WAY
Icahn School of Medicine at Mount Sinai