Women in police departments improves outcomes for women pp. 150 & 191

Wide-ranging noise from deep-sea mining p. 157

Microbial modulation wins NOSTER & *Science* Microbiome Prize p. 164

Science

$15
8 JULY 2022
science.org

SECRET PARTNERS

Hunting the fungi that sustain global ecosystems

p. 142

EDITORIAL

Inclusion doesn't lower standards

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

The cultural wars raging across the United States have sadly found their way into the world of science. Some university science faculty and administrators are resistant to making changes that would allow more students from underrepresented groups to participate and thrive in the sciences. The rationale for this opposition is often that "accommodating" legitimate social and pedagogical needs of marginalized groups will lower the standards of mastery and excellence in these fields. But this concern is just a crutch that protects faculty and institutions from having to do the work of correcting social injustices in higher education.

It's common to hear that improving student diversity in higher education requires lowering the bar to admission and watering down the curriculum so that all students can pass the course of study. I'm not aware of anyone who is advocating such a trade-off. There are known methods of teaching that allow more people from different backgrounds to master scientific material without compromising the quality of education. These include a greater use of active learning methods that engage students with course material through discussions and problem solving (as opposed to passively taking in information). Making such reforms may require faculty to learn new ways of teaching. But isn't that the job—to foster education for everyone?

> "...everyone should have more optimism about who can become a scientist."

Another common refrain is that understanding science requires a high degree of skill in mathematics. I've heard firsthand from faculty that students can't pass their classes unless they have previously achieved a high score on standardized tests in math such as the SAT or ACT. That is a breathtakingly pessimistic view. These high scorers are often students who've had the opportunities and resources to prepare for pre-college exams, which vast numbers of students have difficulty accessing. Isn't the whole point of teaching to provide a pathway to achievement?

Social psychologist Claude Steele and others have demonstrated repeatedly that stereotype threat—in which symbols and signals of exclusion create negative psychological effects—measurably impairs learning. Names of buildings and sports teams that honor racism and racists send a discouraging message to marginalized groups. If removing symbols of exclusion is helpful, why wouldn't a university want to do that?

The most corrosive aspect of faculty members' resistance to change is their response to students' requests for the use of inclusive language and terminology: "Can you believe this student asked me not to say this?" Apparently, faculty cannot be inconvenienced to modify language in a way that creates a supportive learning environment. Rather than mocking the students for asking, educators should be inspired to learn new ways of communicating that make it easier for all students to learn. Institutions need to take responsibility for the needed changes and not put the burden on students. Some proposed solutions have even involved sending all the students to cognitive behavioral therapy to make them less sensitive. Why should they be the ones to do the work needed to fix a dysfunctional world they did not create?

The University of Chicago has engaged in a very effective public relations campaign over the past several years to brand itself as the "free-speech university." It has issued many bold statements from faculty and administrators that welcome all ideas about controversial issues and create a space for conservative ideas to be shared freely. Many faculty opposed to diversity initiatives have praised the stance and assume that the university does not favor diversity initiatives or staff who are devoted to equity and inclusion. But the institution's website for the division of student affairs reveals many identity-based resources, such as a team that helps students report and respond to bias incidents, a center for identity and inclusion, and much more. The University of Chicago understands that all students need support and resources to learn. I don't think anyone would accuse this institution of lowering its standards.

Opening the doors to science for everyone requires that faculty learn the most effective methods for teaching a diverse student body. Yes, it's more work on top of the many other faculty duties, so universities must provide resources to make the adjustments, such as revamping classrooms for active learning, providing time for faculty to redo their curricula, and doing the hard work involved in having the faculty and institution make the cultural changes that students need. And everyone should have more optimism about who can become a scientist.

– H. Holden Thorp

10.1126/science.add7259

PHOTO: CAMERON DAVIDSON

NEWS

> “We just hope we didn’t open a Pandora’s box.”
>
> **Artificial intelligence scientist Almira Osmanovic Thunström**, in *Scientific American*, discussing research articles written by AI algorithms, including one she submitted to a journal.

IN BRIEF

Edited by **Shraddha Chakradhar**

RESEARCH SECURITY

New debate over secrecy

Science has learned that the U.S. National Science Foundation (NSF) has asked the National Academies to take a fresh look this fall at a Cold War-era presidential directive that regards openness in basic research as a boon to both innovation and national security. Advocates of classifying as little information as possible say additional restrictions would harm U.S. research without deterring countries seen as adversaries. China’s aggressive pursuit of several emerging technologies with both commercial and military uses, however, has prompted calls from many lawmakers to cordon off the results of basic research on a range of sensitive technologies. A 2019 report to NSF backed the current policy, issued in 1985 and known as NSDD-189. But NSF officials say the world has changed enough for experts across academia, government, and industry to reexamine the issue.

Court limits EPA’s climate power

CLIMATE POLICY | Siding with a group of Republican-led states and the electric power industry, the U.S. Supreme Court last week limited the Environmental Protection Agency’s (EPA’s) ability to regulate carbon emissions from existing power plants. Analysts say the 6-3 decision in *West Virginia* v. *EPA* will hamper President Joe Biden’s efforts to achieve his goal of substantially reducing U.S. carbon emissions and give more power to states. Notably, however, the court’s conservative majority did not use the case to challenge EPA’s underlying authority to regulate greenhouse gases, especially carbon dioxide, the primary planet-warming gas. But federal courts are still considering other lawsuits that would further reduce the agency’s options for regulating carbon dioxide emissions.

New water lily species named

ECOLOGY | Researchers have found a new species of water lily, and its leaves—as wide as a pingpong table is long—are the world’s largest, they report in *Frontiers in Plant Science*. *Victoria boliviana,* so named because it grows the biggest in Bolivia, was shown to be a distinct species with genetic analysis and detailed morphological observations of specimens at the National Herbarium of Bolivia and at the Royal Botanic Gardens Kew in England, where the new species is on display.

Scientists have discovered and named a new species of giant water lily.

PHOTO: ROYAL BOTANIC GARDENS, KEW

Colombia hits conservation goal

CONSERVATION | Colombia is the first Western Hemisphere country to achieve the widely adopted 30x30 ocean conservation goal, Colombian President Iván Duque claimed last week at the United Nations Oceans Conference in Lisbon, Portugal. More than 100 other countries have signed the pledge to protect 30% of both Earth’s ocean area and land area by 2030. To reach that proportion of protected ocean area in Colombia 8 years before the deadline, Duque designated three new marine protected areas and expanded the existing Malpelo Fauna and Flora Sanctuary in the Eastern Pacific Ocean to include a 1400 kilometer underwater ridge that acts as a highway for sharks and other marine life. Duque has also announced a $245 million initiative that aims to protect 32 million hectares of Colombian land and ocean.

Germany military research urged

RESEARCH FUNDING | German universities should undertake more military research, the German national academy of engineering, Acatech, said in a white paper released last month. The Russian attack on Ukraine has prompted a shift in Germany’s restrained approach to defense, predominant since World War II, with the legislature recently approving €100 billion in new military funding. Many German universities have long had a voluntary ban on military and dual-use research, but that position is out of date since the war in Ukraine, the 24 June report

Swallow-tailed hummingbirds are among the species found to have a plethora of new feather colors.

ECOLOGY

Scientists find new hummingbird colors

The plumage of hummingbirds has more color diversity than the feathers of all other birds combined, a recent study finds. Researchers from Yale University collected feathers from specimens of 114 hummingbird species and, using a spectrometer, documented the wavelengths of light they reflected. These wavelengths were then compared with those found in a previous study of 111 other bird species, including penguins and parrots. The researchers were surprised to find new colors in the hummers, which widened the known avian color gamut by 56% and included rarely seen saturated greens and blues, they report in *Communications Biology*. The newfound variation largely includes colors in the ultraviolet scale that are invisible to humans and probably only seen by hummingbirds themselves. Researchers note the variation is likely due to the reflective qualities of nanostructures present in the small barbs that protrude from the end of each hummingbird feather. The new colors were mostly found on the crowns and throats of the birds, suggesting a role in mating displays and communication.

says. It calls for German science to focus equally on security, resilience, and sustainability and for a broad debate about what research is needed for the country's security.

Protections overlook at-risk trees

BIODIVERSITY | Initiatives to conserve trees largely focus on nonthreatened species, according to an analysis presented last week at the World Biodiversity Forum. The work is an update to the *State of the World's Trees* report, originally published in September 2021, that showed one-third of the 60,000 tree species in the world are at risk of extinction, fueled by clearing land for agriculture, logging, and climate change. The new analysis reveals that areas marked for conservation, such as national parks, protect 85% of nonthreatened tree species, compared with only 56% of threatened ones. The same is true for trees in scientific collections, such as botanical gardens and seed banks; only 21% of threatened species are safeguarded there versus 45% of nonthreatened ones. Researchers worry the lack of threatened trees in collections might hinder conservation and restoration programs aimed at growing and preserving at-risk species. Among the most vulnerable trees worldwide are the *Magnolia ekmanii* from Haiti and the *Diospyros egrettarum*, which has fewer than 10 specimens left in its native Mauritius.

Next-gen COVID-19 jabs coming

COVID-19 | Redesigned COVID-19 booster shots are expected this fall, after advisers to the U.S. Food and Drug Administration (FDA) voted 19-2 last week in favor of incorporating an Omicron strain into existing vaccines. FDA concurred, announcing on 30 June that it's asking manufacturers to retool their shots for a spike protein component shared by Omicron BA.4 and BA.5, which are currently gaining ground worldwide. The new vaccines, which Pfizer and Moderna say they can make available by about October, will have the same messenger RNA dose as previous boosters but will target both Omicron and the original coronavirus strain first detected in 2019. Pfizer and Moderna have reported data from ongoing trials of vaccines against an earlier Omicron strain, BA.1, and FDA says it expects to consider these data in evaluating fall boosters.

Airline ends monkey transports

ANIMAL MODELS | Air France will stop transporting nonhuman primates once its current contracts end. The airline was the last major holdout on carrying nonhuman primates as cargo. People for the Ethical Treatment of Animals lauded the decision, but the European Animal Research Association cautions it will restrict research, especially as the European Union prepares to introduce further limits on primate studies later this year. The U.S. already faces shortages of monkeys for research, in part because of a Chinese ban on monkey exports.

BY THE NUMBERS

350,000

Google searches in the United States for "abortion pill" and names of specific abortion medications between 1 May and 8 May, the week the U.S. Supreme Court's *Dobbs* v. *Jackson Women's Health Organization* draft ruling was leaked. (*JAMA Internal Medicine*)

PHOTO: LUIZ CLAUDIO MARIGO/MINDEN PICTURES

IN DEPTH

Demonstrators gathered outside the Supreme Court the morning after the court overturned *Roe v. Wade*, effectively outlawing abortion in many states.

PHOTO: T.J. KIRKPATRICK/THE NEW YORK TIMES/REDUX

WORKFORCE

Scientists decry reversal of U.S. abortion rights

For some, the ruling limits professional mobility and conference attendance

By **Katie Langin**

When the U.S. Supreme Court overturned *Roe* v. *Wade* on 24 June, eliminating the constitutional right to an abortion and handing decisions about access to state legislators, the response across the polarized country was swift and divided. Abortion opponents cheered the culmination of a long quest to restrict the practice. But large swaths of the public, including many scientists, decried the decision as a potentially deadly violation of human rights. "Abortion bans will kill people in lots of different horrible ways," tweeted Amanda Stevenson of the University of Colorado, Boulder, a researcher who studies abortion.

Some scientists also worried about direct effects on the research community. "It's going to really negatively impact science … if we have scientists actively avoiding half of this country, or all of it entirely," says Rosa Lafer-Sousa, a neuroscience postdoc in the Washington, D.C., area who is considering how the ruling will influence her upcoming faculty job search. She and others fear that the lack of an abortion option in many states will create hardship for some aspiring scientists who become pregnant. "I really worry that it's going to affect people's ability to write their own destiny," says an M.D.-Ph.D. student in Texas who requested anonymity.

The Supreme Court's reversal will likely be felt most strongly in groups that are already underrepresented in science, says Nicole Williams, the outreach director for the nonprofit 500 Women Scientists. "Being an African-American woman, and just knowing the stats—that Black birthing persons already experience high levels of pregnancy-related mortality—the overturning of *Roe* versus *Wade* is a death sentence for Black women scientists and birthing persons," including transgender men and nonbinary people.

These concerns have some scientists reconsidering their career plans and where they're willing to live and work. The Texas M.D.-Ph.D. student, for example, might leave the state once she finishes her program because of Texas's strict abortion laws. "It's really hard. … I love Texas," she says. Researchers across genders and career stages shared similar stories on social media, saying they would be leaving or not pursuing professional opportunities in abortion-restricting states.

Those seeking faculty positions face particular challenges, Lafer-Sousa says, because "you don't have a ton of choice about where you end up in the first place, and now there's even less choice if you cut out half the states and say, 'Well, I'm not willing to live there." But ultimately, she doesn't think she'd feel comfortable recruiting trainees to join her in a state that doesn't grant them reproductive autonomy. As a second-year Ph.D. student, she became pregnant unexpectedly after her contraception failed. Her subsequent decision to end the pregnancy was relatively easy, she says, because she wanted to focus on her education—and she wants that same freedom for others. "Being forced to carry an unwanted pregnancy to term during graduate school would have posed a significant burden on me and potentially derailed my career plans," she says.

Faculty members already established in affected states are wrestling with similar questions. "I had to take a mental health day just to process everything that happened and deal with the emotions," says an assistant professor in a biomedical field who is based in a southern state where abortion is now illegal in nearly all cases. Speaking with *Science* on the condition she remain anonymous, she says she's particularly concerned about students at her university who come from dis-

advantaged backgrounds and may not be able to afford to travel to another state if they need an abortion. "Will I ever have to … send a student of mine to a 'conference' in California?" she wonders. "Is that something that I need to start thinking about?"

She's not sure she wants to stick around to find out, but whether to leave is a hard decision. She's the only racial or ethnic minority in her department and she feels she could have more impact where she is than in more liberal states—"both in terms of my votes, but also in the mentoring and support I can provide the underrepresented minority trainees that exist here," she says. "I'm Latina. I'm a woman. And in science, both of those identities are not well represented." But come fall, she'll likely apply for faculty positions elsewhere. "The fear for my safety and my well-being—and even more so than that, that of my students—it weighs really heavily on me."

Some universities have issued statements expressing support for their students' and employees' reproductive rights. The M.D.-Ph.D. student, for instance, says her school administrators sent out an email assuring trainees and faculty they will be supported to the extent possible. But other universities have remained silent, frustrating those who want to know their employer is paying attention. The southern professor, for example, hasn't received any emails or statements from her university. It "pisses me off," she says.

Some scientists have also called on their professional societies to take action, particularly regarding conference locations, with some proposing boycotting meetings in states that ban abortion. Others have pushed back on that idea, saying it will only harm scientists in those states. But safety of attendees is at stake, boycott advocates say. "Any person capable of childbearing could have an emergency related to pregnancy at your conference," Northwestern University neuroscience postdoc Ana Vlasits tweeted. "Your event should not be held in a place where your childbearing colleagues might be put at risk."

Those concerns make sense to Catherine Alves, a social scientist based in Rhode Island who is currently pregnant. Last year, she had a miscarriage and took the drug misoprostol to help her body expel the fetal tissue. Her situation wasn't an emergency, but other pregnancy-related complications that are often treated with abortion drugs, such as ectopic pregnancies, can be fatal if they aren't addressed quickly. It's not clear whether those drugs will continue to be prescribed in emergency situations in all states. "As a pregnant person," Alves says, she wouldn't feel comfortable attending conferences in locations where, if something went awry, "I couldn't get medical care that I needed and that reflects my values." ■

MICROBIOLOGY

Dengue and zika viruses turn people into mosquito bait

To spread, pathogens drive mice, people to make odorant

By **Mitch Leslie**

The viruses that cause Zika and dengue fever can't get from person to person on their own—they need to hitchhike inside a mosquito. A study last week in *Cell* suggests they hail these rides by making their victims smell more attractive to the blood-suckers.

It's "a big advance," says mosquito neuroscientist Laura Duvall of Columbia University. The work shows that "infection with these mosquito-borne viruses can alter the way some people smell … to make them more likely to be bitten."

Malaria parasites were already known to make human hosts smell irresistible to mosquitoes. Whether the viruses that cause Zika fever and dengue fever, which together infect up to 400 million people every year, also meddle with a person's odor was unclear, however. These pathogens travel from person to person in *Aedes aegypti* mosquitoes, which transmit other viruses as well.

> **"These mosquito-borne viruses can alter the way some people smell."**
> **Laura Duvall**, Columbia University

To determine whether the insects are partial to individuals with Zika or dengue, microbiologist Gong Cheng of Tsinghua University and colleagues set up three interconnected cages for a mouse experiment. Into one cage they piped air that had blown across mice that were sick with the Zika virus. A second cage received air that had flowed over healthy mice. The team then added hungry mosquitoes to the third cage and let them choose where to hang out.

Seventy percent of the mosquitoes crowded into the cage receiving air from the Zika-infected rodents. The insects' distribution was equally lopsided when the air came from rodents with dengue fever rather than Zika. However, the mosquitoes didn't favor a particular cage when air from the infected animals' cages went through a filtration apparatus that trapped chemicals, suggesting the odor of the sick mice was key.

Cheng's team also wiped the armpits of healthy people and dengue fever patients with an absorbent material, isolated molecules that could become airborne, and dabbed them onto filter paper. Mosquitoes preferred the bouquet of dengue.

By analyzing molecules emanating from infected rodents, the researchers identified the ingredients in *eau* de Zika or dengue. Mice gave off larger quantities of 11 potential odorants when they got sick, and one, acetophenone, was a mosquito attractant. Rodents that were ill exuded about 10 times more acetophenone than their uninfected counterparts. Dengue patients also emitted more of the molecule than healthy people.

Certain bacteria that dwell on the skin are the main source of acetophenone. Skin cells normally keep their numbers in check with a protein called RELMα that kills the microbes. However, Cheng and colleagues found that mice infected with the Zika or dengue viruses cranked out much less RELMα, which might allow the bacteria to proliferate and change the animals' scent. And mosquitoes were less fond of mice that had consumed isotretinoin, a derivative of vitamin A that increases RELMα synthesis

"It's a very compelling paper," says Ring Cardé of the University of California, Riverside, who studies chemical ecology and insect behavior. But he cautions that other research teams have uncovered numerous odor molecules that draw *A. aegypti* mosquitoes to their victims, including lactic acid and ammonia. "It's not clear how this compound fits in with the known attractants."

Still, the results could "revolutionize" diagnosis of dengue and Zika, says James Logan of the London School of Hygiene & Tropical Medicine, who was part of the team that showed malaria parasites change people's skin chemistry. Instead of a blood test, an electronic "nose" that detects acetophenone could provide a diagnosis much more quickly and easily, Logan suggests. A company he founded is already developing sensors to identify malaria from body odor.

The findings suggest "a novel avenue" to battle these diseases by reducing human attractiveness to mosquitoes, Cheng says. For example, he and colleagues are testing isotretinoin and related compounds in dengue fever patients. ■

advantaged backgrounds and may not be able to afford to travel to another state if they need an abortion. "Will I ever have to … send a student of mine to a 'conference' in California?" she wonders. "Is that something that I need to start thinking about?"

She's not sure she wants to stick around to find out, but whether to leave is a hard decision. She's the only racial or ethnic minority in her department and she feels she could have more impact where she is than in more liberal states—"both in terms of my votes, but also in the mentoring and support I can provide the underrepresented minority trainees that exist here," she says. "I'm Latina. I'm a woman. And in science, both of those identities are not well represented." But come fall, she'll likely apply for faculty positions elsewhere. "The fear for my safety and my well-being—and even more so than that, that of my students—it weighs really heavily on me."

Some universities have issued statements expressing support for their students' and employees' reproductive rights. The M.D.-Ph.D. student, for instance, says her school administrators sent out an email assuring trainees and faculty they will be supported to the extent possible. But other universities have remained silent, frustrating those who want to know their employer is paying attention. The southern professor, for example, hasn't received any emails or statements from her university. It "pisses me off," she says.

Some scientists have also called on their professional societies to take action, particularly regarding conference locations, with some proposing boycotting meetings in states that ban abortion. Others have pushed back on that idea, saying it will only harm scientists in those states. But safety of attendees is at stake, boycott advocates say. "Any person capable of childbearing could have an emergency related to pregnancy at your conference," Northwestern University neuroscience postdoc Ana Vlasits tweeted. "Your event should not be held in a place where your childbearing colleagues might be put at risk."

Those concerns make sense to Catherine Alves, a social scientist based in Rhode Island who is currently pregnant. Last year, she had a miscarriage and took the drug misoprostol to help her body expel the fetal tissue. Her situation wasn't an emergency, but other pregnancy-related complications that are often treated with abortion drugs, such as ectopic pregnancies, can be fatal if they aren't addressed quickly. It's not clear whether those drugs will continue to be prescribed in emergency situations in all states. "As a pregnant person," Alves says, she wouldn't feel comfortable attending conferences in locations where, if something went awry, "I couldn't get medical care that I needed and that reflects my values." ■

MICROBIOLOGY

Dengue and zika viruses turn people into mosquito bait

To spread, pathogens drive mice, people to make odorant

By **Mitch Leslie**

The viruses that cause Zika and dengue fever can't get from person to person on their own—they need to hitchhike inside a mosquito. A study last week in *Cell* suggests they hail these rides by making their victims smell more attractive to the blood-suckers.

It's "a big advance," says mosquito neuroscientist Laura Duvall of Columbia University. The work shows that "infection with these mosquito-borne viruses can alter the way some people smell … to make them more likely to be bitten."

Malaria parasites were already known to make human hosts smell irresistible to mosquitoes. Whether the viruses that cause Zika fever and dengue fever, which together infect up to 400 million people every year, also meddle with a person's odor was unclear, however. These pathogens travel from person to person in *Aedes aegypti* mosquitoes, which transmit other viruses as well.

> **"These mosquito-borne viruses can alter the way some people smell."**
> **Laura Duvall**, Columbia University

To determine whether the insects are partial to individuals with Zika or dengue, microbiologist Gong Cheng of Tsinghua University and colleagues set up three interconnected cages for a mouse experiment. Into one cage they piped air that had blown across mice that were sick with the Zika virus. A second cage received air that had flowed over healthy mice. The team then added hungry mosquitoes to the third cage and let them choose where to hang out.

Seventy percent of the mosquitoes crowded into the cage receiving air from the Zika-infected rodents. The insects' distribution was equally lopsided when the air came from rodents with dengue fever rather than Zika. However, the mosquitoes didn't favor a particular cage when air from the infected animals' cages went through a filtration apparatus that trapped chemicals, suggesting the odor of the sick mice was key.

Cheng's team also wiped the armpits of healthy people and dengue fever patients with an absorbent material, isolated molecules that could become airborne, and dabbed them onto filter paper. Mosquitoes preferred the bouquet of dengue.

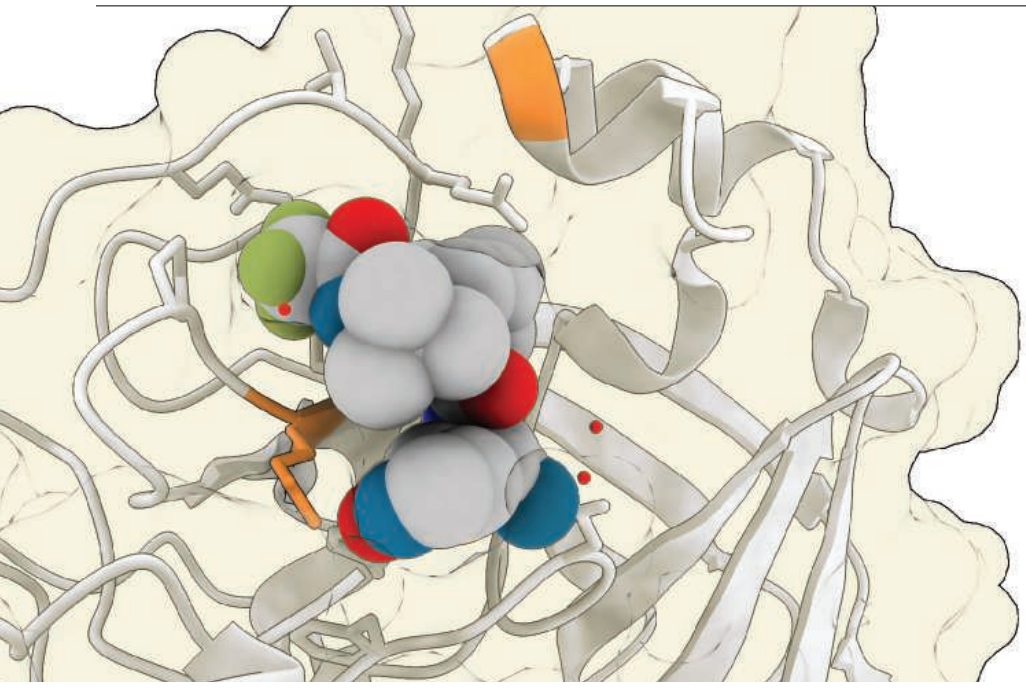
By analyzing molecules emanating from infected rodents, the researchers identified the ingredients in *eau* de Zika or dengue. Mice gave off larger quantities of 11 potential odorants when they got sick, and one, acetophenone, was a mosquito attractant. Rodents that were ill exuded about 10 times more acetophenone than their uninfected counterparts. Dengue patients also emitted more of the molecule than healthy people.

Certain bacteria that dwell on the skin are the main source of acetophenone. Skin cells normally keep their numbers in check with a protein called RELMα that kills the microbes. However, Cheng and colleagues found that mice infected with the Zika or dengue viruses cranked out much less RELMα, which might allow the bacteria to proliferate and change the animals' scent. And mosquitoes were less fond of mice that had consumed isotretinoin, a derivative of vitamin A that increases RELMα synthesis

"It's a very compelling paper," says Ring Cardé of the University of California, Riverside, who studies chemical ecology and insect behavior. But he cautions that other research teams have uncovered numerous odor molecules that draw *A. aegypti* mosquitoes to their victims, including lactic acid and ammonia. "It's not clear how this compound fits in with the known attractants."

Still, the results could "revolutionize" diagnosis of dengue and Zika, says James Logan of the London School of Hygiene & Tropical Medicine, who was part of the team that showed malaria parasites change people's skin chemistry. Instead of a blood test, an electronic "nose" that detects acetophenone could provide a diagnosis much more quickly and easily, Logan suggests. A company he founded is already developing sensors to identify malaria from body odor.

The findings suggest "a novel avenue" to battle these diseases by reducing human attractiveness to mosquitoes, Cheng says. For example, he and colleagues are testing isotretinoin and related compounds in dengue fever patients. ■



Lab studies show that one of the coronavirus' protein-cutting enzymes can mutate at several sites (orange) that make it more resistant to the active ingredient in Paxlovid (shown here as gray and colored spheres).

COVID-19

Bad news for Paxlovid? Resistance may be coming

In lab studies, SARS-CoV-2 finds ways to evade key drug. Some of the viral mutations are already found in people

By **Robert F. Service**

Prescriptions for Pfizer's blockbuster drug Paxlovid have skyrocketed in recent weeks. That's good news for many COVID-19 patients, as the pill has been proven to reduce severe disease. But a bevy of new lab studies show SARS-CoV-2 can mutate in ways that make it less susceptible to the drug, by far the more widely used of the two oral antiviral drugs authorized to treat COVID-19 in the United States. Researchers have found some of those mutations in variants already circulating in infected people, raising fears that physicians could soon lose one of their best tools for fighting the virus.

Taken together, the studies show that "when you put pressure on the virus it escapes," says David Ho, a virologist at Columbia University who was among the first to document drug resistance mutations in HIV some 3 decades ago. Ho was not involved with the new studies but is conducting similar work on SARS-CoV-2. Although the newly identified mutations have yet to become widespread, Ho and many other scientists think it's only a matter of time. "Given the amount of infections out there, it's going to come," Ho says.

The resistance studies come on the heels of other recent concerns about Paxlovid, which in the United States remains restricted to use in people with risk factors making them more likely to develop severe COVID-19. Confirming anecdotal reports widely reported by media, several studies have found a small percentage of infected people who receive the normal 5-day course initially feel better, only to have their symptoms rebound. And questions have grown about whether Paxlovid helps those who aren't at high risk of serious disease—Pfizer earlier this month halted a large trial of the drug in standard risk COVID-19 patients because it was failing to show statistically significant protection against death or hospitalization.

The U.S. Food and Drug Administration (FDA) granted emergency use authorization for Paxlovid in December 2021. The drug consists of nirmatrelvir, the active antiviral, and ritonavir, a compound that slows the breakdown of nirmatrelvir in the body. Because of bottlenecks in manufacturing nirmatrelvir, Paxlovid's rollout was slow—doctors in the United States issued only 40,000 or fewer prescriptions per week through mid-April. Since then, prescriptions have surged to more than 160,000 per week, according to the latest numbers from the Centers for Disease Control and Prevention.

That rise creates selective pressure on the virus, favoring mutations that help it survive in the presence of the drug. And because each infected person makes trillions of copies of SARS-CoV-2, the virus has plenty of opportunities to test out different mutations as it replicates. So far, those mutations don't seem to have interfered with Paxlovid's effectiveness. Nirmatrelvir prevents SARS-CoV-2's main protease (M^{pro}) from cutting a long precursor molecule made by the virus into shorter active proteins, an essential step in SARS-CoV-2's reproduction. In February, Pfizer researchers reported in *JBC Accelerated Communications* that nirmatrelvir remained effective in halting the activity of M^{pro} in multiple SARS-CoV-2 variants, including Alpha, Beta, Delta, Gamma, Lambda, and Omicron, as well as the original strain.

But like many antiviral drugs, Paxlovid may falter as the virus becomes resistant to it. Preprints posted on bioRxiv on 7 June, for example, show that SARS-CoV-2 grown in the lab quickly gains the ability to avoid nirmatrelvir's attack. Two research groups independently cultured the coronavirus with low levels of nirmatrelvir, killing some but not all of the virus. Such tests are meant to simulate what might happen in an infected person who doesn't take the whole regimen of the drug or an immunocompromised patient who has trouble clearing the virus.

One of those studies, led by Dirk Jochmans, a virologist at KU Leuven in Belgium, found that after 12 rounds of nirmatrelvir treatment, SARS-CoV-2 accumulated three mutations—at positions 50, 166, and 167 in the string of amino acids that make up M^{pro}–that reduced the virus' susceptibility to nirmatrelvir 20-fold, as determined by the dose of drug required to kill half the virus in a sample. The other study, led by Judith Margarete Gottwein, a virologist at the University of Copenhagen, also spotted potential resistance-conferring mutations at positions 50 and 166 in M^{pro}. When those mutations occurred together, the virus was 80 times less susceptible to nirmatrelvir. "This tells us what mutations we should be looking for [in patients]," Gottwein says.

CREDITS: (GRAPHIC) V. ALTOUNIAN/*SCIENCE*; (DATA) RCSB PRO

Indeed, some of these mutations are already in coronavirus-infected people, according to work by Adam Godzik, a bioinformatics expert at the University of California, Riverside. Godzik and his colleagues scoured the GISAID database, a catalogue of more than 10 million SARS-CoV-2 genomes sequenced from viruses isolated from infected individuals, searching for amino acid changes at positions in M^{pro} near where nirmatrelvir binds. In a bioRxiv preprint posted on 30 May, they reported that mutations to amino acids 166 and 167—two of the resistance mutations flagged by the Belgian group—were already in circulating viruses. Because these mutations occurred before widespread use of Paxlovid, they likely occurred randomly, Godzik says. However, he adds, they reveal the enzyme has some flexibility at these positions that could help the virus work around the drug.

And the list of potential resistance mutations keeps growing. In a paper posted last week on bioRxiv, Jun Wang, a medicinal chemist at Rutgers University, and colleagues report 66 common mutations to M^{pro} near the nirmatrelvir binding site. Like Godzik's team, they scanned the GISAID database to find altered versions of the protease, but then went a step further. Adding the gene for each of these variants of M^{pro} to *Escherichia coli* bacteria, they created supplies of the mutated enzymes for additional tests: first to determine whether each variant still carried out the essential duties of cutting viral proteins, and second to determine whether the mutations allowed M^{pro} to resist nirmatrelvir. Eleven of the 66 variants retained the protease's function (it was impaired in most of the others), and five of the 11 were resistant to nirmatrelvir, requiring at least a 10-fold increase in the drug to kill half the virus in the sample. One of those variants had a previously seen resistance mutation, at position 166, but the other four had novel workarounds at positions 144, 165, 172, and 192. The bottom line from all this work, Wang says: "It's just a matter of time before we see resistance emerge."

So, why is Paxlovid still working? One possibility is that not enough people have taken the drug yet to force the virus to mutate. Another explanation, Wang says, is that it may take multiple mutations in M^{pro} for the virus to get around Paxlovid while remaining both fully functional and easily transmissible. Thus far, adds Aditya Shah, an infectious disease specialist at the Mayo Clinic, studies show that when symptoms recur in COVID-19 patients on Paxlovid, which happens in just 2% or fewer of those who take the drug, the rebound does not seem to be due to resistance mutations. "It's reassuring," Shah says, but not proof the virus won't eventually find its way around the drug.

Pfizer says its Paxlovid regimen may forestall resistance. Patients only take the drug for a short period and typically get a dose "manyfold higher" than that required to prevent the virus from replicating in cells, thereby minimizing the opportunities for the virus to mutate, says Kit Longley, a company spokesperson.

Giving patients multiple antivirals could help prevent resistance by making it harder for the virus to evolve its way around different compounds at the same time, a strategy that has proved highly effective in treating other viruses, including HIV and hepatitis C, Ho says. Two other SARS-CoV-2 antivirals are authorized in the United States, but they have drawbacks. The other oral drug, molnupiravir, has proven considerably less effective than Paxlovid, and has raised safety concerns because it induces random genetic mutations in the virus—that typically stops it from replicating but could also spawn dangerous new variants, some scientists caution. And remdesivir, which interferes with the ability of the virus to copy its genome, is approved for use in patients with mild to moderate symptoms who are at risk of severe disease, but it must be delivered intravenously. A preprint posted on bioRxiv last week suggests combining molnupiravir and nirmatrelvir is more effective in combating SARS-CoV-2 infections than either antiviral given alone, at least in mice. But the strategy has yet to be widely embraced by doctors.

> **"When you put pressure on the virus, it escapes."**
> **David Ho**, Columbia University

Meanwhile, pharmaceutical companies are racing to complete clinical trials on additional SARS-CoV-2 antivirals, some targeting M^{pro} at different sites. But those aren't available yet. And numerous researchers, including representatives of the nonprofit Drugs for Neglected Diseases Initiative, have complained that Pfizer has not made Paxlovid easily available for trials of combination therapies. The company has said it plans to do those studies itself, although some are skeptical.

Until more antiviral drugs become available, Paxlovid will remain essentially alone, raising fears that sooner or later it will lose its punch. When pressed by a single antiviral, viruses usually find a way around the drug, Gottwein says. "If it can happen it will happen." ■

ECOLOGY

It takes a (microbial) village to make an algal bloom

More than nutrient levels may drive toxic lake growths

By **Elizabeth Pennisi**

Every summer, surges of toxic green muck plague lakes worldwide, sickening hikers who fail to purify drinking water, closing favorite swimming holes, and killing fish. The most feared—and studied—cause of these freshwater "algal" blooms is a genus of cyanobacterium called *Microcystis*. Its explosive summer growth is thought to be spurred by rising levels of phosphorus, nitrogen, and other nutrients, perhaps from fertilizer run off or other pollution sources. But new research, driven by advances in DNA sequencing, suggests other types of microbes also play key roles in these massive overgrowths.

According to one study, viruses killing off a main competitor of toxic *Microcystis* may help pave the way for blooms; another indicates nitrogen fixation by other bacteria may provide the needed boost. The results suggest that reducing nutrients may not be enough to stop these slimy explosions, some scientists say. That doesn't mean curbing pollution is unimportant, they stress, but ecological factors must be considered.

"Interspecies biological interactions help determine blooms," says Kevin Johnson, a marine scientist at the Florida Institute of Technology who was not involved in the work. "The more details we understand of bloom creation, the better our knowledge of how they might be prevented or controlled."

With the warming climate and continuing inflows of pollution, harmful algal blooms are on the rise, becoming more frequent and longer lasting in ever more places across the globe. They are "a pretty wicked problem," says Ariane Peralta, a microbial ecologist at Eastern Carolina University.

In some lakes, reducing fertilizer runoff at first seemed to thwart blooms—then they came back. Similar plans for bloom-choked Lake Erie might backfire, a team of academic microbiologists and water quality experts

Indeed, some of these mutations are already in coronavirus-infected people, according to work by Adam Godzik, a bioinformatics expert at the University of California, Riverside. Godzik and his colleagues scoured the GISAID database, a catalogue of more than 10 million SARS-CoV-2 genomes sequenced from viruses isolated from infected individuals, searching for amino acid changes at positions in M^{pro} near where nirmatrelvir binds. In a bioRxiv preprint posted on 30 May, they reported that mutations to amino acids 166 and 167—two of the resistance mutations flagged by the Belgian group—were already in circulating viruses. Because these mutations occurred before widespread use of Paxlovid, they likely occurred randomly, Godzik says. However, he adds, they reveal the enzyme has some flexibility at these positions that could help the virus work around the drug.

And the list of potential resistance mutations keeps growing. In a paper posted last week on bioRxiv, Jun Wang, a medicinal chemist at Rutgers University, and colleagues report 66 common mutations to M^{pro} near the nirmatrelvir binding site. Like Godzik's team, they scanned the GISAID database to find altered versions of the protease, but then went a step further. Adding the gene for each of these variants of M^{pro} to *Escherichia coli* bacteria, they created supplies of the mutated enzymes for additional tests: first to determine whether each variant still carried out the essential duties of cutting viral proteins, and second to determine whether the mutations allowed M^{pro} to resist nirmatrelvir. Eleven of the 66 variants retained the protease's function (it was impaired in most of the others), and five of the 11 were resistant to nirmatrelvir, requiring at least a 10-fold increase in the drug to kill half the virus in the sample. One of those variants had a previously seen resistance mutation, at position 166, but the other four had novel workarounds at positions 144, 165, 172, and 192. The bottom line from all this work, Wang says: "It's just a matter of time before we see resistance emerge."

So, why is Paxlovid still working? One possibility is that not enough people have taken the drug yet to force the virus to mutate. Another explanation, Wang says, is that it may take multiple mutations in M^{pro} for the virus to get around Paxlovid while remaining both fully functional and easily transmissible. Thus far, adds Aditya Shah, an infectious disease specialist at the Mayo Clinic, studies show that when symptoms recur in COVID-19 patients on Paxlovid, which happens in just 2% or fewer of those who take the drug, the rebound does not seem to be due to resistance mutations. "It's reassuring," Shah says, but not proof the virus won't eventually find its way around the drug.

Pfizer says its Paxlovid regimen may forestall resistance. Patients only take the drug for a short period and typically get a dose "manyfold higher" than that required to prevent the virus from replicating in cells, thereby minimizing the opportunities for the virus to mutate, says Kit Longley, a company spokesperson.

Giving patients multiple antivirals could help prevent resistance by making it harder for the virus to evolve its way around different compounds at the same time, a strategy that has proved highly effective in treating other viruses, including HIV and hepatitis C, Ho says. Two other SARS-CoV-2 antivirals are authorized in the United States, but they have drawbacks. The other oral drug, molnupiravir, has proven considerably less effective than Paxlovid, and has raised safety concerns because it induces random genetic mutations in the virus—that typically stops it from replicating but could also spawn dangerous new variants, some scientists caution. And remdesivir, which interferes with the ability of the virus to copy its genome, is approved for use in patients with mild to moderate symptoms who are at risk of severe disease, but it must be delivered intravenously. A preprint posted on bioRxiv last week suggests combining molnupiravir and nirmatrelvir is more effective in combating SARS-CoV-2 infections than either antiviral given alone, at least in mice. But the strategy has yet to be widely embraced by doctors.

> **"When you put pressure on the virus, it escapes."**
> **David Ho**, Columbia University

Meanwhile, pharmaceutical companies are racing to complete clinical trials on additional SARS-CoV-2 antivirals, some targeting M^{pro} at different sites. But those aren't available yet. And numerous researchers, including representatives of the nonprofit Drugs for Neglected Diseases Initiative, have complained that Pfizer has not made Paxlovid easily available for trials of combination therapies. The company has said it plans to do those studies itself, although some are skeptical.

Until more antiviral drugs become available, Paxlovid will remain essentially alone, raising fears that sooner or later it will lose its punch. When pressed by a single antiviral, viruses usually find a way around the drug, Gottwein says. "If it can happen it will happen." ■

ECOLOGY

It takes a (microbial) village to make an algal bloom

More than nutrient levels may drive toxic lake growths

By **Elizabeth Pennisi**

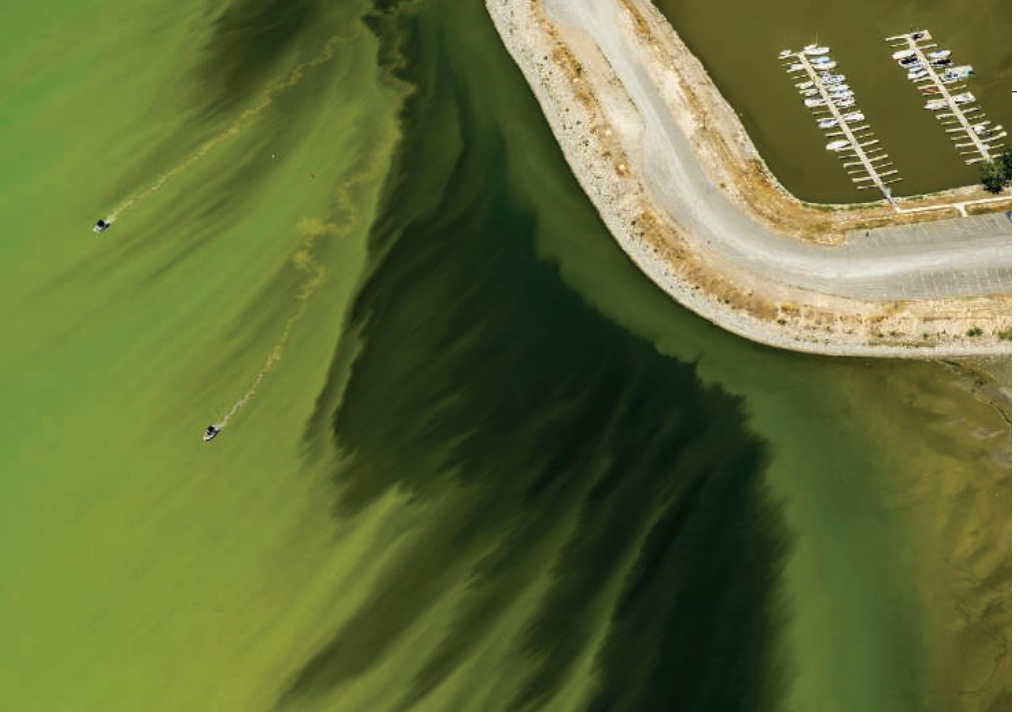
Every summer, surges of toxic green muck plague lakes worldwide, sickening hikers who fail to purify drinking water, closing favorite swimming holes, and killing fish. The most feared—and studied—cause of these freshwater "algal" blooms is a genus of cyanobacterium called *Microcystis*. Its explosive summer growth is thought to be spurred by rising levels of phosphorus, nitrogen, and other nutrients, perhaps from fertilizer run off or other pollution sources. But new research, driven by advances in DNA sequencing, suggests other types of microbes also play key roles in these massive overgrowths.

According to one study, viruses killing off a main competitor of toxic *Microcystis* may help pave the way for blooms; another indicates nitrogen fixation by other bacteria may provide the needed boost. The results suggest that reducing nutrients may not be enough to stop these slimy explosions, some scientists say. That doesn't mean curbing pollution is unimportant, they stress, but ecological factors must be considered.

"Interspecies biological interactions help determine blooms," says Kevin Johnson, a marine scientist at the Florida Institute of Technology who was not involved in the work. "The more details we understand of bloom creation, the better our knowledge of how they might be prevented or controlled."

With the warming climate and continuing inflows of pollution, harmful algal blooms are on the rise, becoming more frequent and longer lasting in ever more places across the globe. They are "a pretty wicked problem," says Ariane Peralta, a microbial ecologist at Eastern Carolina University.

In some lakes, reducing fertilizer runoff at first seemed to thwart blooms—then they came back. Similar plans for bloom-choked Lake Erie might backfire, a team of academic microbiologists and water quality experts



Toxins produced by this algal bloom in a lake in Utah sickened several people.

PHOTO: RICK EGAN/*THE SALT LAKE TRIBUNE*/AP IMAGES

funded by the National Science Foundation and other U.S. agencies reported in May. A 2014 bloom there caused such severe shortages of drinking water in the nearby city of Toledo, Ohio, that Canada and the United States have agreed to cut phosphorus going into the lake by 40%.

But a simulation of that strategy, along with an analysis of more than 100 related scientific papers, led the team to conclude that although limiting phosphorus might shrink Lake Erie blooms, they could also grow more toxic: with lower overall growth of microbes, any photosynthetic *Microcystis* left would receive more sunlight and have more nitrogen available, two conditions that favor an increase in their production of microcystin, a substance that make the blooms toxic (*Science,* 26 May, p. 1001). They suggested the lake's nitrogen should also be curtailed.

That simuation hinted that other microbes can indirectly influence the impact of *Microcystis*. But researchers studying blooms have tended to overlook lakes' many microbial inhabitants, which can include huge numbers of diatoms and other eukaryotes, as well as viruses and various types of bacteria, including smaller than average ones called picocyanobacteria. "Everyone glosses over them as not of managerial concern," says Cody Sheik, a microbial ecologist at the University of Minnesota, Duluth.

Part of the problem has been that it's been difficult to sort out which microbes are doing what in a lake. But Lauren Krausfeldt, a microbiologist at Nova Southeastern University, recently turned to metagenomics, a strategy of sequencing all the DNA in samples of water and other environments, to reconstruct the microbial ecosystem in Florida's Lake Okeechobee. The largest lake in the U.S. southeast, Okeechobee's annual summer blooms have begun to spread down rivers and spill into the Gulf of Mexico and Atlantic Ocean, forcing beaches to close. Between April and September in 2019, the bloom season, Krausfeldt and her colleagues collected multiple water samples at 21 places across the lake. From the fragments of DNA isolated from the samples and sequenced, they pieced together whole genomes belonging to specific species.

The analysis uncovered 30 kinds of cyanobacteria never before detected in the lake, and in some cases new to science, including 13 that could potentially cause blooms, she reported last month at Microbe 2022, the annual meeting of the American Society for Microbiology. "I was surprised at the diversity," Krausfeldt says.

When there was no bloom, the most common organisms were the picocyanobacteria. But as the season progressed, DNA belonging to bacterial viruses, known as phages, that infect the picocyanobacteria rose steeply. Shortly thereafter, the concentration of toxic *Microcystis* began to skyrocket. An analysis of its genome suggested why: *Microcystis* contains several antiviral defenses, such as the system that spawned the genome editor CRISPR, that picocyanobacterial lack. In addition, the bloom-forming cyanobacterium has genes that enable it to store nitrogen, a key nutrient, which may provide another competitive advantage over the many lake microbes that did not.

Krausfeldt suspects the phages lie dormant until some unknown environmental cue activates them. Then, after the viruses start slaying more and more picocyanobacteria, newly available nitrogen, phosphorus, and more light fuel a *Microcystis* bloom, Krausfeldt suggests. The phages' destruction of its hosts' cells may release even more nutrients, playing a key role in enabling algal blooms, she concludes.

Sheik, who says he had not considered phages as a factor in blooms but now wants to explore such viral dynamics, embraces Krausfeldt's ecosystem mindset. "By taking a holistic approach, we can better understand how supporting organisms can help sustain blooms," he says.

Sheik and his colleagues have also added metagenomics, as well as gene activity assessments, to his studies of several small lakes in Minnesota. Those lakes, he reported at the meeting, contain not only some *Microcystis*, but also another bloom-forming cyanobacterium called *Dolichospermum*. In 2020 and 2021, when he and colleagues tracked the microbial dynamics in one lake throughout the summer, they saw *Dolichospermum* become the most abundant microbe only to have its population crash by July. Nitrogen levels in the lake rose and fell in parallel with the microbe, suggesting it was fixing nitrogen and boosting its concentration in the water.

Nitrogen is usually quite scarce in these relatively pristine lakes, yet the nutrient is essential for the production of microcystin. That might explain why Sheik and his colleagues saw levels of *Microcystis* and its toxin rise after the bloom in nitrogen-fixing *Dolichospermum*. *Microcystis* must rely on other members of the freshwater ecosystem to fix nitrogen or to recycle it by breaking down other life forms, Sheik says.

"I'm blown away" by the metagenomic work, says Benjamin Wolfe, a microbiologist at Tufts University, because it can illuminate in great detail the lake's microbial interactions.

The case of *Dolichospermum* illustrates how complicated algal blooms can be. The good news, however, is that unlike in Europe, where this bacterium causes toxic blooms, *Dolichospermum* species in the United States lack the genes to make toxins—at least for now, says Sheik, who plans to keep watching for them in his metagenomic studies.

How the microbial dynamics that drive blooms can be interrupted is still unknown, and the picture is getting more complicated all the time. "We are grappling with understanding what parts of complex microbial communities are changing and what we can change to produce a different outcome," Peralta says. But she's optimistic that in time, "we can figure out what levers we can move." ■

EMPLOYMENT

Mass layoff looms for Japanese researchers

Thousands could see their jobs axed in the wake of labor law adopted a decade ago

By **Dennis Normile**

Thousands of researchers at Japanese institutes and universities may see their jobs disappear by next spring, an unintended result of labor legislation adopted a decade ago that gave researchers who have worked under fixed-term contracts for more than 10 years the right to permanent employment. Japan's science system has many such temporary workers—but rather than fully hire them, institutions are terminating their jobs.

Scientists are trying to head off the layoffs; the union for RIKEN, Japan's network of nationally supported laboratories, filed a protest with a Tokyo labor board last month and may take legal action. Regardless of the outcome, the dispute could create more upheaval in a research system whose global impact is already waning (*Science*, 27 May, p. 903.) "We are on the verge of seeing a possible mass dismissal of researchers this year," Tomoko Tamura, a member of the legislature's upper house, said during a May parliamentary question time on the issue. Tamura's analysis of government data suggests up to 4500 researchers are at risk, which "could have a serious long-term impact on Japan's research and development," she said.

SPring 8, one of the world's most powerful synchrotrons, is among the cutting-edge facilities RIKEN brought online during 3 decades of growth.

PHOTO: RIKEN

Japan's R&D funding grew rapidly in the 1990s and 2000s, but many newly recruited researchers were hired under fixed-term contracts, which offer lower pay, fewer benefits, and less job security than permanent jobs. The scheme gave research institutions more flexibility—but in practice, most fixed-term contracts were renewed indefinitely.

RIKEN is a prime example. Thirty years ago, it had about 400 researchers, most of them permanent employees working on basic physics and chemistry at the main campus near Tokyo. In the mid-1990s, Japan set out to roughly double government spending on research within 5 years, but the National Personnel Authority resisted increasing the number of employees on the national government payroll. Instead, RIKEN used project funds to hire many fixed-term employees. Today RIKEN has programs in brain science, quantum computing, and preventive medicine scattered among 10 branches and campuses, and it runs a powerful synchrotron and a peta-scale supercomputer. But 77% of its 2893 current researchers are fixed-term workers.

Legislation adopted in 2013 and amended in 2014 gave most contract employees the right to request permanent employment after working for the same employer for 5 years; for researchers, the term was set to 10 years. Many employers have responded by making sure contract workers never accumulate that duration of service.

RIKEN took that step in 2016, specifying that the count of years served starts in 2013. That means contract researchers who have already worked for RIKEN for more than 10 years may face termination next year. In an email to *Science*, RIKEN says 203 fixed-term researchers will reach the end of their final contracts before the end of March 2023. The institute is currently screening them and expects to make an unspecified number permanent employees, but many will have to leave. Among the vulnerable scientists are 42 team leaders whose groups will be disbanded if they go, which puts another 177 positions at risk. RIKEN says it hopes those who are forced out "will be able to continue their research activities at universities, research institutes, and private companies in Japan and overseas."

Applying an employment policy adopted in 2016 retroactively to those who have already worked under contract for 10 or more years, is "illegal," says Yasuyuki Kanai, the executive committee chairman of RIKEN's labor union. He says the researchers have a right to continued employment. Unhappy with the way RIKEN negotiated, the union on 20 June formally asked a governmental labor relations board to order the agency to bargain in good faith. With union support, "researchers are now preparing to take the matter to court," Kanai says. The union notes that fresh cohorts of several dozen fixed-term researchers annually will reach the term limit in the years ahead.

Other institutions face similar problems, although few have as many temporary contracts as RIKEN. Some are trying to find ways to retain their workers. The National Institute of Advanced Industrial Science and Technology has converted to permanent status all 245 fixed-term contractors who applied for it, according to press reports. Tohoku University is reportedly screening 275 fixed-term researchers for possible permanent employment. At the University of Tokyo, which has 588 fixed-term employees approaching 10 years of service, some might be moved to new projects, a spokesperson says, without providing details.

The broader problem is a scarcity of opportunities for researchers to change jobs in Japan, says Eisuke Enoki, who heads an Osaka-based organization that studies science policy. "The originally envisioned ideal was for academics to become assistant professors after one or two postdocs and to gain a permanent position if tenure is approved," he says. But a tenure system has never taken hold, and there are few permanent positions, even for team leaders with good track records, Enoki says.

A senior scientist at RIKEN who asked not to be identified agrees. His final contract is ending and it's "very difficult to find a new position," he says: "If I get a job in China, Korea, or Taiwan, I will move." The crisis underscores that for young people, in Japan "being a researcher is not an attractive profession," he adds. ■

FEATURES

Mycologist Toby Kiers marks a fungi sampling plot in Chile's Valdivia National Reserve.

A FUNGAL SAFARI

A new nonprofit has launched an ambitious effort to raise the profile of often invisible soil fungi

By **Gabriel Popkin**, *in Chile's Villarrica National Park*
Photography by **Mateo Barrenengoa L.**

As a motley medley of mycologists climbed the basalt slopes of the Lanín volcano earlier this year, the green foliage at lower elevations gave way to autumnal golds and reds. Chile's famed *Araucaria*—commonly called monkey puzzle trees—soon appeared, their spiny branches curving jauntily upward like so many cats' tails.

Beneath the majestic trees, the scientists were focused on something far less glamorous—indeed, mostly invisible: mycorrhizal fungi, tiny organisms that intertwine with roots of the *Araucaria* and nearly all the other plants in this forest. The multinational research team had come to collect soil samples they hoped would, with help from DNA testing, reveal exactly which fungi live here, and how they support this complex assemblage of flora. By the end of an exhausting day that included bushwhacking through heavy brush, the fungi hunters had filled seven small plastic sacks with dirt from different locations. "I wouldn't be surprised if there are 100 undescribed species" of fungi in each bag, said mycologist Giuliana Furci, founder of the Chilean nonprofit Fungi Foundation and one of the expedition leaders.

The April ascent was also a road test of sorts: the first of many surveys that the Society for the Protection of Underground Networks (SPUN), a new fungus-focused nonprofit, hopes to conduct. It has raised some $3.5 million for an ambitious effort to map the global distribution of mycorrhizal fungi, which can create subterranean networks that are thought to play a key, but often overlooked, role in shaping ecosystems.

"Up to 50% of the living biomass of soils are these networks," says ecologist Toby Kiers of the Free University of Amsterdam, a co-founder of SPUN and one of the leaders of the Chile expedition. "We have to figure out where they are and what they're doing."

SPUN's approach is bold—even bombastic. The project launched last fall with a media-savvy campaign, including slick videos, arguing that society must do more to study and protect fungi to safeguard biodiversity and curb climate change. SPUN researchers describe themselves as "myconauts" heading into the unknown. They wear customized blue jumpsuits emblazoned with "PROTECT THE UNDERGROUND" for publicity photos and when working in the field. ("I really love jumpsuits," Kiers says.) A documentary video crew followed SPUN scientists on their first expedition, into Chile. Celebrities such as primatologist Jane Goodall and best-selling author Michael Pollan have signed on as SPUN advisers.

> **"Up to 50% of the living biomass of soils are these networks"**
> **Toby Kiers**, Free University of Amsterdam

Some researchers, however, harbor doubts that SPUN's mapping effort will have much practical impact, noting that conservationists are already protecting forests and other ecosystems that harbor fungi and store planet-warming carbon. Others question whether the surveys will appreciably add to what scientists already know, in part because SPUN is only studying one segment of the fungal community: those that form associations with plant roots.

Mapping "a single class of microorganism seems to me too limited to come to an understanding of the big picture," says Heribert Hirt, a plant scientist at the King Abdullah University of Science and Technology. "I am rather skeptical that we will really learn a lot from this big science project."

But SPUN's effort to make soil fungi more visible is being welcomed by most mycologists, who often feel as overlooked as the organisms they study. "I don't think anything like this has ever happened before," says Kabir Peay, a mycologist at Stanford Univer-

sity who advises SPUN. "It's amazing there are philanthropists that have the vision and interest to support this type of activity."

APPRECIATED OR NOT, fungi are integral to Earth's ecosystems. They evolved hundreds of millions of years before land plants and animals. By breaking down rock and freeing up nutrients, they helped plants colonize land some 500 million years ago. To this day, most land plants access water and nutrients in part by partnering with mycorrhizal fungi that grow on—and often into—their roots. (Roughly speaking, "mycorrhiza" means "fungal root.") Some plants get up to 80% of their phosphorus—a vital nutrient—from fungi. And some fungi construct intricate underground webs known as mycelium that can stretch for kilometers. "Wherever there are roots," Kiers says, "there are fungi."

Despite their ubiquity and importance, however, fungi challenge biological paradigms and have defied easy description. The mushrooms that many people think of as classic fungi, for example, are just the spore-forming appendages of larger organisms that are typically hidden from view. Early biologists lumped fungi with plants, even though they don't photosynthesize. Only in 1969 did scientists recognize fungi as a separate kingdom of life. (Furci prefers "queendom.") It's also hard to define an individual fungus: "One" mycelium can hold many cell nuclei that don't always share the same DNA. So is the mycelium an individual, or is each nucleus?

Fungi are often relegated to second-class status within global scientific and conservation agendas. Whereas visually striking and charismatic species such as tigers, whales, and orchids have grabbed attention, fewer than 600 fungal species have had their conservation status assessed. "Fungi are seen as the subservient group" to plants, says Greg Mueller, chief scientist at the Chicago Botanic Garden. But some researchers have responded by flipping the script: Maybe, Mueller says, a bit tongue-in-cheek, "plants just exist to feed fungi."

Kiers, for one, has adopted a fungi-first worldview. She grew up in the United States and fell for fungi during a stay at the Smithsonian Institution's tropical research station in Panama. In her lab in Amsterdam, she uses tools such as microscopy and fluorescence to reveal how nutrients flow through mycelial networks. In one highly cited paper, Kiers' team showed one type of mycorrhizal fungus could reward individual plants that provided it with plentiful sugars by directing other nutrients to those plants' roots, while "punishing" stingier plants by withholding nutrients. Such

Three kinds of ectomycorrhizal fungi found in Chile hint at the immense diversity of these organisms, which associate with trees. From top to bottom: *Ramaria flava*, *Cortinarius lebre*, and *Laccaria tetraspora*.

PHOTOS: MATEO BARRENENGOA L.

results, Kiers says, demonstrate that fungi can wield real power and agency.

But such experiments drastically simplify the unruly networks that shuttle water and nutrients through natural ecosystems. That disconnect frustrated Kiers. "You're thinking: 'My God, how different is that from what's happening in the real world?'"

That question was in the air when, in September 2020, Kiers met American eco-logist Colin Averill over Zoom. Averill works in the Zürich-based lab of ecologist Thomas Crowther, which specializes in mapping the global distributions of organisms such as trees, nematodes, and mycorrhizal fungi, using computer algorithms to "fill in" areas lacking field data (*Science*, 25 October 2019, p. 412). Kiers was keen to connect insights from her idealized lab experiments to the global scale at which Averill works. The two eventually pitched the Boston-based Grantham Foundation on a global effort to discover and map mycorrhizal fungi, and harness them as a climate solution (see sidebar, p. 147). In November 2021, the foundation gave the researchers $3.5 million to launch SPUN.

THE GROUP'S TASK is daunting. Vast realms of the underground world have hardly been sampled, and scientists estimate that fewer than 10% of fungal species have been formally described—indeed, it could be as little as 1%, according to a recent paper in the journal *Fungal Diversity*.

In part, that's because studying fungi is really hard. They live mostly underground, and many are microscopic. Their most visible and familiar manifestation—the mushroom—is only made by certain types of fungi. By contrast, arbuscular mycorrhizal fungi, which associate with most of the world's plant species, nestle within the cell walls of plant roots. They are so hard to find and identify that scientists have described fewer than 300 kinds. "We can't even talk about diversity in a way that makes sense yet," Kiers says.

Some researchers have been chipping away at the problem. In 2014, a team led by mycologist Leho Tedersoo of the University of Tarfu in Estonia reported in *Science* on an analysis of soil samples from 365 sites on every continent except Antarctica—at the time, an unprecedented sampling effort that one news outlet described as "staggering." The research revealed, among other things, that fungal diversity does not always mirror that of plants. In other words, protecting just the richest aboveground ecosystems might fail to safeguard the full diversity of belowground life.

That survey effort "was just the very tip of the iceberg," says Tedersoo, who now advises SPUN. He subsequently launched a Global Soil Mycobiome consortium and started to email colleagues, begging them to dig up fungi, dry them, and send them to him. The result was a paper published late last year that analyzed samples from 3200 sites holding more than 700,000 "operational taxonomic units"—DNA sequences that could represent fungal species.

Even with that accomplishment, Peay says, "If you took the total volume of soil sampled by all fungal ecologists, it's still teeny." SPUN aims to paint a fuller picture by quickly doubling the number of fungal samples collected from documented locations. The group's leaders will collect some of the material themselves, but they also plan to fund and train a far-flung network of myconauts to sample their own regions.

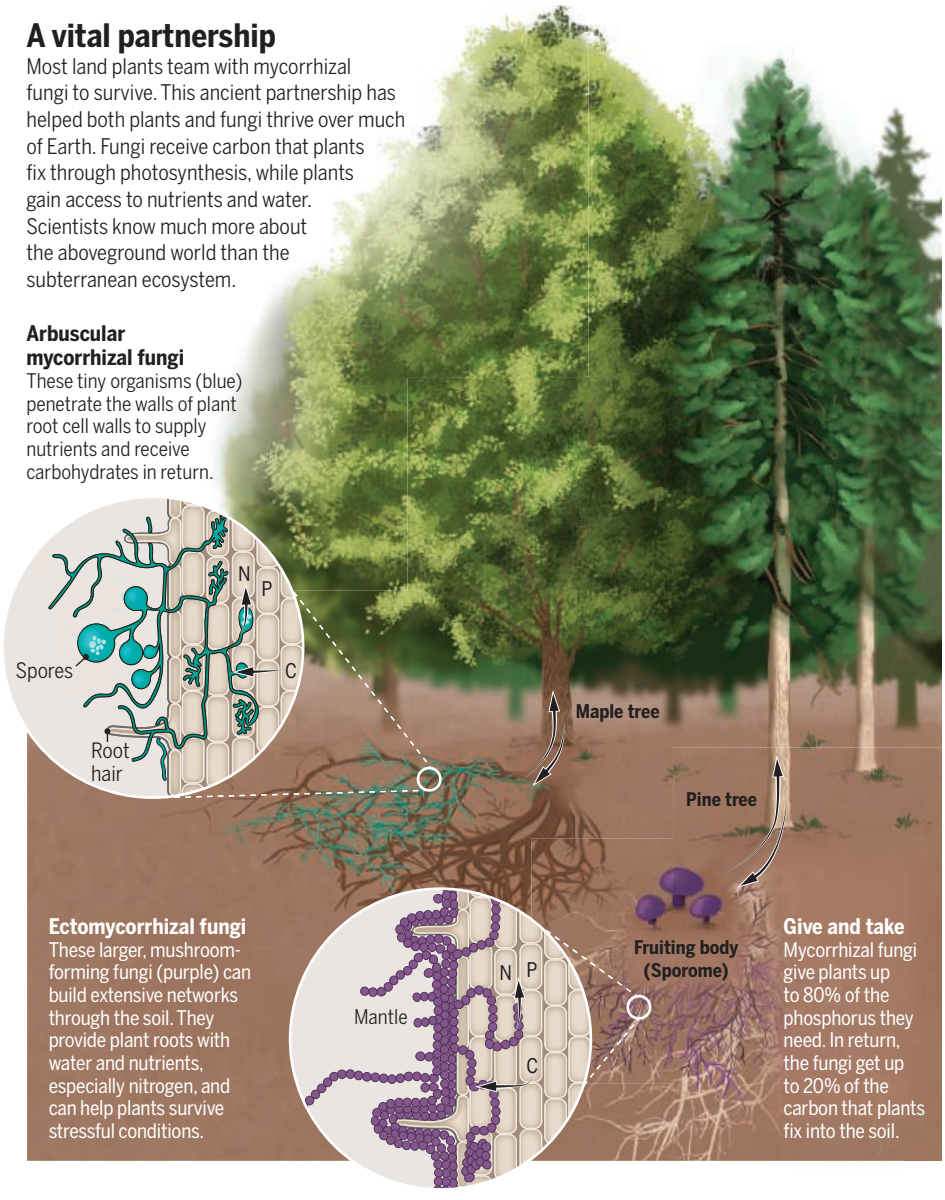
The SPUN data, which will be combined with Tedersoo's in an open repository, will help fill a key gap, Averill says: "There's no systematic baseline" data on the global distribution and diversity of soil fungi. "We want to build that baseline."

To guide the sampling, Kiers and Averill have harnessed machine learning algorithms developed by Johan van den Hoogen, a researcher in the Crowther lab. The software uses some 10,000 existing fungal records and a bevy of environmental data sets to tease out subtle correlations between where fungi live and variables such as aboveground vegetation, temperature, and rainfall. Those correlations can identify places with conditions favorable to fungi that have not been surveyed using modern DNA analysis.

The potential hot spots identified by SPUN include places such as the high steppes of Mongolia and the lowlands of the Congo River Basin. The modeling also highlighted another candidate: the old-growth forests of Chile.

A vital partnership

Most land plants team with mycorrhizal fungi to survive. This ancient partnership has helped both plants and fungi thrive over much of Earth. Fungi receive carbon that plants fix through photosynthesis, while plants gain access to nutrients and water. Scientists know much more about the aboveground world than the subterranean ecosystem.



GRAPHIC: N. DESAI/*SCIENCE*

Mycolgist Giuliana Furci, here with a long-lived Alerce tree in Chile, helped make the South American nation one of the first in the world to legally protect fungi.

SPUN LEADERS CHOSE the South American nation for their inaugural expedition for two reasons. One is that Chile holds ancient and globally unique forests. The Villarrica foothills, for example, are dominated by a mix of *Araucaria* and southern beech trees, whose roots could host unique assemblages of mycorrhizal fungi.

The other is that Chile is home to Furci, who has extensively surveyed the country's mushrooms. She is also the founder of the Fungi Foundation, one of the world's first charities devoted to protecting the underground kingdom. "A sense of injustice" spurred her to start the organization in 2012, Furci says. "At the time, there was nobody advocating for fungi in Chile."

For the trip, Kiers assembled a team that included Furci, independent mycologist Merlin Sheldrake, and his brother Cosmo Sheldrake, a U.K. musician and sound recorder. (Furci and Merlin are both on SPUN's advisory board.) The morning after the researchers trekked across the volcano, they clambered into a van and bumped up a dirt road toward a constellation of sampling spots identified by SPUN's modeling. Pulling over near a gap in a barbed wire fence, the team plunged into a grove of eucalyptus trees—an exotic species imported from Australia—that hid their destination. Every step required negotiating thickets of bamboo and thorny blackberries that caught on hair and clothes. "She emerges from the deep," Furci joked as Kiers used a GPS to navigate through a particularly treacherous stretch. "F*** the computer!" Kiers retorted.

To gather their treasure, the researchers donned blue plastic gloves to avoid contaminating the soil and swung into action. Merlin pounded a roughly quarter-meter-long metal cylinder into the ground with a rubber mallet, then extracted it. Once the dirt core was safely deposited into a plastic ziplock bag, he and Kiers trekked further to collect eight additional samples, forming a three-by-three grid 30 meters on a side.

Once analyzed, these samples could confirm or refute a hypothesis based on the computer modeling: that the eucalyptus grove is a "cold spot" with relatively few fungal species. But the researchers were eager to collect the soil anyway, because they wanted to see whether the roots of the tall, almost comically skinny eucalyptus trees hosted foreign fungi that had hitchhiked in with the Australian trees. Understanding how human-altered ecosystems like this one function is as important as understanding more natural forests, Kiers says.

Mycologists Toby Kiers and Merlin Sheldrake sample along the rocky coast of Chile earlier this year.

The expedition featured some unconventional episodes. Before taking samples on the volcano, for example, the researchers asked Cosmo Sheldrake to play a song, inspired by an Indigenous elder they had met the day before who advised offering music to fungi. The group fell quiet as he produced a penny whistle and played a haunting, ethereal melody followed by *Cooley's Reel*, a popular Celtic tune.

PHOTOS: MATEO BARRENENGOA L.

At another stop, Furci charmed a couple into allowing sampling in their front yard. By the time Kiers and Merlin were done, Furci was handing out licorice and hugs were exchanged all around. She even offered the couple a handful of dirt to smell. "I love it," the man said rapturously.

The next morning, the team headed west toward the Pacific coast to sample in Alerce Costero National Park, home to another old-growth forest dominated by a rare tree species, the alerce. In all, the researchers collected some 30 soil samples in just over a week of fieldwork. They handed them over to César Marín, a mycologist at Santo Tomás University in Chile, for DNA analysis. The results—including some from additional samples collected by Marín—will be fed back into SPUN's modeling in order to improve its predictions.

SPUN'S EFFORTS to map fungi in understudied places reflect a "pragmatic" approach that could help scientists better understand how complex ecosystems like forests work, says mycologist Justine Karst of the University of Alberta, Edmonton. She also lauds the team's efforts to excite the public—something she says most scientists don't prioritize. "I'm looking forward to seeing what they produce."

SPUN, meanwhile, hopes its analyses will also help bolster emerging efforts to protect fungi—and Furci's advocacy work in Chile could provide a template. Largely because of her lobbying, Chilean lawmakers a decade ago passed the world's first law formally protecting fungi nationwide.

Elsewhere, fungi are also beginning to capture official attention. Several European countries have taken steps to protect endangered fungal species, and Estonia has created a small preserve where several red-listed fungi live. (In the United States, by contrast, fungi have relatively low status; just two fungi—both lichens—are protected under the Endangered Species Act.)

Last fall, the International Union for the Conservation of Nature (IUCN), which maintains the global "red list" of threatened and endangered species, explicitly called for fungi to be given the same consideration as plants and animals, as did the conservation organization Re:wild. And in December 2021, the United Nations's Food and Agriculture Organization launched an International Network on Soil Biodiversity that includes research on fungi.

Mueller, who chairs IUCN's fungal conservation committee, and other advocates are now working to have the Convention on Biological Diversity—a multinational pact that helps set the global conservation agenda—explicitly embrace the goal of protecting fungi. And although SPUN "is not the only game in town" when it comes to lobbying for fungi, Mueller believes its well-funded public outreach could aid such efforts. "Having this bright light shining on the issue," he says, "will move a lot of initiatives forward."

Some influential groups, however, have yet to afford fungi as much consideration as flora and fauna. The Nature Conservancy—one of the world's largest conservation groups—"does not specifically target fungi" when prioritizing ecosystems for protection, says David Banks, the group's chief conservation officer. "But because we're working in these bigger [ecosystems], we can capture them." (The conservancy is nevertheless funding a SPUN-led trip to sample mycorrhizal networks on a remote Pacific island.)

Even SPUN supporters acknowledge that a more complete picture of fungi won't automatically lead to better real-world outcomes. "There is a gap between mapping and demonstrating patterns of biodiversity, and demonstrating a connection to conservation," Peay says.

Still, SPUN's leaders believe they can add a new, subterranean dimension to global conservation efforts. Kiers points to recent research suggesting that, over nearly one-third of Earth's land area, aboveground biological diversity doesn't match what's found beneath the surface. That means habitats often seen as relatively species-poor—such as boreal forests and drylands—could harbor far greater underground diversity than is currently recognized. Better maps of fungal diversity, she adds, could aid forestry, farming, and efforts to curb climate change.

Kiers recognizes, however, that the onus is now on SPUN to show its value before its first flush of funding runs out. "There has to be a demonstrable benefit to building an organization dedicated to protecting underground ecosystems," she says. "The clock is ticking." ■

By aiding trees, fungi might help curb warming

Could adding certain kinds of fungi to soil help curb climate change by enabling trees to grow faster and suck more carbon dioxide from the atmosphere? That's a question that researchers with the new Society for the Protection of Underground Networks (SPUN) are trying to answer.

In the spring of 2021, in an abandoned pasture in southwestern Wales, a forestry company planted 25,000 trees over 11 hectares for an experiment designed by mycologist Colin Averill, a co-founder of SPUN (see main story, p. 142). The plantations include sitka spruce, a common timber tree in the United Kingdom, and a mix of native deciduous trees. To half of the seedlings' roots, researchers added mycorrhizal fungi—microorganisms that help provide plants with nutrients by associating their roots—that were sourced from mature forests of the same type. Now, the researchers are waiting to see whether the treated trees grow faster and absorb more carbon than those that didn't get the treatment. A similar trial is ongoing in Yucatán, and a third will be planted this fall in Ireland.

The effort builds on research suggesting the right fungi can give arboreal growth a powerful boost. In a January paper published in *The ISME Journal*, Averill and colleagues reported growth rates of trees in forests across Europe vary by up to a factor of three depending on their fungal partners.

But just how much fungi can supercharge saplings is an open question. Companies have long sold mycorrhizae that tree planters can dab onto roots. But such commercial mixes are not adapted to specific tree species or locations, Averill says, and there is little evidence they help. For success, he believes "you need to get the right organisms in the right place."

The Wales experiment is testing that hypothesis. Averill visited the site in April to measure the trees and found the inoculated saplings are growing faster than ones grown without added fungi. He's awaiting a second year of data before publishing them, "but the effect sizes are large enough that I'd be surprised if they go away."

Even if bespoke fungi enhance tree growth in field studies, however, it's not clear whether they can be deployed cheaply and conveniently enough to broadly appeal to forest owners. Averill will investigate that this fall through a company he's created, Funga, that will inoculate pines being commercially planted in the southeastern United States.

The idea of tailoring fungi to boost tree growth "is something many of us write about in our grants," says Kabir Peay, a mycologist at Stanford University who advises SPUN. "It's nice to see that someone's actually out there trying. —*G.P.*

At another stop, Furci charmed a couple into allowing sampling in their front yard. By the time Kiers and Merlin were done, Furci was handing out licorice and hugs were exchanged all around. She even offered the couple a handful of dirt to smell. "I love it," the man said rapturously.

The next morning, the team headed west toward the Pacific coast to sample in Alerce Costero National Park, home to another old-growth forest dominated by a rare tree species, the alerce. In all, the researchers collected some 30 soil samples in just over a week of fieldwork. They handed them over to César Marín, a mycologist at Santo Tomás University in Chile, for DNA analysis. The results—including some from additional samples collected by Marín—will be fed back into SPUN's modeling in order to improve its predictions.

SPUN'S EFFORTS to map fungi in understudied places reflect a "pragmatic" approach that could help scientists better understand how complex ecosystems like forests work, says mycologist Justine Karst of the University of Alberta, Edmonton. She also lauds the team's efforts to excite the public—something she says most scientists don't prioritize. "I'm looking forward to seeing what they produce."

SPUN, meanwhile, hopes its analyses will also help bolster emerging efforts to protect fungi—and Furci's advocacy work in Chile could provide a template. Largely because of her lobbying, Chilean lawmakers a decade ago passed the world's first law formally protecting fungi nationwide.

Elsewhere, fungi are also beginning to capture official attention. Several European countries have taken steps to protect endangered fungal species, and Estonia has created a small preserve where several red-listed fungi live. (In the United States, by contrast, fungi have relatively low status; just two fungi—both lichens—are protected under the Endangered Species Act.)

Last fall, the International Union for the Conservation of Nature (IUCN), which maintains the global "red list" of threatened and endangered species, explicitly called for fungi to be given the same consideration as plants and animals, as did the conservation organization Re:wild. And in December 2021, the United Nations's Food and Agriculture Organization launched an International Network on Soil Biodiversity that includes research on fungi.

Mueller, who chairs IUCN's fungal conservation committee, and other advocates are now working to have the Convention on Biological Diversity—a multinational pact that helps set the global conservation agenda—explicitly embrace the goal of protecting fungi. And although SPUN "is not the only game in town" when it comes to lobbying for fungi, Mueller believes its well-funded public outreach could aid such efforts. "Having this bright light shining on the issue," he says, "will move a lot of initiatives forward."

Some influential groups, however, have yet to afford fungi as much consideration as flora and fauna. The Nature Conservancy—one of the world's largest conservation groups—"does not specifically target fungi" when prioritizing ecosystems for protection, says David Banks, the group's chief conservation officer. "But because we're working in these bigger [ecosystems], we can capture them." (The conservancy is nevertheless funding a SPUN-led trip to sample mycorrhizal networks on a remote Pacific island.)

Even SPUN supporters acknowledge that a more complete picture of fungi won't automatically lead to better real-world outcomes. "There is a gap between mapping and demonstrating patterns of biodiversity, and demonstrating a connection to conservation," Peay says.

Still, SPUN's leaders believe they can add a new, subterranean dimension to global conservation efforts. Kiers points to recent research suggesting that, over nearly one-third of Earth's land area, aboveground biological diversity doesn't match what's found beneath the surface. That means habitats often seen as relatively species-poor—such as boreal forests and drylands—could harbor far greater underground diversity than is currently recognized. Better maps of fungal diversity, she adds, could aid forestry, farming, and efforts to curb climate change.

Kiers recognizes, however, that the onus is now on SPUN to show its value before its first flush of funding runs out. "There has to be a demonstrable benefit to building an organization dedicated to protecting underground ecosystems," she says. "The clock is ticking." ■

By aiding trees, fungi might help curb warming

Could adding certain kinds of fungi to soil help curb climate change by enabling trees to grow faster and suck more carbon dioxide from the atmosphere? That's a question that researchers with the new Society for the Protection of Underground Networks (SPUN) are trying to answer.

In the spring of 2021, in an abandoned pasture in southwestern Wales, a forestry company planted 25,000 trees over 11 hectares for an experiment designed by mycologist Colin Averill, a co-founder of SPUN (see main story, p. 142). The plantations include sitka spruce, a common timber tree in the United Kingdom, and a mix of native deciduous trees. To half of the seedlings' roots, researchers added mycorrhizal fungi—microorganisms that help provide plants with nutrients by associating their roots—that were sourced from mature forests of the same type. Now, the researchers are waiting to see whether the treated trees grow faster and absorb more carbon than those that didn't get the treatment. A similar trial is ongoing in Yucatán, and a third will be planted this fall in Ireland.

The effort builds on research suggesting the right fungi can give arboreal growth a powerful boost. In a January paper published in *The ISME Journal*, Averill and colleagues reported growth rates of trees in forests across Europe vary by up to a factor of three depending on their fungal partners.

But just how much fungi can supercharge saplings is an open question. Companies have long sold mycorrhizae that tree planters can dab onto roots. But such commercial mixes are not adapted to specific tree species or locations, Averill says, and there is little evidence they help. For success, he believes "you need to get the right organisms in the right place."

The Wales experiment is testing that hypothesis. Averill visited the site in April to measure the trees and found the inoculated saplings are growing faster than ones grown without added fungi. He's awaiting a second year of data before publishing them, "but the effect sizes are large enough that I'd be surprised if they go away."

Even if bespoke fungi enhance tree growth in field studies, however, it's not clear whether they can be deployed cheaply and conveniently enough to broadly appeal to forest owners. Averill will investigate that this fall through a company he's created, Funga, that will inoculate pines being commercially planted in the southeastern United States.

The idea of tailoring fungi to boost tree growth "is something many of us write about in our grants," says Kabir Peay, a mycologist at Stanford University who advises SPUN. "It's nice to see that someone's actually out there trying. —*G.P.*

A model of a full-sized human heart was created using specific arrangements of gelatin microfibers.

BIOENGINEERING

Hearts by design

Scalable biofabrication of heart helical tissue pattern augments pumping function

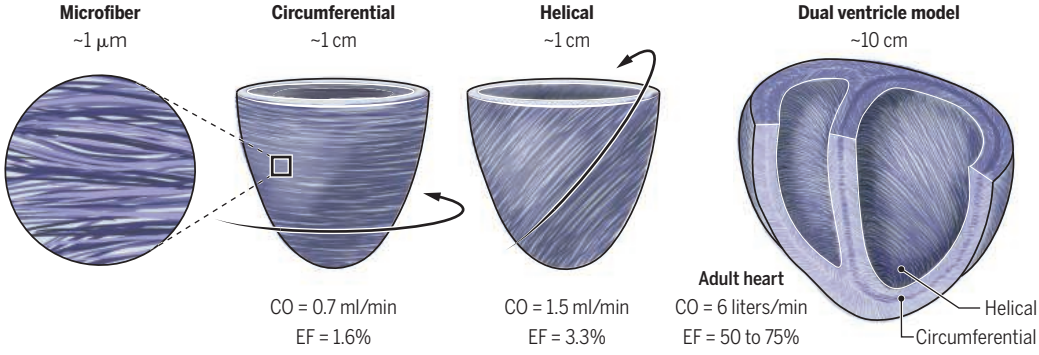
By **Michael V. Sefton**[1] *and* **Craig A. Simmons**[2]

The pumping action of the heart comes from cardiomyocytes, the micrometer-sized muscle cells of the heart, which generate centimeter-sized mechanical contraction and pumping of 70 ml from a ventricle against a pressure of 100 mmHg. Cardiomyocytes are organized as helical fibers that enwrap the ventricles. Orchestrated, electrically paced contraction of the helical fibers causes constriction of the heart chamber to eject blood, but also twisting (torsion). From a mechanical perspective, heart failure is a consequence of inadequate pumping or the heart being too stiff to fill. Regenerative medicine approaches to fix damaged hearts are promising (*1*), but design principles are needed to recapitulate native

[1]Medicine by Design, Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada. [2]Ted Rogers Centre for Heart Research, Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada. Email: michael.sefton@utoronto.ca

Building a heart fiber by fiber

Gelatin microfibers created with focused rotary jet spinning (FRJS) are formed circumferentially or helically in heart ventricle–shaped molds. These are seeded with cardiomyocytes, which align along the spun fibers and produce immature beating heart muscle. Testing revealed the benefit of torsion created by the helical ventricle structure on ejection fraction (EF) and cardiac output (CO), and this arrangement was used to create a full-sized dual ventricle model with three layers.



heart function and address the global heart failure epidemic. On page 180 of this issue, Chang *et al.* (*2*) introduce a new manufacturing process, focused rotary jet spinning (FRJS), to create 3-dimensional (3D) human heart structures with prescribed microscale polymer fiber alignment.

As early as the mid-17th century, the functional importance of torsion to the heart's pumping efficiency (analogous to wringing of a wet cloth) was speculated (*3*). Further circumstantial evidence for the importance of the helical structure comes from failing hearts in which structural abnormalities due to congenital defects, hypertension, ischemia, or fibrosis alter left ventricle twist and are associated with hindered pumping capacity (*4*). However, it is difficult to specifically assess the extent to which the heart's helical structure contributes to its function (or dysfunction) because heart failure is a multifaceted disease, and medical treatments manage symptoms rather than fixing underlying deficits. The limited ability of the heart to regenerate leads to tissue engineering solutions. To that end, understanding and replicating the heart's helical structure-function relationship is thought to be an important step.

FRJS is based on rotary jet spinning, in which centrifugal force expels polymer solution through a fine nozzle to form free-floating fibers that solidify as the polymer solvent evaporates (*5*). In an innovation reported by Chang *et al.*, air jet streams align and focus the extruded fibers to deposit them on collecting structures (molds) (see the second photo). By controlling the shape and rotation of the mold, complex organ-scale 3D scaffolds with specified micrometer-scale fiber orientations were rapidly formed and cells then seeded on the structures. Because FRJS requires molds with convex surfaces, a full human heart could not be produced as a single structure but required assembly and fusion of four individual chambers (see the first photo). Monolithic hearts can be bioprinted by means of 3D extrusion (*6*), but this does not offer the throughput or spatial resolution of FRJS. FRJS also improves on electrospinning and other fiber-spinning methods that cannot create 3D structures with the same geometric complexity or fabrication speed.

Chang *et al.* capitalized on the distinct capabilities of FRJS to revisit whether helical muscle architecture improves ventricle function. They fabricated ellipsoid-shaped ventricles with a single layer of helically or circumferentially aligned gelatin fibers. Fiber orientations were based on an analytical model that predicted greater ejection fractions (fraction of blood ejected from the ventricle with each beat) with helical fibers (*7*). Cardiomyocytes from rat neonates or derived from human induced pluripotent stem cells seeded on the scaffolds generally aligned with the spun fibers, producing oriented, beating, yet developing heart muscle. When electrically paced, engineered ventricles with helical fibers showed higher conduction velocities along the ventricle long axis, more uniform deformation, greater twist, and increased ventricle output compared with those with circumferential fibers.

The authors also fabricated dual-ventricle models with three layers, each with distinct helical arrangements that better represent the native human heart than do single-layer ellipsoid models (see the figure). Although electromechanical coupling, which is necessary to avoid arrhythmias, was not demonstrated, the dual-ventricle models seeded with cardiomyocytes contracted and demonstrated calcium wave propagation in the direction of the surface fibers. This experimentally demonstrates the dependency of electrical wave propagation and mechanical contraction on myocardial fiber orientation that was predicted and incorporated into computational models of the heart (*8*).

Although the study of Chang *et al.* contributes important advances to tissue biofabrication and understanding of cardiac structure-function relationships, these proof-of-concept models fall short of native heart function. In particular, ejection fractions of the engineered single ventricles are much less (at best, ~6% of native) than those of healthy human hearts. Efficient heart contraction involves interplay between the multiscale architecture of the heart, the mechanical properties of the tissue, and the contraction of embedded cardiomyocytes, all of which were likely limiting factors in this study. Despite their helical architecture, the engineered ventricles twisted and shortened less than in native hearts. Greater torsional contraction may be possible with further optimization of the fiber architecture, guided by greater understanding

Microfibers are precisely deposited onto a rotating heart ventricle mold by using focused rotary jet spinning.

CREDITS: (PHOTO, OPPOSITE PAGE) CHANG *ET AL.* (*2*); (RIGHT) MICHAEL ROSNACH, JOHN F. ZIMMERMAN, HUIBIN CHANG, MICHAEL M. PETERS; (GRAPHIC) N. DEESA/*SCIENCE*

of native ventricular microanatomy from diffusion tensor magnetic resonance and other imaging methods (*9*).

The mechanical behavior of the spun scaffolds in this study also did not match that of native myocardium. The myocardium consists of muscle fibers sheathed in a crisscross network of crimped collagen fibers that bear little load at low strains (*10*). As such, myocardium is soft at low strains and stiffens at high strains as the collagen fibers uncoil (*11*). By contrast, FRJS produces straight, bundled fibers that are constantly engaged and stiff when loaded because they do not replicate the nanoscale configuration of collagen fibers. Stiff scaffolds both resist constriction and reduce the contractile forces generated by cardiomyocytes (*12*). This effect was exacerbated by using relatively few immature cardiomyocytes. Emerging strategies to mature cardiomyocytes through specialized media (*13*) or culture substrates (*14*) would increase their contractility, but a concomitant softening of the scaffold would still likely be needed for physiological function. Alternative bioprinting methods such as melt electrowriting can produce soft tissue–like mechanics through precise control of single-fiber microarchitecture (*15*), but not in complex 3D tissue geometries or with the speed of FRJS. This remains a challenge: Can hierarchical organization over spatial scales beyond that demonstrated by Chang *et al.* be achieved while maintaining the versatility and throughput of FRJS?

To achieve a bioengineered heart, a blood-compatible vasculature, an electrical conduction system, and means for avoiding immune responses are still needed. Large metabolically active, cell-dense organs need lots of cells (of different types) and lots of oxygen. The heart is more than a simple pump. Nonetheless, Chang *et al.* have taken a step toward making biomechanically equivalent, functional structures. ■

REFERENCES AND NOTES

1. R. Z. Zhuang, R. Lock, B. Liu, G. Vunjak-Novakovic, *Nat. Biomed. Eng.* **6**, 327 (2022).
2. H. Chang *et al.*, *Science* **377**, 180 (2022)
3. J. I. E. Hoffman, *Physiol. Rep.* **5**, e13404 (2017).
4. A. M. S. Omar, S. Vallabhajosyula, P. P. Sengupta, *Circ. Cardiovasc. Imaging*. **8**, e003029 (2015).
5. M. R. Badrossamay, H. A. McIlwee, J. A. Goss, K. K. Parker, *Nano Lett.* **10**, 2257 (2010).
6. A. Lee *et al.*, *Science* **365**, 482 (2019).
7. E. A. Sallin, *Biophys. J.* **9**, 954 (1969).
8. P. H. M. Bovendeerd *et al.*, *J. Biomech.* **25**, 1129 (1992).
9. S. Nielles-Vallespin *et al.*, *J. Magn. Reson. Imaging*. **52**, 348 (2020).
10. R. Avazmohammadi *et al.*, *Annu. Rev. Biomed. Eng.* **21**, 417 (2019).
11. G. Sommer *et al.*, *Acta Biomater.* **24**, 172 (2015).
12. J. G. Jacot, A. D. McCulloch, J. H. Omens, *Biophys. J.* **95**, 3479 (2008).
13. D. A. M. Feyen *et al.*, *Cell Rep.* **32**, 107925 (2020).
14. W. Dhahri *et al.*, *Circulation*. **145**, 1412 (2022).
15. N. T. Saidy *et al.*, *Small*. **15**, 1900873 (2019).

10.1126/science.add0829

PHOTO: REBECCA CONWAY/AFP/GETTY IMAGES

POLICING

Accessing justice for survivors of violence against women

A police reform experiment in India shows mixed results

By **Graeme Blair**[1] *and* **Nirvikar Jassal**[2]

One in three women has survived physical or sexual violence in her lifetime. In a wave of reforms designed in part to increase women's access to justice for such crimes, governments around the world instituted gender quotas in police hiring (see the photo), policewomen-run counseling centers, women-only stations, and legal mandates that women officers exclusively handle cases of gender-based violence. On page 191 of this issue, Sukhtankar *et al.* (*1*) report results of the first randomized trial on these reforms. The researchers partnered with the state police in Madhya Pradesh, India, in the Hindi heartland with a reputation for deep-rooted patriarchy, to randomize the introduction of "women's help desks"—spaces within police stations where women officers can interface with women complainants. The results are mixed: More incident reports were filed and some police officer attitudes toward violence against women changed, but women were no more likely to report crimes and the arrest rate was unaffected.

Gender-based reforms in policing share similar goals but differ in design and underlying motivation. They can be viewed as falling on a spectrum from integration to separation. On the integrated end, quotas and affirmative action increase the representation of women in the police. Typically, women officers are then assigned the same tasks and roles as men officers. This may shift norms and behaviors of officers and citizens by promoting contact between policemen and policewomen, as well as through a role-model effect. On the separated end, governments establish women-only police stations. This may empower policewomen by reducing contact with sexist policemen, and encourage complainants to report in spaces supposedly removed from patriarchal norms. Such "enclaves" also imply occupational separation: Policewomen are tasked solely with the complaints of other women, often based on essentialist assumptions that they are innately suited for such roles. Separated institutions are an implicit acknowledgment of the limits of integration: if gender norms are unlikely to change in the short term, the state must resign itself to the segregation of women officers and complainants.

Previous studies have examined the extremes of the reform spectrum, with more evidence about separation than integration. Research on separated institutions finds mixed results. Women-only police stations improved police legitimacy and attitudes toward violence against women in Brazil (*2*). "Women's justice centers" in Peru that combine police services with legal and medical aid increased police filings and arrests (*3*). However, women-only stations in India led policemen in regular police stations to refer survivors to alternate sites, forcing victims to travel longer distances to access justice (*4*, *5*). Women officers staffing these sites may not necessarily be more accommodating to women complainants either, because policewomen are not immune from exhibiting gender bias (*6*, *7*). By contrast, the integration of women officers into mainstream law enforcement through affirmative action

[1]Department of Political Science, University of California, Los Angeles, Los Angeles, CA, USA. [2]Department of Government, London School of Economics and Political Science, London, UK. Email: graeme.blair@gmail.com; njassal@stanford.edu

of native ventricular microanatomy from diffusion tensor magnetic resonance and other imaging methods (*9*).

The mechanical behavior of the spun scaffolds in this study also did not match that of native myocardium. The myocardium consists of muscle fibers sheathed in a crisscross network of crimped collagen fibers that bear little load at low strains (*10*). As such, myocardium is soft at low strains and stiffens at high strains as the collagen fibers uncoil (*11*). By contrast, FRJS produces straight, bundled fibers that are constantly engaged and stiff when loaded because they do not replicate the nanoscale configuration of collagen fibers. Stiff scaffolds both resist constriction and reduce the contractile forces generated by cardiomyocytes (*12*). This effect was exacerbated by using relatively few immature cardiomyocytes. Emerging strategies to mature cardiomyocytes through specialized media (*13*) or culture substrates (*14*) would increase their contractility, but a concomitant softening of the scaffold would still likely be needed for physiological function. Alternative bioprinting methods such as melt electrowriting can produce soft tissue–like mechanics through precise control of single-fiber microarchitecture (*15*), but not in complex 3D tissue geometries or with the speed of FRJS. This remains a challenge: Can hierarchical organization over spatial scales beyond that demonstrated by Chang *et al.* be achieved while maintaining the versatility and throughput of FRJS?

To achieve a bioengineered heart, a blood-compatible vasculature, an electrical conduction system, and means for avoiding immune responses are still needed. Large metabolically active, cell-dense organs need lots of cells (of different types) and lots of oxygen. The heart is more than a simple pump. Nonetheless, Chang *et al.* have taken a step toward making biomechanically equivalent, functional structures. ■

REFERENCES AND NOTES

1. R. Z. Zhuang, R. Lock, B. Liu, G. Vunjak-Novakovic, *Nat. Biomed. Eng.* **6**, 327 (2022).
2. H. Chang *et al.*, *Science* **377**, 180 (2022)
3. J. I. E. Hoffman, *Physiol. Rep.* **5**, e13404 (2017).
4. A. M. S. Omar, S. Vallabhajosyula, P. P. Sengupta, *Circ. Cardiovasc. Imaging*. **8**, e003029 (2015).
5. M. R. Badrossamay, H. A. McIlwee, J. A. Goss, K. K. Parker, *Nano Lett.* **10**, 2257 (2010).
6. A. Lee *et al.*, *Science* **365**, 482 (2019).
7. E. A. Sallin, *Biophys. J.* **9**, 954 (1969).
8. P. H. M. Bovendeerd *et al.*, *J. Biomech.* **25**, 1129 (1992).
9. S. Nielles-Vallespin *et al.*, *J. Magn. Reson. Imaging*. **52**, 348 (2020).
10. R. Avazmohammadi *et al.*, *Annu. Rev. Biomed. Eng.* **21**, 417 (2019).
11. G. Sommer *et al.*, *Acta Biomater.* **24**, 172 (2015).
12. J. G. Jacot, A. D. McCulloch, J. H. Omens, *Biophys. J.* **95**, 3479 (2008).
13. D. A. M. Feyen *et al.*, *Cell Rep.* **32**, 107925 (2020).
14. W. Dhahri *et al.*, *Circulation*. **145**, 1412 (2022).
15. N. T. Saidy *et al.*, *Small*. **15**, 1900873 (2019).

10.1126/science.add0829

PHOTO: REBECCA CONWAY/AFP/GETTY IMAGES

POLICING

Accessing justice for survivors of violence against women

A police reform experiment in India shows mixed results

By **Graeme Blair**[1] *and* **Nirvikar Jassal**[2]

One in three women has survived physical or sexual violence in her lifetime. In a wave of reforms designed in part to increase women's access to justice for such crimes, governments around the world instituted gender quotas in police hiring (see the photo), policewomen-run counseling centers, women-only stations, and legal mandates that women officers exclusively handle cases of gender-based violence. On page 191 of this issue, Sukhtankar *et al.* (*1*) report results of the first randomized trial on these reforms. The researchers partnered with the state police in Madhya Pradesh, India, in the Hindi heartland with a reputation for deep-rooted patriarchy, to randomize the introduction of "women's help desks"—spaces within police stations where women officers can interface with women complainants. The results are mixed: More incident reports were filed and some police officer attitudes toward violence against women changed, but women were no more likely to report crimes and the arrest rate was unaffected.

Gender-based reforms in policing share similar goals but differ in design and underlying motivation. They can be viewed as falling on a spectrum from integration to separation. On the integrated end, quotas and affirmative action increase the representation of women in the police. Typically, women officers are then assigned the same tasks and roles as men officers. This may shift norms and behaviors of officers and citizens by promoting contact between policemen and policewomen, as well as through a role-model effect. On the separated end, governments establish women-only police stations. This may empower policewomen by reducing contact with sexist policemen, and encourage complainants to report in spaces supposedly removed from patriarchal norms. Such "enclaves" also imply occupational separation: Policewomen are tasked solely with the complaints of other women, often based on essentialist assumptions that they are innately suited for such roles. Separated institutions are an implicit acknowledgment of the limits of integration: if gender norms are unlikely to change in the short term, the state must resign itself to the segregation of women officers and complainants.

Previous studies have examined the extremes of the reform spectrum, with more evidence about separation than integration. Research on separated institutions finds mixed results. Women-only police stations improved police legitimacy and attitudes toward violence against women in Brazil (*2*). "Women's justice centers" in Peru that combine police services with legal and medical aid increased police filings and arrests (*3*). However, women-only stations in India led policemen in regular police stations to refer survivors to alternate sites, forcing victims to travel longer distances to access justice (*4*, *5*). Women officers staffing these sites may not necessarily be more accommodating to women complainants either, because policewomen are not immune from exhibiting gender bias (*6*, *7*). By contrast, the integration of women officers into mainstream law enforcement through affirmative action

[1]Department of Political Science, University of California, Los Angeles, Los Angeles, CA, USA. [2]Department of Government, London School of Economics and Political Science, London, UK. Email: graeme.blair@gmail.com; njassal@stanford.edu

Women police officers serve at a station house in Delhi. Several of India's states and all its union territories, including Delhi, have installed a 33% gender quota in police hiring to increase women's comfort with law enforcement.

or quotas not only increased registrations of violence against women in the US but also improved police efficacy (*8*). Measures that integrate women officers in security agencies by enabling them to carry out the same tasks as men can diminish stereotypes (*9*).

The Sukhtankar *et al.* experiment, one of the largest of its kind, provides important new evidence on the middle of the theoretical reform spectrum. Women's help desks are a form of separation—within regular police stations as opposed to fully segregated sites—that also retain occupational divides between administrators staffing the institutions and other officers. The authors found that the intervention did not increase the likelihood that women report violence to the police, but it did increase the filing of incident reports that do not lead automatically to criminal investigations. They found only weak effects on formal registrations that warrant mandatory investigation. The help desks did not increase arrests of abusers. Although the intervention increased officers' knowledge of the law, it did not generally alter gender norms. Policewomen's perceptions that women report false or exaggerated claims fell slightly, but policemen's attitudes were unchanged.

The null effect on arrests admits several interpretations. For domestic violence in particular, women may wish abuse to stop but not for the perpetrator to be arrested. But because information on what police do in response to a report was not captured, the null may also reflect police inaction, either because women request investigations and are ignored or because the police act but the inquiries do not result in arrests. Police also may not have completed investigations by the time data were collected, 2 months after the intervention activities ended. In addition, law enforcement may have acted, but not by making arrests. Women's help desks may have provided restorative or rehabilitative justice instead of standard police services, e.g., informal couples' counseling for crimes such as dowry-based harassment and marital rape.

Future work should disaggregate the types and gradations of crime affected by police reforms. Institutions like help desks may improve registrations of crimes like dowry-based harassment, but police may not take action on other crimes like gang rape or acid attacks, which they may face pressure to downplay in official statistics. If these institutions emphasize alternative modes of justice, future researchers should investigate whether such measures lead to better outcomes for survivors or are merely window dressing. For instance, does officer-mediated reconciliation of victims with abusers inside station-houses empower complainants, or are such measures simply a mechanism by which the police avoid taking formal action against abuse? Punitive justice may discourage reporting because victims may not necessarily desire the arrest of intimate partners; then again, rehabilitative justice may appear toothless and fail to deter violence.

A key challenge in studies of access to justice is distinguishing whether effects are driven by an increase in citizens trying to report or officers recording those complaints. Sukhtankar *et al.* counted citizens entering police stations from video surveillance camera footage to distinguish between reporting and registration. Future research should go further by linking police registrations to court records, allowing cases to be traced across successive stages of investigation, trial, and verdict. Improvements in data availability in India and China now make this data linkage a possibility (*10*). This would allow researchers to probe whether policies like women's help desks mostly affect the first step of accessing justice (case filings) or also change later outcomes, such as convictions of suspects. Linking multiple stages of the justice system would reveal where and when discrimination creeps into the process.

The short-term outcome measurement by Sukhtankar *et al.* leaves open the possibility that police behavior does change but only in the long run, and conversely that the increases in case filing may not endure. Effects may dissipate for many reasons. For example, when partnerships with academics end, police leaders' prioritization of the reform may wane. Maintaining buy-in from police leaders may be a key driver of successful implementation (*11*). The duration of implementation (distinct from how long after implementation measurement takes place) may also be important. Because the Sukhtankar *et al.* intervention lasted only 13 months, new studies with longer time horizons might find different effects.

More research is needed to understand the implications of gender-based police reform on personnel dynamics and gender relations within police forces. Future scholarship could probe whether officers appreciate being deployed to institutions like women's help desks, or whether they feel that that deployment prevents them from handling diverse cases. If policewomen are disproportionately or exclusively tasked with investigating other women's complaints, might this diminish professionalization and induce type-casting, including among women complainants and police colleagues?

Another important inquiry concerns the impact of police reforms on actual rates of violence against women. Early evidence is mixed (*12*). Subsequent scholarship should examine how institutional reforms operate in conjunction with other tools, including media and educational campaigns (*13*), increasing women's representation in elected office (*14*), limiting behaviors such as alcohol abuse, or resource transfers targeted to women (*15*). In addition, researchers should explore whether reforms targeted at women can also affect gender-based violence against men and LGBTQ citizens and their access to justice for such crimes.

Recent studies have introduced new standards for transparent ethical decision-making, including scrutinizing the human rights records of police partners and monitoring harm "red lines" (*11*). Sukhtankar *et al.* push further. For instance, to mitigate a potentially unintended consequence of their scheme in assigning policewomen to women's help desks, they worked with law enforcement to avoid reassigning women officers from other stations, which might have displaced personnel and undermined access to justice for women in untreated areas. Subsequent scholarship should follow their example in transparently describing harm reduction practices and consider preregistering ethical decision-making criteria. With these measures, researchers should make the case, as Sukhtankar *et al.* persuasively do, that the risks of conducting randomized interventions must be balanced against the value of building more evidence about which police reforms work—and which do not. ■

REFERENCES AND NOTES

1. S. Sukhtankar *et al.*, *Science* **377**, 191 (2022).
2. A. Córdova, H. Kras, *Comp. Polit. Stud.* **53**, 775 (2020).
3. M. M. Sviatschi, I. Trako, "Gender Violence, Enforcement, and Human Capital: Evidence from Women's Justice Centers in Peru," Policy Research Working Papers, The World Bank, 12 April 2021.
4. N. Jassal, *Am. Polit. Sci. Rev.* **114**, 1035 (2020).
5. S. Amaral, S. Bhalotra, N. Prakash, "Gender, Crime and Punishment: Evidence from Women Police Stations in India," Working paper, ifo Center for Labour and Demographic Economics, 2018.
6. S. Karim *et al.*, *Int. Stud. Q.* **62**, 618 (2018).
7. C. Santos, *Latin Am. Res. Rev.* **39**, 29 (2004).
8. A. Miller, C. Segal, *Rev. Econ. Stud.* **86**, 2220 (2019).
9. G. B. Dahl *et al.*, *Q. J. Econ.* **136**, 987 (2021).
10. E. Ash *et al.*, "Measuring Gender and Religious Bias in the Indian Judiciary," Center for Law and Economics Working Paper Series 3, 2021.
11. G. Blair *et al.*, *Science* **374**, eabd3446 (2021).
12. E. Perova, S. A. Reynolds, *Soc. Sci. Med.* **174**, 188 (2017).
13. D. P. Green *et al.*, *Comp. Polit. Stud.* **53**, 2283 (2020).
14. L. Iyer *et al.*, *Am. Econ. J. Appl. Econ.* **4**, 165 (2012).
15. M. Hidrobo *et al.*, *Am. Econ. J. Appl. Econ.* **8**, 284 (2016).

ACKNOWLEDGMENTS

We thank R. Blair, D. Hausman, M. Shah, and A. Wilke for helpful comments.

10.1126/science.abp9542

MEMBRANES

Bridging membrane transport models

Analysis of transport processes in swollen polymer networks bridges two classical models

By **Geoffrey M. Geise**

The movement of small molecules (e.g., water, organic solvents, gases) through polymeric membrane materials is critical for a wide range of applications including water desalination and filtration, crude oil refining, and some controlled-release devices (*1*–*4*). For decades, researchers have used two classical models to describe fluid passage through membranes: solution diffusion and pore flow (*1*). Although each model is generally accepted to describe certain materials, the mechanism of how solvent passes through materials whose structure falls between the two models, such as swollen polymers, has been debated since the 1960s. On page 186 of this issue, Hegde *et al.* (*5*) present a two-phase fluid-solid model and perturbation analysis to demonstrate that both the solution-diffusion model and the pore-flow model ultimately describe the same driving force for solvent passage in swollen polymers.

Transport in dense polymer gas separation membranes can be described appropriately using the solution-diffusion model, where gas molecules enter the material on one side of the membrane and diffuse through the polymer before emerging on the other side. Alternatively, micro- or ultrafiltration membranes are often porous polymers, where the polymer supports a rigid tortuous network of pores that works more like a sieve, and this mechanism of transport can be described using the pore-flow model.

An intermediate case arises when the membrane is a swollen polymer. Central to the debate of how to model these intermediate cases are the different mechanisms described by each model for how molecules pass through the material. The pore-flow model describes solvent passage as driven by a hydraulic pressure, whereas the solution-diffusion model describes solvent passage as driven by a diffusive process. Hegde *et al.* resolve the apparent mechanistic differences associated with these two models for swollen polymers and ultimately extend transport modeling through an application of conservation of mass and momentum for solvent and the polymer membrane. This approach, coupled with perturbation analysis, satisfyingly results in a quantitative agreement between the two modeling approaches and resolves apparent mechanistic discrepancies between the two classical models.

One of the core differences between the solution-diffusion and pore-flow models is the treatment of concentration and pressure within the membrane. The solution-diffusion model is based on the existence of a concentration gradient, whereas the pore-flow model is based on a pressure gradient. The pressure difference–induced concentration gradient, suggested by the solution-diffusion model, was experimentally observed decades ago (*6*), but the application of this observation to highly swollen polymers has been questioned. An argument can be made that highly swollen polymers (e.g., hydrogels) have continuous solvent pathways through which transport can occur by means of hydraulic pressure–driven flow, in the absence of a solvent concentration gradient. This argument is resolved by Hegde *et al.* by recognizing that the use of either model does not negate the other and that the resulting solvent flux relationships are equivalent.

"...the solution-diffusion model and the pore-flow model ultimately describe the same driving force for solvent passage..."

Solution-diffusion transport in gas separation membranes is considered to be gated by polymer chain motions needed for gases to execute diffusional jumps—that is, the steps that a molecule takes while navigating through a material (*7*). In other words, the polymer must move out of the way for transport to occur. As solvent concentration increases and becomes a continuous fluid inside the swollen polymer, the mechanistic view of polymer chain–gated diffusion has been questioned because of the presence of continuous solvent pathways. The ability of the polymer to deform, however, is an important extension embodied in the fluid-solid model presented by Hegde *et al.* Recent experiments have supported the notion that hydrated polymer segmental dynamics may control the selective transport of water versus salt (*8*, *9*). The combination of polymer and solvent momentum balances in the fluid-solid model and the continuity equation permits the description of solvent transport, to an order of approximation that is reasonable for many membrane applications, in terms of either a solvent concentration or pressure gradient.

Hegde *et al.* also address the different meanings of "pressure" in the two classical models. The pore-flow model considers pressure to be the hydrodynamic pressure of the solvent phase. The solution-diffusion model, by contrast, considers the solvent and polymer mixture as a single phase. Therefore, pressure in the solution-diffusion context includes all of the mechanical stresses associated with the swollen material. Mechanical equilibrium in the fluid-solid model is consistent with the appropriate analogs in the classical models, and the extension provided by Hegde *et al.* provides clarity to inform the usage of the term "pressure."

The debate over the mechanism of transport in swollen polymers has been suggested to have implications for the design of advanced materials. The realization that both the solution-diffusion and pore-flow approaches are complementary through the extension provided by the fluid-solid model suggests interesting new ways to characterize and understand materials. This advance in modeling transport in swollen polymers supports and expands a more holistic view of materials design with a more productive focus on uncovering structure–property relationships needed to engineer advanced membranes. ■

Department of Chemical Engineering, University of Virginia, Charlottesville, VA, USA. Email: geise@virginia.edu

REFERENCES AND NOTES

1. J. G. Wijmans, R. W. Baker, *J. Membr. Sci.* **107**, 1 (1995).
2. G. M. Geise, D. R. Paul, B. D. Freeman, *Prog. Polym. Sci.* **39**, 1 (2014).
3. K. P. Bye, M. Galizia, *J. Membr. Sci.* **603**, 118020 (2020).
4. R. W. Baker, *Membrane Technology and Applications* (Wiley, ed. 3, 2012).
5. V. H. Hegde, M. F. Doherty, T. M. Squires, *Science* **377**, 186 (2022).
6. S. Rosenbaum, O. Cotton, *J. Polym. Sci. A1* **7**, 101 (1969).
7. B. D. Freeman, *Macromolecules* **32**, 375 (1999).
8. K. Chang *et al.*, *Macromolecules* **51**, 9222 (2018).
9. A. G. Korovich, K. Chang, G. M. Geise, L. A. Madsen, *Macromolecules* **54**, 11187 (2021).

10.1126/science.abn5485

SOFT MATTER

Less is more when forming gels by dilution

Molecular self-assembly yields soft materials arising from the liquid state when diluted

By **Matthew J. Webber**

Soft materials that change form or function in response to environmental or user-applied stimuli have a wide range of biomedical applications (*1*). Gels can form in water from weakly interacting molecules but can return to the state of a flowing liquid suspension, known as a sol, upon changes in the concentration of the molecules or the applied temperature. This behavior is known as a reentrant phase transition. A gel-to-sol phase transition typically arises from a reduction in concentration, meaning that a gel becomes a sol upon dilution and a sol becomes a gel with increased concentration. On page 213 of this issue, Su *et al.* (*2*) demonstrate a system that exhibits a sol-to-gel transition when diluted, inverting the common behavior of gels. Their observations offer insight into systems that undergo reentrant phase transitions in biology.

Self-assembly is a process in which molecules form specific structures either as a consequence of their properties or the conditions to which they are exposed. Several self-assembly motifs have been used to prepare soft materials. One such motif leverages molecules with certain supramolecular interactions, such as directional hydrogen bonding units, to form supramolecular polymers. This type of filamentous assembly is reminiscent of high persistence-length polymers (*3*). Another common motif is that of simple surfactants, which assemble into spherical objects known as micelles through hydrophobic association. Surfactants have broad applications as detergents and interface stabilizers (*4*) and in lipid nanoparticle vaccines.

When molecules prepared from different motifs are mixed, self-assembly typically obeys a thermodynamic preference for self-sorting, meaning that each species forms its own segregated structures. However, realizing pure self-sorting can be a challenge because of common driving forces shared by the different molecular motifs (*5*). Interesting phenomena emerge from multicomponent co-assembly when underlying driving forces, such as hydrophobic association, are conserved between different motifs. In this case, minor adjustments in the strength of the interactions between molecules can dictate the emergence of different self-assembled structures (*6*). Because these same interactions are governed by parameters such as concentration, temperature, and salt or pH levels of the solution, the resulting assemblies can be tuned by controlling these parameters to create responsive material platforms.

The concentration-dependent interaction of two self-assembly motifs—here being those that assemble through disparate mechanisms of supramolecular polymerization or surfactant micelle formation—is key to the phenomena reported by Su *et al.* (see the figure). The primary supramolecular polymer motif is a disk-like molecule known as BTA-EG_4 (*7*), which is prone to stack in water into elongated nanoscale filaments that can entangle to form a gel. However, when this supramolecular motif is combined with a surfactant, the formation of supramolecular filaments is inhibited as the surfactant disrupts the stacking. Upon dilution, however, the unexpected sol-to-gel transition is observed as interactions between the surfactant and BTA-EG_4 are weakened, allowing BTA-EG_4 to resume its preferred filamentous assembly. The sol and the gel phases both exist at the same relative composition of these two orthogonal motifs and require only a change in the total overall concentration for the system to oscillate between its sol and gel states. This sol-to-gel transition proved to be a robust phenomenon, arising when combining a suite of different common surfactants with BTA-EG_4. The gel phase that forms upon dilution is a transient state. Further dilution crosses the critical gelation threshold of the BTA-EG_4 supramolecular filaments and restores the sol.

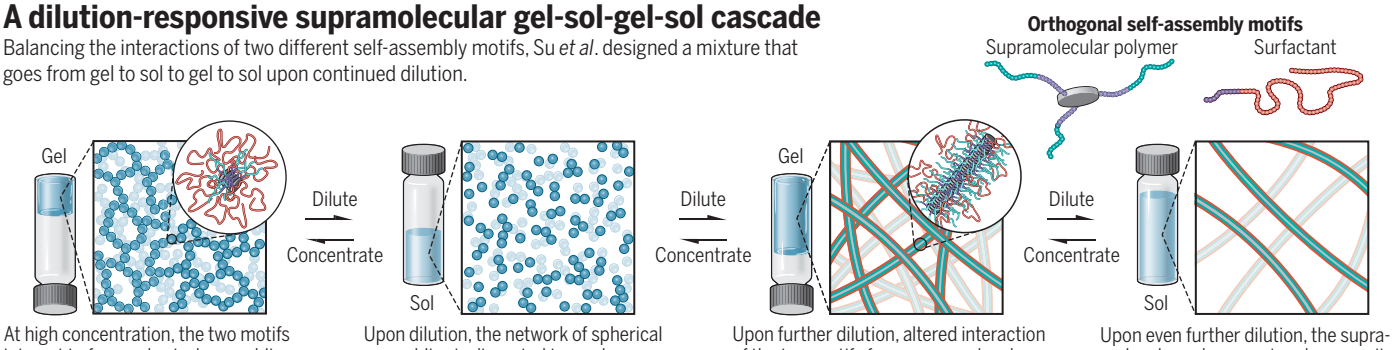
In demonstrating even more complex gel-sol-gel-sol dilution cascades, Su *et al.* prepared two different protocols based on multicomponent and orthogonal self-assemblies. The system prepared from BTA-EG_4 and surfactant can begin from the state of a robust gel at a very high starting concentration at a molar ratio of 2 to 3 for supramolecular motif to surfactant. At this high initial concentration, spherical assemblies consisting of both motifs become fused to form a network. Upon dilution, a sol emerges that is composed of similar spherical assemblies but at a concentration below that necessary to form the fused network. Continued dilution from this point restores the gel state through formation of the supramolecular filaments preferred by the BTA-EG_4 motif. Yet additional dilution reduces entanglement of these filaments below the threshold for gel formation, restoring the sol once again.

An alternate protocol enabling a different gel-sol-gel-sol cascade is also demonstrated by Su *et al.* by including a third molecule capable of forming another orthogonal su-

Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA.
Email: mwebber@nd.edu

A dilution-responsive supramolecular gel-sol-gel-sol cascade

Balancing the interactions of two different self-assembly motifs, Su *et al*. designed a mixture that goes from gel to sol to gel to sol upon continued dilution.



At high concentration, the two motifs interact to form spherical assemblies that fuse into a gel network.

Upon dilution, the network of spherical assemblies is disrupted to produce a flowing colloidal suspension (sol).

Upon further dilution, altered interaction of the two motifs form supramolecular polymers that entangle into a gel network.

Upon even further dilution, the supramolecular polymer network passes its gel point to restore the sol.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE* BASED ON M. J. WEBBER

pramolecular network. This component comprises a ureidopyrimidinone (UPy) supramolecular motif that forms long fiber bundles (*8*). Although the UPy motif also interacts with the surfactant, the assembled UPy fiber bundles are not evidently disturbed by the presence of surfactant. By this approach, the system of BTA-EG_4 and surfactant, at concentrations that previously formed a sol consisting of spherical assemblies, can instead form a gel by adding a high concentration of UPy fibrillar structures. Dilution from this initial point surpasses the gelation capacity of the UPy network while not yet reaching the point of inducing BTA-EG_4 filament formation, yielding a sol. Continued dilution activates BTA-EG_4 filament formation, which in combination with the remaining UPy fiber bundles restores a gel state comprising two orthogonal networks of supramolecular polymers: the BTA-EG_4 and UPy networks. Upon further dilution, the system transitions back to a sol once again.

The tunable nature of molecular-scale self-assembly in these materials offers simple synthetic analogues of more complex phenomena observed in nature. For example, membraneless organelles—distinct compartments within a cell that are not enclosed by a traditional lipid membrane—are thought to arise from liquid-liquid phase separation because of concentration gradients of associating multicomponent systems forming these assemblies in a water environment (*9*). The roles of membraneless organelles in biological signaling during both normal and diseased states are increasingly appreciated (*10*). The behavior of the simple systems described by Su *et al.* is therefore reminiscent of more complex self-assembly phenomena in biology, illuminating the importance of subtle thermodynamic driving forces that give rise to concentration-dependent phase separation. This new paradigm in self-assembled materials consisting of highly adaptive and dilution-triggered hydrogels may furthermore lead to the design of stimuli-responsive material platforms for in situ modulation of function in therapeutic biomedicine. ■

REFERENCES AND NOTES

1. M. A. Stuart *et al.*, *Nat. Mater.* **9**, 101 (2010).
2. L. Su *et al.*, *Science* **377**, 213 (2022).
3. T. Aida, E. W. Meijer, S. I. Stupp, *Science* **335**, 813 (2012).
4. M. J. Rosen, J. T. Kunjappu, *Surfactants and Interfacial Phenomena* (Wiley, ed. 4, 2012).
5. K. L. Morris *et al.*, *Nat. Commun.* **4**, 1480 (2013).
6. W. M. Jacobs, D. Frenkel, *Biophys. J.* **112**, 683 (2017).
7. C. M. Leenders *et al.*, *Chem. Commun.* **49**, 1963 (2013).
8. S. I. S. Hendrikse *et al.*, *Chem. Commun.* **53**, 2279 (2017).
9. J. A. Riback *et al.*, *Nature* **581**, 209 (2020).
10. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).

ACKNOWLEDGMENTS

M.J.W. acknowledges funding from the National Institutes of Health (R35GM137987) and the National Science Foundation (BMAT, 1944875).

10.1126/science.abo7656

EVOLUTION

"Arch"-etyping vertebrates

Cellular details of gill arches in Cambrian fossils reignite a centuries-old debate

By **Tetsuto Miyashita**

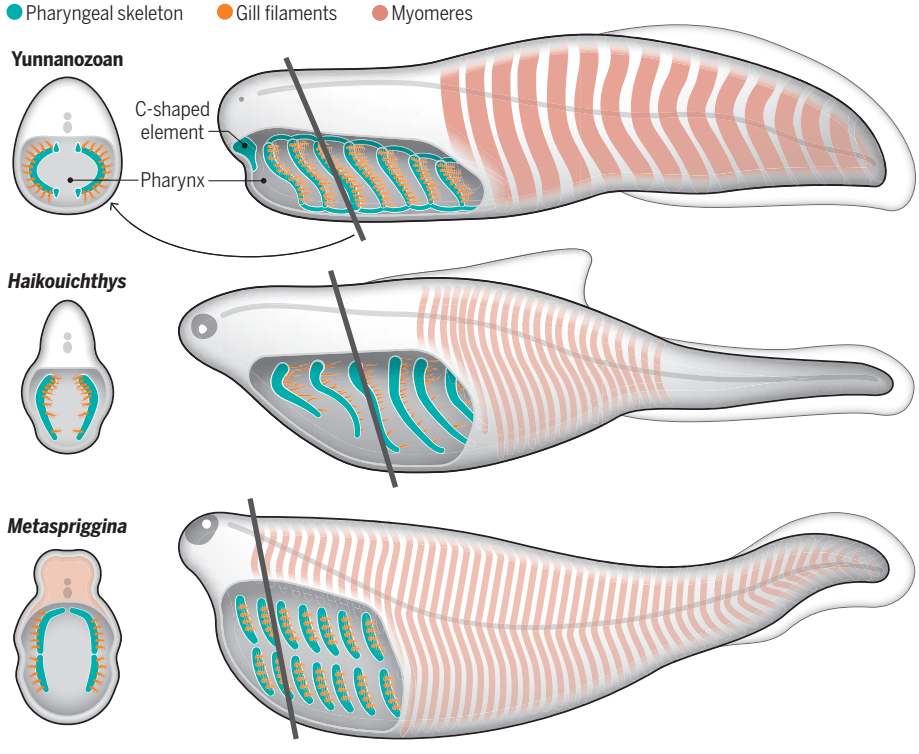
Scientists have long been searching for fossils of distant vertebrate ancestors. In the 1990s, mysterious fishlike forms—now known as yunnanozoans—were discovered at a 520-million-year-old Cambrian fossil site in the Yunnan province of China (*1*–*3*). More fishlike forms (e.g., *Haikouichthys* and *Myllokunmingia*) were reported from the same locality shortly thereafter (*4*, *5*), while the 508-million-year-old Burgess Shale in the Canadian Rockies yielded *Metaspriggina* (*6*). Having eyes and a brain at the front end of an otherwise wormlike soft body, these animals appear to have branched off the phylogenetic tree before the last common ancestor of all living vertebrates. However, there is ongoing controversy about precisely how close to vertebrates these Cambrian forms were. On page 218 of this issue, Tian *et al.* (*7*) present compelling evidence in yunnanozoans for an unmistakable vertebrate trait—a pharyngeal skeleton made of cellular cartilage.

Interpreting organic stains on a shale slab is both a science and an art. Wielding scanning electron microscopy and computed microtomography scans to yield unprecedented details, Tian *et al.* reveal cellular and subcellular structures of the skeletal bars that best compare to cartilaginous gill arches of modern vertebrates. These bars in yunnanozoans are patterned in a series, each associated with gill filaments and connected by horizontal rods. The morphology closely approximates various predictions for vertebrate ancestors.

Of all the internal structures of the earliest vertebrates, pharyngeal skeletons perhaps stand the best chance for fossilization given their robustness. Nonetheless, the complex evolutionary history of the pharynx has

Of gills and jaws

Cambrian vertebrates each evolved distinct pharyngeal anatomy with a series of gill-supporting skeletons. The yunnanozoan has a cartilaginous basket and the *Haikouichthys* has unjoined bars, whereas the *Metaspriggina* has upper and lower rods. However, it remains an open question whether their gill anatomies represent any evolutionary link to the jaws of modern vertebrates.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

pramolecular network. This component comprises a ureidopyrimidinone (UPy) supramolecular motif that forms long fiber bundles (*8*). Although the UPy motif also interacts with the surfactant, the assembled UPy fiber bundles are not evidently disturbed by the presence of surfactant. By this approach, the system of BTA-EG_4 and surfactant, at concentrations that previously formed a sol consisting of spherical assemblies, can instead form a gel by adding a high concentration of UPy fibrillar structures. Dilution from this initial point surpasses the gelation capacity of the UPy network while not yet reaching the point of inducing BTA-EG_4 filament formation, yielding a sol. Continued dilution activates BTA-EG_4 filament formation, which in combination with the remaining UPy fiber bundles restores a gel state comprising two orthogonal networks of supramolecular polymers: the BTA-EG_4 and UPy networks. Upon further dilution, the system transitions back to a sol once again.

The tunable nature of molecular-scale self-assembly in these materials offers simple synthetic analogues of more complex phenomena observed in nature. For example, membraneless organelles—distinct compartments within a cell that are not enclosed by a traditional lipid membrane—are thought to arise from liquid-liquid phase separation because of concentration gradients of associating multicomponent systems forming these assemblies in a water environment (*9*). The roles of membraneless organelles in biological signaling during both normal and diseased states are increasingly appreciated (*10*). The behavior of the simple systems described by Su *et al.* is therefore reminiscent of more complex self-assembly phenomena in biology, illuminating the importance of subtle thermodynamic driving forces that give rise to concentration-dependent phase separation. This new paradigm in self-assembled materials consisting of highly adaptive and dilution-triggered hydrogels may furthermore lead to the design of stimuli-responsive material platforms for in situ modulation of function in therapeutic biomedicine. ■

REFERENCES AND NOTES

1. M. A. Stuart *et al.*, *Nat. Mater.* **9**, 101 (2010).
2. L. Su *et al.*, *Science* **377**, 213 (2022).
3. T. Aida, E. W. Meijer, S. I. Stupp, *Science* **335**, 813 (2012).
4. M. J. Rosen, J. T. Kunjappu, *Surfactants and Interfacial Phenomena* (Wiley, ed. 4, 2012).
5. K. L. Morris *et al.*, *Nat. Commun.* **4**, 1480 (2013).
6. W. M. Jacobs, D. Frenkel, *Biophys. J.* **112**, 683 (2017).
7. C. M. Leenders *et al.*, *Chem. Commun.* **49**, 1963 (2013).
8. S. I. S. Hendrikse *et al.*, *Chem. Commun.* **53**, 2279 (2017).
9. J. A. Riback *et al.*, *Nature* **581**, 209 (2020).
10. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).

ACKNOWLEDGMENTS

M.J.W. acknowledges funding from the National Institutes of Health (R35GM137987) and the National Science Foundation (BMAT, 1944875).

10.1126/science.abo7656

EVOLUTION

"Arch"-etyping vertebrates

Cellular details of gill arches in Cambrian fossils reignite a centuries-old debate

By **Tetsuto Miyashita**

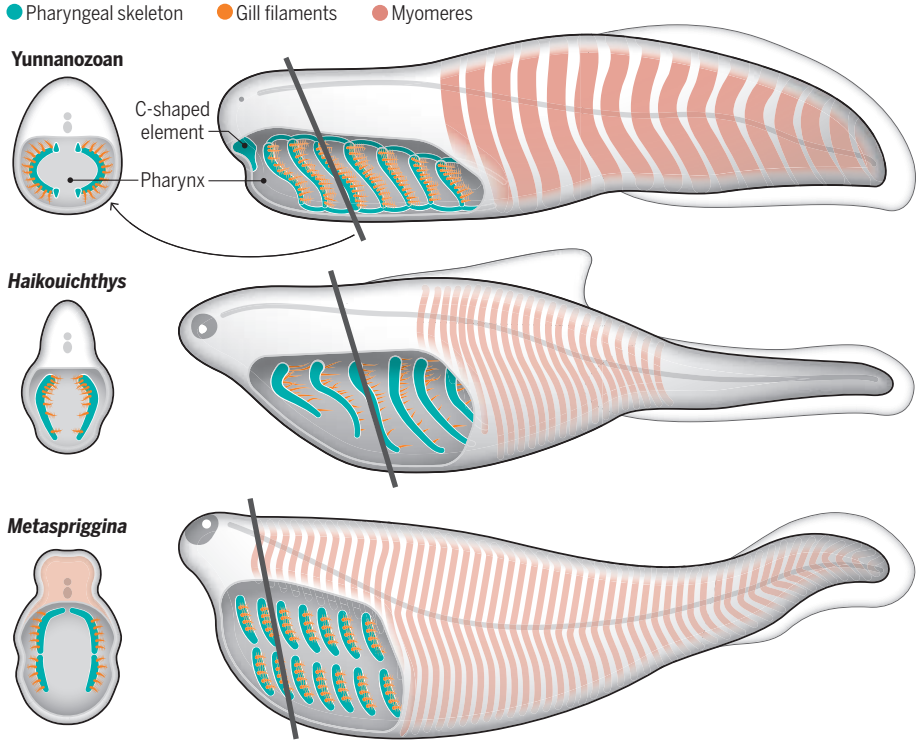
Scientists have long been searching for fossils of distant vertebrate ancestors. In the 1990s, mysterious fishlike forms—now known as yunnanozoans—were discovered at a 520-million-year-old Cambrian fossil site in the Yunnan province of China (*1*–*3*). More fishlike forms (e.g., *Haikouichthys* and *Myllokunmingia*) were reported from the same locality shortly thereafter (*4*, *5*), while the 508-million-year-old Burgess Shale in the Canadian Rockies yielded *Metaspriggina* (*6*). Having eyes and a brain at the front end of an otherwise wormlike soft body, these animals appear to have branched off the phylogenetic tree before the last common ancestor of all living vertebrates. However, there is ongoing controversy about precisely how close to vertebrates these Cambrian forms were. On page 218 of this issue, Tian *et al.* (*7*) present compelling evidence in yunnanozoans for an unmistakable vertebrate trait—a pharyngeal skeleton made of cellular cartilage.

Interpreting organic stains on a shale slab is both a science and an art. Wielding scanning electron microscopy and computed microtomography scans to yield unprecedented details, Tian *et al.* reveal cellular and subcellular structures of the skeletal bars that best compare to cartilaginous gill arches of modern vertebrates. These bars in yunnanozoans are patterned in a series, each associated with gill filaments and connected by horizontal rods. The morphology closely approximates various predictions for vertebrate ancestors.

Of all the internal structures of the earliest vertebrates, pharyngeal skeletons perhaps stand the best chance for fossilization given their robustness. Nonetheless, the complex evolutionary history of the pharynx has

Of gills and jaws

Cambrian vertebrates each evolved distinct pharyngeal anatomy with a series of gill-supporting skeletons. The yunnanozoan has a cartilaginous basket and the *Haikouichthys* has unjoined bars, whereas the *Metaspriggina* has upper and lower rods. However, it remains an open question whether their gill anatomies represent any evolutionary link to the jaws of modern vertebrates.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

strained attempts to interpret fossil imprints. Pharyngeal slits and skeletons long precede the origins of vertebrates, and cellular cartilage has been found in the lips of a developing invertebrate chordate (*8*). Vertebrates develop the entire pharyngeal skeleton from cellular cartilage in the embryonic pharyngeal arches. This cartilage originates in the vertebrate-specific cell lineage called the neural crest. To complicate matters further, the pharynx is regionally and phylogenetically differentiated among vertebrates. Depending on whether similarities or differences are emphasized, the pharyngeal arch I in which jaws develop, that is, the mandibular arch, may (*9*–*11*) or may not (*12*–*14*) share an evolutionary origin with gill-supporting skeletons. New information from yunnanozoans presents an opportunity to clarify these issues.

Tian *et al*. avoided jumping to conclusions, as they did not explicitly identify which gill bar in a yunnanozoan corresponds to which arch in a modern vertebrate. But they did signal their favored interpretation that yunnanozoans, with each gill bar identical to the next, represent the ancestral vertebrate condition. This view is in line with the belief that all pharyngeal arches originally supported gills and that one of them evolved into a jaw (*9*–*11*). At face value, yunnanozoans could serve as long-awaited evidence for the gill-arch hypothesis of jaw origins.

However, this jaw-origin narrative relies solely on how one chooses to identify the gill bars of yunnanozoans—that is, whether the first gill bar in yunnanozoans corresponds to the mandibular arch in jawed vertebrates. Historically, one or more additional arches were postulated in front of the mandibular arch for vertebrate ancestors (*11*). Other views posit no such extra arches and consider the mandibular arch as distinct from the gill arches (*12*–*14*). Yunnanozoan morphology excludes none of these ideas. Their first gill bar could be the elusive arch that was later lost in vertebrates, or, conversely, the mandibular region might not have formed a full skeletal arch. Consistent with the latter scenario, the snout and lips either appear diminutive or are absent in yunnanozoans and other Cambrian forms (*5*–*7*). Specifically, one C-shaped oral structure, as identified by Tian *et al*., may represent what might be a slim mandibular arch of yunnanozoans. If this is true, then a prominent snout and lips, which arise from the neural crest, are a later innovation of the living vertebrate group.

To discriminate between different hypotheses, unequivocal correlates of arch identities are needed. A mandibular arch is not defined by being the most anterior pharyngeal arch. Rather, its identity is predicated on the presence of a specific stream of neural crest cells, a fifth cranial nerve, and specialized mouth structures for pumping water and feeding. Without these markers, and with variations observed among different vertebrate lineages, overall positions of the gills help little to determine arch homology in yunnanozoans.

Tian *et al*. offer an emerging scope of diversity in pharyngeal anatomy of early vertebrates (see the figure). Among other Cambrian fishlike forms, *Haikouichthys* seems to have skeletal rods that support the gills, whereas only the gill pouches are described for *Myllokunmingia* (*4*). *Metaspriggina* has skeletal bars segmented into upper and lower halves (*6*). Hagfish and lampreys evolved from the common ancestor with a cartilaginous basket around the gill pouches and a specialized oral skeleton (*15*). Similar pharyngeal skeletons also occur in successive out-groups of jawed vertebrates (*12*). And jawed vertebrates have a series of jointed skeletal arches, the first of which differentiate as jaws. These different patterns are as anatomically disconnected from each other as they are phylogenetically distant.

Given such a complex distribution of characters, it seems premature to assume any single form as the ancestral phenotype on a linear path toward modern vertebrates. In the phylogenetic analysis by Tian *et al*., yunnanozoans are an out-group to all other vertebrate branches. This suggests differential evolution of pharyngeal patterning among early vertebrate lineages. By the time yunnanozoans were sloshing about in the Cambrian sea, other primitive fishes had evolved to slurp food differently with their uniquely derived pharyngeal anatomy. Although evolutionary biologists have been busy chasing the mythical ancestor that explains everything about the vertebrate body plan, perhaps the opposite is a sensible approach. In other words, the meandering journey toward modern vertebrates may be best understood by populating the family tree with divergent and discontinuous anatomical forms, guided by phylogenetic inference rather than by theory. ■

REFERENCES AND NOTES

1. J.-Y. Chen *et al*., *Nature* **377**, 720 (1995).
2. J.-Y. Chen *et al*., *Nature* **402**, 518 (1999).
3. D. Shu *et al*., *Science* **299**, 1380 (2003).
4. D.-G. Shu *et al*., *Nature* **402**, 42 (1999).
5. D.-G. Shu *et al*., *Nature* **421**, 526 (2003).
6. S. Conway Morris, J.-B. Caron, *Nature* **512**, 419 (2014).
7. Q. Tian *et al*., *Science* **377**, 218 (2022).
8. D. Jandzik *et al*., *Nature* **518**, 534 (2015).
9. J. A. Gillis *et al*., *Nat. Commun.* **4**, 1436 (2013).
10. C. Hirschberger *et al*., *Mol. Biol. Evol.* **38**, 4187 (2021).
11. J. Mallatt, *Zoolog. J. Linn. Soc.* **117**, 329 (2008).
12. P. Janvier, *Early Vertebrates* (Oxford Monographs on Geology and Geophysics, Clarendon Press, 1996).
13. S. Kuratani, *Evol. Dev.* **14**, 76 (2012).
14. T. Miyashita, *Biol. Rev. Camb. Philos. Soc.* **91**, 611 (2016).
15. T. Miyashita, *Can. J. Zool.* **98**, 850 (2020).

Canadian Museum of Nature, Ottawa, Ontario K1P 6P4, Canada. Email: tmiyashita@nature.ca

10.1126/science.adc9198

NEUROSCIENCE

Sounding out pain

A circuit for sound-induced analgesia has been found in the mouse brain

By **Rohini Kuner**[1] *and* **Thomas Kuner**[2]

The perception of physical pain is subject to variation depending on the context and which other sensory inputs are being received, including sound. The emerging field of music therapy (*1*)—which is applied to control postoperative, pediatric, postpartum, and cancer pain and is being increasingly tested in chronic pain disorders—capitalizes on the interactions between sound and pain perception to attenuate pain. Given that music and natural sounds can positively affect mood, relieve stress, and relax the body, it is not unreasonable to think that these factors underlie pain relief. On page 198 of this issue, Zhou *et al*. (*2*) demonstrate that pain relief by sound is not purely attributable to stress reduction and distraction. They interrogate neural circuits to unravel a specific pathway for sound-induced analgesia in the brains of mice.

Using rodents to study how music and sound are related to pain presents major challenges, not least because it is unknown how animals perceive music. Zhou *et al*. carried out behavioral tests addressing pain sensitivity and found that mice did not show differential responses to melodic classical music (consonant sounds), dissonant music, or white noise. Notably, they found that the decisive factor in eliciting pain relief is a 5-dB increase in sound intensity in any of these three types of sound relative to ambient sound levels, whereas 10-, 15-, or 20-dB increases were ineffective. In mouse models, a 5-dB increase in sound intensity led to inhibition of both sensory-discriminative aspects of pain, such as evoked responses aiding escape from noxious stimuli (nociception), and affective behaviors that are linked to suffering and negative emotions associated with acute and chronic pain. Therapeutically relevant findings were that repetitive application of 5-dB sound over ambient levels

[1]Institute of Pharmacology, Heidelberg University, Heidelberg, Germany. [2]Department of Functional Neuroanatomy, Institute for Anatomy and Cell Biology, Heidelberg University, Heidelberg, Germany. Email: rohini.kuner@pharma.uni-heidelberg.de

strained attempts to interpret fossil imprints. Pharyngeal slits and skeletons long precede the origins of vertebrates, and cellular cartilage has been found in the lips of a developing invertebrate chordate (*8*). Vertebrates develop the entire pharyngeal skeleton from cellular cartilage in the embryonic pharyngeal arches. This cartilage originates in the vertebrate-specific cell lineage called the neural crest. To complicate matters further, the pharynx is regionally and phylogenetically differentiated among vertebrates. Depending on whether similarities or differences are emphasized, the pharyngeal arch I in which jaws develop, that is, the mandibular arch, may (*9*–*11*) or may not (*12*–*14*) share an evolutionary origin with gill-supporting skeletons. New information from yunnanozoans presents an opportunity to clarify these issues.

Tian *et al*. avoided jumping to conclusions, as they did not explicitly identify which gill bar in a yunnanozoan corresponds to which arch in a modern vertebrate. But they did signal their favored interpretation that yunnanozoans, with each gill bar identical to the next, represent the ancestral vertebrate condition. This view is in line with the belief that all pharyngeal arches originally supported gills and that one of them evolved into a jaw (*9*–*11*). At face value, yunnanozoans could serve as long-awaited evidence for the gill-arch hypothesis of jaw origins.

However, this jaw-origin narrative relies solely on how one chooses to identify the gill bars of yunnanozoans—that is, whether the first gill bar in yunnanozoans corresponds to the mandibular arch in jawed vertebrates. Historically, one or more additional arches were postulated in front of the mandibular arch for vertebrate ancestors (*11*). Other views posit no such extra arches and consider the mandibular arch as distinct from the gill arches (*12*–*14*). Yunnanozoan morphology excludes none of these ideas. Their first gill bar could be the elusive arch that was later lost in vertebrates, or, conversely, the mandibular region might not have formed a full skeletal arch. Consistent with the latter scenario, the snout and lips either appear diminutive or are absent in yunnanozoans and other Cambrian forms (*5*–*7*). Specifically, one C-shaped oral structure, as identified by Tian *et al*., may represent what might be a slim mandibular arch of yunnanozoans. If this is true, then a prominent snout and lips, which arise from the neural crest, are a later innovation of the living vertebrate group.

To discriminate between different hypotheses, unequivocal correlates of arch identities are needed. A mandibular arch is not defined by being the most anterior pharyngeal arch. Rather, its identity is predicated on the presence of a specific stream of neural crest cells, a fifth cranial nerve, and specialized mouth structures for pumping water and feeding. Without these markers, and with variations observed among different vertebrate lineages, overall positions of the gills help little to determine arch homology in yunnanozoans.

Tian *et al*. offer an emerging scope of diversity in pharyngeal anatomy of early vertebrates (see the figure). Among other Cambrian fishlike forms, *Haikouichthys* seems to have skeletal rods that support the gills, whereas only the gill pouches are described for *Myllokunmingia* (*4*). *Metaspriggina* has skeletal bars segmented into upper and lower halves (*6*). Hagfish and lampreys evolved from the common ancestor with a cartilaginous basket around the gill pouches and a specialized oral skeleton (*15*). Similar pharyngeal skeletons also occur in successive out-groups of jawed vertebrates (*12*). And jawed vertebrates have a series of jointed skeletal arches, the first of which differentiate as jaws. These different patterns are as anatomically disconnected from each other as they are phylogenetically distant.

Given such a complex distribution of characters, it seems premature to assume any single form as the ancestral phenotype on a linear path toward modern vertebrates. In the phylogenetic analysis by Tian *et al*., yunnanozoans are an out-group to all other vertebrate branches. This suggests differential evolution of pharyngeal patterning among early vertebrate lineages. By the time yunnanozoans were sloshing about in the Cambrian sea, other primitive fishes had evolved to slurp food differently with their uniquely derived pharyngeal anatomy. Although evolutionary biologists have been busy chasing the mythical ancestor that explains everything about the vertebrate body plan, perhaps the opposite is a sensible approach. In other words, the meandering journey toward modern vertebrates may be best understood by populating the family tree with divergent and discontinuous anatomical forms, guided by phylogenetic inference rather than by theory. ■

REFERENCES AND NOTES

1. J.-Y. Chen *et al*., *Nature* **377**, 720 (1995).
2. J.-Y. Chen *et al*., *Nature* **402**, 518 (1999).
3. D. Shu *et al*., *Science* **299**, 1380 (2003).
4. D.-G. Shu *et al*., *Nature* **402**, 42 (1999).
5. D.-G. Shu *et al*., *Nature* **421**, 526 (2003).
6. S. Conway Morris, J.-B. Caron, *Nature* **512**, 419 (2014).
7. Q. Tian *et al*., *Science* **377**, 218 (2022).
8. D. Jandzik *et al*., *Nature* **518**, 534 (2015).
9. J. A. Gillis *et al*., *Nat. Commun.* **4**, 1436 (2013).
10. C. Hirschberger *et al*., *Mol. Biol. Evol.* **38**, 4187 (2021).
11. J. Mallatt, *Zoolog. J. Linn. Soc.* **117**, 329 (2008).
12. P. Janvier, *Early Vertebrates* (Oxford Monographs on Geology and Geophysics, Clarendon Press, 1996).
13. S. Kuratani, *Evol. Dev.* **14**, 76 (2012).
14. T. Miyashita, *Biol. Rev. Camb. Philos. Soc.* **91**, 611 (2016).
15. T. Miyashita, *Can. J. Zool.* **98**, 850 (2020).

Canadian Museum of Nature, Ottawa, Ontario K1P 6P4, Canada. Email: tmiyashita@nature.ca

10.1126/science.adc9198

NEUROSCIENCE

Sounding out pain

A circuit for sound-induced analgesia has been found in the mouse brain

By **Rohini Kuner**[1] *and* **Thomas Kuner**[2]

The perception of physical pain is subject to variation depending on the context and which other sensory inputs are being received, including sound. The emerging field of music therapy (*1*)—which is applied to control postoperative, pediatric, postpartum, and cancer pain and is being increasingly tested in chronic pain disorders—capitalizes on the interactions between sound and pain perception to attenuate pain. Given that music and natural sounds can positively affect mood, relieve stress, and relax the body, it is not unreasonable to think that these factors underlie pain relief. On page 198 of this issue, Zhou *et al*. (*2*) demonstrate that pain relief by sound is not purely attributable to stress reduction and distraction. They interrogate neural circuits to unravel a specific pathway for sound-induced analgesia in the brains of mice.

Using rodents to study how music and sound are related to pain presents major challenges, not least because it is unknown how animals perceive music. Zhou *et al*. carried out behavioral tests addressing pain sensitivity and found that mice did not show differential responses to melodic classical music (consonant sounds), dissonant music, or white noise. Notably, they found that the decisive factor in eliciting pain relief is a 5-dB increase in sound intensity in any of these three types of sound relative to ambient sound levels, whereas 10-, 15-, or 20-dB increases were ineffective. In mouse models, a 5-dB increase in sound intensity led to inhibition of both sensory-discriminative aspects of pain, such as evoked responses aiding escape from noxious stimuli (nociception), and affective behaviors that are linked to suffering and negative emotions associated with acute and chronic pain. Therapeutically relevant findings were that repetitive application of 5-dB sound over ambient levels

[1]Institute of Pharmacology, Heidelberg University, Heidelberg, Germany. [2]Department of Functional Neuroanatomy, Institute for Anatomy and Cell Biology, Heidelberg University, Heidelberg, Germany. Email: rohini.kuner@pharma.uni-heidelberg.de

led to analgesia lasting multiple days and that ongoing pain and inflammatory hypersensitivity were suppressed. Although this experimental paradigm is not equivalent to using music and pleasant sounds to evoke analgesia in humans, the study of Zhou *et al.* opens up new directions for research on sound-induced analgesia by creating a model in which the mechanistic underpinnings can be investigated.

Neuroimaging studies on the human brain suggest that a variety of pain-associated regions show changes in activity during music or sound therapy, such as the prefrontal, cingulate, and insular cortices; the thalamic relay nuclei; and the brainstem centers gating descending pathways (*3*). Moreover, recent studies point to a key role for the insular and prefrontal cortices in integrating auditory inputs with pleasant or negative affective emotions (*4*). Analyses of how sound is encoded have recently revealed a rapid transition of brain activity from the primary auditory cortex to higher-order cortical association areas—including the insula cortex—within 220 ms of the onset of the auditory stimulus, suggesting direct connectivity (*5*). However, contrary to the expected corticocortical interaction, Zhou *et al.* found that a corticothalamic pathway mediates sound-induced analgesia (see the figure). This was achieved using high-resolution tracing of anatomical connections originating from the auditory cortex and live monitoring of cellular activity in their projection areas in awake mice. When Zhou *et al.* systematically manipulated efferent projections from the auditory cortex to diverse brain regions using light-induced inhibition of neuronal activity (optogenetic silencing), they found that only the connectivity between the auditory cortex and the thalamic posterior nucleus (PO) and ventral posterior nucleus (VP) is functionally important for sound-induced analgesia. Conversely, selective activation of this pathway in the absence of sound was sufficient to recapitulate sound-induced analgesia.

What makes the thalamus such an important target for audio-somatosensory interactions? All somatosensory inputs—i.e., those leading to the perception of touch, pressure, vibration, temperature, and pain—get relayed to the cortex through excitatory projection neurons of the VP, rendering them a prime site to modulate the magnitude of the forwarded nociceptive signal (*6*). In general, such modulatory influences on thalamic relay neurons are known, but a descending corticothalamic input from layer 5 neurons of the auditory cortex to the somatosensory thalamus is surprising. Furthermore, the involvement of the higher-order PO, which receives a strong corticothalamic input from the primary somatosensory cortex (*6*), is unexpected because it mostly contributes to corticothalamocortical processing and, to a lesser extent, receives input from the spinal cord (*6*). However, the auditory cortex could influence pain processing by modulating the relay of nociceptive signals from the periphery as well as by affecting corticothalamocortical processing of somatosensory signals, a system that is currently poorly understood. Burst and tonic firing modes of thalamic relay neurons are important in the modulation of pain, and deciphering the cellular mechanisms through which auditory inputs affect these will help further elucidate how sound-induced analgesia comes about.

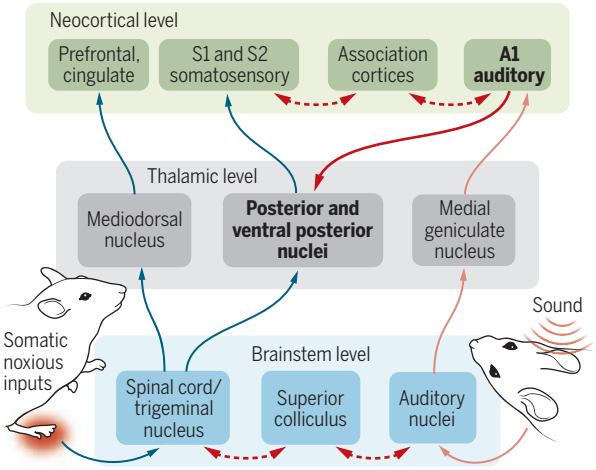
What is special about sound that is 5 dB above ambient noise? This is an intriguing question that might be best studied in an interdisciplinary manner. Detecting a small increase over ambient noise would permit animals to engage innate defense behaviors, which have been shown to be driven through corticofugal projections arising from the auditory cortex (*7*). From an evolutionary perspective, it makes sense that sound-driven defense behaviors are acutely accompanied by pain suppression. Along these lines, it will be insightful to study areas in the brain that play a role in multisensory integration, such as the superior colliculus, or association areas where gustatory (taste), olfactory, visual, auditory, and somatosensory inputs converge, such as the insula cortex.

Previous concepts on using music and sounds for pain relief attributed their effects to distraction-associated analgesia and reduction of anxiety (*8*). Although sound likely contributes to distraction, the study by Zhou *et al.* reveals that sound-induced analgesia is a specific mechanistic entity in its own right, supported by the observation that suppression of pain outlasted the application of sound by several days. Similarly, neither a 5-dB change over ambient noise nor manipulations of the auditory cortex–thalamic pathway led to any changes in anxiety or stress levels. The lack of redundancy of sound-induced analgesia with mechanisms of analgesia elicited by distraction, stress, or anxiolysis, as well as by opioids, have important clinical implications. In particular, sound-induced analgesia could be implemented in parallel to these approaches to additively, and perhaps even synergistically, enhance pain relief and to reduce opioid dosage. These attractive possibilities need rigorous experimental testing and translation to humans.

The interaction between sound and pain is a double-edged sword, so caution in interpreting these data is warranted. Certain sounds can trigger pain, such as headaches, or intensify the pain experience, such as phonophobia, which is common to migraines. Nonpreferred music associated with increased pain ratings has been reported to further enhance gamma band neural oscillations that are evoked by painful stimuli in the somatosensory cortex, whereas this was not true for preferred music (*9*). The study by Zhou *et al.* lays a path for testing the neurobiological basis of these interesting associations in rodent circuitry and manipulating them for optimizing pain relief, which may be translated to humans. ■

How can sound affect pain perception?

Sounds reach the brain via the auditory pathway, signaling through the thalamus into the auditory cortex (A1; light red arrows). External pain stimuli activate a subset of peripheral sensory neurons, creating signals that are processed in, and travel along, the spinal cord and thalamic nuclei before being conveyed to neocortical areas that generate the perception of pain (blue arrows). Zhou *et al.* identified a circuit (red arrow) that mediates sound-induced analgesia. Other possible circuitry is shown using dashed arrows.



REFERENCES AND NOTES

1. L. C. Cole, G. LoBiondo-Wood, *Pain Manag. Nurs.* **15**, 406 (2014).
2. W. Zhuo *et al.*, *Science* **377**, 198 (2022).
3. C. E. Dobek *et al.*, *J. Pain* **15**, 1057 (2014).
4. M. E. Sachs *et al.*, *Neuroimage* **218**, 116512 (2020).
5. L. Bonetti *et al.*, *Neuroimage* **245**, 118735 (2021).
6. R. Kuner, T. Kuner, *Physiol. Rev.* **101**, 213 (2021).
7. H. Wang *et al.*, *PLOS Biol.* **17**, e3000417 (2019).
8. C. R. Millet, L. F. Gooding, *J. Music Ther.* **54**, 460 (2018).
9. M. Hauck *et al.*, *Pain* **154**, 539 (2013).

ACKNOWLEDGMENTS

The authors acknowledge funding by Deutsche Forschungsgemeinschaft (SFB1158, project nos. B01, B06, and B08).

10.1126/science.add0640

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

POLICY FORUM

MARINE POLICY

Noise from deep-sea mining may span vast ocean areas

Potential harm is understudied and largely overlooked

By **Rob Williams[1], Christine Erbe[2], Alec Duncan[2], Kimberly Nielsen[1], Travis Washburn[3], Craig Smith[4]**

Despite increasing interest in deep-sea mining, there are long-standing concerns about environmental impacts on vulnerable and poorly understood ecosystems (*1*–*3*). These concerns took on new urgency in June 2021, when the Republic of Nauru notified the International Seabed Authority [ISA, the intergovernmental body erected by the UN Convention on the Law of the Sea (UNCLOS) responsible for managing seabed resources in areas beyond national jurisdiction] of intent to sponsor an exploitation application for polymetallic nodule mining in the Pacific in 2 years. This triggered a provision in the 1994 Agreement related to UNCLOS, leaving 2 years for the ISA to adopt its first set of exploitation regulations or, failing that, consider Nauru's application under existing international law. We argue that a critical source of potential environmental harm is understudied and largely overlooked: underwater noise generated by mining activities, which can disrupt ecosystems in many ways.

Environmental concerns about deep-sea mining include removal and burial of critical seafloor habitat and large suspended sediment plumes, but underwater noise generated by mining activities has received little study, with limited modeling and measurement in environmental impact assessments (*4*). For example, although sound was considered in the environmental impact statement for the Solwara 1 project (*5*), information was so scarce 15 years ago that only the topside vessel was modeled. Sound is the primary modality for many marine organisms to probe their environment and to communicate. Noise from human activities can affect organisms ranging from plankton to whales through, e.g., acoustic masking, behavior disruption, stress, and hearing loss (*4*). Here we focus on extraction of polymetallic nodules in the Clarion-Clipperton Zone (CCZ), the abyssal region where Nauru's and 16 other nodule-mining exploration contract areas are concentrated, to illustrate the broad issue of potential ocean noise from mining. Many cetacean species occur throughout the CCZ, including endangered migratory baleen whales and deep-diving toothed whales (*6*).

In the absence of empirical data on sound-source characteristics, we used reasonable analogs to estimate potential noise levels from nodule mining equipment. For mining tools that remove nodules from the seabed, we used coastal dredges that remove seafloor gravel and sediment (181 dB re 1 μPa m, 20 Hz to 20 kHz) (*7*). Noise generated by each booster pump along the riser was approximated by pump-out operations from coastal dredging (183 dB re 1 μPa m, 20 Hz to 20 kHz) (*7*). For the mining vessel, we took recordings from floating production storage and offloading (FPSO) vessels used by the oil and gas industry (188 dB re 1 μPa m, 20 Hz to 2 kHz) (*8*). We then applied common sound propagation models (see supplementary materials) to predict how noisy full-scale operations may be. Our predictions are expected to be underestimates. Shallow-water surrogates likely missed acoustic energy below 100 Hz that will be particularly important for heavier machinery required by seabed mining. Sonars, support vessels, and additional pipes and pumps for discharge of dewatering plumes represent additional sound sources that we have not yet considered.

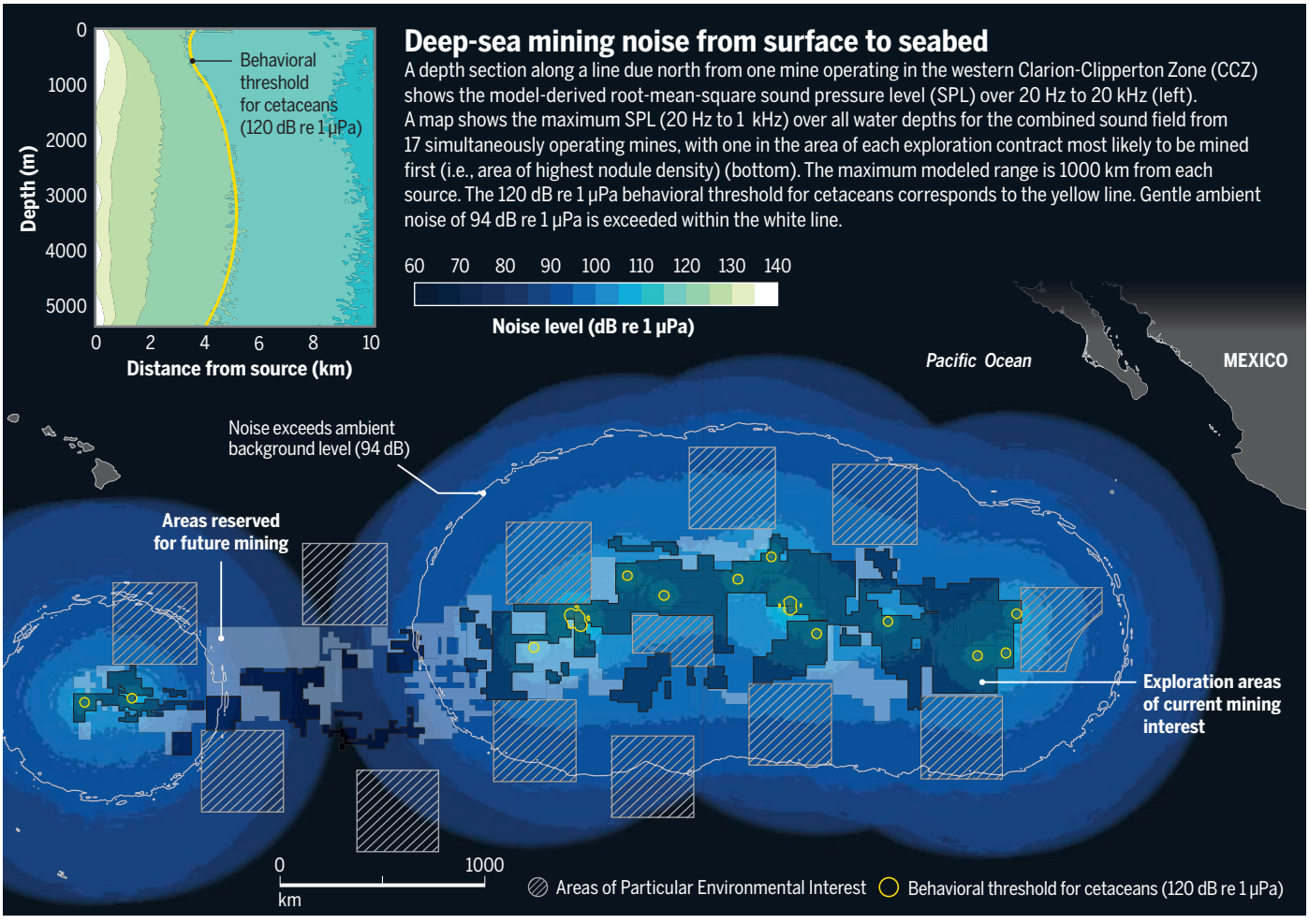
Our predictive model shows that the acoustic environment surrounding a mining operation in the CCZ will likely be substantially altered by even the incomplete list of activities modeled. Noise will be distributed throughout the water column from ships at the surface, risers with pumps every ~1 km in depth, and mining tools on the seabed, ultimately creating a cylinder of sound gradually decreasing in amplitude while propagating outward. Mid-water pumps will couple with the Sound Fixing And Ranging (SOFAR) channel at ~1-km depth. Here, water temperature, salinity, and pressure create a minimum in sound speed. Propagating sound bends toward, and may become trapped in, low-speed water layers and thus efficiently travel over great distances. Our model predicts that mining noise from a single operation will exceed ambient noise levels experienced in gentle weather conditions (~94 dB re 1 μPa) to a range of ~500 km (*9*) (see the figure).

If each contractor operated one mining system, a radius of 4 to 6 km about each mine would exceed 120 dB re 1 μPa, a US National Marine Fisheries Service threshold denoting the possibility of behavioral impacts, generically, to marine mammals. Roughly 5.5 million km^2 would be ensonified above gentle-weather ambient conditions (see the figure) (*9*). Some acoustically sensitive whales show behavioral responses to low-frequency continuous noise <120 dB re 1 μPa (*10*). Most species in the deep sea have yet to be described, and sensitivities to noise have not been studied for even the most common deep-sea species, so no information is available on their responses to noise. However, without sunlight, many species use sound and vibrations in their ecology and thus are likely to be relatively vulnerable to noise from human activities (*4*, *11*).

Our prediction that noise from one mine will exceed gentle-weather background levels to distances of hundreds of kilometers appears to be problematic under the requirements of UNCLOS and for the ISA's draft exploitation guidelines to assess environmental impacts of nodule mining. The UNCLOS framework requires sufficient protection against harmful effects through precautionary management. Under ISA guidelines, ecological metrics in areas affected by mining will be compared to similar metrics in zones within contractor areas required to be unimpacted, called preservation reference zones (PRZs). However, to fit within contractor areas, PRZs are expected to be within 100 to 300 km of mining sites, causing PRZs to be ensonified by mining activities and thus of limited value as controls for understanding impacts of mining noise (*3*, *12*) (see the figure). To be within ambient levels, it may be necessary to place PRZs outside the entire CCZ; however, this would likely cause the PRZs to be ecologically unrepresentative of the areas mined (*13*).

Given the costs and logistical challenges to answering basic questions about sound impacts on deep-sea ecosystems, it seems unlikely that adequate data to assess ecological risks from mining noise will be collected within ~18 months, or even within a 10-year mining moratorium proposed in 2019 (*12*, *14*). Risk assessments will be even more challenging without access to empiri-

[1]Oceans Initiative, Seattle, WA, USA. [2]Centre for Marine Science and Technology, Curtin University, Perth, WA, Australia. [3]National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan. [4]Department of Oceanography, University of Hawai'i at Mānoa, Honolulu, HI, USA. Email: rob@oceansinitiative.org



cal data on sound-source characteristics of nodule-mining equipment. Testing of reduced-scale, incomplete, prototype mining machines is underway; however, contractors, thus far, have not published sound-source characteristics from internal risk assessments or pilot studies. We urge contractors to release in a timely manner information on sound-source characteristics of all seabed-mining components (e.g., from lifting pumps on riser pipes, and rattling of nodules in lifting pipes and on barges).

Transparency from industry, agencies, and scientists related to the acoustic energy produced from mining is urgent. We need to learn about mining systems and their impacts iteratively, through real-world experience (*15*). As such, the regulatory regime should start with cautious thresholds for noise, enable rapid management responses should thresholds be approached, and include regular review of thresholds as more data about noise impacts are reported. Our first approximation of noise generated by polymetallic nodule mining highlights both the potential extent and uncertainty of this noise and its ecological impacts.

Moving forward with nodule mining applications within the next 18 months without data transparency and rigorous standards and guidelines in place would represent the start of a large-scale, uncontrolled experiment. If this industry does proceed through Nauru's claim or elsewhere, we recommend application of the precautionary principle (in keeping with the ISA's duty to provide for effective environmental protection) such that the ISA allows only one or two mining operations to proceed until the long-term cumulative environmental impacts from noise (and other stressors) are well documented (*13*).

The ISA is expected to require that noise impact be considered in environmental impact assessments for future exploitation contracts, but noise impacts will be very difficult to assess without substantially more data on mining noise and ecosystem sensitivities. The ISA is tasked with protecting the marine environment from "serious harm" from seabed mining, but it has yet to define what constitutes serious harm, including unacceptable noise levels. Thresholds should be set at precautionary levels, and contractors should be required to collect environmental and mining noise data and make them publicly available. We recommend that mining should be coupled with intensive, independent studies of regional baselines and environmental impacts, as well as with the ability to alter or halt operations quickly if newly acquired data indicate notable unexpected effects. ■

REFERENCES AND NOTES

1. J. Halfar, R. M. Fujita, *Science* **316**, 987 (2007).
2. K. J. Mengerink *et al.*, *Science* **344**, 696 (2014).
3. C. R. Smith *et al.*, *Trends Ecol. Evol.* **35**, 853 (2020).
4. C. Martin, L. Weilgart, D. J. Amon, J. Müller, Deep-Sea Mining: A noisy affair (2021); https://www.oceancare.org/wp-content/uploads/2021/11/DeepSeaMining_a-noisy-affair_report_OceanCare_2021.pdf.
5. Coffey Natural Systems: Environmental Impact Statement: Nautilus Minerals Niugini Limited, Solwara 1 Project, Vol. A, Coffey Natural Systems, Queensland, Australia (2008).
6. R. Williams *et al.*, *Ecography* **37**, 212 (2014).
7. K. J. Reine, D. Clarke, C. Dickerson, *J. Acoust. Soc. Am.* **135**, 3280 (2014).
8. C. Erbe *et al.*, *J. Acoust. Soc. Am.* **133**, EL465 (2013).
9. F. Niu *et al.*, *Mar. Pollut. Bull.* **173**, 113057 (2021).
10. B. L. Southall *et al.*, *Aquat. Mamm.* **33**, 411 (2007).
11. M. A. Wale, R. A. Briers, K. Diele, *Mar. Pollut. Bull.* **173**, 112958 (2021).
12. D. J. Amon *et al.*, *Mar. Policy* **138**, 105006 (2022).
13. C. R. Smith, M. R. Clark, E. Goetze, A. G. Glover, K. L. Howell, *Front. Mar. Sci.* **8**, **797516** (2021).
14. A. Kung *et al.*, *J. Environ. Manage.* **279**, 111593 (2021).
15. C. S. Holling, *Adaptive Environmental Assessment and Management* (Wiley, 1978).

ACKNOWLEDGMENTS

R.W. and C.E. contributed equally to this work.

10.1126/science.abo2804

GRAPHIC: K. FRANKLIN/*SCIENCE*

BOOKS *et al.*

Robotic spacecraft like the Mars 2020 rover are still reliant on human operators.

PHOTO: NASA/JPL-CALTECH

PLANETARY STUDIES

Interrogating the final frontier from afar

Robots are better equipped for space exploration than humans, argue a pair of authors

By **Matt Shindell**

At the dawn of the space age, it was assumed that humans would be necessary for planetary exploration. Computers were large and limited, robots stiff and clunky. Humans, meanwhile, were versatile and creative manipulators of their environment whose eyes and brains would be needed to make discoveries in situ. Early popular depictions of space exploration featured ships filled with people performing space heroics. Robots, in television shows like Star Trek, were few and far between; even the starship Enterprise's computer, as powerful as it was, required a human interface.

These visions of spaceflight may yet come true—humans may eventually start taking to space in large numbers, building larger ships and space stations and living entire lifetimes in space. We have continuously occupied low Earth orbit throughout the entire 21st century, thanks to the International Space Station (ISS), the space shuttle, and the various Russian and commercial crew capsules that have delivered astronauts and cosmonauts to the station. With China now operating a station of its own, and commercial firms vying to build new stations, these activities show no sign of slowing. NASA now plans to return humans to the Moon, as do other national programs. Even so, the history of space exploration has largely left humans safely in control rooms on Earth, operating the robotic craft that explore the Solar System.

That humans should leave the exploration of space to robots is the central argument of science writer Donald Goldsmith and astronomer Martin Rees in *The End of Astronauts*. We may want to see astronauts exploring new worlds, and we may wish to be among them, they acknowledge, but humans—at least physically—do not make ideal space explorers. The human body is far too vulnerable to the hazards of space and the traumas of spaceflight. And as much as spacecraft and space suits have improved, they remain insufficient to keep humans safe during long journeys.

Much of the research done by humans in space thus far has been related to human spaceflight, but not all. Astronauts have performed or participated in scientific research, including the 2018 production of Bose-Einstein condensate in the ISS Cold Atom Lab. But they do so at great expense and at great risk to themselves. Having humans embark on longer missions farther into the Solar System only increases cost and risk and brings diminishing scientific returns, maintain Goldsmith and Rees.

The authors expand their argument over the course of nine chapters, exploring the challenges, hazards, and expenses of sending humans into orbit or to the Moon, Mars, asteroids, or space colonies. Humans, they argue, may be able to do more science quickly, but robotic exploration will, over time, yield more science per dollar and with less risk to human life. And the robots we use will only become progressively more capable.

Computer technologies have grown not only more capable and reliable in recent years but also much smaller and lighter—a phenomenon that pervades our daily lives as much as it does spacecraft design. The same is true for imaging technologies and electronics in general. Rovers have gotten larger over time, but the instruments they carry use increasingly miniaturized technologies.

For a book that is ostensibly designed to sell the reader on a future of robotic exploration, there is a surprising absence of any detailed history of the development of robotic exploration, systems engineering, the experts who make robotics possible, or the sites where this expertise is produced or practiced. While the authors do provide examples of robotic spacecraft accomplishments, they offer only vague projections about how advancements in artificial intelligence (AI) will continue to improve robotic capabilities.

The End of Astronauts
Donald Goldsmith and Martin Rees
Belknap Press, 2022. 192 pp.

Another notable absence is any discussion of the human work behind robotic exploration. Aside from the design and production of spacecraft, teams of humans operate these craft, while other teams set science goals and analyze returned data. The AI integrated into robotic spacecraft today is minimal. People operate spacecraft, even as spacecraft have facilitated the emergence of the interdisciplinary fields of planetary science and robotic exploration.

In leaving the people out, the authors make it seem as though the spacecraft can—or someday will—do everything on their own. But space exploration, no matter how it is conducted, is human exploration. ■

10.1126/science.abq7394

The reviewer is a space history curator at the Smithsonian's National Air and Space Museum, Washington, DC, USA. Email: shindellm@si.edu

PSYCHOLOGY

The cultural landscape of emotions

Social context plays an important role in shaping how we feel

By **Asifa Majid**

In the animated film *Inside Out*, 11-year-old Riley's emotions are personified as brightly colored internal figures that drive her behaviors. The same five emotions—anger, fear, disgust, sadness, and joy—appear in every other character's head as well, functioning in much the same way in each individual. In Western cultures, this is the case, argues psychologist Batja Mesquita in *Between Us*. Emotions in such contexts, she writes, are considered "MINE," or "Mental, INside the person, and Essentialist," the latter defined in the book as always having the same properties.

This conception of emotion is not universal, however. Emotions elsewhere, she argues, are thought of as "OURS"—"OUtside the person, Relational, and Situated." Using this distinction, Mesquita sets about contrasting emotions in "the West," where the individual is the top concern, with "the Rest," where community is prioritized.

Mesquita describes *amae* as a central emotion in Japanese culture, where it builds interdependence by fostering nurturing indulgence, and *hasham*—which encompasses shame, embarrassment, and social respectability—as a fundamental emotion for Egyptian Bedouins. Such observations provide a backdrop for her to explore a range of issues, including childhood socialization, the nature of friendship, the role of language in shaping emotions, and cross-cultural communication in a globalized world.

Despite Mesquita's emphasis on cross-cultural emotions, there is little discussion of whether the MINE-OURS dichotomy accurately articulates global cultural variation. Other scholars have noted, for example, that hunter-gatherer societies simultaneously emphasize both individual autonomy and social cooperation (*1*, *2*). And in an apparent contradiction to her earlier arguments, Mesquita herself ultimately concludes that Westerners have OURS emotions.

Taken as a whole, however, the book contributes much to the discussion of the origins of emotions, presenting a dazzling array of cross-cultural studies interspersed with personal anecdotes about immigrants' struggles to reconcile diverse emotional and social worlds. In chapter 8, for example, Mesquita describes an incident where she—a Dutch native living in the United States—bumped into the eminent American psychologist Hazel Markus at a conference Markus helped organize. Wishing to express empathy with Markus's workload, Mesquita declared "You look a little tired." The remark appeared to fluster Markus but was intended as an expression of sympathy—to sympathize in Dutch is to acknowledge suffering, not offer a silver lining as in the US. Navigating emotional worlds across cultures is full of hidden pitfalls, we learn.

Sociocultural influences underlie our emotions, argues Mesquita.

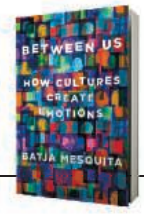
The book's take-home message is radical: There are no natural emotions, no innate emotions, no universal emotions. Mesquita argues that emotions are "meaning making" and "a preparation for action" and that the idea of "emotions as inner states" is a Western construct. Instead, she suggests that emotions are a "dance" cocreated between people who live in a specific cultural context at a particular historical moment.

By pushing a strongly social constructionist account of emotions, Mesquita raises but leaves hanging some questions. If emotional worlds are truly constructed on the fly according to individual histories and identities, for example, then how is intersubjectivity possible? Mesquita claims that emotion lexicons differ radically and are not amenable to direct translation, so language cannot serve as a bridge between cultures.

But differences in words do not necessarily entail differences in emotion concepts. There can be shared prototypes despite variations around the periphery, distinct nonlinguistic concepts represented by a single label, and mental content where there are lexical gaps. Mesquita's theory also fails to account for the possibility that we can be wrong about our own emotions—for example, I thought I was scared, but now I realize I was actually anxious—because, in her view, there is no emotional truth to be known, only one to be constructed. And it is unclear how to incorporate evidence from prelinguistic infants or nonhuman primates into this particular theory of emotions, or how to account for the recurrent emotional motifs that appear across cultures—grief at the loss of a loved one, for example, or disgust upon smelling something rancid.

Sociocultural and biological factors both contribute to emotions, but exactly how large a role is played by either remains controversial. Mesquita emphasizes the importance of the larger social context in shaping emotions and urges empathy and understanding of cultural differences. These are valuable lessons, whatever theory of emotion one subscribes to. ■

Between Us: How Cultures Create Emotions
Batja Mesquita
Norton, 2022. 304 pp.



The reviewer is at the Department of Experimental Psychology, University of Oxford, Oxford OX3 9DU, UK. Email: asifa.majid@psy.ox.ac.uk

REFERENCES AND NOTES

1. K. M. Endicott, in *Anarchic Solidarity: Autonomy, Equality, and Fellowship in Southeast Asia*, T. Gibson, K. Sillander, Eds. (Yale Univ. Southeast Asia Studies, 2011), pp. 62–87.
2. E. Wnuk, Y. Ito, *Cogn. Linguist.* **32**, 195 (2021).

10.1126/science.abq8098

LETTERS

Immigration and population growth are crucial to making Sweden's inbred wolf population sustainable.

Edited by **Jennifer Sills**

Planned cull endangers Swedish wolf population

In May, the Swedish Parliament announced a goal to reduce the Swedish wolf population from about 400 to about 200 individuals (*1*). This action further threatens this highly endangered population, which is genetically isolated and inbred. Scientific advice for improvements has not been implemented (*2*, *3*).

The Swedish Parliament proposed this drastic cull at a time when biodiversity is a global focus. The 50-year anniversary of the first UN conference on the environment was celebrated in June, and the UN Convention on Biological Diversity (CBD) will soon finalize its global biodiversity framework for 2020 to 2050. Sweden's actions are inconsistent with the country's obligations under the CBD and European Union law.

Few wild populations are as well studied as the Scandinavian wolf. Genetic monitoring has provided a full pedigree since the population was reestablished in the 1980s after extinction, and the data confirm persisting genetic isolation (*4*–*6*). Hunting, conducted both legally and illegally, has prevented population expansion and the influx of genetic variation.

Three founders comprised the population's genetic origin until 2007, and only three more wolves have subsequently contributed genetically to the present population (*6*). The genetic base is thus extremely narrow, and genomic erosion has been confirmed (*7*, *8*). The average level of inbreeding is similar to the level found in the offspring of two full siblings (*6*). Inbreeding in this population has been shown to reduce litter size (*4*). Also, high frequencies of anatomical defects (*9*) and male reproductive disorders (*10*) have been observed.

To make this population viable, population size and immigration must increase. So far, the population has been too small, and limited immigration followed by inbreeding could lead to extinction, similar to the Isle Royale wolf population (*11*). The goal should be to recreate a well-connected metapopulation spanning Scandinavia and Finland with a genetically effective population size of over 500, in line with the proposed CBD indicator (*12*). Considerably more genetic exchange than the current one-migrant-per-generation aim is needed (*3*).

Linda Laikre[1]*, Fred W. Allendorf[2], Jouni Aspi[3], Carlos Carroll[4], Love Dalén[1,5,6], Richard Fredrickson[7], Christina Hansen Wheat[8], Philip Hedrick[9], Kerstin Johannesson[10], Marty Kardos[11], Rolf O. Peterson[12], Mike Phillips[13], Nils Ryman[1], Jannikke Räikkönen[14], Carles Vilà[15], Christopher W. Wheat[1], Cristiano Vernesi[16], John A.Vucetich[12]

[1]Department of Zoology, Stockholm University, SE 10691 Stockholm, Sweden. [2]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA. [3]Ecology and Genetics Research Unit, University of Oulu, FIN-90014 Oulu, Finland. [4]Klamath Center for Conservation Research, Orleans, CA 95556, USA. [5]Centre for Palaeogenetics, SE 10691 Stockholm, Sweden. [6]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE 10405 Stockholm, Sweden. [7]Missoula, MT 59802, USA. [8]Department of Biology, Lund University, SE 22362 Lund, Sweden. [9]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA. [10]Department of Marine Sciences, University of Gothenburg, Tjärnö, Sweden. [11]National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA 98112, USA. [12]College of Forest Resources & Environmental Science, Michigan Technological University, Houghton, MI 49931, USA. [13]Turner Endangered Species Fund, Bozeman, MT 59718, USA. [14]Department of Environmental Research and Monitoring, Swedish Museum of Natural History, SE 10405 Stockholm, Sweden. [15]Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana–Consejo Superior de Investigaciones Científicas, 41092 Sevilla, Spain. [16]Forest Ecology Unit, Research and Innovation Centre–Fondazione Edmund Mach, San Michele all'Adige 38010 (TN), Italy.
*Corresponding author.
Email: linda.laikre@popgen.su.se

REFERENCES AND NOTES

1. Sveriges Riksdag, "Committees want to see changes in species protection and in the management of wolves" (2022); www.riksdagen.se/sv/aktuellt/2022/maj/5/utskott-vill-se-andringar-i-artskyddet-och-i-forvaltningen-av-varg/ [in Swedish].
2. M. M. Hansen *et al.*, "Evaluation of the conservation genetic basis of management of grey wolves in Sweden" (Swedish Government Investigation SOU 2011:37, 2011).
3. L. Laikre *et al.*, *Heredity* **117**, 279 (2016).
4. O. Liberg *et al.*, *Biol. Lett.* **1**, 17 (2005).
5. M. Åkesson *et al.*, *Conserv. Genet.* **23**, 359 (2022).
6. M. Åkesson, L. Svensson, "The pedigree of the Scandinavian wolf population up until 2021: Report from the Swedish Wildlife Damage Centre 2022–3" (Swedish University of Agricultural Sciences, 2022) [in Swedish].
7. M. Kardos *et al.*, *Nat. Ecol. Evol.* **2**, 124 (2018).
8. A. Viluma *et al.*, *Genome Res.* **32**, 1 (2022).
9. J. Räikkönen *et al.*, *PLOS ONE* **8**, e67218 (2013).
10. Swedish National Veterinary Institute (SVA), "License hunting of wolves 2021, report 65/2021" (2021) [in Swedish].
11. P. W. Hedrick *et al.*, *Anim. Conserv.* **22**, 302 (2019).
12. CBD, "Recommendation adopted by the Subsidiary Body on Scientific, Technical, and Technological Advice" (2021), p. 10, Headline Indicator A.0.4; www.cbd.int/doc/recommendations/sbstta-24/sbstta-24-rec-02-en.pdf.

COMPETING INTERESTS

This work is the view of the authors and not necessarily that of the agencies they represent.

10.1126/science.add5299

End animal testing for biosimilar approval

Drug toxicology testing in animals has long been a standard requirement for establishing the safety of both new drugs (*1*) and copies of biological drugs coming off patent, known as biosimilars (*2*). Recently, the international community has acknowledged that this type of test may not be necessary or useful. Although policies for new drug approval are in the process of changing, biosimilar approval policies have been overlooked. Regulatory agencies should update these policies to streamline the biosimilars approval process and to prevent unnecessary, and thus unethical, animal testing.

Policies requiring animal toxicology studies to test biosimilars often stipulate the

PHOTO: RONALD KAMPHUIS/MINDEN PICTURES

use of a dose multiple times as high as the human dose (*3*). This strategy fails to recognize that higher doses lead to nonlinear responses, which invalidates the results. In addition, the animal species used in many studies do not have the binding receptors that the drugs target in humans, or they have similar receptors that bind to drugs at concentrations different from those in humans. Given the differences in receptors, either animals cannot respond or develop side effects to the drug, or their response lacks relevance to the human response (*4*).

The US Food and Drug Administration (FDA) is discussing alternative models of testing for new drugs (*5*, *6*) but still requires animal toxicology studies for biosimilars (*2*). The requirement remains even though the FDA has discarded most of these studies, which are often conducted by the dozens for a single biosimilar drug (*7*, *8*). The World Health Organization and other regulatory agencies [e.g., (*3*)] also still require animal toxicology studies for biosimilar approval (*9*). Only the European Medicines Agency (*10*) has updated its policy to reflect that animal pharmacology and toxicology studies are irrelevant to the evaluation of biosimilars.

More than 120 biosimilars have been approved in the United States and the European Union (*7*), and no results from the required animal studies have yielded useful information about responses or side effects in humans. It is time for regulatory agencies to forbid animal testing for biosimilar approval for the sake of animal welfare, cost, and time.

Sarfaraz K. Niazi
Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois, Chicago, IL 60612, USA. Email: sniazi3@uic.edu

REFERENCES AND NOTES

1. S. W. Junod, "FDA and clinical drug trials: A short history" (FDA, 2014); www.fda.gov/media/110437/download.
2. C. B. Rangel, "H.R. 3590—111th Congress (2009–2010): Patient Protection and Affordable Care Act" (2010), www.congress.gov/bill/111th-congress/house-bill/3590.
3. "Guidelines on similar biologics: Regulatory requirements for marketing authorization in India" (2016); https://cdsco.gov.in/opencms/resources/UploadCDSCOWeb/2018/UploadAlertsFiles/BiosimilarGuideline2016.pdf.
4. P. van Meer *et al.*, *Drug Discov. Today* **20**, 483 (2015).
5. FDA, "Publications co-authored by FDA on alternative methods" (2022); www.fda.gov/science-research/about-science-research-fda/publications-co-authored-fda-alternative-methods.
6. A. M. Avila *et al.*, *Reg. Toxicol. Pharm.* **114**, 104662 (2020).
7. S. K. Niazi, *Biologics* **2**, 107 (2022).
8. S. K. Niazi, *Expert Opin. Biol. Ther.* **22**, 149 (2022).
9. World Health Organization (WHO), "Guidelines on evaluation of biosimilars" (2022); https://cdn.who.int/media/docs/default-source/biologicals/who-guidelines-on-evaluation-of-biosimilars---4-nov-2021.pdf?sfvrsn=f17799ae_5.
10. European Medicines Agency, "Tailored scientific advice for biosimilar development" (2021); www.ema.europa.eu/en/documents/report/tailored-scientific-advice-biosimilar-development-report-experience-pilot-2017-2020_en.pdf.

10.1126/science.add4664

LIFE IN SCIENCE

Danger in the desert

Deep in Chile's Atacama Desert, a landscape often compared to Mars, our team of four female microbiologists watched as a car full of curious men pulled up beside us. Because we were strangers in a desolate place, our minds immediately jumped to ways we could protect ourselves. Defensively, our Chilean compatriot hefted the sturdy shovel she'd been using to dig up plant roots. The rest of us tried to look braver than we felt.

We had come to this desert to conduct DNA studies on giant horsetails, fern-related plants that somehow flourish in one of Earth's driest places. We were searching for plants in the most remote locations, where they would be unaffected by human activities such as mining and agriculture.

We'd been warned that the trip could be dangerous. Because we were traveling so far from fuel sources, we were told to take along a can of gasoline. Our destination was at the end of a tortuous single-lane dirt road lined with burned-out vehicles that had not successfully negotiated the steep descent. Our sample site was near a village, and the people, we were told, might not respond positively to us. We were instructed to report our travel plans at the nearest police station so that search parties would know where to look for us if we disappeared.

We had found the amazing plants; their bright green stems towered over our heads, evoking thoughts of ancient Carboniferous swamp vegetation. The men approached as we finished collecting our samples. We waited tensely as a man exited the car and walked toward us. To our surprise and relief, he politely invited us to visit their village—they wanted to show us a lovely church of which they were justifiably proud. That day, we learned about more than the microbiomes that help desert plants thrive. We also met a welcoming community who had likewise beautifully adapted to their challenging home.

Call for submissions
Life in Science is an occasional feature highlighting some of the humorous or unusual day-to-day realities that face our readers. Can you top this? Submit your story to www.submit2science.org.

Linda Graham
Department of Botany, University of Wisconsin–Madison, Madison, WI 53706, USA.
Email: lkgraham@wisc.edu

10.1126/science.add3926

The author's team negotiated remote roads deep in Chile's Atacama Desert.

PHOTO: MARIE TREST

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PHOTO: APPIO STUDIOS/SHUTTERSTOCK

ANTHROPOLOGY

The emergence of cumulative culture

Experimental research has found that individuals with no prior knowledge spontaneously produce the types of stone tools associated with early hominins. One of the hallmarks of modern humans is the capacity for cumulative culture: the ability for knowledge to accumulate across generations through cultural transmission. Archaeologists have thought that this capacity first emerged with the appearance of the Oldowan stone tool industry about 2.6 million years ago. Snyder *et al.* tested this supposition by presenting naïve individuals with a puzzle that could be solved by making simple core and flake tools, finding that individuals taught themselves to make artifacts that match Oldowan industry examples. These results suggest that the capacity for cumulative culture is more recent than the Oldowan industry. —SGO

Sci. Adv. 10.1126/sciadv.abo2894 (2022).

Experiments with modern humans suggest that specialized training was not necessary for early hominins to craft Oldowan industry tools.

NEUROSCIENCE

Boosting brain power with omega-3s

Activation of the nuclear receptor peroxisome proliferator–activated receptor α (PPARα) in mice is neuroprotective and enhances memory. Liu *et al*. found an endogenous ligand for PPARα in the brain that may promote these effects. 7(*S*)-HDHA, a hydroxylated derivative of the omega-3 fatty acid DHA, bound to PPARα with high affinity. Treating primary cortical neurons with 7(*S*)-HDHA promoted dendritic outgrowth and arborization, as well as the expression of PPARα target genes that are associated with synaptic plasticity. Supplementing the diet of young male rats with DHA increased the amount of 7(*S*)-HDHA in the cortex. —LKF

Sci. Signal. **15**, eabo1857 (2022).

ORGANIC CHEMISTRY

Toward automation of common chemistry

Most experiments in modern organic chemistry require prior preparation of the reactants using procedures found in the literature. Automating the most common such procedures could save the research community a great deal of time while also facilitating crowd-sourced optimization. Rohrbach *et al*. translated more than 100 popular protocols into formats executable on a synthesis machine, half of which they then validated experimentally. The authors also set up an open database to which new executable protocols can be added. —JSY

Science, abo0058, this issue p. 172

ACTUATORS

Designing better ways to stretch

Dielectric elastomers are a type of electroactive polymer that can transform electrical energy into mechanical work. They are often made from silicone or acrylic elastomers. Silicone materials are easier to process but have lower strains, whereas acrylics require prestretching to achieve high strains. Shi *et al*. designed a dielectric elastomer based on acrylates in which the selection of monomers, the

use of both a short and long cross-linker molecules, and the ability to control hydrogen bonding enabled fine tuning of the stress-strain responses and viscoelasticity. The authors further developed a method to dry stack the elastomeric sheets without a loss of properties that is more compatible with current large-scale fabrication processes. —MSL

Science, abn0099, this issue p. 228

STRUCTURAL BIOLOGY

Dynamic control of GPCRs

Arrestins are a group of proteins that regulate signaling through G protein–coupled receptors (GPCRs). They are best known as an off switch in signaling through G proteins, but they also coordinate G protein–independent signaling. Kleist *et al*. took advantage of an intrinsically β-arrestin–biased GPCR, atypical chemokine receptor 3 (ACKR3), to study β-arrestin recruitment. Nuclear magnetic resonance spectroscopy experiments support a role for conformational selection. The inactive state shows conformational heterogeneity at the ligand-binding pocket. Ligand binding can cause stabilization of an active state that in turn tunes the dynamics at the intracellular region to allow β-arrestin recruitment. —VV

Science, abj4922, this issue p. 222

GENDER-BASED VIOLENCE

Tools to improve police responsiveness

Violence against women is a problem of global concern, and inadequate police responsiveness in many settings is a factor in its persistence. Sukhtankar *et al*. undertook a randomized controlled trial of police reform measures in India that was designed to improve reporting rates of incidences of violence against women in a context where such violence is widespread and where getting the police to take action has been historically difficult (see the Perspective by Blair and Jassal). The authors found that the intervention, which was composed of training, outreach, and the assignment of female officers across 180 police stations, significantly increased police registration of cases of violence against women. These findings carry policy implications for efforts to improve police responsiveness to abused women more widely. —AMS

Science, abm7387, this issue p. 191; see also abp9542, p. 150

CATALYSIS

Better on the outside

One way to maximize the activity of scarce and costly noble metals in catalysis is to use small cluster or even single atoms on metal oxide supports. Bifunctional platinum and zeolite catalysts are used industrially in alkane hydroconversion. Cheng *et al*. show that the amount of platinum used could be reduced and isomer selectivity improved when the catalyst preparation created platinum clusters on the surface of zeolite or its alumina binder, rather than in the interior of the zeolite channels. Preparations that would have formed single-atom catalysts were less active because the platinum ions adsorbed on the alumina surface were more difficult to reduce. —PDS

Science, abn8289, this issue p. 204

THERMOELECTRICS

Disordered but efficient

Thermoelectric materials, many having relative simple compositions, convert heat into electricity. However, Jiang *et al*. found that adding more cations into a germanium tellurium–based material stabilized a phase with excellent thermoelectric properties. This high-entropy material has low thermal conductivity due to the cation disordering but improved symmetry that helps to maintain good electrical properties. The material was used in several devices that showed a high thermoelectric efficiency. —BG

Science, abq5815, this issue p. 208

IN OTHER JOURNALS

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

DEVELOPMENT

Tissue fusion aided by migration

During development, tissues sometimes need to fuse to build structure. Failure of fusion in the human lip and palate can result in cleft lip and/or palate, one of the most common congenital abnormalities. The palate forms as precursor tissue grows to meet at the midline, above the tongue. Edges of the growing tissues are covered by squamous periderm cells that prevent premature fusion. Teng *et al.* show that apoptosis plays a part in the removal of the periderm cells but is dispensable; removal of fusion-resistant surface cells depends crucially on epithelial migration. The epithelial surface breaks up as the cells converge into migrating streams that sweep away the surface of the developing palate precursors before midline fusion. —PJH *Development* **149**, dev200181 (2022).

Cell migration, rather than apoptosis, is crucial for correct antenatal closure of the palate.

CELL BIOLOGY

Organizing actin bundles

Plastins are actin-bundling proteins that are involved in cellular processes from cell migration to endocytosis. In the extensions of migrating cells called filipodia, actin filaments form aligned bundles, but in the endocytic machinery, they form branched networks. Schwebach *et al.* suggest that the ability of plastin to remodel actin into a range of structures comes from a tunable interaction between the two actin-binding domains, ABD1 and ABD2. Initial binding to actin filaments weakens the interaction between ABD1 and ABD2, and this may be sufficient to allow cross-linking of aligned actin filaments. Phosphorylation further weakens the ABD1–ABD2 interaction, which may be required to cross-link meshed networks of actin filaments. A structure of ABD1 bound to F actin supports this model of a tunable interaction between ABD1 and ABD2

PHOTO: MECHIR/ALAMY STOCK PHOTO

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

IMMUNOMETABOLISM

Metabolic rewiring helps wounds heal

Rapid healing after skin injury is essential for protecting our internal organs from external threats. Epithelial cells must cope with a wound environment low in oxygen while simultaneously multiplying and migrating to quickly repair damage. Such adaptations to oxygen deprivation have long been thought to rely on direct cellular sensing and responsiveness by activating hypoxia-inducible factor 1α (HIF1α). Konieczny *et al*. found instead that the cytokine interleukin-17A (IL-17A), which is produced by skin-resident γδT cells, facilitates the adaptation of epithelium to hypoxia through HIF1α. This IL-17A–HIF1α axis activates a program of glycolytic metabolism in epithelial cells that fuels repair. These findings underscore a role for immune-derived secondary signals in expediting repair by driving cellular adaptation to chronic hypoxia. —STS

Science, abg9302, this issue p. 170

BIOENGINEERING

Replicating the structure of a heart

The orientation of matrix fibers in many tissues, including those in the heart, can influence proper function or dysfunction. Chang *et al*. developed focused rotary jet spinning that enables the deposition of polymeric materials in the form of long, thin fibers with a preferred orientation at the micrometer scale (see the Perspective by Sefton and Simmons). Using this method, the authors were able to fabricate artificial hearts that have properties similar to normal human hearts, but they could also deliberately misalign the fiber orientation to build models of diseased hearts. Once fully constructed, contractile cells are seeded onto the scaffold so that the composite form takes on some of the active properties of a natural heart. This model can also show how the misorientation of specific fibers affects functioning. —MSL

Science, abl6395, this issue p. 180; see also add0829, p. 148

MEMBRANES

Unite and expand for better models

Transport through swollen polymer membranes has traditionally been described by either a pore flow or a solution-diffusion model. The former works best for porous membranes, in which transport is controlled by size exclusion and electrostatic effects, whereas the latter works best for nonporous membranes, in which transport is driven by solubility and diffusivity. Hegde *et al*. show that is possible to unite these two classical models using a fluid-solid model that treats the solvent and the matrix as separate phases (see the Perspective by Geise). This approach captures the predictions of the earlier models and allows for the determination of both the intramembrane pressure and the concentration profiles. It also allows for the modeling of mechanically heterogeneous membranes, and thus may be a useful tool in membrane design. —MSL

Science, abm7192, this issue p. 186; see also abn5485, p. 152

NEUROSCIENCE

The pain-reducing effects of music

That sound can effectively suppress pain has been known for some time. However, it is still unclear what drives the analgesic effect induced by music or noise. Zhou *et al*. used a range of methods to demonstrate in mice that the auditory cortex is functionally connected to regions involve52 (see the Perspective by Kuner and Kuner). The neuronal circuits depend on the physical location of the pain. Whereas the analgesic effect of a 5-decibel signal-to-noise ratio white noise on the hindpaws involved projections from the auditory cortex to the posterior thalamic nuclei, on the forepaws, it involved projections from the auditory cortex to the ventral posterior nuclei. Distinct thalamic nuclei are thus involved in the processing of nociceptive information perceived at distinct physical locations. —PRS

Science, abn4663, this issue p. 198; see also add0640, p. 155

CHEMISTRY

Dilution-induced ordering

Many molecules, such as surfactants, can form ordered structures when placed in solution. Typically, the systems become more ordered and the structures change from spheres to elongated shapes as the concentration is increased. Su *et al*. studied a system of benzene-1,3,5-tricarboxamide) (BTA-EG4) with the cationic surfactant octyltrimethylammonium bromide (OTAB), in water (see the Perspective by Webber). BTA-EG4 undergoes supramolecular polymerization in water and will form hydrogels at higher concentrations, whereas OTAB will form small aggregates. However, when combined, the OTAB initially disrupts the BTA-EG4 hydrogels, but these can be reestablished upon dilution because this lessens the effect of the surfactant. With careful engineering, this can be expanded to a gel-sol-gel-sol system as a function of concentration. —MSL

Science, abn3438, this issue p. 213; see also abo7656, p. 153

PALEONTOLOGY

Vertebrate ancestor?

Yunnanozoans are Cambrian animals with a taxonomic position that has long been debated with regard to whether they are ancestral chordates. Tian *et al*. use new imaging approaches on new yunnanozoan specimens and found evidence that their branchial arches are composed of cartilage within an extracellular matrix of micofibrils (see the Perspective by Miyashita). This combination of tissue types has been considered vertebrate specific, suggesting that this group of animals are indeed basal vertebrates. This relationship allows insight into the evolution of the pharyngeal skeleton (including the jaw and cranium) to be obtained from examination of this ancient progenitor. —SNV

Science, abm2708, this issue p. 218; see also adc9198, p. 154

PHYSICS

Understanding an exotic phase

The nature of strange metals, a metallic phase with unconventional transport properties that appears in phase diagrams of materials such as cuprate superconductors, remains one of the major puzzles in condensed matter physics. Phillips *et al*. reviewed experimental and theoretical progress toward understanding this phase. The authors examined its relationship to quantum criticality, Planckian dissipation, and quantum gravity. —JS

Science, abh4273, this issue p. 169

NEUROSCIENCE

The difference between human and mouse

Over the past few decades, the mouse has become a model organism for brain research. Because of the close evolutionary similarity of ion channels, synaptic receptors, and other key molecular constituents of the mouse brain to that of humans, corresponding similarity has been assumed for cortical neuronal circuits.

However, comparative synaptic-resolution connectomic studies are required to determine the degree to which circuit structure has evolved between species. Using three-dimensional electron microscopy, Loomba *et al*. compared mouse and human/macaque cortex synaptic connectivity. Although human cells are much larger compared with mouse neurons and are more numerous, on average, they do not receive more synapses. And, even though there are three times more interneurons in the human cortex than in the mouse, the excitation-to-inhibition ratio is similar between the species. —PRS

Science, abo0924, this issue p. 171

DRUG DEVELOPMENT

Screening to save sarcomeres

Loss-of-function mutations in *BAG3* are associated with dilated cardiomyopathy. Yang *et al.* performed high-content phenotypic screening in human induced pluripotent stem cell–derived cardiomyocytes (iPSC-CMs) lacking *BAG3* expression. Using deep learning, the authors identified histone deacetylase (HDAC) and microtubule inhibitors as potential cardioprotective agents. HDAC6 inhibition reduced sarcomere damage in cells and protected against heart failure in BAG3 cardiomyocyte-specific knockout mice and in mice harboring a particular mutation in *BAG3*. An HDAC6-specific inhibitor reduced sarcomere damage and apoptosis and improved mitochondrial function in mice and human iPSC-CMs, supporting the translational potential of HDAC6 inhibition for dilated cardiomyopathy. —CC

Sci. Transl. Med. **14**, eabl5654 (2022).

COSMOLOGY

Dark matter isn't hiding here

The existence of dark matter is inferred from a wide range of astrophysical observations, and some models suggest that it makes up 85% of the matter in the universe. However, very little is known about the nature of dark matter. Researchers have used searches by mass, motivated by different models of dark matter composition and interactions with radiation, to try to constrain what is possible. All such constraints to date have been negative, including the current work by Quiskamp *et al.*, which eliminated a specific mass range and thus also eliminated one well-motivated model. The search goes on until, with luck, someone hits on the right place to look. —WW

Sci. Adv. 10.1126/sciadv.abq3765 (2022).

MACROPHAGES

Nature over nurture

Alveolar macrophages derived from fetal monocytes (FeMo-AMs) are crucial for defense against respiratory viral infections. During influenza A virus infection, FeMo-AMs are depleted, but it is unclear whether this population is replenished by self-renewing FeMo-AMs or by infiltration of bone marrow–derived (BMo)-AMs. Li *et al.* used multiple mouse models to investigate the long-term phenotype, fate, and function of these two cell types after recurrent influenza virus infection. AMs were initially replaced by self-renewal of FeMo-AMs but, over time, were outcompeted by BMo-AMs, which had superior glycolytic and proliferative capacity. Replacement of FeMo-AMs by naive, inflammation-experienced, or aged BMo-AMs promoted influenza pathogenesis and mortality. These findings demonstrate that the origin of AMs rather than previous experience defined their long-term function in recurrent viral infection. —HMI

Sci. Immunol. **7**, eabj5761 (2022).

use of both a short and long cross-linker molecules, and the ability to control hydrogen bonding enabled fine tuning of the stress-strain responses and viscoelasticity. The authors further developed a method to dry stack the elastomeric sheets without a loss of properties that is more compatible with current large-scale fabrication processes. —MSL

Science, abn0099, this issue p. 228

STRUCTURAL BIOLOGY

Dynamic control of GPCRs

Arrestins are a group of proteins that regulate signaling through G protein–coupled receptors (GPCRs). They are best known as an off switch in signaling through G proteins, but they also coordinate G protein–independent signaling. Kleist *et al*. took advantage of an intrinsically β-arrestin–biased GPCR, atypical chemokine receptor 3 (ACKR3), to study β-arrestin recruitment. Nuclear magnetic resonance spectroscopy experiments support a role for conformational selection. The inactive state shows conformational heterogeneity at the ligand-binding pocket. Ligand binding can cause stabilization of an active state that in turn tunes the dynamics at the intracellular region to allow β-arrestin recruitment. —VV

Science, abj4922, this issue p. 222

GENDER-BASED VIOLENCE

Tools to improve police responsiveness

Violence against women is a problem of global concern, and inadequate police responsiveness in many settings is a factor in its persistence. Sukhtankar *et al*. undertook a randomized controlled trial of police reform measures in India that was designed to improve reporting rates of incidences of violence against women in a context where such violence is widespread and where getting the police to take action has been historically difficult (see the Perspective by Blair and Jassal). The authors found that the intervention, which was composed of training, outreach, and the assignment of female officers across 180 police stations, significantly increased police registration of cases of violence against women. These findings carry policy implications for efforts to improve police responsiveness to abused women more widely. —AMS

Science, abm7387, this issue p. 191; see also abp9542, p. 150

CATALYSIS

Better on the outside

One way to maximize the activity of scarce and costly noble metals in catalysis is to use small cluster or even single atoms on metal oxide supports. Bifunctional platinum and zeolite catalysts are used industrially in alkane hydroconversion. Cheng *et al*. show that the amount of platinum used could be reduced and isomer selectivity improved when the catalyst preparation created platinum clusters on the surface of zeolite or its alumina binder, rather than in the interior of the zeolite channels. Preparations that would have formed single-atom catalysts were less active because the platinum ions adsorbed on the alumina surface were more difficult to reduce. —PDS

Science, abn8289, this issue p. 204

THERMOELECTRICS

Disordered but efficient

Thermoelectric materials, many having relative simple compositions, convert heat into electricity. However, Jiang *et al*. found that adding more cations into a germanium tellurium–based material stabilized a phase with excellent thermoelectric properties. This high-entropy material has low thermal conductivity due to the cation disordering but improved symmetry that helps to maintain good electrical properties. The material was used in several devices that showed a high thermoelectric efficiency. —BG

Science, abq5815, this issue p. 208

IN OTHER JOURNALS

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

DEVELOPMENT

Tissue fusion aided by migration

During development, tissues sometimes need to fuse to build structure. Failure of fusion in the human lip and palate can result in cleft lip and/or palate, one of the most common congenital abnormalities. The palate forms as precursor tissue grows to meet at the midline, above the tongue. Edges of the growing tissues are covered by squamous periderm cells that prevent premature fusion. Teng *et al.* show that apoptosis plays a part in the removal of the periderm cells but is dispensable; removal of fusion-resistant surface cells depends crucially on epithelial migration. The epithelial surface breaks up as the cells converge into migrating streams that sweep away the surface of the developing palate precursors before midline fusion. —PJH *Development* **149**, dev200181 (2022).

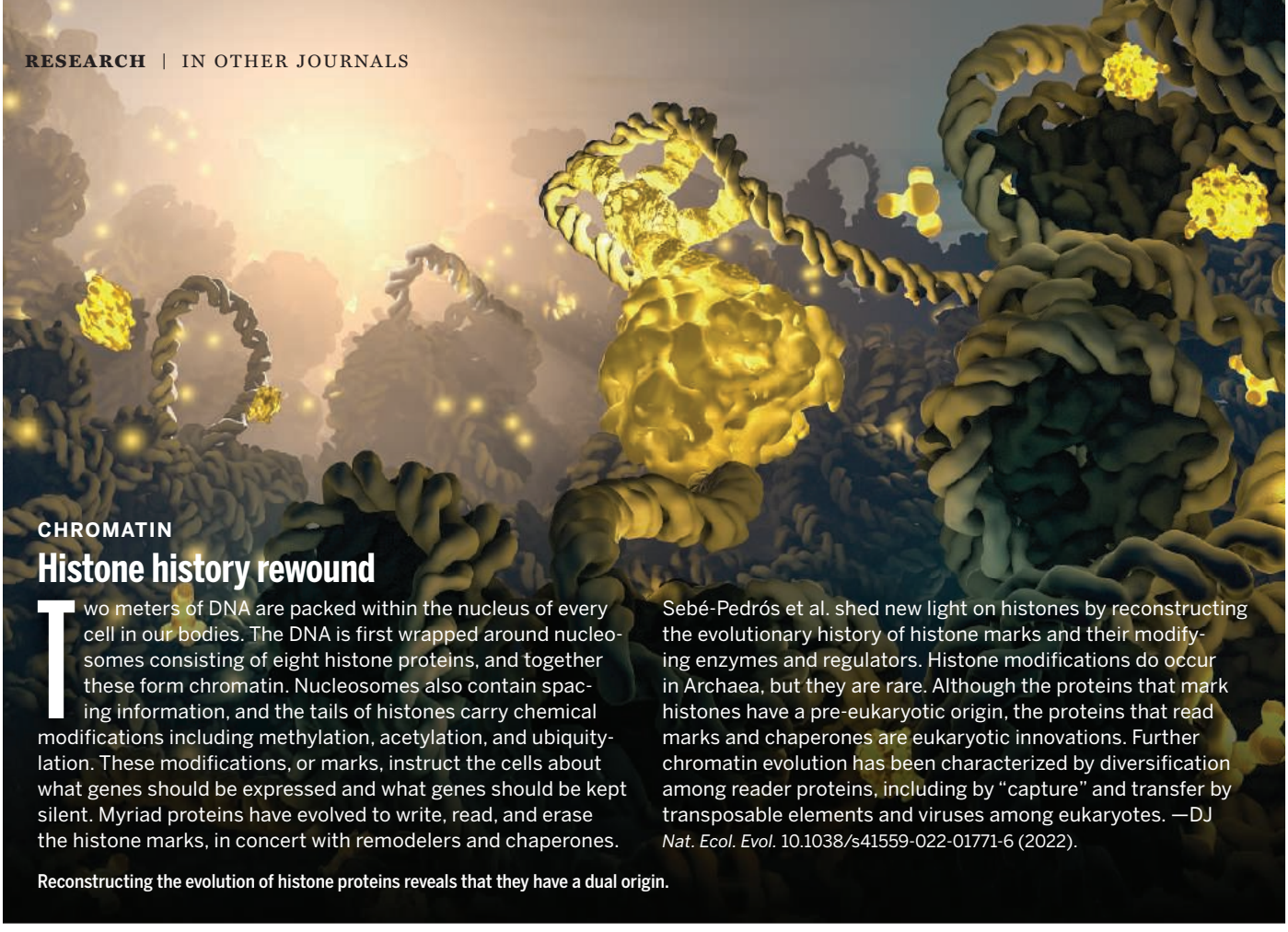
Cell migration, rather than apoptosis, is crucial for correct antenatal closure of the palate.

CELL BIOLOGY

Organizing actin bundles

Plastins are actin-bundling proteins that are involved in cellular processes from cell migration to endocytosis. In the extensions of migrating cells called filipodia, actin filaments form aligned bundles, but in the endocytic machinery, they form branched networks. Schwebach *et al.* suggest that the ability of plastin to remodel actin into a range of structures comes from a tunable interaction between the two actin-binding domains, ABD1 and ABD2. Initial binding to actin filaments weakens the interaction between ABD1 and ABD2, and this may be sufficient to allow cross-linking of aligned actin filaments. Phosphorylation further weakens the ABD1–ABD2 interaction, which may be required to cross-link meshed networks of actin filaments. A structure of ABD1 bound to F actin supports this model of a tunable interaction between ABD1 and ABD2

PHOTO: MECHIR/ALAMY STOCK PHOTO



CHROMATIN

Histone history rewound

Two meters of DNA are packed within the nucleus of every cell in our bodies. The DNA is first wrapped around nucleosomes consisting of eight histone proteins, and together these form chromatin. Nucleosomes also contain spacing information, and the tails of histones carry chemical modifications including methylation, acetylation, and ubiquitylation. These modifications, or marks, instruct the cells about what genes should be expressed and what genes should be kept silent. Myriad proteins have evolved to write, read, and erase the histone marks, in concert with remodelers and chaperones. Sebé-Pedrós et al. shed new light on histones by reconstructing the evolutionary history of histone marks and their modifying enzymes and regulators. Histone modifications do occur in Archaea, but they are rare. Although the proteins that mark histones have a pre-eukaryotic origin, the proteins that read marks and chaperones are eukaryotic innovations. Further chromatin evolution has been characterized by diversification among reader proteins, including by "capture" and transfer by transposable elements and viruses among eukaryotes. —DJ

Nat. Ecol. Evol. 10.1038/s41559-022-01771-6 (2022).

Reconstructing the evolution of histone proteins reveals that they have a dual origin.

that allows plastins to modulate the strength of actin cross-linking. —VV

Nat. Struct. Mol. Biol. **29**, 519 (2022).

QUANTUM OPTICS

Designed for separation

Thin film lithium niobate is a versatile platform across a range of classical and quantum optoelectronic applications. One requirement for quantum optics is the need for an efficient source of heralded single photons. However, the spontaneous parametric down-conversion process that occurs in lithium niobate typically results in the emission of entangled photon pairs, and untangling them compromises the purity of the single photons. Xin *et al.* demonstrate that by engineering the dispersion and phase-matching conditions of the lithium niobate thin film, they could produce photon pairs that are spectrally separable. The high purity of the separated photons suggests that this technique can be used as a source of heralded single photons in quantum networks.—ISO

Opt. Lett. **47**, 2830 (2022).

SOCIAL POLICY

Welfare payments, crime-reduction payback

Removing low-income young adults with disabilities from a US cash assistance program when they turned 18 increased the number of criminal charges against them over the next 20 years by 20%. Deshpande and Mueller-Smith show that as a result of those charges, primarily for crimes to generate income, the annual likelihood of incarceration in these subjects increased by 60%. The per-removal costs of courts, police, and incarceration alone, about $41,000, nearly eliminated the savings from removing these young adults from the program (about $49,000). When also factoring in the per-victim costs of the crimes themselves at about $85,000, the assistance program provided $16 of value for every dollar spent by the government. —BW

Q. J. Econ. 10.1093/qje/qjac017 (2022).

PLANT BIOLOGY

Rapid cell expansion

When root cells grow, they often need to extend rapidly. This is done by expanding the vacuole, which means that the volume of the cytosol does not need to expand at the same rate. Dunser *et al.* developed a small molecule that inhibits vacuole membrane, or tonoplast, expansion. The authors used this small molecule to show that endocytosis-derived vesicles contribute to the rapid expansion of the tonoplast during root cell expansion. Endocytosis is required for plasma membrane and cell wall remodeling. A potential change in membrane recycling after endocytosis, as well as increased secretion of new membrane during plasma membrane and cell wall biosynthesis, offers an efficient mechanism that enables rapid cell expansion. —SMH

eLife **11**, e75945 (2022).

ELECTRON MICROSCOPY

Nanoscale vibrational mapping

Complex microscopic structures have been developed to modulate thermal transport for a variety of applications. However, vibrational mapping for understanding this transport has spatial resolution limits. Gadre *et al.* show how to circumvent this issue by using electron energy loss spectroscopy for phonon mapping of silicon-germanium quantum dots. The mapping allowed the authors to get a clear picture of how atomic structure and interfaces affect phonon dynamics and thermal transport. This strategy can be used to better understand and develop thermoelectrics, thermal barriers, phase-change memory, and other materials. —BG

Nature **606**, 292 (2022).

IMAGE: V. ALTOUNIAN/*SCIENCE*

RESEARCH ARTICLE SUMMARY

IMMUNOMETABOLISM

Interleukin-17 governs hypoxic adaptation of injured epithelium

Piotr Konieczny†, Yue Xing*†, Ikjot Sidhu, Ipsita Subudhi, Kody P. Mansfield, Brandon Hsieh, Douglas E. Biancur, Samantha B. Larsen, Michael Cammer, Dongqing Li, Ning Xu Landén, Cynthia Loomis, Adriana Heguy, Anastasia N. Tikhonova, Aristotelis Tsirigos, Shruti Naik*

INTRODUCTION: Injury compromises our epithelial barriers, leaving them vulnerable to external threats. Organismal survival thus impinges on rapid barrier re-establishment after damage. Immunocompromised individuals and immunodeficient animals have profound defects in epithelial repair. However, the precise mechanisms and consequences of immune-epithelial cross-talk after injury remain incompletely defined.

RATIONALE: Epithelial repair requires cellular adaptation to the hypoxic wound microenvironment through evolutionarily conserved hypoxia-inducible factors (HIFs). This response has long been thought to depend upon epithelial cell-autonomous sensing of oxygen deprivation and responsiveness through HIFs. However, whether and how supportive microenvironmental signals from immune cells intersect with ancient hypoxia responses during epithelial repair have yet to be explored. To address these questions, we were drawn to the skin, a primary epithelial barrier that is surveilled by immune cells and has evolved sophisticated repair mechanisms.

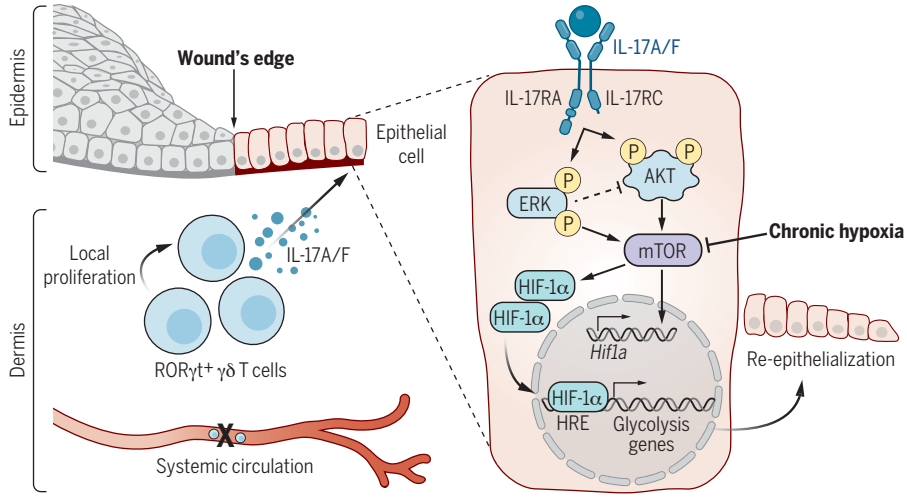
RESULTS: Using unbiased profiling of repair-associated lymphocytes, we uncovered the enrichment of heterogeneous populations of type 17 lymphocytes. Homozygous *Rorgt*-EGFP (GFI-KI) mice deficient in RORγt$^+$ cells exhibited a significant defect in epithelial migration and wound re-epithelialization. After damage, we observed a rapid local expansion of preexisting skin-resident RORγt$^+$ cells that was sufficient to drive repair. Of the myriad wound RORγt$^+$ populations, spatial transcriptomics and functional studies revealed that innate-like γδ T cells directed tissue repair by localizing to the wound front and boosting epithelial migration. These cells dominantly produced IL-17A/F in the wound microenvironment, which signaled directly to epithelia through IL-17RC to induce a migratory program.

Comparative spatial transcriptomics of RORγt$^+$ cell-deficient and wild-type (WT) wound edges revealed differences in HIF1α signaling. WT migrating epithelia had robust nuclear HIF1α activation that was notably missing from RORγt$^+$ cell– and epithelial IL-17RC–deficient (*Il17rc*^{EKO}) animals. Accordingly, epithelial-specific loss of HIF1α (*Hif1a*^{EKO}) phenocopied the wound-healing defect of GFP-KI and *Il17rc*^{EKO} mice. However, in contrast to GFP-KI wounds, recombinant IL-17A was unable to augment repair in *Hif1a*^{EKO} mice, underscoring the necessity of this transcriptional effector mediating re-epithelialization downstream of IL-17A signaling.

RORγt$^+$ cell–deficient and control animals had comparable levels of wound hypoxia, indicating that the failure of RORγt$^+$ cell–deficient animals to induce HIF1α was independent of oxygen sensing. Instead, IL-17A was sufficient to activate epithelial HIF1α both in vivo and in epithelial organoids. IL-17RC signaling rapidly induced activation of both ERK and AKT, which augmented HIF1α protein and transcript levels, but not stability, through mammalian target of rapamycin (mTOR). Acute hypoxia potently activated epithelial HIF1α in vitro, and this response was further boosted by IL-17A. By contrast, we uncovered a striking loss of mTOR and HIF1α in chronic hypoxia. Exogenous IL-17A rescued mTOR, and consequently HIF1α, under conditions of chronic hypoxia, indicating that IL-17A acts as a second signal to enable cell adaptation to low-oxygen environments.

IL-17A induced a transcriptional and functional program of glycolytic metabolism in epithelia, which was dependent upon mTOR and HIF1α. The IL-17A–HIF1α-dependent program of glycolysis fueled epithelial migration, and pharmacological inhibition of glycolysis impaired wound re-epithelialization.

CONCLUSION: Our findings upend a long-held view that hypoxia is sufficient to cell-autonomously induce HIF1α-mediated metabolic remodeling. We illustrate that IL-17A supplied by RORγt$^+$ γδ T cells is necessary for optimal HIF1α activation in the wound-edge epithelium. The IL-17A–HIF1α axis directed the metabolic rewiring of damaged epithelium toward a program of glycolysis to fuel migration. HIF1α and glycolysis are drivers of tumor progression and metastasis, raising the possibility that IL-17A or other immune-derived signals could drive these pathways in cancer. IL-17A is also central to the pathology of many autoimmune conditions, including psoriasis and inflammatory bowel disease. Thus, the IL-17A–HIF1α axis unveiled by our study may provide therapeutic opportunities for a range of epithelial inflammatory and metastatic diseases. ■



IL-17 drives epithelial HIF1α to fuel wound repair through glycolysis. IL-17A/F supplied by expanding skin-resident RORγt$^+$ γδ T cells is necessary for optimal HIF1α activation in the wound-edge epithelium, even in the presence of hypoxia. IL-17A signaling through the IL-17RC receptor rapidly induces ERK/AKT/mTOR, which augments HIF1α transcripts and protein. The IL-17–HIF1α axis directs a transcriptional and functional program of glycolysis to promote migration.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: shruti.naik@nylangone.org (S.N.); yue.xing@nyulangone.org (Y.X.)
†These authors contributed equally to this work.
Cite this article as P. Konieczny *et al.*, *Science* **377**, eabg9302 (2022). DOI: 10.1126/science.abg9302

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg9302

RESEARCH ARTICLE

IMMUNOMETABOLISM

Interleukin-17 governs hypoxic adaptation of injured epithelium

Piotr Konieczny[1]†, Yue Xing[1]*†, Ikjot Sidhu[1,2], Ipsita Subudhi[1], Kody P. Mansfield[1], Brandon Hsieh[1], Douglas E. Biancur[3], Samantha B. Larsen[4], Michael Cammer[5], Dongqing Li[6], Ning Xu Landén[6], Cynthia Loomis[7], Adriana Heguy[8], Anastasia N. Tikhonova[9], Aristotelis Tsirigos[1,2], Shruti Naik[1,10]*

Mammalian cells autonomously activate hypoxia-inducible transcription factors (HIFs) to ensure survival in low-oxygen environments. We report here that injury-induced hypoxia is insufficient to trigger HIF1α in damaged epithelium. Instead, multimodal single-cell and spatial transcriptomics analyses and functional studies reveal that retinoic acid–related orphan receptor γt$^+$ (RORγt$^+$) γδ T cell–derived interleukin-17A (IL-17A) is necessary and sufficient to activate HIF1α. Protein kinase B (AKT) and extracellular signal–regulated kinase 1/2 (ERK1/2) signaling proximal of IL-17 receptor C (IL-17RC) activates mammalian target of rapamycin (mTOR) and consequently HIF1α. The IL-17A–HIF1α axis drives glycolysis in wound front epithelia. Epithelial-specific loss of IL-17RC, HIF1α, or blockade of glycolysis derails repair. Our findings underscore the coupling of inflammatory, metabolic, and migratory programs to expedite epithelial healing and illuminate the immune cell–derived inputs in cellular adaptation to hypoxic stress during repair.

Epithelial barriers limit the penetration of disease-causing environmental agents. Consequently, organismal survival impinges on the rapid repair and restoration of barrier integrity after injury. Distress signals from wounded microenvironments (i.e., hypoxic, genotoxic, and oxidative stress) evoke ancient cell-intrinsic responses to ensure epithelial cell adaptation to hostile conditions (*1*). Concomitantly, extrinsic "secondary signals" generated by accessory cells such as growth factors and cytokines direct re-epithelialization of the denuded surface (*2*). If and how extrinsic microenvironmental signals intersect with primitive cellular stress responses during tissue repair has yet to be explored.

Epithelia are surveilled by an arsenal of lymphocytes that rapidly respond to injury (*3*). Comprising innate, innate-like, and adaptive lymphocytes, these "first responders" react to alarmins secreted by damaged epithelium (*3*, *4*). Alarmin-induced lymphocyte activation triggers the production of interleukin-17A (IL-17A), IL-17F, and IL-22, cytokines with mitogenic effects that are well documented (*5*). Epithelial proliferation, however, is dispensable for wound re-epithelialization, a process requiring epithelial cell migration across the breach (*6*). Thus, defining the migratory programs downstream of lymphocyte–epithelial interactions is of utmost importance.

Mobilized epithelial cells must cope with the harsh wound microenvironment, including low oxygen and nutrient availability, reactive oxygen species, and cellular and microbial debris (*2*). Precisely how adaptations to these severe conditions occur and whether they are tied to factors emitted by repair-associated lymphocytes is unclear. To tackle this question, we were drawn to the skin epithelium, a primary barrier that has evolved sophisticated repair mechanisms and a prototypic site for lymphocyte-mediated immunosurveillance (*7*). Using single-cell RNA-sequencing (scRNA-seq) and spatial transcriptomics (*8*), imaging, and mouse genetic strategies to track communications between lymphocytes and wound edge epithelium (*9*), we identified retinoic acid–related orphan receptor γt$^+$ (RORγt$^+$) γδ T cell–derived IL-17A as an essential mediator of this dialogue. IL-17A signaling through epithelial IL-17 receptor C (IL-17RC) orchestrates a program of migration and hypoxic adaptation by activating hypoxia-inducible factor 1α (HIF1α). The IL-17A–HIF1α transcriptional axis induces glycolysis, which is necessary for wound re-epithelialization. Thus, we challenge a long-held view that hypoxia is sufficient to cell-autonomously induce HIF1α-mediated metabolic remodeling (*10*). Moreover, we illuminate the role of microenvironmental signals, such as lymphocyte-derived IL-17A, in facilitating this adaptive tissue response.

Results

Skin-resident RORγt$^+$ cells direct wound re-epithelialization

To define the composition of first responder lymphocytes in an unbiased manner, we performed cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) (*11*). CD45$^+$CD90$^+$ lymphocytes were purified and profiled from unwounded skin and at day 3 (D3) and D5 after full-thickness wounding (fig. S1, A and B). Principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) plots with unsupervised analysis of 4923 cells revealed 12 clusters with a clear segregation of effector and regulatory T (T_{reg}) lymphocytes (fig. S1C). Because lymphocytes are notorious for low-abundance transcripts, we used surface protein epitopes alongside transcripts for definitive cluster annotations (fig. S2, A to C). T helper and cytotoxic T (T_H/T_C) cell clusters 1 to 3, γδ T cell and mucosal-associated invariant T (γδ T/MAIT) cell cluster 1, the natural killer and natural killer T (NK2/NKT) cell cluster, and T_{reg} cell clusters 2 and 3 were enriched in wounded skin compared with unwounded controls (Fig. 1A and fig. S2D). Thus, contrary to expectations, a dominant repair-associated lymphocyte population does not emerge. Rather, many different immune cell populations are substantially remodeled in acute injury.

We next examined the expression of cytokines that typify type 1, type 2, type 17, and T_{reg} archetypes (*12*). Of these, the type 17 cytokines *Il17a* and *Il17f* and the type 1 cytokine *Ifng* were enriched in wounded samples (Fig. 1B). Type 17 immunity encompasses different cell types controlled by a master transcription factor RORγt, which has also been implicated in repair (*13*, *14*). Indeed, RORγt$^+$ lymphocytes were enriched in both the human and murine epithelial wound front (Fig. 1, C and D, and fig. S3A). *Rorgt*-EGFP mice have enhanced green fluorescent protein (EGFP) knocked into the *Rorgt* locus, rendering it nonfunctional (*15*). Using homozygous *Rorgt*-EGFP mice (hereafter called GFP-KI), we measured gross wound closure in the absence of RORγt. Both genetic loss of RORγt in GFP-KI mice and topical administration of an RORγt inhibitor, digoxin, resulted in a 40 to 50% reduction in wound closure rate compared with respective controls (fig. S3, B and C) (*16*).

Human skin heals exclusively by re-epithelialization, a process reliant on epithelial proliferation and subsequent migration

[1]Department of Pathology, New York University Langone Health, New York, NY 10016, USA. [2]Applied Bioinformatics Laboratory, New York University Langone Health, New York, NY 10016, USA. [3]Department of Radiation Oncology and Perlmutter Cancer Center, New York University Langone Health, New York, NY 10016, USA. [4]Neuroscience Institute, New York University Langone Health, New York, NY 10016, USA. [5]Microscopy Laboratory, New York University Langone Health, New York, NY 10016, USA. [6]Dermatology and Venereology Division, Department of Medicine, Solna Center for Molecular Medicine, Ming Wai Lau Centre for Reparative Medicine, Karolinska Institute, 17176 Stockholm, Sweden. [7]Experimental Pathology Research Laboratory, New York University Langone Health, New York, NY 10016, USA. [8]Genome Technology Center, New York University Langone Health, New York, NY 10016, USA. [9]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada. [10]Department of Medicine, Ronald O. Perelman Department of Dermatology, and Perlmutter Cancer Center, New York University Langone Health, New York, NY 10016, USA.
*Corresponding author. Email: shruti.naik@nylangone.org (S.N.); yue.xing@nyulangone.org (Y.X.)
†These authors contributed equally to this work.

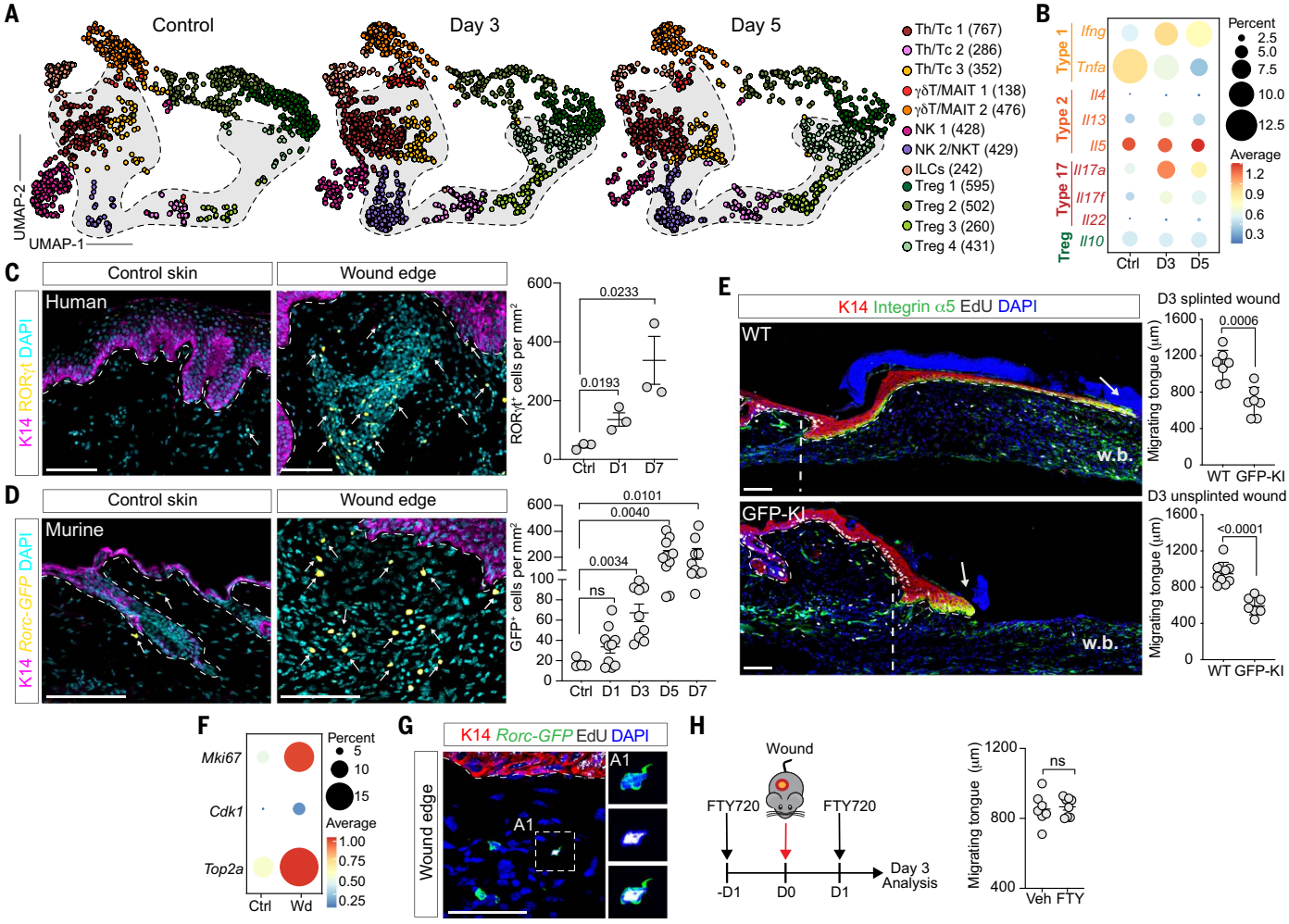


Fig. 1. Skin-resident RORγt cells direct wound re-epithelialization. (**A**) UMAP visualization of individual samples (control, n = 1443 cells; D3, n = 1997 cells; D5, n = 1486 cells). (**B**) Dot plot of cytokine expression in control and wounded lymphocytes. Shown is the frequency of cells expressing gene (percentage) and average expression per cluster (average). RORγt$^+$ cells (yellow) are enriched at the edge of human (**C**) (n = 3, N = 1) and murine (**D**) ($n \geq 4$, N = 3) acute wounds. Magenta, keratin (K)14; cyan, DAPI nuclei. White arrows mark RORγt cells. (**E**) Impaired re-epithelialization in RORγt-deficient (GFP-KI) compared with WT animals. Shown are images and quantifications of integrin α5$^+$ (green), K14$^+$ (red) migrating epidermal tongue from full-thickness silicone-splinted (top) and unsplinted (bottom) D3 wounds. The dashed lines and arrows mark the beginning and end of migrating tongue, respectively. White, EdU; blue, DAPI nuclei; w.b., wound bed ($n \geq 7$, N = 3). (**F**) Wound lymphocytes have a higher frequency (percentage) and expression of proliferation genes (*Mki67*, *Cdk1*, and *Top2a*) than control skin. (**G**) Wound RORγt cells actively proliferate. Shown is an immunofluorescence image of D3 wounded *Rorc*-EGFP^{Tg} mice after a 24-hour EdU pulse. White, EdU; blue, DAPI nuclei; green, RORγt$^+$ cells; red, keratin (K)14 (n = 3, N = 3). (**H**) FTY720 treatment does not alter wound re-epithelialization ($n \geq 7$, N = 3). In (C), (D), and (G), white dashed lines denote dermo-epidermal borders and yellow boxes define magnified areas. Scale bars in (C), (D), and (E), 100 μm; scale bar in (G), 50 μm. Significance was determined using a two-tailed t test and a 95% confidence interval.

into the wound bed. By contrast, mice heal by re-epithelialization and myofibroblast-mediated dermal contraction (*17*). We therefore secured murine wounds with silicone splints to eliminate dermal contraction. Microscopic visualization revealed a specific impairment of the keratin 14$^+$ (K14$^+$) integrin α5$^+$ migrating epithelial tongue, but not epithelial proliferation, in GFP-KI splinted wounds (Fig. 1E and fig. S3D). Similarly, epithelial migration was markedly compromised in unsplinted GFP-KI wounds relative to controls (Fig. 1E).

Given the rapid enrichment of RORγt lymphocytes in wounds, we postulated that resident RORγt lymphocytes locally expand to fuel repair. Indeed, wound RORγt$^+$ cells had heightened expression of cell-cycle genes (*Mki67*, *Cdk1*, and *Top2a*) and 5-ethynyl-2′-deoxyuridine (EdU) incorporation relative to unwounded skin and to their systemic lymph node counterparts (Fig. 1, F and G, and fig. S3E). To determine whether skin RORγt$^+$ lymphocytes were sufficient to instigate repair, we treated mice with FTY720, a sphingosine-1 phosphate receptor agonist that inhibits lymphocyte egress into blood by "trapping" them within lymph nodes (fig. S3F) (*18*). FTY720 blockade did not alter the frequency of wound RORγt cells or epithelial migration (Fig. 1H and fig. S3G). Thus, the expansion of pre-existing skin RORγt$^+$ cells is sufficient to drive re-epithelialization.

IL-17RC–mediated γδ T cell–epithelial cross-talk controls re-epithelialization

We mapped the dynamics of skin RORγt subsets over the course of repair using multiparameter flow cytometry. Innate-like γδ T cells, MAIT (MR1 tetramer$^+$), and invariant NKT (iNKT) cells (CD1d tetramer$^+$) expanded first at D1 after wounding (fig. S4, A to C). Adaptive T_H17, T_C17, and RORγt$^+$ T_{reg} cells

subsequently increased between D3 and D5. Consistent with our CITE-seq analysis, the proportions of RORγt$^+$ type 3 innate lymphoid cells (ILC3s) were unchanged by injury, suggesting that in the presence of adaptive immunity, ILC3s may not play a dominant role in repair (figs. S2D and S4, A to C).

To pinpoint which RORγt subset(s) was positioned at the re-epithelializing wound front, we turned to spatial transcriptomics (ST), a technology that enables spatially resolved transcriptional profiling of tissues at 50-μm resolution (*8*) (fig. S5, A to C, and table S4). PCA and UMAP projections distinguished seven clusters (tissue regions) in unwounded skin and six clusters in wounded skin (Fig. 2, A and B). The identified clusters were consistent with histological regions and genes that typified tissue architecture (Fig. 2, A and B, and fig. S5, D and E). Wound-edge epithelium was discernable by *Krt14* and *Krt17* expression (fig. S6A). Nevertheless, to confirm the identity of epidermal wound edge clusters and validate our ST data, we performed multimodal intersectional analysis (MIA) (*8*), a method that can infer cell-type enrichment in tissue regions. Integrating an independent scRNA-seq dataset (*19*), which defined wounded epithelium as growth arrest cells and wound distal epithelium as *Col17a1*$^+$ cells, with our ST data revealed a significant degree of overlap between *Col17a1*$^+$ cells and wound distal epithelial cluster 3 ($P = 2.3 \times 10^{-16}$) and growth arrest cells and wound-edge epithelial cluster 4 ($P = 5.4 \times 10^{-35}$) (fig. S6, B and C).

We next leveraged MIA to resolve the location of wound lymphocyte populations identified by our CITE-seq data (Fig. 1A). MIA of the top 300 up-regulated genes (≥0.25 log FC) from each effector lymphocyte CITE-seq cluster and each ST cluster simultaneously located all of our lymphocyte populations in wounded skin and revealed a specific enrichment of γδ T/MAIT cluster 1 ($P = 7.9 \times 10^{-6}$) and cluster 2 ($P = 1.9 \times 10^{-5}$) within wound-edge epithelial cluster 4 (Fig. 2C and fig. S6D).

Both γδ T cells and MAIT cells have been implicated in wound repair (*20*). However, systematic comparisons between these and other RORγt$^+$ cell subsets in the context of re-epithelialization are lacking. We therefore analyzed mice specifically lacking RORγt$^+$ γδ T cells (*Tcrd*^{CreER};*Rorc*$^{fl/fl}$), MAIT cells (*Mr1*$^{-/-}$), and T_H17/T_c17/MAIT/iNKT cells (*Cd4*^{Cre};*Rorc*$^{fl/fl}$) (fig. S6E). After tamoxifen-induced depletion, *Tcrd*^{CreER};*Rorc*$^{fl/fl}$ mice exhibited a shorter migrating tongue than control animals, indicating a nonredundant role for RORγt$^+$ γδ T cells in wound re-epithelialization (Fig. 2D). Wounds from both *Mr1*$^{-/-}$ and *Cd4*^{Cre};*Rorc*$^{fl/fl}$ mice had migrating tongues comparable to their respective controls, underscoring the unicity of RORγt$^+$ γδ T cells in directing re-epithelialization (Fig. 2, E and F).

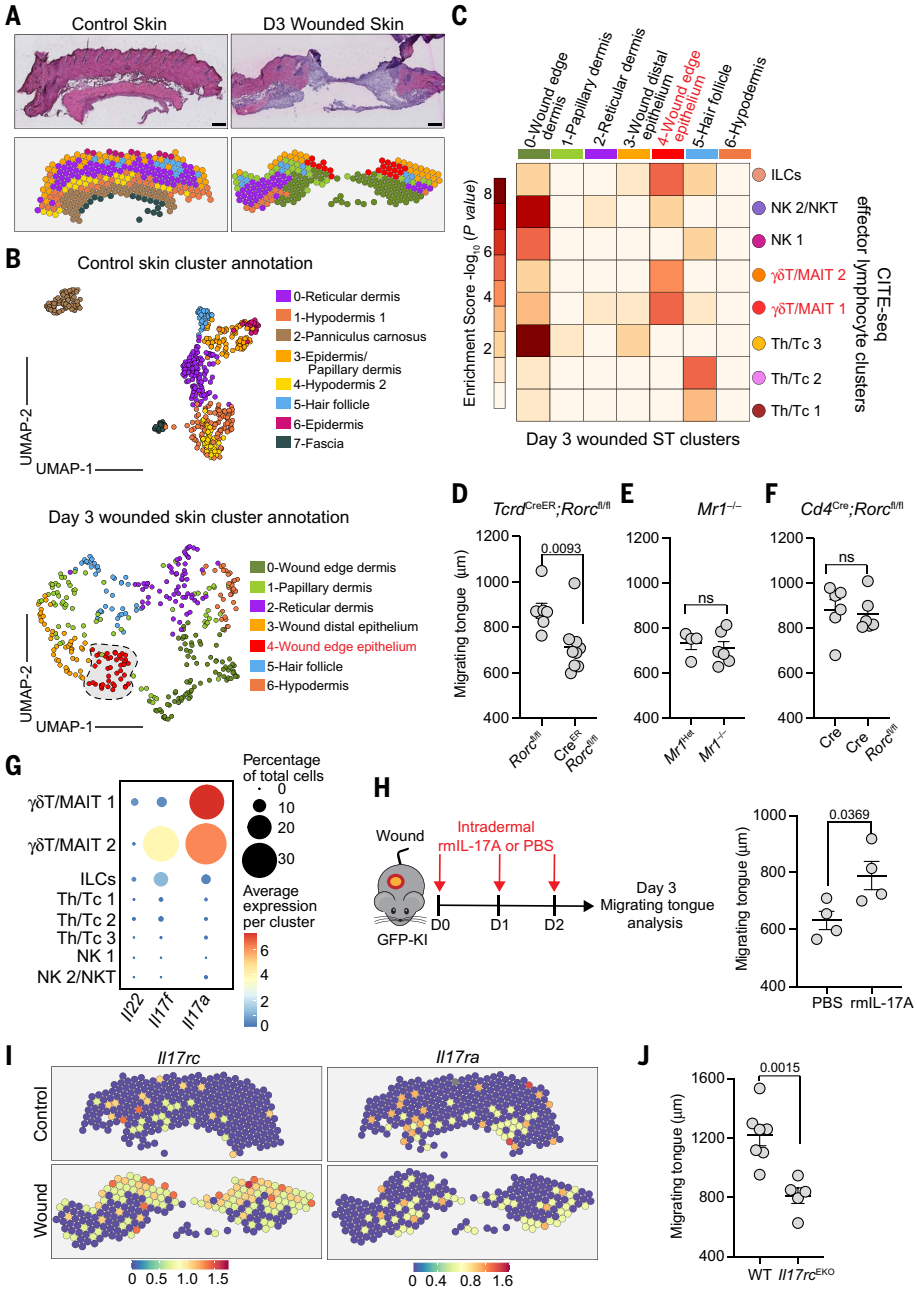


Fig. 2. RORγt$^+$ γδT cells drive re-epithelialization through epidermal IL-17RC. (**A**) Hematoxylin and eosin–stained tissue image (top) and ST plot of microarray spots (bottom) of unwounded and D3 wounded skin (unwounded, $n = 1$; wounded, $n = 2$). Scale bars, 200 μm. (**B**) UMAP plot and cluster annotation of ST spots based on marker genes in fig. S5. (**C**) γδ T/MAIT clusters are enriched in wound edge epithelium. *P*-value score table from MIA of D3 wounded ST clusters and CITE-seq effector lymphocyte clusters. Integrin α5$^+$, K14$^+$ migrating epidermal tongue length in unsplinted D3 wounds from *Tcrd*^{CreER};*Rorc*$^{fl/fl}$ and *Rorc*$^{fl/fl}$ (**D**) ($n \geq 7$, $N = 3$). *Mr1*$^{-/-}$ and *Mr1*^{Het} (**E**) ($n \geq 4$, $N = 3$). *Cd4*$^{Cre/-}$;*Rorc*$^{fl/fl}$ and *Cd4*$^{Cre/-}$ (**F**) ($n = 5$, $N = 3$). (**G**) γδ T/MAIT cell clusters dominantly express *Il17a* and *Il17f*. Dot plot of genes from D3 wounded skin CITE-seq analysis. Frequency of cells expressing stated gene (percent) and average expression per cluster (average). (**H**) rmIL-17A administration augments the re-epithelialization in GFP-KI mice. Experimental schematic (left) and quantification of D3 migrating tongue length of GFP-KI skin intradermally injected with rmIL-17A or PBS (right) ($n = 4$, $N = 2$). (**I**) Spatial plots of *Il17rc* and *Il17ra* reveal an up-regulation of these receptors at the wound's edge. (**J**) IL-17RC expression on epithelial cells drives re-epithelialization. Quantification of migrating tongue length from splinted D3 wounds *Krt14*^{Cre}; *Il17rc*$^{fl/fl}$ mice ($n \geq 5$, $N = 2$). Significance was determined using a two-tailed *t* test and a 95% confidence interval.

RORγt$^+$ γδ T cells exert potent effector function by producing the cytokines IL-17A and IL-17F, which provide instructive signals to the surrounding parenchyma (*14*, *21*). Of all the effector lymphocytes analyzed by CITE-seq, γδ T/MAIT cell clusters expressed the highest level of these cytokines at D3 after wounding, hinting at their role in the RORγt$^+$ γδ T cell–epithelial cross-talk at the wound's edge (Fig. 2G). Accordingly, intradermal recombinant murine IL-17A (rmIL-17A) administration reversed the re-epithelization defect of GFP-KI mice (Fig. 2H).

IL-17A/F signaling is mediated by an IL-17 receptor A (IL-17RA)–IL-17RC heterodimer, whereas IL-17F can also bind an IL-17RC/RC homodimeric complex (*14*, *22*). ST plots showed increased *Il17ra* and *Il17rc* transcripts at the wound's edge, with the most prominent expression being *Il17rc* (Fig. 2I). We therefore generated and wounded epithelial-specific IL-17RC–deficient mice (*Krt14*^{Cre};*Il17rc*$^{fl/fl}$ labeled *Il17rc*^{EKO}) (fig. S6F). Compared with IL-17RC–sufficient mice, epithelial-specific loss of *Il17rc* mirrored the re-epithelialization defect of GFP-KI and *Tcrd*^{CreER};*Rorc*$^{fl/fl}$ mice (Fig. 2J). Thus, IL-17A/F supplied by RORγt$^+$ γδ T cells directly engages epithelial IL-17RC to induce migration and re-epithelialization.

HIF1α is a re-epithelialization factor controlled by RORγt$^+$ cells

PCA and UMAP plots integrating wild-type (WT) and GFP-KI tissue sections from D3 wounds discerned 15 clusters with high reproducibility between biological replicates (Fig. 3, A and B, and fig. S7, A and B). We focused on wound edge clusters 7 and 8 on the basis of histology and expression of the re-epithelialization factors *Gjb2*, *Lamc2*, *Mmp9*, *Mmp13*, and *Stfa1* (fig. S7A). We further delineated cluster 7 as wound-edge basal epithelium (*Krt5*, *Krt6*, *Krt14*, and *Krt17*) and cluster 8 as wound-edge differentiated epithelium (*Krt10*, *Lce1a1*, and *Lor*). Pathway analysis of WT and GFP-KI cluster 7 differentially expressed genes revealed enrichment of IL-17 signaling, skin/epidermal development, epidermal differentiation, epithelial cell/tissue migration, and T cell activation pathways in WT (Fig. 3C and fig. S8A). HIF1 signaling, which is typically associated with tissue hypoxia, also emerged as a differentially expressed pathway (Fig. 3C) (*10*), with *Hif1a* transcripts higher in WT compared with GFP-KI wound edges (Fig. 3D).

To validate this finding, we analyzed the transcriptomes of integrin α*5* and integrin α*6* double-positive wound-edge epithelium (migrating tongue) and integrin α*6*$^+$-only wound distal epithelium, from WT and GFP-KI wounds (fig. S8B). We observed very modest differences (62 differentially expressed genes, adjusted $P < 0.1$) between WT and GFP-KI wound distal epithelium (fig. S8C). By contrast, WT and GFP-KI migrating tongue cells had 1548 differentially expressed transcripts (adjusted $P < 0.1$) (fig. S8D). HIF1α signaling and associated glycolysis pathways were significantly enriched in WT samples (fig. S8E).

Consistent with the enrichment of *Hif1a* transcripts, WT migrating epithelial tongue robustly expressed nuclear HIF1α, which was absent from GFP-KI wound fronts (Fig. 3E and fig. S9, A and B). GFP-KI mice did not compensate for the loss of HIF1α by up-regulating HIF2α (fig. S9D). Furthermore, *Il17rc*^{EKO} mice phenocopied the HIF1α defect observed in GFP-KI wounds (Fig. 3F). Nuclear HIF1α was also robustly expressed in the migrating epithelial front of human wounds, suggesting a highly conserved role for this transcription factor in re-epithelialization (fig. S9C).

HIFs ubiquitously mediate cellular adaptation to low-oxygen environments (*23*). Therefore, we measured hypoxia in WT and GFP-KI wounds using both a fluorescent hypoxia marker, pimonidazole (PIM), and a fiber probe to directly measure tissue O_2. WT and GFP-KI wounds had equivalent PIM signals and O_2 levels (~10 to 12 mmHg or 2% O_2) (Fig. 3, G and H). Moreover, WT and GFP-KI epithelia expressed comparable levels of prolyl hydroxylases (*Egln1*, *Egln2*, and *Egln3*) and *Vhl* ubiquitin ligase (fig. S9D). Thus, wound-edge epithelium failed to sustain HIF1α without RORγt$^+$ cell–derived signals even in the presence of tissue hypoxia and while having the necessary regulatory machinery.

We next generated and wounded epithelial-specific HIF1α-deficient mice (*Krt14*^{Cre}; *Hif1a*$^{fl/fl}$ labeled *Hif1a*^{EKO}) (fig. S9E). Loss of epithelial HIF1α impaired the development of migrating tongue (Fig. 3I). Additionally, unlike GFP-KI mice, rmIL-17A failed to rescue the re-epithelialization defect of *Hif1a*^{EKO} mice (Fig. 2H and fig. S9F). Thus, epithelial HIF1α is necessary to respond to IL-17A signals and activate a program of migration.

IL-17RC signaling induces HIF1α through extracellular signal–regulated kinase 1/2 (ERK 1/2)-AKT–mammalian target of rapamycin (mTOR)

To test our hypothesis that γδ T cell–derived IL-17 serves as a second signal to activate HIF1α, we intradermally injected rmIL-17A, rmIL-17F, or rmIL-22 into unwounded skin. Administration of rmIL-17A and, to a lesser degree, rmIL-17F was sufficient to induce epidermal HIF1α activation and hyperplasia (Fig. 4A and fig. S10, A and B). IL-22 marginally increased epidermal thickness but did not up-regulate HIF1α (fig. S10, A and B). Skin epithelial organoids cultured for 5 days in the presence of rmIL-17A were significantly larger than controls and robustly expressed HIF1α (Fig. 4B and fig. S10, C to F). Other damage-associated cytokines [rmIL-1, rmIL-6, and recombinant murine tumor necrosis factor α (rmTNFα)] did not activate HIF1α (fig. S10G). Epidermal growth factor also failed to augment HIF1α to the levels observed after rmIL-17A treatment, underscoring the specificity of IL-17 in amplifying this pathway (fig. S10H).

IL-17RA and IL-7RC were robustly expressed in our epithelial organoids and wound-edge integrin α*5*$^+$ cells (fig. S10I) (*14*, *22*). Epithelial-specific loss of *Il17rc* abrogated the ability of rmIL-17A to activate HIF1α both in vivo and in vitro, excluding a role for accessory cells in this cross-talk (Fig. 4, C and D, and fig. S10, J and K). To delineate how epithelial IL-17RC signaling resulted in HIF1α activation, we returned to our comparative GFP-KI and WT wound transcriptomics (Fig. 3, A to C). mTOR signaling, which is implicated in HIF1α regulation, was salient as a key differentially expressed pathway (*24*, *25*). We therefore evaluated ribosomal protein S6$^{S240/244}$ phosphorylation, a major downstream substrate of mTOR complex 1. rmIL-17A robustly increased epithelial pS6$^{S240/244}$ in WT but not *Il17rc*^{EKO} mice and organoids (Fig. 4, E and F, and fig. S10L). rmIL-17A–treated organoids also showed enhanced activation of the upstream kinase phosphorylated S6 kinase^{T389} (p-S6K^{T389}) and phosphorylated mTOR^{S2448} (p-mTOR^{S2448}) (Fig. 5, A and B). Inhibition of mTOR with rapamycin significantly reduced HIF1α protein downstream of IL-17A (Fig. 5, A to C, and fig. S11, A and B). Additionally, *Hif1a*^{EKO} and control organoids had comparable sizes and mTOR activation (p-S6$^{S240/244}$) after rmIL-17A stimulation, confirming that mTOR is upstream of HIF1α (fig. S11, C and D).

In addition to augmenting HIF1α protein, rmIL-17A stimulation also induced *Hif1a* gene expression in an mTOR-dependent manner (Fig. 5D). Increased *Hif1a* transcripts in rmIL-17A–treated organoids were not caused by enhanced mRNA stability, and HIF1α protein was more rapidly degraded in the presence of rmIL-17A (fig. S11, E and F). Thus, mTOR controls HIF1α transcription and protein levels, but not stability, downstream of IL-17A.

To determine whether and how IL-17RC proximal signaling links to mTOR, we next measured early (≤1 hour) mTOR activation after rmIL-17A stimulation and found enrichment of p-S6K^{T389} and p-S6$^{S240/244}$ (fig. S11G). Working our way upstream, we detected activation of p-ERK1/2$^{T202/Y204}$, a canonical mediator of IL-17RC signaling, and protein kinase B (p-AKT ^{S473}), an upstream mTOR kinase, after IL-17A stimulation (Fig. 5E). Inhibition of either ERK1/2 (U0126) or AKT (MK-2206) after long- and short-term rmIL-17A treatment (5 days and 30 min, respectively), diminished p-S6K^{T389}, p-S6$^{S240/244}$, and HIF1α expression (Fig. 5E and fig. S11H). However,

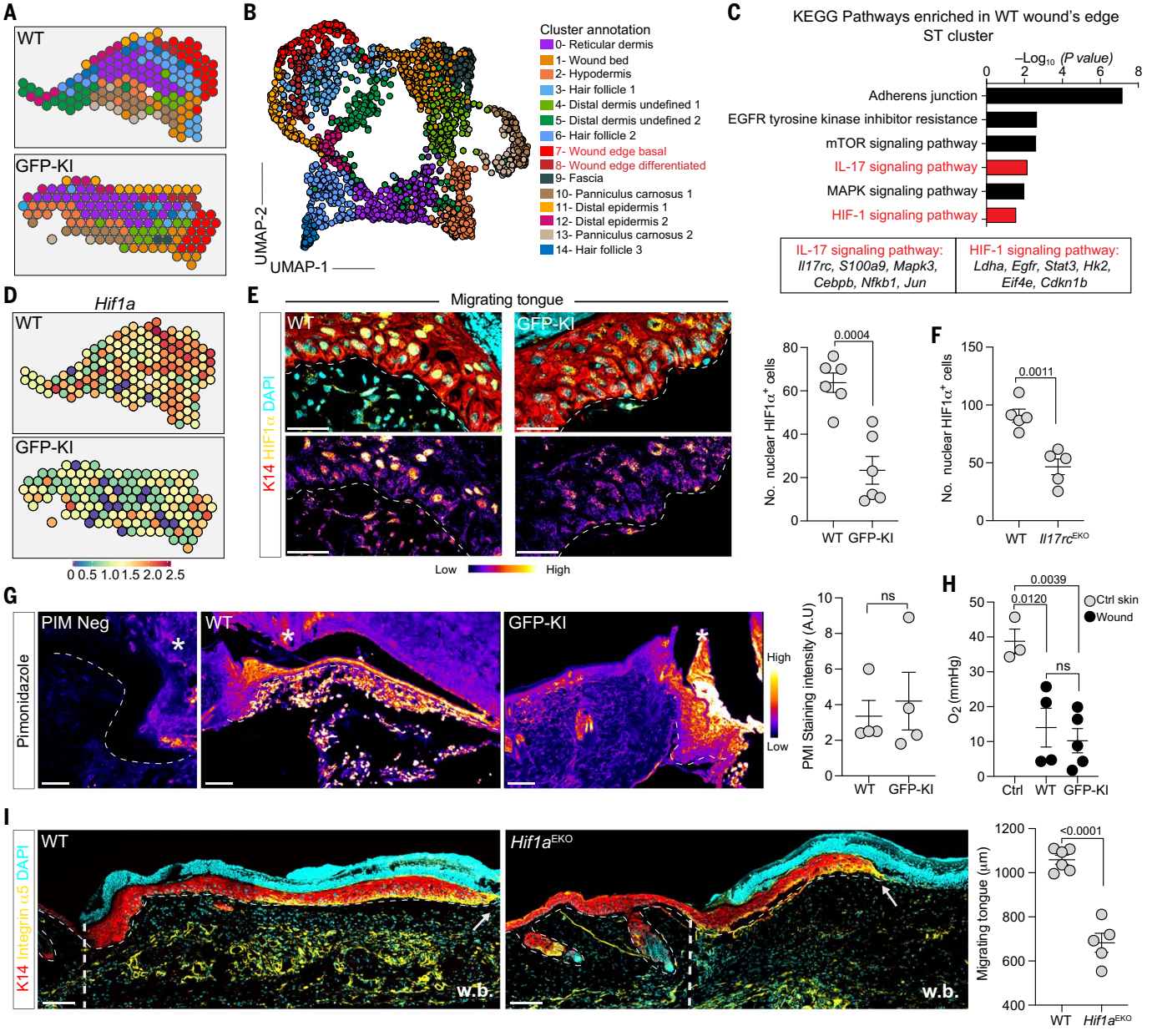


Fig. 3. RORγt^+ cells control wound edge epithelial HIF1α. ST spots (**A**) and UMAP plot (**B**) of annotated clusters (marker genes in fig. S7) from WT and GFP-KI D3 wounded skin samples (WT, n = 2; GFP-KI, n = 3). (**C**) KEGG pathways enriched in WT versus GFP-KI cluster 7 spots. (**D**) Spatial plots of *Hif1a* expression in D3 WT and GFP-KI wounds. (**E**) WT migrating tongues have higher HIF1α expression than GFP-KI migrating tongues. Red, top, K14; yellow, top, HIF1α (pseudocolor fire, bottom); cyan, top, DAPI nuclei. Quantification of HIF1α in integrin $\alpha5^+$ cells from GFP-KI and WT D3 wounds (n = 6, N = 3). (**F**) Quantification of HIF1α in integrin $\alpha5^+$cells from WT and *Krt14*^{Cre}; *Il17rc*$^{fl/fl}$ D3 wounds (n = 5, N = 2). (**G**) Pseudocolor fire view images and quantification of PIM staining in D3 WT and GFP-KI wounds (n = 4, N = 2). (**H**) Fiber probe measurement of unwounded and D3 wounded tissue oxygen ($n \geq 3$, N = 2). (**I**) Impaired re-epithelialization in epithelial HIF1α-deficient animals (*Hif1a*^{EKO}) compared with WT. Confocal images and corresponding quantifications of integrin $\alpha5^+$ (yellow), K14$^+$ (red) migrating epidermal tongue from splinted D3 wounds. The dashed line and arrows mark the beginning and end of migrating tongue, respectively. Cyan, DAPI nuclei; w.b., wound bed ($n \geq 5$, N = 3). Scale bars in (E), 50 μm; scale bars in (G) and (I), 100 μm. For (E), (G), and (I), two-tailed t test at a 95% confidence interval was used. For (H), one-way ANOVA multiple-comparisons test was used.

blocking ERK also resulted in a compensatory increase in p-AKT^{S473}. Dual inhibition of ERK and AKT ablated the mTOR-HIF1α response downstream of rmIL-17A stimulation (Fig. 5E and fig. S11H). Thus, IL-17RC–proximal AKT and ERK signaling directly activates mTOR and consequently induces HIF1α.

Given the prominence of hypoxia at the wound edge, we next examined the IL-17A–mTOR–HIF1α axis in organoids cultured under acute (5 hours) or chronic (5 days) hypoxic conditions (2% O_2). Acute hypoxia was sufficient to stabilize HIF1α, and IL-17A further augmented this response (Fig. 5F). By contrast, and in agreement with our in vivo wounding data, p-S6$^{S240/244}$ and HIF1α levels were markedly reduced by long-term hypoxic exposure and rescued by the presence of rmIL-17A (Fig. 5F). Rapamycin treatment abrogated the ability of IL-17A to augment HIF1α in chronic hypoxic cultures (Fig. 5G and fig. S11, I and J). Thus,

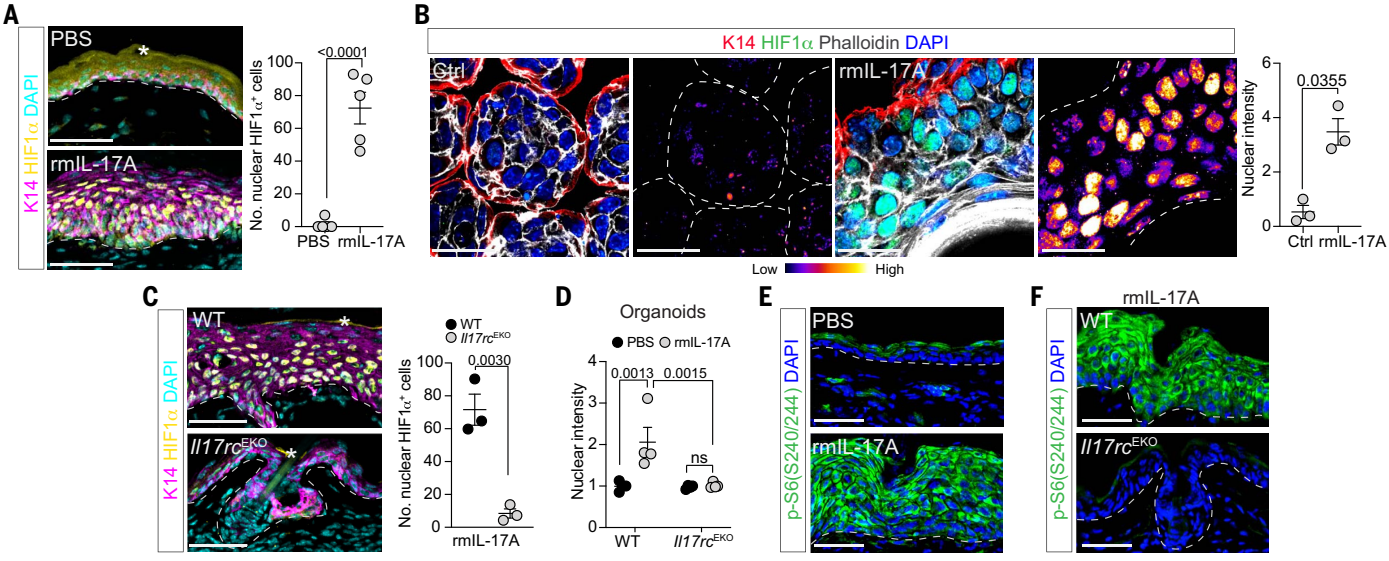


Fig. 4. IL-17A induces HIF1α through epidermal IL-17RC. (**A**) Intradermal injection of rmIL-17A in unwounded skin induces epidermal hyperplasia and HIF1α. Magenta, K14; yellow, HIF1α; cyan, DAPI nuclei (n = 5, N = 3). (**B**) Epidermal organoids cultured for 6 days at 21% O_2, ± rmIL-17A stimulation for 5 days (see experimental schematic in fig. S10C) robustly up-regulate HIF1α. Quantification of HIF1α staining normalized to controls. Red, K14; green, HIF1α; white, phalloidin F actin; blue, DAPI nuclei (n = 3, N = 3). (**C**) rmIL-17A induced nuclear HIF1α in WT mice but not epidermal IL-17RC–deficient (*Il17rc*^{EKO}) mice (n = 3, N = 2). Staining and color scheme are the same as in (A). (**D**) IL-17A–induced HIF1α in organoids is dependent on epidermal IL-17RC (n = 4, N = 2). (**E**) Epidermal response to rmIL-17A is accompanied by increased phosphorylation of S6 (p-S6$^{Ser240/244}$, green) (n = 3, N = 2). (**F**) rmIL-17A–induced p-S6$^{Ser240/244}$ (green) requires epidermal IL-17RC (n = 3, N = 2). In (E) and (F), blue, DAPI nuclei. In (A) and (C), white asterisks label autofluorescence. In (A) to (C) and (E) and (F), white dashed lines demarcate dermo-epidermal junction or organoid boundaries. Scale bars, 50 μm. For (A) to (C), two-tailed t test at a 95% confidence interval was used. For (D), two-way ANOVA multiple-comparisons test was used.

IL-17A directly and potently induces mTOR to up-regulate HIF1α under both normoxic and hypoxic conditions (Fig. 5H).

IL-17A–dependent metabolic reprogramming is vital for re-epithelialization

Migration, a cardinal feature of re-epithelialization, requires cells to expend energy. However, exactly how migrating epithelial cells meet their energy demands in a hypoxic wound microenvironment is unclear (Fig. 3, G and H). Glycolysis and oxidative phosphorylation are the two major cellular energy–producing pathways (*26*). Of these, HIF1α is known to control a transcriptional program of glycolysis (*10*). Accordingly, the WT wound front expressed higher levels of glycolysis-associated genes [hexokinase 2 (*hk2*), phosphoglycerate kinase 1 (*Pgk1*), and glucose transporter protein 1 (*Slc2a1*)] than did that of GFP-KI (Fig. 6A).

rmIL-17A treatment significantly increased glucose transporter 1 (*Slc2a1*) and glycolytic enzyme expression in epithelial organoids (Fig. 6B and fig. S12, A and B). rmIL-17A–treated organoids had heightened glycolysis activity and glycolytic capacity and secreted higher levels of lactate, an end product of glycolysis, compared with controls (Fig. 6C and fig. S12, C and D). Consistent with a role for mTOR in regulating HIF1α, rapamycin-treated or *Hif1a*^{EKO} organoids expressed lower levels of glycolysis enzymes than did controls (Fig. 6, D and E, and fig. S12, E and F). GKP-KI, *Il17rc*^{EKO}, and *Hif1a*^{EKO} wound-edge epithelia expressed significantly lower glucose transporter 1 levels than their respective controls (Fig. 6, F and G, and fig. S12G). rmIL-17A administration rescued glucose transporter 1 levels in GKP-KI wounds (fig. S12H). Thus, the IL-17A–HIF1α axis induces a program of glycolytic metabolism in epithelial cells.

Our findings raised the possibility that the IL-17A–HIF1α–dependent program of glucose metabolism fuels epithelial migration. We topically treated mice with 2-deoxy-D-glucose (2-DG), a glucose analog that inhibits glycolysis. 2-DG–treated mice had significantly shorter integrin α5$^+$ migrating tongues than their control counterparts (fig. S12I). Blocking glycolysis did not alter epithelial cell proliferation, indicating that glycolysis is dispensable for expanding the epithelial cell pool at the wound front (fig. S12J). 2-DG treatment also led to a reduction in wound-edge type 17 cells (fig. S12K). We therefore turned to an in vitro epithelial scratch assay to control for the immune effects of topical 2-DG treatment. 2-DG or HIF1α inhibitor (BAY872243) hindered the ability of human epithelial cells to migrate and seal the in vitro "wound" both in the presence and absence of recombinant human IL-17A (rhIL-17A) (Fig. 6H). Thus, IL-17A augments glycolysis, an essential bioenergetic pathway powering epithelial cell migration, to expedite healing.

Discussion

Collectively, our results illustrate that IL-17A supplied by RORγt$^+$ γδ T cells is necessary for optimal HIF1α activation in the wound-edge epithelium. HIFs are an ancient class of transcription factors that are present in all metazoan cells and are stabilized in response to low microenvironmental oxygen (*27*). Previous studies attributed the increased expression of HIF1α in injured skin and intestinal epithelium to tissue hypoxia (*28*, *29*). Thus, our discovery that the chronic hypoxic environment of the wound edge is insufficient to activate HIF1α prompts investigation into IL-17, and other secondary activators of HIF1α, in nonhealing wounds and conditions involving chronic hypoxia.

Consistent with published reports, we found that acute hypoxia robustly induces HIF1α (*25*). By contrast, exposure to chronic hypoxia (reminiscent of the wound microenvironment) leads to a reduction in mTOR and consequently HIF1α. The reasons for down-modulation of mTOR in chronic hypoxia are unclear and warrant further investigation. Nevertheless, IL-17A acts as a "second signal" to activate mTOR in this context through ERK1/2 and AKT kinases. We found that IL-17A–induced mTOR controls HIF1α gene transcription and protein levels but not stability. Thus, our results suggest that adaptation to hypoxia in complex tissue microenvironments requires second

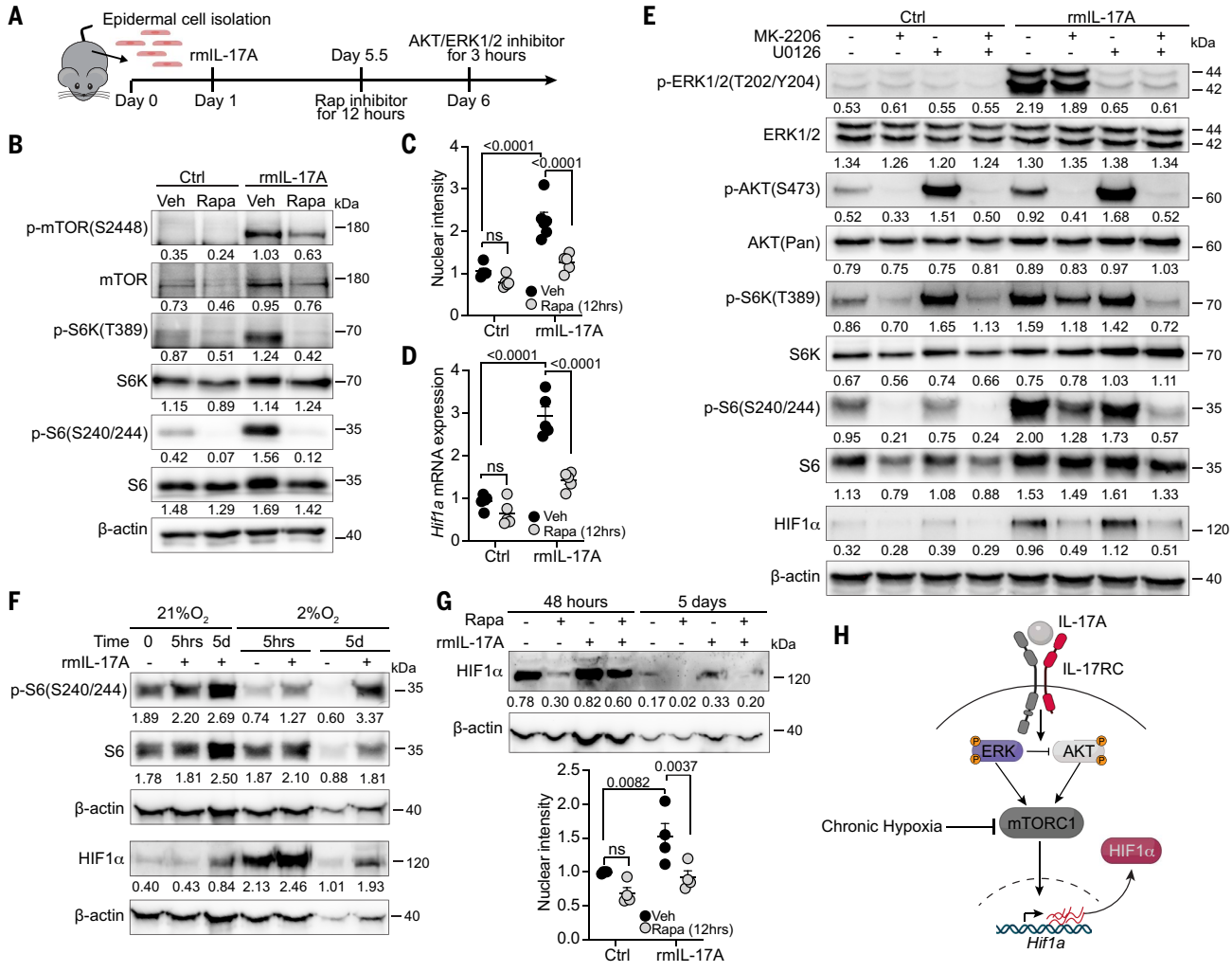


Fig. 5. ERK/AKT signaling downstream of IL-17RC controls mTOR and HIF1α. (**A**) Experimental schematic of inhibitor treatment in control and rmIL-17A–stimulated organoids. (**B**) rmIL-17A–treated organoids (D5) have increased pmTOR^{S2448}, p-S6K^{T389}, and p-S6$^{S240/244}$ that is diminished after 12 hours of rapamycin (Rapa) treatment. Vehicle (Veh), *N* = 2. (**C**) Rapamycin abrogates rmIL-17A–mediated HIF1α expression in normoxia (21% O_2). Quantification of HIF1α staining normalized to controls ($n \geq 5$, *N* = 4). (**D**) IL-17A up-regulates *Hif1a* transcripts through mTOR (*n* = 5, *N* = 2). (**E**) Inhibiting AKT (MK-2206) and/or ERK (U0126) for 3 hours blocks rmIL-17A–induced p-S6K^{T389}, p-S6$^{S240/244}$, and HIF1α (*N* = 2). (**F**) IL-17A stimulates p-S6$^{S240/244}$ and HIF1α in organoids cultured under normoxic or hypoxic (2% O_2) conditions (*N* = 3). (**G**) Rapamycin abrogates IL-17A–mediated HIF1α expression in hypoxia (*n* = 4, *N* = 3). (**H**) Schematic of proposed mechanism illustrating that AKT and ERK activation proximal of IL-17RC induces mTOR and HIFα. Chronic hypoxia inhibits mTOR and consequently HIF1α. Significance was determined using a two-way ANOVA multiple-comparisons test. In (B), (E), (F), and (G), protein quantifications are relative to the presented internal β-actin controls.

signals from accessory cells in addition to cell-autonomous oxygen monitoring.

IL-17A directed the metabolic rewiring of wound-edge epithelium toward a program of glycolysis. In contrast to recent work highlighting a role for IL-17A–induced glycolysis in fibroblast survival and proliferation, wound-edge epithelia did not require glycolysis to proliferate (*21*). Instead, this mechanism of energy production was vital for epithelial migration at the wound front. HIF1α and glycolysis are drivers of tumor progression and metastasis, raising the possibility that IL-17A or other immune-derived signals could drive these pathways in cancer (*30*, *31*). IL-17A is also central to the pathology of many autoimmune conditions, including psoriasis, inflammatory bowel disease, and spondyloarthropathies (*14*). In addition to glycolysis, other HIF1α transcriptional targets may fuel IL-17–mediated epithelial pathologies. Thus, the IL-17A–HIF1α axis that we uncovered here may provide therapeutic opportunities for a wide range of inflammatory and metastatic diseases.

Materials and Methods

Animals

The following mouse strains were purchased from The Jackson Laboratory: C57BL/6J, B6.129P2(Cg)-*Rorc*^{tm2Litt}/J (*Rorgt*-EGFP), B6N.Cg-Tg(KRT14-cre)1Amc/J, B6.Cg-*Il17rc*$^{tm1.1Koll}$/J, B6.Cg-Tg(Cd4-cre)1Cwi/BfluJ, B6.129S-*Tcrd*$^{tm1.1(cre/ERT2)Zhu}$/J, B6(Cg)-*Rorc*^{tm3Litt}/J, B6.129-*Hif1a*^{tm3Rsjo}/J. *Mr1*$^{tm1(KOMP)Vlcg}$ mice were generated from the KOMP ES cell line *Mr1*$^{tm1(KOMP)Vlcg}$, RRID:MMRRC_058774-UCD, obtained from the Mutant Mouse Resource and Research Center (MMRRC) at University of California, Davis. Tg(Rorc-EGFP)1Ebe mice were a gift from Dr. G. Eberl (Institut Pasteur). All animal studies were approved by the institutional animal care and use committee (IACUC). Mice were bred and maintained under specific-pathogen-free conditions at the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)–accredited facilities at the New York University Langone Health Center and housed in accordance with the

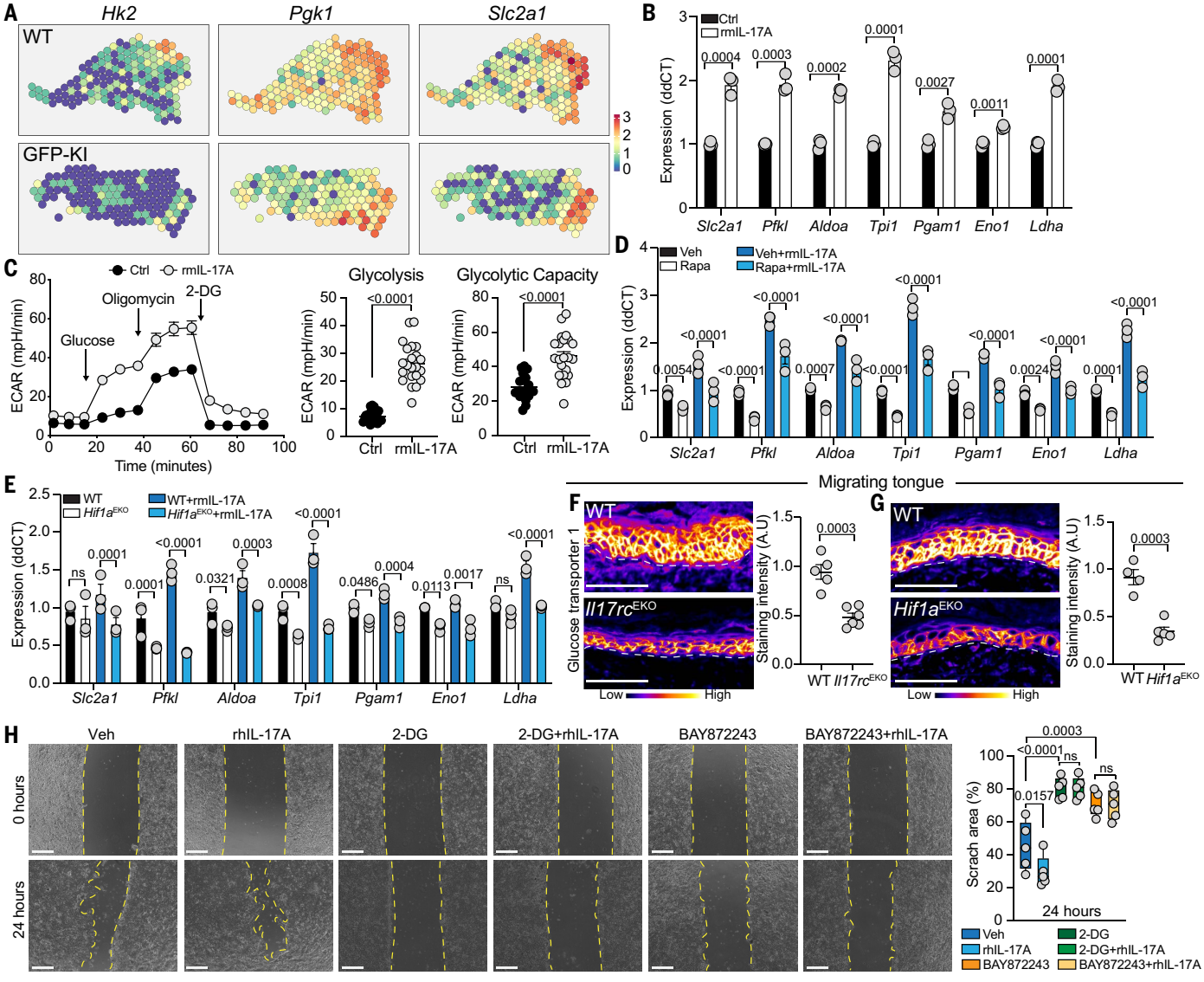


Fig. 6. IL-17A–HIF1α-mediated metabolic reprograming fuels re-epithelialization. (**A**) ST plots of *hk2*, *Pgk1*, and *Slc2a1* expression in GFP-KI (n = 3) and WT (n = 2) wounds. (**B**) Increased expression of glycolytic enzymes after rmIL-17A treatment (normalized to control) (n = 6, N = 4). (**C**) rmIL-17A stimulation results in functional enhancement of glycolysis. Representative ECAR of organoids cultured in the presence or absence of rmIL-17A. There were 22 technical replicates. (**D** and **E**) IL-17A–induced expression of glycolytic enzymes is abrogated in rapamycin-treated (D) and *Hif1a*^{EKO} (E) organoids ($n \geq 5$, N = 2). (**F** and **G**) pseudocolor fire images and quantifications of glucose transporter 1 staining in D3 WT and *Il17rc*^{EKO} ($n \geq 5$, N = 2) (F) and WT and *Hif1a*^{EKO} ($n \geq 4$, N = 2) (G) migrating tongue. Scale bars, 100 μm. (**H**) Inhibition of HIF1α (BAY872243) and glycolysis (2-DG) impairs the migration of primary human keratinocytes in the presence and absence of rhIL-17A. Yellow dashed line marks scratch wound edges. Scratched areas were quantified as a percentage relative to the start area at 0 hours (n = 5, N = 3). Scale bars, 500 μm. In (B), (C), (F), and (G), two-tailed t test at a 95% confidence interval was used. In (D) and (E), two-way ANOVA multiple-comparisons test was used. In (H), one-way ANOVA was used.

procedures outlined in the National Institutes of Health's *Guide for the Care and Use of Laboratory Animals*. Experiments were performed with IACUC-approved protocols. Age- and sex-matched controls were used in all experiments presented.

Human wound samples

Normal and wounded human skin samples were collected from three healthy donors (a 48-year-old female, a 24-year-old female, and a 27-year-old male) at Karolinska University Hospital, Stockholm, Sweden. After local lidocaine injection, two full-thickness excisional wounds were created using a 4-mm biopsy punch at the upper buttock area of each donor. Central excised skin was used as unwounded control. Wound-edge skin was collected using a 6-mm biopsy punch 1 and 7 days later. All donors consented to the collection and use of clinical samples. The study was approved by the Stockholm Regional Ethics Committee and conducted according to the Declaration of Helsinki's principles.

Skin wound-healing models

Full-thickness wounds

Seven- to 8-week-old mice in the telogen (resting) phase of the hair cycle were used for wounding studies. After hair removal, a 4-mm biopsy punch (Miltex) was used to make full-thickness wounds on dorsal skin. Wound closure was assessed macroscopically with an

engineer's caliper, and the wound-healing area and rate were calculated as previously described (*9*).

Splinted full-thickness wounds

Four-millimeter full-thickness wounds were stented open by a 10-mm-diameter silicone ring and secured with skin glue and eight simple, interrupted 6-0 sutures. Wounds were dressed immediately after splinting and dressings were changed every other day thereafter. Animals were singly housed to limit disruption to splint.

Skin tissue digest

Unwounded or 0.25 mm of skin around the wound edge was excised and digested in Liberase TL (Sigma-Aldrich) as previously described (*9*).

Flow cytometry and fluorescence-activated cell sorting (FACS)

Single-cell suspensions were preincubated with anti-CD16/32 before staining with surface fluorescent conjugated antibodies and/or oligo-tagged antibodies (table S1) at predetermined concentrations in a 100 μl of staining buffer [phosphate-buffered saline (PBS) containing 5% fetal bovine serum (FBS) and 1% HEPES] per 10^7 cells. Stained cells were resuspended in 4′,6-diamidino-2-phenylindole (DAPI) in FACS buffer before purification or analysis. For intracellular staining, fixable Live/Dead dye was added before surface antibody incubation, which was performed according to the manufacturer's instructions for the Foxp3/Transcription Factor Staining Buffer Set (BD Biosciences). Flow cytometry data were acquired on LSRII Analyzers (BD Biosciences) and then analyzed with FlowJo program. FACS was performed using FACSAria Cell Sorters (BD Biosciences).

Immunofluorescence and image analysis

Tissue was fixed, processed, and stained as previously described (*9*). EdU reactions were performed according to the manufacturer's directions (Life Technologies). Migrating tongues were determined by measuring the length of $K14^+$ integrin $\alpha5^+$ cells. The number of wound-edge proliferating epithelial cells was determined by counting EdU^+ $K14^+$ cells from the first hair follicle adjacent to the wound until the wound's leading edge. To detect PIM, slides were pretreated with a mouse-on-mouse immunodetection kit (Vector Laboratories) and then stained with biotin–anti-PIM adduct antibodies per the manufacturer's guidelines (Hydroxyprobe). ImageJ was used to quantify PIM and glucose transporter 1 integrated density, a well-established method of measuring fluorescence intensity that accounts for differences in the area of the signal in $K14^+$ and integrin $\alpha5^+$ migrating tongue. Images are shown with the pseudocolor Fire from ImageJ after they were contrasted equally and background was uniformly removed.

Epithelial organoids were released from basement membrane matrix (BME) with organoid harvesting solution (Cultrex), fixed in 4% paraformaldehyde for 2 hours at 4°C, and then incubated in acetone at −20°C. Organoids were stained with conditions similar to tissue sections and embedded in agarose for imaging. Data were analyzed using ImageJ. Organoid HIF1α intensity was determined using the Single Target Translocation Assay from HSC studio (Thermo Fisher). For the complete list of immunofluorescence antibodies and concentrations used, see table S2.

Microscopy and image processing

Images were acquired with a Zeiss Axioplan2 using a Plan-Apochromat 20×/0.8 air objective or Zeiss LSM 710 laser scanning confocal microscope using Plan-Apochromat 20×/0.75 air and Plan-Apochromat 63×/1.40 oil objectives. Tiled and stitched images of sagittal sections were collected using a Plan-Apochromat 20×/0.75 air objective, controlled by Zen software (Zeiss).

Organoids were imaged using Cell Insight CX7-LZR (Thermo Scientific), and images were processed with HCS Studio Cell Analysis software (Thermo Scientific). Maximal projection *z*-stacks are presented, and co-localization was interpreted only in single *z*-stacks. *z*-stacks were projected using ImageJ. Images were assembled in Adobe Illustrator CC2015.3.

Animal drug treatments

EdU pulse

Mice were injected intraperitoneally with EdU (50 μg/g) (Thermo Scientific) either 3 hours (epithelial cells) or 24 hours (lymphocytes) before analysis.

FTY720

Mice were intraperitoneally injected with 2 mg/kg of FTY720 or vehicle control for 3 consecutive days before wounding and then throughout the wound repair response. Ten microliters of blood were collected through the jugular vein to test FTY70 treatment efficacy before wound analysis.

Tamoxifen

To activate Cre recombinase, 7- to 8-week-old mice were treated with 7.5 mg of 4-hydroxytamoxifen (Sigma-Aldrich) dissolved in 250 μl of corn oil every other day by oral gavage three times before wounding and continuously during wound repair.

PIM HCl

PIM HCl (hypoxyprobe) was intraperitoneally injected into mice at the dose of 60 mg/kg 2 hours before sacrifice.

Recombinant cytokines

The wound edge was intradermally injected with either rmIL-17A (500 ng in 20 μl of PBS, carrier-free; PeproTech) or PBS control immediately after a 4-mm biopsy. The following cytokine concentrations were used for daily intradermal injections for 7 days in unwounded ear pinnae and compared with the PBS-only vehicle control: IL-17A (500 ng in 20 μl of PBS, carrier-free; PeproTech), IL-17F (1 μg in 20 μl of PBS, carrier-free; PeproTech), or IL-22 (500 ng in 20 μl of PBS, carrier-free; PeproTech).

2-DG

2-DG (40 mg/ml; Cayman Chemical) was first dissolved in PBS and then diluted in a water-based gel. This gel was topically applied to dorsal skin 2 days before wounding and then at the wound perimeter daily until analysis.

CITE-seq

FACS-purified live $CD45.2^+CD90.2^+$ $TCRV\gamma3^-$ cells from unwounded skin at D3 and D5 after wounding were prelabeled with surface-epitope-marking oligo-tagged antibodies and sample-specific oligo-tagged Totalseq-A antibodies (BioLegend; see table S1). Hashed samples were pooled at a 1:1.5:1 ratio of control: D3:D5 before library preparation (Chromium Single Cell 3′ Library, 10X Genomics) and sequenced on an Illumina HiSeq 4000 as 150-bp paired-end reads. Sequencing results were demultiplexed and converted to FASTQ format using Illumina bcl2fastq software. The Cell Ranger Single-Cell Software Suite was used to perform sample demultiplexing, barcode processing, and single-cell 3′ gene counting. The cDNA insert was aligned to the mm10/GRCm38 reference genome. Only confidently mapped, non–polymerase chain reaction (PCR) duplicates with valid barcodes and unique molecular identifiers (UMIs) were used to generate the gene-barcode matrix. Cell Ranger output was further analyzed in R using the Seurat package (*32*). Surface epitope oligo sequences were merged with cell transcriptome data by matching the cell barcode IDs.

Sample demultiplexing and quality controls

Further analysis including quality filtering, the identification of highly variable genes, dimensionality reduction, standard unsupervised clustering algorithms, and the discovery of differentially expressed genes was performed using the Seurat R package. Samples were demultiplexed to filter out multiplets (cells mapping to multiple hashtags) and negative cells (cells missing hashtags) with a positive quantile threshold of 0.99 between samples. Individual samples were further processed to remove cells with >20% mitochondrial gene expression. To exclude low-quality cells and remaining multiplets or cells that were extreme outliers, the distribution of

total genes per cell was calculated. Then, control parameters were applied to filter out cells with <200 detected genes and >3800 detected genes. After removing unwanted cells from the dataset, we normalized the data by the total expression, multiplied by a scale factor of 10,000, and log-transformed the result.

Cell clustering and annotation for CITE-seq data

Clustering was performed using the Seurat R package. PCA was performed to generate graph-based cell clusters, and the top seven PCs with a resolution of 0.6 were used. This yielded 16 unique cell clusters, which were further plotted on the UMAP. Next, contaminating nonlymphoid clusters were removed: macrophages (*Cd74*, *Fcgr1a*, and *Mrc1*), fibroblasts (*Col1a2*, *Pdgfra*, and *Thy1*), and dendritic cells (*Cd74*, *Cd207*, and *Itgax*). Dendritic epidermal T cells ($Cd3e^{hi}$, $Tcrd^{hi}$), which are unique to mice and are not found in human skin, were also excluded. PCA was performed again, and the top seven PCs with a resolution of 0.5 were used to generate a UMAP with 12 unique clusters. Each of these clusters was then defined and labeled for cell types using marker genes obtained by performing differential gene expression analysis between the clusters using Wilcoxon rank sum test with an adjusted *P* value < 0.1 and natural log (fold-change) >0.25 as cutoff and parallel analysis of surface epitope enrichment. Violin plots of marker genes and surface epitopes in fig. S1 generated using the scanpy and seaborn packages in Python (*33*).

Spatial transcriptomics

Sample processing

Tissues were embedded in optimal cutting temperature medium and frozen in liquid nitrogen–chilled isopentane within 15 min of harvesting. Ten-micrometer cryosections were mounted onto the ST arrays (10X Genomics Visium) and stored at −80°C until use. Tissue sections were fixed in methanol at −20°C and then stained with hematoxylin and eosin. Bright-field images were taken on a Leica SCN400F slide scanner at 20× resolution. Slides were permeabilized with permeabilization enzyme for 5 min, as determined by the tissue optimization protocol. Polyadenylated RNAs captured on the underlying arrays were resuspended in 1.2 ml of 0.1 N HCl for 5 min and reverse transcribed at 53°C for 45 min, followed by second-strand synthesis at 65°C for 5 min. After library preparation, samples were sequenced on a Novaseq 600 (Illumina)

Spot selection and image alignment

After probe cleavage, fluorescent images were taken on a Hamamatsu NanoZoomer whole-slide fluorescence scanner. Bright-field images of the tissue and fluorescent images were manually aligned with Adobe Photoshop CS6 to identify the array of spots beneath the tissue.

Sequence alignment and annotation

Sequencing output and the histology images were processed using space ranger software (10X Genomics). The space ranger mkfastq function was used for sample demultiplexing and to converting spatial barcodes and reads into FASTQ format. The space ranger count function was used to align reads from FATSQ files to the mouse genome (mm10/GRCm38) and then align the microscopic slide image and transcriptome to generate barcode/UMI counts, feature spot matrices, cluster data, and perform gene-expression analysis.

Clustering analysis of ST data

The data were then further analyzed using the Seurat R package. To account for variance in sequencing depth across tissue spots (biological variance driven by heterogeneity in cell density and transcriptional activity) while controlling for technical artifacts, scTransform was used for normalization of gene expression by spots (*34*). Dimensionality reduction and clustering of sequencing data were run using the same workflow as CITE-seq. Before clustering analysis, scab regions were filtered out from each ST sample using subset function in the Seurat R package.

For the WT control (unwounded) skin sample, PCA was performed and the top 10 PCs with a resolution of 1 were used to identify spatial clusters. For the WT-wounded D3 skin sample, PCA was performed and the top 20 PCs with a resolution of 1 were used to identify spatial clusters. UMAP dimensionality reduction was performed (Fig. 2B) and then overlaid on a histology image for spatial interpretation (Fig. 2A). Clusters were annotated on the basis of top differentially expressed genes and localization in tissue as revealed by histology. To combine and batch align multiple ST samples, the scTranform pipeline from the Seurat vignette was used. After batch alignment, PCA was performed and the top 18 PCs with resolution of 0.9 were used to identify spatial clusters, followed by UMAP dimensionality reduction (Fig. 3, A and B). The data were normalized using scTransform after batch alignment to regress for replicates, and the number of genes per spot and the spatial feature expression visualizations were performed using the normalized data.

Differential expression analysis for ST data

The markers genes were determined using FindAllMarkers function in Seurat, and the Wilcoxon rank-sum test was used for differential testing. The differential expression analysis between the wounded GFP-KI and wounded WT group was also performed using Wilcoxon rank sum test and a non-batch-transformed spatial assay was used for this test. The results were filtered to only include genes with adjusted *P* values <0.1 and natural log (fold-change) >0.25 for up-regulated genes. The pathway analysis was performed using the cluster Profiler package in R (*35*).

Multimodal integration analysis

The significance of the overlap between ST genes and cell type marker genes from CITE-seq and published scRNA-seq data [Gene Expression Omnibus (GEO) accession number GSE142471] were queried using the hypergeometric cumulative distribution, with all genes as the background to compute the *P* value. All genes were calculated as the number of common genes detected between ST and CITE-seq in raw data.

Low-input bulk-RNA sequencing of wound-edge epithelium

A ring width of 0.25 mm skin around the wound edge from control and *Rorc*-deficient GFP-KI (two or three mice pooled per group per sequencing run) was excised and digested as described in the skin tissue digest section above. Would-edge epithelium was FACS purified from live $Epcam^+$ cells on the basis of integrin α6 and integrin α5 expression while excluding $CD45^+$ immune cells, $CD31^+$ endothelial cells, and $CD117^+$ melanocytes. Total RNA was extracted from FACS-purified integrin $\alpha6^+$ integrin $\alpha5^+$ (migrating tongue) and integrin $\alpha6^+$ integrin $\alpha5^-$ (wound edge) samples using the RNeasy Plus Micro Kit (Life Technologies). Libraries were generated with the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing and the SMARTer ThruPLEX DNA-Seq Kit (Takara Bio). Libraries were sequenced on Nova 6000 system with 50-bp paired-end reads.

Sequencing results were demultiplexed and converted to FASTQ format using Illumina bcl2fastq software. The reads were trimmed for adapter sequences using trimmomatic (version 0.36) in paired end mode with the minimal read length option set to 35 base pairs, trailing set to 5, and the sliding window set to 4:15. The sequencing reads were aligned to the mouse genome (mm10/GRCm38) using the splice-aware STAR aligner (*36*). The feature Counts program was used to generate counts for each gene on the basis of how many aligned reads overlap its exons (*37*). The raw counts were further converted to TPM counts to filter for genes with a minimal TPM value ≥10 before differential expression analysis. The raw gene counts were then subset using these TPM-filtered genes (n = 9412) and used to test for differential expression using negative binomial generalized linear models implemented by the DESeq2 R package (*38*). Volcano plots of differentially expressed genes (log_2-fold change

≥0.25, adjusted *P* value <0.1) were generated using the Enhanced Volcano package in R. Pathways analysis of differentially expressed and up-regulated genes (log_2-fold change ≥0.25, adjusted *P* value <0.1) was performed using the cluster Profiler package in R with the Gene Ontology (GO) (*39*) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (*40*) databases.

In vivo oxygen measurements

Tissue oxygen levels were measured using OxyLite probes (Oxford Optronix). The OxyLite NXpO_2 Barer Fiber Sensor with a ~250-μm tip was directly inserted into wounded or healthy skin of an anesthetized mouse to continuously monitor oxygen pressure (pO_2) for 15 to 20 min. pO_2 levels were recorded and analyzed with a data-acquisition system (LabChart Reader, Chart version 8 for Windows; AD Instruments)

Organoid culture

Epithelial organoids were generated as previously described from the dorsal skin of C56BL/6 mice (ages postnatal days 46 to 50) (*41*). One day after culturing in normoxia (21% O_2) or hypoxia (2% O_2), rmIL-17A (500 ng/ml, Peprotech), TNFα (100 ng/ml, Peprotech) (*42*), IL-1β (100 ng/ml, Peprotech) (*43*), IL-6 (100 ng/ml, Peprotech) (*44*), epidermal growth factor (500 ng/ml, Peprotech) (*41*), or PBS was added to the organoid culture medium and replenished on D3. Organoids were harvested on D6 using a nonenzymatic cultrex organoid harvesting solution (R&D Systems) for downstream analyses. For inhibitor studies, organoids were treated 12 hours before harvest with 60 nM rapamycin (LC Laboratories) (*45*) and vehicle control or 3 hours with U0126-ERK1/2 inhibitor (10 μM, MedChemExpress) (*46*) or MK-2206-AKT inhibitor (5 μM, MedChemExpress) (*47*) and vehicle control. Inhibitor concentrations and time course were chosen on the basis of published studies, efficacy of inhibition, and conditions that minimized organoid cell death in our cultures. For protein stability studies, cycloheximide (50 μg/ml, Cayman Chemical) was added to the culture for 5, 10, and 15 min. For mRNA stability studies, actinomycin D (10 μg/ml, Cayman Chemical) was added to the culture for 2, 5, and 8 hours.

Organoid assays

Quantitative PCR

Total RNA was extracted from organoids using RNeasy Plus Micro Kit, and equal amounts of RNA were reverse-transcribed using the superscript VILO cDNA synthesis kit (Invitrogen). Analytes were normalized using the housekeeping gene *Actb*. For the complete list of quantitative PCR primers, refer to table S3.

Immunoblot

Organoids were lysed with RIPA buffer [62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% glycerol, 5% β-mercaptoethanol supplemented with the Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher)] and dissociated by passing through a 30-gauge needle. Organoids cultured in a hypoxia chamber were lysed directly in the wells without releasing with organoid-harvesting solution. After incubation at 100°C for 10 min, protein concentration was measured with the DC Protein Assay Kit (Bio-Rad). XT sample buffer was added to the protein lysate (Bio-Rad) and incubated in 95°C for 5 min before being resolved in SDS–polyacrylamide gel electrophoresis and analyzed with immunoblot. Antibodies used for immunoblots are listed in table S2. Figures present representative immunoblots quantified with ImageJ; integrated density for each band was normalized with internal β-actin.

Seahorse extracellular acidification rate analysis

Extracellular acidification rate (ECAR) measurements were performed using an XFe96 analyzer (Agilent Technologies). Seahorse XFe96 plates (Agilent Technologies) were coated with a 10× dilution of BME in Dulbecco's modified Eagle's medium (DMEM; Sigma-Aldrich). IL-17A–treated or control organoids were collected with organoid harvest solution, washed once in DMEM basic medium, resuspended in DMEM basic medium with 2 mM glutamine, and seeded onto the coated plates at 175 μl per well. Respiratory rates were measured in response to sequential injections of glucose (10 mM), oligomycin (2 μM), and 2-DG (50 mM) (all Cayman Chemical). Values were normalized to a total number of cells in organoid culture well using the CyQUANT Cell Proliferation Assay (Thermo Fisher).

Lactate secretion assay

Organoid culture supernatant was harvested and centrifuged at 1000*g* to pellet debris. Lactate was measured using the colorimetric L-Lactate Assay Kit (Abcam). Values were normalized to a total number of cells in each organoid culture well using the CyQUANT Cell Proliferation Assay (Thermo Fisher).

Human primary keratinocyte scratch assay

Primary human keratinocytes were isolated from the foreskin and then cultured in serum-free Keratinocytes Growth Medium KGM-Gold (Lonza, Basel, Switzerland) supplemented with bovine pituitary extract, human endothelial growth factor, bovine insulin, hydrocortisone, gentamicin-amphotericin B (GA-1000), epinephrine, and transferrin.

For scratch migration assays, keratinocytes were allowed to reach confluency. Cells were pretreated with mitomycin C (8 μg/ml) 2 hours before and with 2-DG (40 mM) and BAY872243 (10 μM) 1 hour before scratch assay. Scratches were created by manual scraping of the cell monolayer with a 1-ml pipette tip. Cells were treated with rhIL-17A (100 ng/ml, Peprotech). The scratch area was imaged at 0 and 24 hours using EVOS (Thermo Fisher) and quantified using ImageJ.

Statistics

Data are presented as mean ± SEM. Group sizes were determined on the basis of the results of preliminary experiments. Mice were assigned at random to groups. Experiments were not performed in a blinded fashion. Statistical significance for each experiment was determined as shown in the figure legends, where *n* = the number of independent biological replicates (animals, unless noted as cells) per group, and *N* = the number of independent experimental replicates. Statistical significance of bulk RNA-seq data was calculated using an adjusted *P* value cutoff <0.1 in the DESeq2 R package. Statistical analyses were performed in Prism (GraphPad), DESeq2, or R. No data points were excluded.

REFERENCES AND NOTES

1. R. Chovatiya, R. Medzhitov, Stress, inflammation, and defense of homeostasis. *Mol. Cell* **54**, 281–288 (2014). doi: 10.1016/j.molcel.2014.03.030; pmid: 24766892
2. S. A. Eming, P. Martin, M. Tomic-Canic, Wound repair and regeneration: Mechanisms, signaling, and translation. *Sci. Transl. Med.* **6**, 265sr6 (2014). doi: 10.1126/scitranslmed.3009337; pmid: 25473038
3. Y. Belkaid, O. J. Harrison, Homeostatic immunity and the microbiota. *Immunity* **46**, 562–576 (2017). doi: 10.1016/j.immuni.2017.04.008; pmid: 28423337
4. O. J. Harrison *et al.*, Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury. *Science* **363**, eaat6280 (2019). doi: 10.1126/science.aat6280; pmid: 30523076
5. S. Naik, S. B. Larsen, C. J. Cowley, E. Fuchs, Two to Tango: Dialog between Immunity and Stem Cells in Health and Disease. *Cell* **175**, 908–920 (2018). doi: 10.1016/j.cell.2018.08.071; pmid: 30388451
6. M. Aragona *et al.*, Defining stem cell dynamics and migration during wound healing in mouse skin epidermis. *Nat. Commun.* **8**, 14684 (2017). doi: 10.1038/ncomms14684; pmid: 28248284
7. S. Bukhari, A. F. Mertz, S. Naik, Eavesdropping on the conversation between immune cells and the skin epithelium. *Int. Immunol.* **31**, 415–422 (2019). doi: 10.1093/intimm/dxy088; pmid: 30721971
8. R. Moncada *et al.*, Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020). doi: 10.1038/s41587-019-0392-8; pmid: 31932730
9. S. Naik *et al.*, Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* **550**, 475–480 (2017). doi: 10.1038/nature24271; pmid: 29045388
10. G. L. Semenza, Hypoxia-inducible factors in physiology and medicine. *Cell* **148**, 399–408 (2012). doi: 10.1016/j.cell.2012.01.021; pmid: 22304911
11. M. Stoeckius *et al.*, Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017). doi: 10.1038/nmeth.4380; pmid: 28759029
12. J. J. O'Shea, W. E. Paul, Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* **327**, 1098–1102 (2010). doi: 10.1126/science.1178334; pmid: 20185720
13. G. Eberl, RORγt, a multitask nuclear receptor at mucosal surfaces. *Mucosal Immunol.* **10**, 27–34 (2017). doi: 10.1038/mi.2016.86; pmid: 27706126
14. M. J. McGeachy, D. J. Cua, S. L. Gaffen, The IL-17 family of cytokines in health and disease. *Immunity* **50**, 892–906 (2019). doi: 10.1016/j.immuni.2019.03.021; pmid: 30995505
15. G. Eberl *et al.*, An essential function for the nuclear receptor RORgamma(t) in the generation of fetal lymphoid tissue

inducer cells. *Nat. Immunol.* **5**, 64–73 (2004). doi: 10.1038/ni1022; pmid: 14691482
16. J. R. Huh *et al*., Digoxin and its derivatives suppress TH17 cell differentiation by antagonizing RORγt activity. *Nature* **472**, 486–490 (2011). doi: 10.1038/nature09978; pmid: 21441909
17. I. Pastar *et al*., Epithelialization in wound healing: A comprehensive review. *Adv. Wound Care* **3**, 445–464 (2014). doi: 10.1089/wound.2013.0473; pmid: 25032064
18. A. Baeyens, V. Fang, C. Chen, S. R. Schwab, Exit strategies: S1P signaling and T cell migration. *Trends Immunol.* **36**, 778–787 (2015). doi: 10.1016/j.it.2015.10.005; pmid: 26596799
19. D. Haensel *et al*., Defining epidermal basal cell states during skin homeostasis and wound healing using single-cell transcriptomics. *Cell Rep.* **30**, 3932–3947.e6 (2020). doi: 10.1016/j.celrep.2020.02.091; pmid: 32187560
20. M. G. Constantinides *et al*., MAIT cells are imprinted by the microbiota in early life and promote tissue repair. *Science* **366**, eaax6624 (2019). doi: 10.1126/science.aax6624; pmid: 31649166
21. S. Majumder *et al*., IL-17 metabolically reprograms activated fibroblastic reticular cells for proliferation and survival. *Nat. Immunol.* **20**, 534–545 (2019). doi: 10.1038/s41590-019-0367-4; pmid: 30962593
22. X. Chen *et al*., IL-17R-EGFR axis links wound healing to tumorigenesis in Lrig1$^+$ stem cells. *J. Exp. Med.* **216**, 195–214 (2019). doi: 10.1084/jem.20171849; pmid: 30573323
23. A. J. Majmundar, W. J. Wong, M. C. Simon, Hypoxia-inducible factors and the response to hypoxic stress. *Mol. Cell* **40**, 294–309 (2010). doi: 10.1016/j.molcel.2010.09.022; pmid: 20965423
24. M. Laplante, D. M. Sabatini, mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012). doi: 10.1016/j.cell.2012.03.017; pmid: 22500797
25. H. Harada *et al*., The Akt/mTOR pathway assures the synthesis of HIF-1alpha protein in a glucose- and reoxygenation-dependent manner in irradiated tumors. *J. Biol. Chem.* **284**, 5332–5342 (2009). doi: 10.1074/jbc.M806653200; pmid: 19098000
26. J. Zhu, C. B. Thompson, Metabolic regulation of cell growth and proliferation. *Nat. Rev. Mol. Cell Biol.* **20**, 436–450 (2019). doi: 10.1038/s41580-019-0123-5; pmid: 30976106
27. C. T. Taylor, J. C. McElwain, Ancient atmospheres and the evolution of oxygen sensing via the hypoxia-inducible factor in metazoans. *Physiology* **25**, 272–279 (2010). doi: 10.1152/physiol.00029.2010; pmid: 20940432
28. H. R. Rezvani *et al*., Loss of epidermal hypoxia-inducible factor-1α accelerates epidermal aging and affects re-epithelialization in human and mouse. *J. Cell Sci.* **124**, 4172–4183 (2011). doi: 10.1242/jcs.082370; pmid: 22193962
29. Y. Wang *et al*., Long-term culture captures injury-repair cycles of colonic stem cells. *Cell* **179**, 1144–1159.e15 (2019). doi: 10.1016/j.cell.2019.10.015; pmid: 31708126
30. G. L. Semenza, Targeting HIF-1 for cancer therapy. *Nat. Rev. Cancer* **3**, 721–732 (2003). doi: 10.1038/nrc1187; pmid: 13130303
31. T. Shiraishi *et al*., Glycolysis is the primary bioenergetic pathway for cell motility and cytoskeletal remodeling in human prostate and breast cancer cells. *Oncotarget* **6**, 130–143 (2015). doi: 10.18632/oncotarget.2766; pmid: 25426557
32. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi: 10.1038/nbt.4096; pmid: 29608179
33. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi: 10.1186/s13059-017-1382-0; pmid: 29409532
34. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019). doi: 10.1186/s13059-019-1874-1; pmid: 31870423
35. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012). doi: 10.1089/omi.2011.0118; pmid: 22455463
36. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
37. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi: 10.1093/bioinformatics/btt656; pmid: 24227677
38. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
39. M. Ashburner *et al*., Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000). doi: 10.1038/75556; pmid: 10802651
40. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49** (D1), D545–D551 (2021). doi: 10.1093/nar/gkaa970; pmid: 33125081
41. K. E. Boonekamp *et al*., Long-term expansion and differentiation of adult murine epidermal stem cells in 3D organoid cultures. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14630–14638 (2019). doi: 10.1073/pnas.1715272116; pmid: 31253707
42. T. Grabinger *et al*., Ex vivo culture of intestinal crypt organoids as a model system for assessing cell death induction in intestinal epithelial cells and enteropathy. *Cell Death Dis.* **5**, e1228 (2014). doi: 10.1038/cddis.2014.183; pmid: 24832600
43. H. F. Farin *et al*., Paneth cell extrusion and release of antimicrobial products is directly controlled by immune cell-derived IFN-γ. *J. Exp. Med.* **211**, 1393–1405 (2014). doi: 10.1084/jem.20130753; pmid: 24980747
44. M. Soutto *et al*., Activation of STAT3 signaling is mediated by TFF1 silencing in gastric neoplasia. *Nat. Commun.* **10**, 3039 (2019). doi: 10.1038/s41467-019-11011-4; pmid: 31292446
45. K. B. Jung *et al*., Interleukin-2 induces the in vitro maturation of human pluripotent stem cell-derived intestinal organoids. *Nat. Commun.* **9**, 3039 (2018). doi: 10.1038/s41467-018-05450-8; pmid: 30072687
46. Y. H. Tsai *et al*., *In vitro* patterning of pluripotent stem cell-derived intestine recapitulates *in vivo* human development. *Development* **144**, 1045–1055 (2017). pmid: 27927684
47. Y. Li *et al*., Induction of expansion and folding in human cerebral organoids. *Cell Stem Cell* **20**, 385–396.e3 (2017). doi: 10.1016/j.stem.2016.11.017; pmid: 28041895

ACKNOWLEDGMENTS

We thank D. R. Littman, D. Rosenblum, L. Guenin-Mace, B. G. Neel, and M. R. Phillips for helpful discussions, advice, and/or critical reading of this manuscript. The following core facilities enabled our study: NYULMC High-Performance Computing, Flow Cytometry, Genome Technology Center, Histopathology Core, the Microscopy Laboratory, and Applied Bioinformatics Laboratory. **Funding:** This work was supported by Cancer Center Support Grant P30CA016087 from the Laura and Isaac Perlmutter Cancer Center (core facility subsidies), the National Psoriasis Foundation (Early Career Research Grant to P.K.), a Bernard Levine Postdoctoral Fellowship in Immunology to P.K., the American Association of Immunologists (Intersect Fellowship to Y.X.), the Howard Hughes Medical Institute (HHMI Gilliam Fellowship GT13641 to K.P.M), a Pew Stewert Scholar Award (00034119 to S.N.), and the National Institutes of Health (NIH grants 1K22AI135099-01, 1DP2AR079173-01, and R01-AI168462 to S.N.). S.N. is a NYSCF Robertson Stem Cell Investigator and a Packard Fellow. **Author contributions:** S.N, P.K., and Y.X. conceptualized the study, designed experiments, and wrote the manuscript. P.K., Y.X. and I.S. performed animal, flow cytometry, microscopy, organoid, and genomic experiments. I.S. and A.T. performed computational analysis for CITE-seq, ST, and low-input RNA-seq. P.K. K.P.M, C.L., and A.H. performed ST. B.H. and K.P.M aided in animal studies and colony maintenance. D.E.B. assisted with the seahorse glycolysis stress test. M.C. aided in fluorescent image acquisition and analysis. S.B.L. assisted with rapamycin studies. A.N.T. assisted with low-input RNA-seq. D.L and N.X.L provided human acute wound samples. **Competing interests:** S.N. is on the SAB of Seed Inc., is a consultant for BiomX, and receives funding from Takeda Pharmaceuticals. A.T. is a scientific advisor to Intelligencia.ai. The remaining authors declare no competing interests. **Data and materials availability:** All data to support the conclusions in this manuscript can be found in the main text or the supplemental materials. Genomic data are publicly available at the GEO (accession number GSE166950). Raw data for in vivo measurements of wound repair and image quantifications from our study may be requested from the corresponding authors.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg9302
Figs. S1 to S14
Tables S1 to S4
MDAR Reproducibility Checklist

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

Submitted 4 February 2021; resubmitted 19 October 2021
Accepted 31 May 2022
10.1126/science.abg9302

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Connectomic comparison of mouse and human cortex

Sahil Loomba†, Jakob Straehle†, Vijayan Gangadharan†, Natalie Heike†, Abdelrahman Khalifa†, Alessandro Motta†, Niansheng Ju, Meike Sievers, Jens Gempt, Hanno S. Meyer, Moritz Helmstaedter*

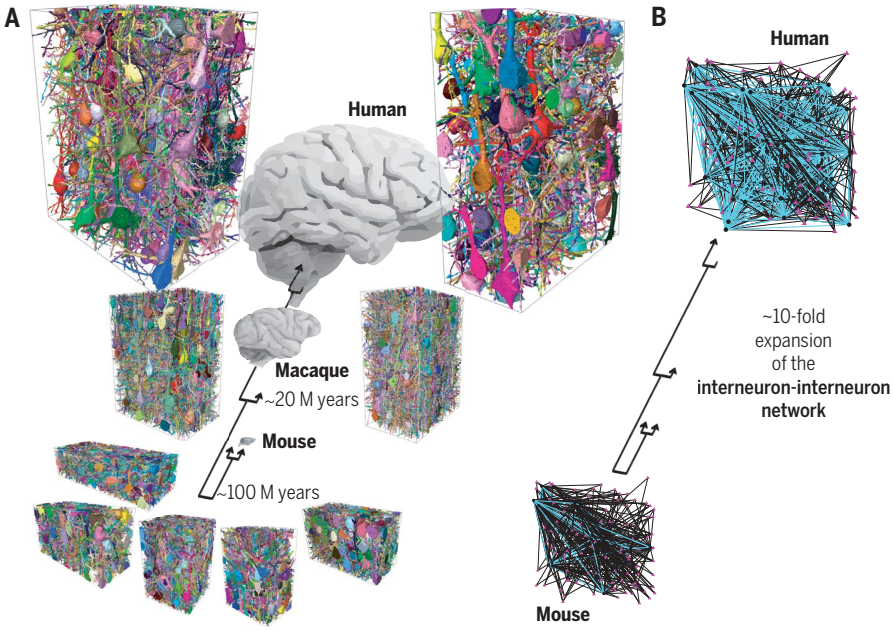
INTRODUCTION: The analysis of the human brain is a central goal of neuroscience, but for methodological reasons, research has focused on model organisms, the mouse in particular. Because substantial homology was found at the level of ion channels, transcriptional programs, and basic neuronal types, a strong similarity of neuronal circuits across species has also been assumed. However, a rigorous test of the configuration of local neuronal circuitry in mouse versus human—in particular, in the gray matter of the cerebral cortex—is missing.

The about 1000-fold increase in number of neurons is the most obvious evolutionary change of neuronal network properties from mouse to human. Whether the structure of the local cortical circuitry has changed as well is, however, unclear. Recent data from transcriptomic analyses has indicated an increase in the proportion of inhibitory interneurons from mouse to human. But what the effect of such a change is on the circuit configurations found in the human cerebral cortex is not known. This is, however, of particular interest also to the study of neuropsychiatric disorders because in these, the alteration of inhibitory-to-excitatory synaptic balance has been identified as one possible mechanistic underpinning.

RATIONALE: We used recent methodological improvements in connectomics to acquire data from one macaque and two human individuals, using biopsies of the temporal, parietal, and frontal cortex. Human tissue was obtained from neurosurgical interventions related to tumor removal, in which access path tissue was harvested that was not primarily affected by the underlying disease. A key concern in the analysis of human patient tissue has been the relation to epilepsy surgery, when the underlying disease has required often year-long treatment with pharmaceuticals, plausibly altering synaptic connectivity. Therefore, the analysis of nonepileptic surgery tissue seemed of particular importance. We also included data from one macaque individual, who was not known to have any brain-related pathology.

RESULTS: We acquired three-dimensional electron microscopy data from temporal and frontal cortex of human and temporal and parietal cortex of macaque. From these, we obtained connectomic reconstructions and compared these with five connectomes from mouse cortex. On the basis of these data, we were able to determine the effect of the about 2.5-fold expansion of the interneuron pool in macaque and human cortex compared with that of mouse. Contrary to expectation, the inhibitory-to-excitatory synaptic balance on pyramidal neurons in macaque and human cortex was not substantially altered. Rather, the interneuron pool was selectively expanded for bipolar-type interneurons, which prefer the innervation of other interneurons, and which further increased their preference for interneuron innervation from mouse to human. These changes were each multifold, yielding in effect an about 10-fold expanded interneuron-to-interneuron network in the human cortex that is only sparsely present in mouse. The total amount of synaptic input to pyramidal neurons, however, did not change according to the threefold thickening of the cortex; rather, a modest increase from about 12,000 synaptic inputs in mouse to about 15,000 in human was found.

CONCLUSION: The principal cells of the cerebral cortex, pyramidal neurons, maintain almost constant inhibitory-to-excitatory input balance and total synaptic input across 100 million years of evolutionary divergence, which is particularly noteworthy with the concomitant 1000-fold expansion of the neuronal network size and the 2.5-fold increase of inhibitory interneurons from mouse to human. Rather, the key network change from mouse to human is an expansion of almost an order of magnitude of an interneuron-to-interneuron network that is virtually absent in mouse but constitutes a substantial part of the human cortical network. Whether this new network is primarily created through the expansion of existing neuronal types, or is related to the creation of new interneuron subtypes, requires further study. The discovery of this network component in human cortex encourages detailed analysis of its function in health and disease. ■



Connectomic screening across mammalian species: Comparison of five mouse, two macaque, and two human connectomic datasets from the cerebral cortex. (**A**) Automated reconstructions of all neurons with their cell bodies in the volume shown, using random colors. The analyzed connectomes comprised a total of ~1.6 million synapses. Arrows indicate evolutionary divergence: the last common ancestor between human and mouse, approximately 100 million years ago, and the last common ancestor between human and macaque, about 20 million years ago. (**B**) Illustration of the about 10-fold expansion of the interneuron-to-interneuron network from mouse to human.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mh@brain.mpg.de
†These authors contributed equally to this work.
Cite this article as S. Loomba *et al.*, *Science* **377**, eabo0924 (2022). DOI: 10.1126/science.abo0924

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abo0924

RESEARCH ARTICLE

NEUROSCIENCE

Connectomic comparison of mouse and human cortex

Sahil Loomba[1,2]†, Jakob Straehle[1]†‡§, Vijayan Gangadharan[1]†¶, Natalie Heike[1]†, Abdelrahman Khalifa[1]†, Alessandro Motta[1]†, Niansheng Ju[1], Meike Sievers[1,2], Jens Gempt[3], Hanno S. Meyer[3], Moritz Helmstaedter[1]*

The human cerebral cortex houses 1000 times more neurons than that of the cerebral cortex of a mouse, but the possible differences in synaptic circuits between these species are still poorly understood. We used three-dimensional electron microscopy of mouse, macaque, and human cortical samples to study their cell type composition and synaptic circuit architecture. The 2.5-fold increase in interneurons in humans compared with mice was compensated by a change in axonal connection probabilities and therefore did not yield a commensurate increase in inhibitory-versus-excitatory synaptic input balance on human pyramidal cells. Rather, increased inhibition created an expanded interneuron-to-interneuron network, driven by an expansion of interneuron-targeting interneuron types and an increase in their synaptic selectivity for interneuron innervation. These constitute key neuronal network alterations in the human cortex.

The human brain, with its 86 billion nerve cells (*1*) forming a network of unparalleled complexity, is of special interest for neuroscience. Yet over the past 50 years, rodents (in particular, the mouse) have emerged as key research subjects, offering methodological opportunities not available for the study of the human brain. Because at the molecular level evolutionary homology is substantial for ion channels, synaptic receptors, and other key molecular constituents of the brain, similar homology has been assumed for neuronal circuits, especially in the cerebral cortex. However, comparative synaptic-resolution connectomic studies of mouse and human cortex are required to determine the degree to which circuit structure may have evolved between species.

Detailed studies of the human cerebral cortex have provided data on cellular composition of the neuropil (*1*–*14*), synaptic properties (*15*–*25*), and neuronal morphology (*14*, *26*–*28*), yielding a comparative description with reference to rodents that indicates a larger number of glial cells (*2*), larger synapses (*20*, *29*, *30*), and more inhibitory interneurons (INs) (*4*, *9*) to form the neuronal network of the human cortex. This latter comparison had been confounded by rodent data reporting an IN fraction of 8 to 25%, (*3*, *4*, *31*–*37*), thus potentially on par with data from primates (range 15 to 37%) (*38*–*42*). Recent transcriptomic data substantiates an at least twofold increase (*4*, *9*), suggesting the balance between inhibitory and excitatory synapses to be substantially shifted toward inhibition. Furthermore, on the basis of the about threefold larger extent of the dendritic trees of human pyramidal cells, it has been assumed that human cortical neurons receive substantially more synapses than those of rodents (for example, 10,000 in rodents versus 30,000 in human) (*29*, *43*–*46*).

However, a circuit-level analysis of the human cortex that addresses the potential effect of multifold increased inhibitory circuit elements is still missing. Would, as a result, the inhibitory-to-excitatory synaptic balance be increased in the human cortex? Because inhibitory-to-excitatory synaptic balance has emerged as a key set point that might be altered in neuropsychiatric diseases [studied primarily in mouse models (*47*, *48*)], this question is relevant for the healthy and diseased human brain.

We used three-dimensional (3D) electron microscopy (EM) followed by sparse and dense circuit reconstruction (*49*) to map the synaptic and cellular composition of layer 2/3 (L2/3) in mouse, macaque, and human cortex. Sampling from multiple individuals and cortical regions, we screened for key connectomic alterations between mouse and human cortex, which a single dataset from a given human individual and disease condition would not have allowed (*10*). Although our human data was from diseased individuals who underwent neurosurgical interventions, the patients had not undergone yearlong treatment for epileptic disorders that could have substantially altered the synaptic network. By including additional data from macaque in our study, we furthermore could ensure that none of the reported effects was exclusively attributable to a diseased state of the human brain. We determined the circuit effects of an expanded inhibitory neuron population in the human cortex and obtained a quantitative picture of the human pyramidal cell synaptic input architecture.

Results

We obtained tissue samples (Fig. 1) from the cerebral cortex of two human individuals (a 69-year-old female and an 80-year-old male) who underwent neurosurgical operations (access tissue that had to be removed during surgery and would have been discarded otherwise) (Materials and methods) and of one rhesus macaque (Macaca mulatta, 15.6-year-old male). Individual ages corresponded to 83, 102, and 52% of statistical life expectancy, respectively.

From these tissue samples, we acquired 3D EM image datasets in cortical L2/3 using either serial blockface EM (SBEM) (*50*) [macaque primary somatosensory cortex (S1) dataset sized 176 by 224 by 100 μm; superior temporal gyrus (STG) sized 175 by 225 by 108 μm; human (H5) STG sized 166 by 216 by 112 μm, and human (H6) inferior frontal gyrus (IFG) sized 170 by 215 by 79 μm, all at a voxel size of 11.24 by 11.24 by (28 to 30) nm] or automated tape-collecting ultramicrotome multiple scanning EM (ATUM-multiSEM) (*51*–*53*) [dataset human (H5) STG spanning all cortical layers, sized 1.7 mm by 2.1 mm by 28.3 μm and dataset from the same sample focused on L2/3 sized 1.1 mm by 1.1 mm by 82 μm (total 0.41 PB of data) both at a voxel size of 4 by 4 by (35 to 40) nm]. For comparison with mouse cortex, we densely reconstructed (*49*, *54*) previously published 3D EM datasets from L2/3 of S1, secondary visual (V2), posterior parietal (PPC), and anterior cingulate (ACC) cortical areas (*55*) and acquired an additional dataset from mouse secondary auditory cortex (A2) sized 115 by 175 by 109 μm, approximately corresponding to the location of the STG datasets from macaque and human. In the following, we report effects that were multifold between mouse (n = 5 individuals, n = 5 cortical regions) and macaque or human (n = 3 individuals, n = 4 cortical regions).

Inhibitory versus excitatory neuronal composition

The fraction of nerve cells that were INs increased 2.5-fold between mouse and macaque or human [n = 88 of 734 versus 278 of 914 neurons from n = 5 mouse and n = 4 macaque and human datasets pooled, $P < 0.001$ (Materials and methods)] (Fig. 2, A and B), which is consistent with recent reports from transcriptomic cell type analyses (*4*, *9*). This change in the neuronal composition of cortical tissue could have profound effects on the synaptic input to pyramidal (excitatory) neurons (ExNs)

[1]Department of Connectomics, Max Planck Institute for Brain Research, Frankfurt, Germany. [2]Faculty of Science, Radboud University, Nijmegen, Netherlands. [3]Department of Neurosurgery, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Germany.
*Corresponding author. Email: mh@brain.mpg.de
†These authors contributed equally to this work.
‡Present address: Department of Neurosurgery, University Freiburg, Freiburg, Germany.
§Present address: Berta-Ottenstein Clinician Scientist Program, Faculty of Medicine, University Freiburg, Freiburg, Germany.
¶Present address: Institute for Vascular and Islet Cell Biology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

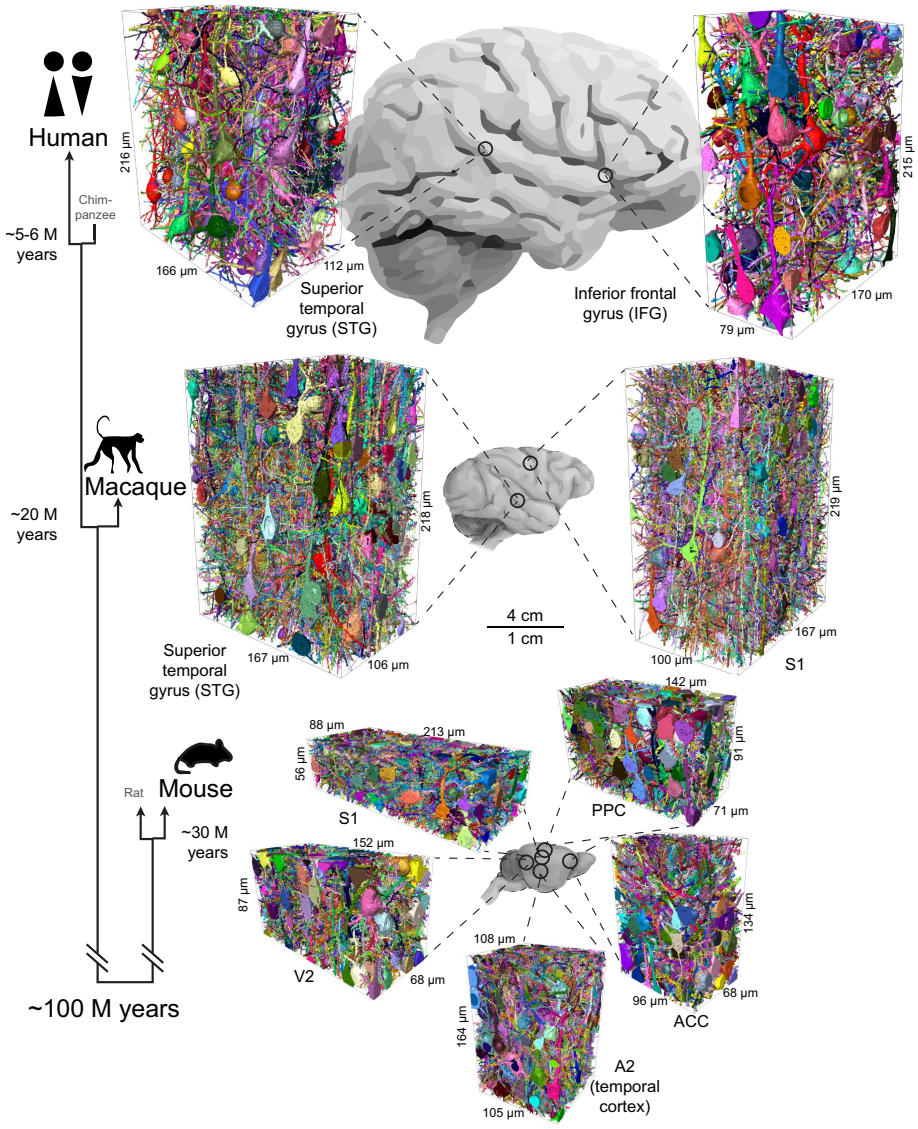


Fig. 1. Comparative connectomic analysis of mouse, macaque, and human cortex. Dense connectomic reconstructions from L2/3 of five cortical areas of mouse (bottom, n = 5 individuals) and from four cortical areas of macaque and human (n = 3 individuals). There are matched cortical areas (A2 and STG) across all three species and paired samples from S1 (mouse and macaque). A total of 202,954 axons and 1,618,129 synapses were analyzed (Materials and methods). The raw 3D EM data of mouse datasets S1, V2, PPC, and ACC were previously published (*55*), but their dense reconstruction have not. (Left) Simplified phylogenetic tree [based on (*100*)] indicating time to last common ancestor between human (*Homo sapiens*), rhesus macaque (*Macaca mulatta*), and mouse (*Mus musculus*). Scale bars apply to the brain sketches. S1, primary somatosensory cortex; A2, secondary auditory cortex; V2, secondary visual cortex; PPC, posterior parietal cortex; ACC, anterior cingulate cortex; STG, superior temporal gyrus; IFG, inferior frontal gyrus.

(Fig. 2C): If all other circuit properties were unchanged between mouse and human, the inhibitory-to-excitatory balance (i/e balance) would also be 2.5-fold shifted toward inhibition compared with that in mouse.

Synaptic inputs to pyramidal cell dendrites

We therefore first analyzed the synaptic input to pyramidal cell dendrites in mouse, macaque, and human (Fig. 2, D to H). Although the density of synapses onto the dendritic shaft of pyramidal cells remained largely constant from mouse to macaque and only slightly increased in human [0.16 ± 0.13 per micrometer of dendritic shaft length versus 0.17 ± 0.11 and 0.21 ± 0.08; n = 46, 36, and 49; P = 0.6 and P = 0.003 for mouse to macaque and mouse to human, respectively; Kolmogorov–Smirnov (KS) test; mean ± SD] (Fig. 2, D and E), synapses onto dendritic spines were 2.7-fold sparser in macaque and human [1.99 ± 0.65 per micrometer of dendritic shaft length (mouse) versus 0.86 ± 0.36 (macaque) and 0.67 ± 0.25 (human); mean ± SD; n = 46, 36, and 49, respectively; $P < 10^{-18}$, KS test mouse versus macaque and human] (Fig. 2, D and E). As a result, the fraction of synapses made onto the dendritic shafts of pyramidal cells (out of all input synapses to these dendrites) was increased 2.7-fold (n = 1111 synapses, mouse and n = 1638 synapses macaque and human, P < 0.001, KS test) (Fig. 2, F and H). Synaptic inputs to spine necks, stubs, or double innervation of dendritic spines were rare but slightly increased in macaque and human (Fig. 2G; statistics are provided in the caption).

In mouse, synapses onto dendritic spines have been reported to be predominantly excitatory (*49*, *56*) and synapses onto dendritic shafts predominantly inhibitory (*55*, *57*, *58*). Therefore, the finding of a 2.7-fold increased fraction of shaft synapses could indicate a substantial shift in the set point for the i/e balance in human [as recently claimed for nonhuman primate (*59*)]. For this to be correct, however, the properties of excitatory and inhibitory axons—in particular, their preference for establishing synapses onto dendritic shafts and spines—would have to remain unaltered between mouse and macaque and human. We investigated this next (Figs. 3 and 4).

Synaptic properties of excitatory and inhibitory neurons

We first reconstructed axons of pyramidal cells and INs in mouse and human (Fig. 3, A to C) to determine their synaptic preferences. The rate of spine innervation showed the well-documented dichotomy in mouse (Fig. 3, A and B) (*49*, *55*, *60*): Pyramidal cell axons targeted dendritic spines with 61.47 ± 9.26% (mean ± SD) of their output synapses (n = 11 axons, n = 367 output synapses), whereas INs only very rarely made an output synapses as the only synapse onto a dendritic spine (n = 1 of 263 IN output synapses in mouse), with no overlap between the populations. In human, however, axons reconstructed from the cell bodies of pyramidal cells made only 30.37 ± 16.16% (mean ± SD) of their output synapses onto dendritic spines (n = 15 axons, n = 1126 output synapses), whereas INs maintained their almost complete exclusion of single spine innervation (Fig. 3, B and C) [n = 12 axons, n = 345 output synapses, of which n = 1 were single spine synapses; IN axons did, however, innervate dendritic spines as double innervations, in all species, as reported for subtypes of INs (fig. S3) (*61*–*63*); only the single spine innervation rate was used for the following analyses]. Thus, the distributions of synaptic target properties between pyramidal and IN axons reconstructed from their cell body of origin were less separated in human (Fig. 3B) and indicated a change in axonal synaptic preference from mouse to human.

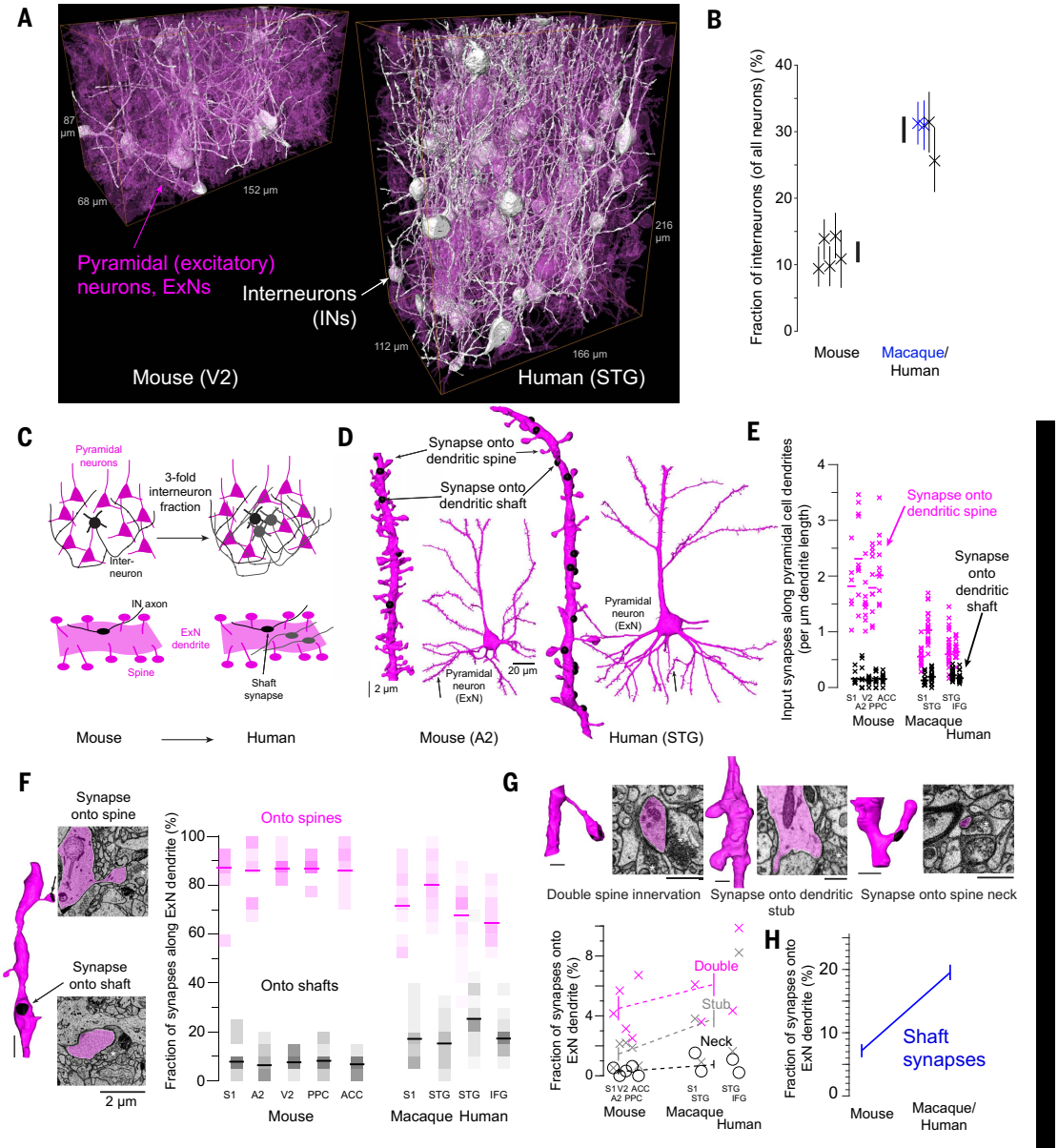


Fig. 2. Multifold changes in cellular composition and synaptic input of mouse versus macaque and human cortex. (**A** and **B**) Reconstruction of ExNs (magenta) and inhibitory INs (white) shows 2.54-fold expansion of the IN pool from mouse to macaque and human, which (**C**) would predict a similar-fold increase in the inhibitory input onto ExNs, which would substantially alter the set point for the inhibitory versus excitatory synaptic input balance in human compared with mouse (illustrated for threefold increase). (**D** to **F**) Mapping of (D) the synaptic input to excitatory neuron dendrites showed (E) an unaltered rate of shaft input synapses but a 2.65-fold decrease of spine input synapses, resulting in (F) a 2.72-fold increase in the fraction of input synapses made onto dendritic shafts. (**G**) Other dendritic inputs were rare but slightly increased from mouse to macaque and human: doubly innervated spines, 4.49 ± 0.01 versus 4.68 ± 0.01 and 6.91 ± 0.01%; stub synapses, 1.44 ± 0.00 versus 2.17 ± 0.01 and 4.71 ± 0.01%; spine neck innervations, 0.26 ± 0.00 versus 0.84 ± 0.00 and 0.67 ± 0.00%; n = 1111, 598, and 1040 total synapses, respectively. [(F) and (G)] (Insets) EM images of example synapses from (F) human STG multiSEM-imaged and (G) (left to right) macaque STG, human IFG, and macaque S1, respectively. (**H**) Concomitant increase of shaft synapse input could support the altered i/e balance model in (C), if axonal properties remain unchanged from mouse to macaque and human (analyzed in Figs. 3 and 4). Scale bars, 1 µm unless indicated otherwise. Data in (A) are from automated reconstructions, and data in (B) and (D) to (H) are from expert reconstructions.

In the dense cortical neuropil, only a fraction of axons originates from neurons whose cell body is located in close proximity, whereas most have more distal origins. To study the target properties of all axons in the neuropil, also those without their cell body of origin in the dataset, we next analyzed the dense population of axons in mouse versus macaque and human that we obtained from the dense automated reconstructions (n = 202,954 axons total with n = 1,618,129 output synapses) (Fig. 3, D to F). Mouse axons predominantly innervated spines (Fig. 3E), as expected from the soma-based reconstructions (Fig. 3B) and the small fraction of INs (Fig. 2B). Densely sampled macaque and human axons, however, did not reproduce the soma-based axonal properties but showed a broader distribution of spine innervation from 0 to 80% spine preference (here and in the following, spine preference implies fraction of an axon's synapses established as single spine innervations, not considering doubly innervated spines).

What could yield this difference in axonal properties between soma-based and dense reconstructions in macaque and human? In particular, we needed to understand the origin of axons with >40% spine targeting in macaque and human (Fig. 3E) that we had rarely observed in the soma-based reconstructions (Fig. 3B). Soma-based reconstructions in datasets smaller than the complete axonal arbor bias axonal reconstructions toward local (not distally located) neurons and to the proximal parts of these neurons' axons. When analyzing the distance-dependence of axonal targeting properties for pyramidal axons from human (Fig. 3, G to I), we found evidence for a substantial change from proximal axonal preference of shaft innervation (*64*–*66*) to distal preference for spine innervation along the axon's path [see (*67*) for a first report of such a path length–dependent synaptic sorting phenomenon along axons in the mammalian cerebral cortex, and (*68*, *69*) for earlier reports on the bird brain]. We had to take this effect (Fig. 3, B, E, and I) into account when comparing the axonal properties in these species (Fig. 4).

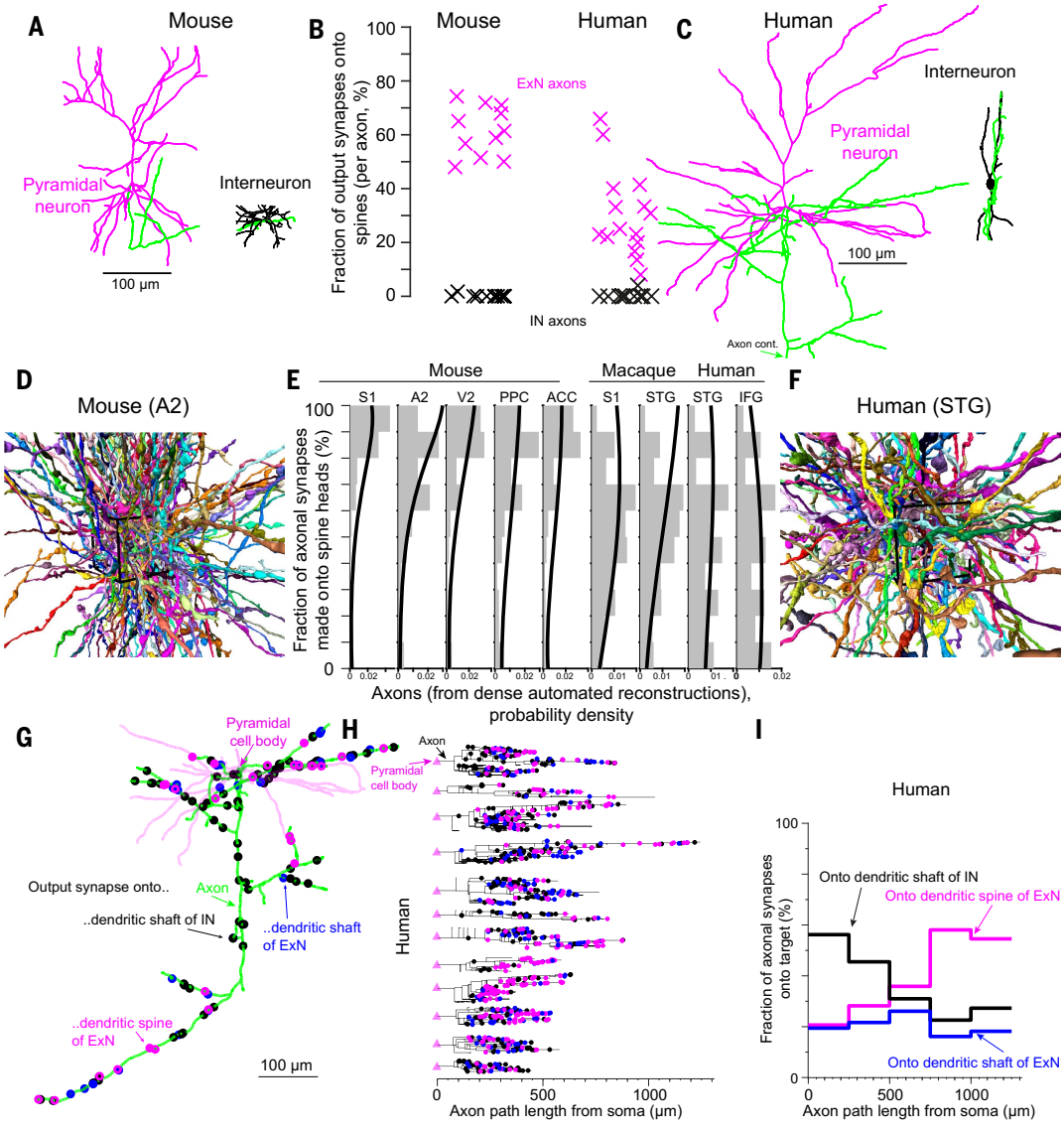


Fig. 3. Synaptic target properties of excitatory and inhibitory neurons in mouse and human. (**A** to **C**) Axonal target properties from identified pyramidal cells and INs (n = 50 neurons, n = 2101 output synapses, cross-validated expert reconstructions). All reconstructions are available at https://wklink.org/7881 and https://wklink.org/3103 (mouse), and https://wklink.org/9448 and https://wklink.org/2204 (human). (B) Absence of single spine innervation for INs in mouse and human, but shift in spine innervation of excitatory axons from mouse to human, yields a less separated synaptic preference of excitatory versus inhibitory axons. (**D** to **F**) Axonal target properties from dense axonal reconstructions in mouse, macaque, and human. Examples of dense axon populations from (D) mouse A2 and (F) human STG, all axons traversing boxes of 10 by 10 by 10 µm size shown (n = 120 and n = 89, respectively). Quantifications in (E) are based on all axons in the respective datasets (n = 202,954 axons, with 5 to 25 output synapses each; n = 1,618,129 synapses). Gray indicates distribution of spine target fraction for all axons in the datasets, uncorrected. Black lines indicate average likelihood function of true spine target fraction under consideration of error rates (Materials and methods). There is a broadening of axon target property distribution toward lower spine targeting in macaque and human compared with mouse that is, however, less pronounced than in (B) soma-based axon reconstructions. (**G** to **I**) Path length–dependent axonal synapse sorting (PLASS) as a possible origin of broader axonal target property distributions in human. (G) Example of pyramidal cell axon from human STG [path length, 3.74 mm, n = 132 output synapses; same neuron as shown in (C)]. Colors indicate synaptic target of axonal output synapses. [(H) and (I)] Distribution of axonal output synapses along pyramidal cell axons from human STG (n = 15 axons, 1126 output synapses, expert reconstructions, n = 12 reconstructions from mSEM dataset shown). All axons from mSEM and SBEM datasets are available at https://wklink.org/9448 and https://wklink.org/2204. Synapse symbols are as in (G). (I) Substantial increase in targeting of ExN spines over axonal path length. Data in (A) to (C) and (G) to (I) are from expert reconstructions, and data from (D) to (F) are from dense reconstructions.

Excitatory versus inhibitory synaptic input balance

We first identified synapses onto a dendritic shaft of distal pyramidal cell dendrites and reconstructed the presynaptic axon from these synapses ("shaft-seeded" axons). Then we determined these axons' synaptic target properties on the basis of their other output synapses (Fig. 4A). Whereas in mouse, as expected, axons were clearly identifiable as excitatory versus inhibitory on the basis of their dichotomy in spine targeting preference (Fig. 4A), the distributions of axonal properties were much less distinct in human, recapitulating the properties of densely reconstructed axons in macaque and human (Fig. 4A; compared with Fig. 3, B and E). Because this data did not allow the simple threshold-based classification of axons into inhibitory versus excitatory that can be used in mouse, we needed a more rigorous approach for axon type determination (Fig. 4, C to E).

We used the notion that the various types of axons and dendrites and their synaptic output and input properties in dense cortical neuropil are ultimately constrained by the synaptic composition of this neuropil volume (Fig. 4B). The volumetric density of synapses onto dendritic spines in the neuropil, for example, has to be composed by the intermixing of the spine preference of the various axon types present in the neuropil and similarly for other types of synapses. Because dense 3D EM data allowed us to concomitantly measure the properties of extended stretches of axons and dendrites together with the dense volumetric synaptic composition of the neuropil, we could determine the occurrence and target properties of excitatory and inhibitory axons without having to make prior assumptions about their relative prevalence (Materials and methods).

We built a model in which the synaptic properties of excitatory and inhibitory axons

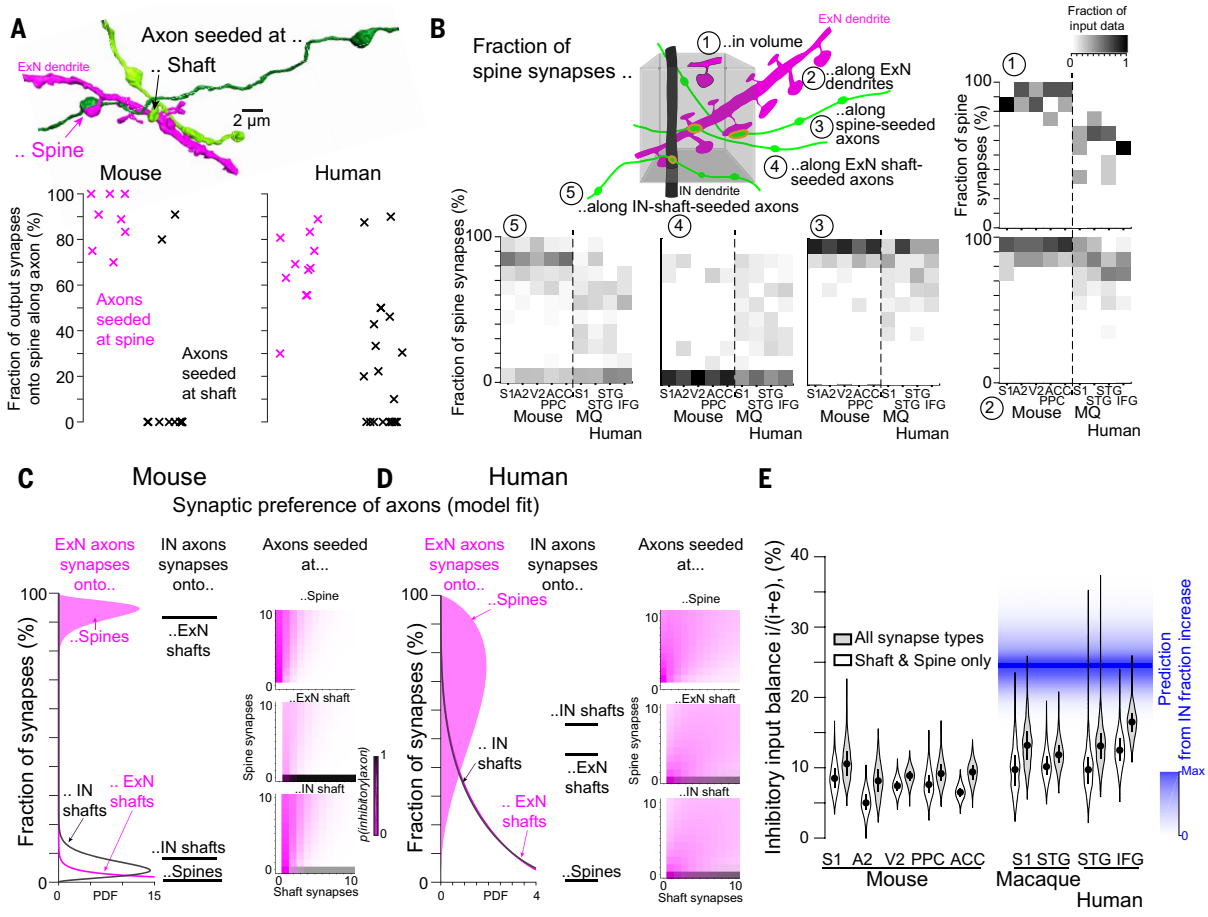


Fig. 4. Detailed analysis of i/e balance onto ExN dendrites. (**A**) Target properties of axons seeded at dendritic spines and shafts of ExN dendrites in mouse and human. (Top) Example reconstruction (macaque STG). (Bottom) Data from expert reconstructions in mouse S1 and human STG (n = 53 axons, n = 626 output synapses) showing shift of axonal targeting analogous to densely reconstructed axons in mouse versus human (compared with Fig. 3E). (**B** to **D**) Model incorporating dense volumetric synapse densities together with dendritic and axonal targeting properties, but not using any assumption about excitatory versus inhibitory synapses or axons. (B) Input data from expert annotations (n = 754 axons, n = 11,308 synapses total) that [(C) and (D)] fully constrain the model (Results and Materials and methods) (model validation is provided in fig. S2). Shaded magenta curve indicates distribution of spine targeting by ExN axons; magenta line indicates ExN shaft targeting by ExN axons; and black line indicates IN shaft targeting by ExN axons. Shown is a broadening of ExN axons' spine targeting fraction and shift of excitatory and inhibitory shaft targeting from mouse to human. MQ, macaque. (**E**) Resulting estimates of inhibitory input fraction [i/(i+e) balance] onto ExN dendrites in mouse, macaque, and human. Violin plots indicate expected inhibitory input synapse fraction along ExN dendrites (distribution across n = 1000 bootstrap samples per dataset). Open shading indicates only shaft and single spine inputs considered; gray shading indicates that multiply innervated spines and other inputs are included (Fig. 2G). Synaptic input balance does not approach the inhibitory bias predicted by the increased fraction of INs in macaque and human. Blue shading indicates prediction from mouse to macaque and human (24.9 ± 3.2%, mean ± SD; P < 0.001, by bootstrap sampling) (Materials and methods). All data are from expert reconstructions.

and their relative prevalence were determined (Fig. 4, C and D): Inhibitory axons were modeled with a multinomial distribution based on point estimates for their synaptic target preferences, whereas for excitatory axons, a distribution of target preference was modeled with a Dirichlet-multinomial to account for the effects of broadened synaptic preference shown in Fig. 3 (Materials and methods). The model was fully constrained by the following input data, which we obtained from carefully curated expert reconstructions in all datasets (n = 11,308 synapses annotated) (Fig. 4B): (i) volumetric fraction of spine synapses, (ii) fraction of input synapses onto spines along spiny (ExN) dendrites, (iii) fraction of output synapses onto spines along axons seeded from ExN dendritic spines, (iv) fraction of output synapses onto spines along axons seeded from ExN dendritic shafts, and (v) fraction of output synapses onto spines along axons seeded from IN (smooth) dendritic shafts. These input data did not require a pre hoc classification of the synapses or axons as inhibitory versus excitatory and did not contain the axons reconstructed from identified cell bodies (Fig. 3, A to C). With this input data, we obtained fits of the synaptic target properties of excitatory and inhibitory axons in each dataset (Fig. 4, C and D, shown for mouse S1 and human STG datasets): point estimates for the fraction of synapses made onto ExN spines, ExN shafts, and IN shafts for inhibitory axons (Fig. 4, C and D) and distributions of these target fractions for excitatory axons. Together with the modeled relative prevalence of inhibitory versus excitatory synapses, for any axonal stretch with any combination of x synapses onto shaft and y synapses onto spine, we obtained a probability of this axon to be excitatory versus inhibitory (Fig. 4, C and D, right). For validation of the model, we used the soma-based axon reconstructions (Fig. 3, A to C) for which we had certainty about their excitatory versus inhibitory property, sampled local stretches

from these ground truth axons, and asked the model to predict their excitatory versus inhibitory character (E→I and I→E misclassification rates: 0 and 0% for mouse, 6.9 and 8.4% for primates, respectively) (fig. S2 and supplementary materials, materials and methods).

We then applied the model to determine the expected inhibitory versus excitatory synaptic properties of ExN dendrites in all datasets (bootstrapped to account for sample size in input data and controlled for initial conditions; Materials and methods) (Fig. 4E). We found that the inhibitory input balance increased only moderately from 9.3 ± 0.8% (mean ± SD) in mouse to 13.8 ± 1.4% (P = 0.001, bootstrap sampling) in macaque and human, revoking a setting in which the inhibitory-to-excitatory input balance were to change in proportion to the 2.5-fold increase in IN fraction from mouse to macaque and human (Fig. 4E; compared with Fig. 3C) (P = 0.003; when also considering all other types of input synapses, P < 0.001; bootstrap sampling; Materials and methods). Rather, the increased fraction of shaft input synapses coincided with a change in axonal targeting properties: Excitatory axons made 0.7 ± 0.3% of their synapses onto ExN shafts in mouse, but 12.0 ± 1.1% in macaque and human (P < 0.001, bootstrap), and their IN shaft targeting changed from 4.4 ± 0.7% to 17.3 ± 1.7% (P < 0.001, bootstrap), which is consistent with the automated axon reconstructions.

This lack of enhanced inhibition onto ExN dendrites was also found for the proximal input domains of pyramidal cells [axon initial segment (AIS), soma, and proximal dendrites] (fig. S1), excluding the possibility that inhibitory synapses had been redistributed toward the perisomatic domains.

These data yielded the question of where the expanded inhibitory population in human is establishing its synapses.

Properties of the expanded inhibitory neuronal network in human

To determine the inhibitory network properties in mouse versus human, we first applied a simple IN classification based on the configuration of IN dendrites as multipolar (MP) versus non-MP [which included bipolar, bitufted, and vertically oriented dendritic trees (Fig. 5, A to D), labeled as "bipolar" (BP) for simplicity]. The IN pool changed from dominated by MP INs in mouse (about 70 to 82% MP versus 18 to 30% BP, n = 52 versus 17) (Fig. 5D) to a majority of BP INs in macaque and human (53% BP; n = 122 MP versus n = 135 BP, P < 0.01; a test is described in Materials and methods) (Fig. 5D). When mapping the synaptic target properties of these classes of INs in mouse versus human (Fig. 5, E and F), we found a fraction of BP INs with almost exclusive innervation of other INs in human that was not found in mouse. IN targeting by BP IN axons increased from 56% (Fig. 5, F and G) (70 IN targets out of 126 synapses, n = 4 axons) to 72% (Fig. 5G) (225 IN targets out of 314 synapses, n = 9 axons, P = 0.018) (Materials and methods) (*63*). The changed IN pool composition together with the changed distribution of IN targeting properties would predict a substantial increase of IN-preferring inhibitory axons in the dense neuropil of human versus mouse (Fig. 5F). When analyzing the target preferences of inhibitory axons for smooth (IN) dendrites in mouse versus macaque and human from the dense automated reconstructions (Fig. 5H), we found a substantial shift toward IN-to-IN connectivity that can account for the inhibitory synapses contributed by the expanded IN pool in macaque and human (average smooth dendrite targeting probability of inhibitory axons, 8.0 ± 15.0% in mouse versus 21.4 ± 29.0% in human; n = 6565 versus n = 2048 axons; $P = 2.2 \times 10^{-104}$, one-sided KS test). Together with the increased IN fraction (Fig. 2B), this would estimate the IN-to-IN network to expand 6.7-fold from mouse to human. On the basis of the model (Fig. 4), we estimated the expansion to be 8.6-fold (from 1.0 ± 0.2% IN-IN connectivity in mouse to 8.6 ± 1.4% in macaque and human) (fig. S2F and supplementary materials, materials and methods).

To better understand the contributions of types of INs to this enhanced network, we then analyzed the synaptic input balance onto IN dendrites in MP and BP INs in mouse, macaque, and human (Fig. 5, I to K). To our surprise, already in mouse, MP INs and BP INs have different inhibitory input balance (Fig. 5, I and K): Whereas MP INs receive 7.6 ± 2.5% inhibitory input (bootstrapped mean ± SD; n = 105 axons from n = 5 mouse datasets, n = 977 synapses total), BP INs receive 26.2 ± 5.0% (n = 88 axons, n = 694 synapses, P = 0.001). In macaque and human, MP INs receive inhibitory input that is commensurate to the fraction of INs in the cortex (24.2 ± 6.4%, n = 69 axons from n = 4 datasets, n = 514 synapses; P = 0.6 for scaling by IN fraction increase), whereas BP INs receive further enhanced inhibition (44.3 ± 7.8%, n = 71 axons, n = 506 synapses, P = 0.026). These data indicate differential inhibitory targeting across species and may imply separate IN-to-IN circuits to be substantially enhanced in macaque and human.

Excitatory synaptic network

Last, we wanted to determine the magnitude of the synaptic input to pyramidal cells in human cortex. With the expanded size of pyramidal cell dendritic trees, it has been assumed that human pyramidal cells receive a larger number of input synapses than that of mouse pyramidal cells (from about 10,000 to about 30,000) (*15*, *29*, *43*, *44*, *46*). Because we found a strongly reduced spine-targeting synapse density in the human cortex (Fig. 2E), we wondered what the quantitative input structure of human pyramidal cells would be (Fig. 6, A and B, and figs. S1 and S4).

We used our large-scale 3D EM datasets obtained in one human individual that spanned the entire depth of cortical gray matter to reconstruct a substantial portion of L2/3 pyramidal cells. On the basis of these reconstructions, we extrapolated to the full extent of the dendritic tree. The large dataset extent in-plane allowed us to map some pyramidal cell dendrites in their full extent from soma to dendritic tip and to use the properties of these completely mapped dendrites for estimating the total path length of pyramidal cell dendrites in human (Materials and methods). The resulting estimates of total dendritic path length (about 9 to 20 mm path length) (Fig. 6B and fig. S4) were consistent with light-microscopic reconstructions (*43*, *45*). Because we measured synaptic input density in parallel, we could exclude a compensation of lower dendritic path length by higher spine-targeting synapse density for individual pyramidal cells. Together, total synaptic input to pyramidal cells was 12,000 to 17,000 in human L2/3, which is far below an increase corresponding to the about threefold thicker cerebral cortex yielding about threefold larger pyramidal cell dendritic trees compared with those of mouse (Fig. 6B).

Discussion

The comparative analysis of mouse, macaque, and human cortical neuronal networks revealed that the most substantial changes in neuronal architecture—the increase of pyramidal cell size and the numeric expansion of the IN pool—have not resulted in the most immediate possible circuit changes: neither an increase of total synaptic input on pyramidal cells nor an overall shift of their synaptic input balance toward inhibition. Rather, we found an at least six- to eightfold expanded IN-to-IN network in the macaque and human cortex (Fig. 6C and fig. S2F). These circuit alterations point toward IN-to-IN connectivity as a key evolutionary change from mouse to primates, including human.

Dependence of spine densities on age

Our finding of an almost constant total synaptic input to human pyramidal cells when compared with those of mouse may be affected by a reported age-dependent decline of dendritic spines in cortex (*70*–*72*), which could amount to a reduction in spine density of 45 to 48% during puberty (*72*). The human samples were from individuals in the upper quarter of expected life length. However, we found low spine densities in macaque as well, at about 50% life expectancy. Similarly, rodents of 50 to 80% life expectancy show spine densities comparable with those reported here for younger mice (*55*, *67*). It is therefore unlikely that

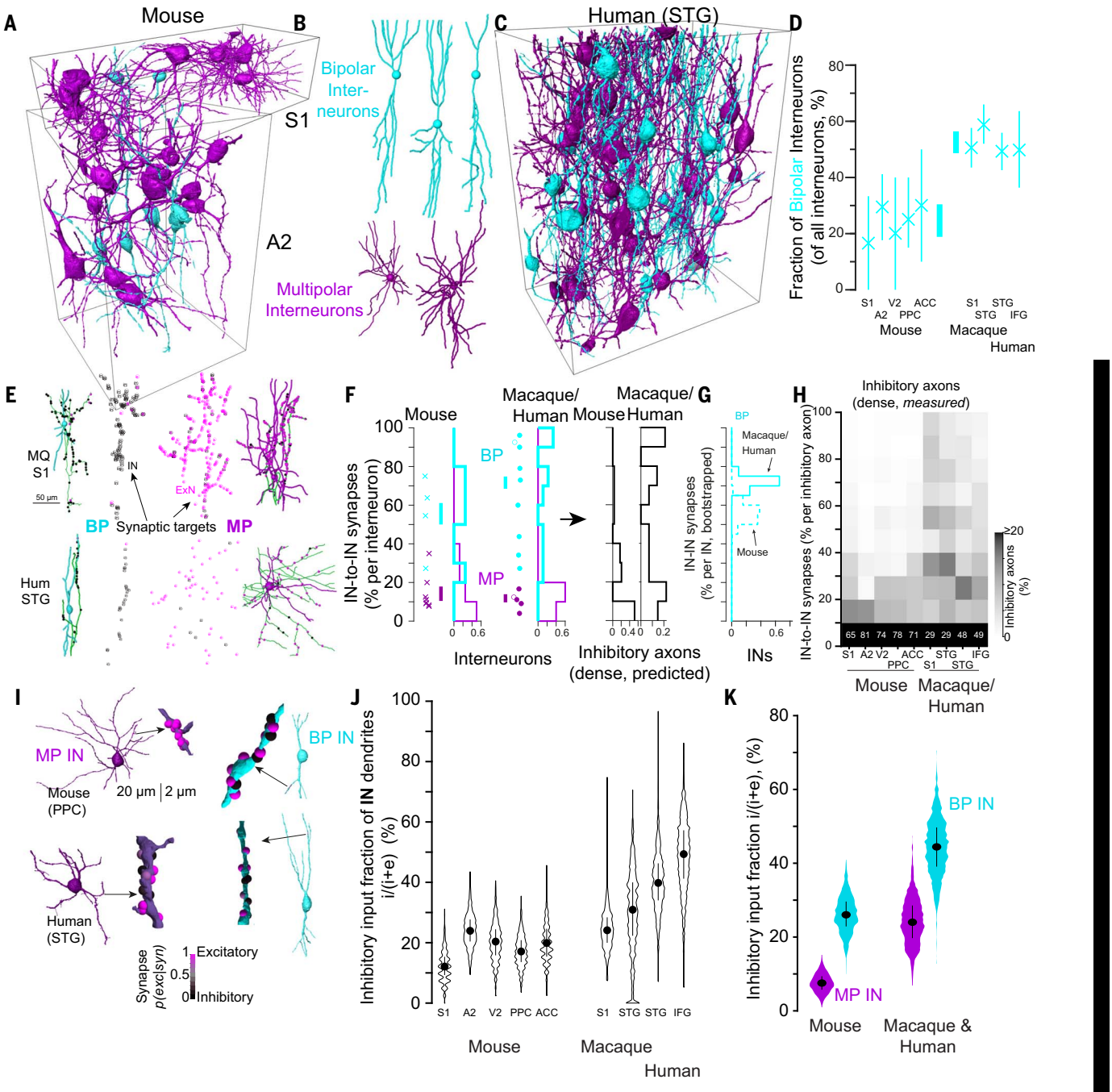


Fig. 5. Selective expansion of IN types and its effect on the inhibitory-to-inhibitory network. (**A** to **C**) Dense reconstruction of all INs in (A) mouse and (C) human with (B) labeling of MP (purple) versus BP (cyan) INs showing sparsity of BP INs in (A) mouse compared with (C) human. (**D**) Proportion of BP INs 2.3-fold expanded from mouse to macaque and human. (**E**) Synaptic targets of BP versus MP INs in macaque and human show selectivity for (left) inhibitory (93% of synaptic targets) versus (right) excitatory (87% of synaptic targets) postsynaptic partners. (**F**) Broader distribution of IN innervation selectivity in macaque and human versus mouse from soma-based axon reconstructions. This predicts expansion of inhibitory-to-inhibitory connectivity in macaque and human versus mouse (right). (**G**) Within the BP INs, IN targeting is further enhanced (from 56.63 ± 4.46% to 71.67 ± 2.48%; $P < 0.05$, bootstrapped from 126 and 314 synapses, respectively, for mouse versus macaque and human). (**H**) Analysis of densely reconstructed inhibitory axons and their targeting of IN dendrites (n = 94,391 synapses, n = 11,384 axons) by inference of the most likely smooth dendrite targeting probability under consideration of the error rates of automated synapse detection. (F) Expansion of IN-to-IN connectivity as predicted from soma-based reconstructions. (**I** to **K**) Detailed analysis of inhibitory input balance to MP versus BP IN subtypes across species. (I) Example reconstructions of input synapses onto IN dendrites. (J) Inhibitory input determined from the model in Fig. 4 for IN dendrites. The 2.35-fold increase from mouse to human (18.9 ± 2.5% to 44.5 ± 7.3%, mean ± SD) is consistent with the prediction from IN expansion for IN dendrites (P = 0.662). [(J) and (K)] Separate analysis of inhibitory input to MP versus BP INs reveals a difference in inhibitory input already in mouse (7.6 versus 26.2%; P = 0.001) that is further enhanced in macaque and human (24.2 versus 44.3%; P = 0.026). Data in (H) are from dense reconstructions; all other data are from expert reconstructions.

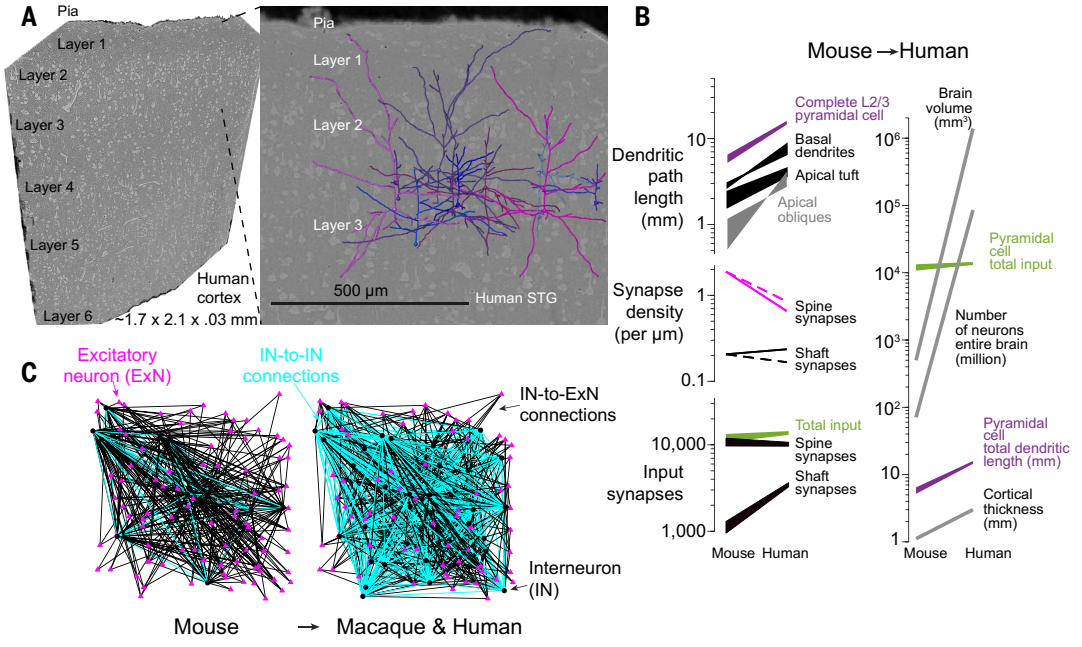


Fig. 6. Scaling of pyramidal cell synaptic input and network properties from mouse to human. (**A**) Large-scale 3D EM dataset spanning all cortical layers from human cortex (STG) for reconstruction of dendritic arbors of L2/3 pyramidal neurons. (**B**) With about threefold reduced spine density from mouse to human, the about threefold increase in pyramidal cell dendritic path length yields only moderately increased number of input synapses for human pyramidal cells compared with that in mouse. Dashed lines indicate macaque. (**C**) Expanded IN-to-IN network, illustrated according to model fit (Fig. 4; fig. S2F; and supplementary materials, materials and methods). There is an 8.6-fold expansion of IN-to-IN connectivity from mouse to macaque and human. Data in (A) and (B) are from expert reconstructions. In (B), right, lower limits of pyramidal cell total dendritic length and total input are based on (*45*); upper limits are based on data in this study. Brain volume is based on (*101*), and the number of neurons in the entire brain is based on (*1*).

age-related effects have dominated the finding of reduced spine densities in human compared with mouse. Also, although temperature-dependent changes in spine densities have been reported (*73*), these are unlikely the cause of our measured spine densities (Fig. 2E and supplementary materials) (*29*, *44*, *46*, *55*, *66*, *74*–*77*).

Synaptic strength versus synaptic number

Our data indicate the maintenance of the relative number of inhibitory versus excitatory input synapses on the dendrites of pyramidal cells in human versus mouse. The effective balance between inhibition and excitation could be altered by different strengths of synapses. Evidence for larger unitary excitatory postsynaptic potentials and for larger synapses and presynaptic vesicle pools has been found in experiments on human cortical slices (*16*, *30*, *78*, *79*). At the same time, the impact of single inhibitory synapses is enhanced (*19*, *20*). Whether synaptic number is a proper indicator of overall synaptic impact is a matter of investigation, with evidence in ferret (*80*) pointing toward a dominant effect of synapse number, as we quantified in this work.

Increased complexity of inhibitory networks

The IN-to-IN network expansion found in macaque and human could have a multitude of dynamic effects. Although the most immediate consequence could be a more evolved disinhibitory network capability—for example, for the gating of otherwise inhibited excitatory activity (*81*, *82*)—theoretical studies have also indicated possible effects on the maintenance of working memory through enhanced IN-to-IN connectivity and the ensuing network dynamics (*83*). Our data indicate that a detailed investigation of such phenomena is required for an understanding of the human cortex. In particular, alterations in IN-to-IN connectivity should become a focus of study in the context of possible pathological alterations of human cortex.

Materials and methods

Animal experiments and human tissue samples

All animal-related experimental procedures were performed according to the law of animal experimentation issued by the German Federal Government under the supervision of local ethics committees and according to the guidelines of the Max Planck Society. Experimental procedures were approved by Regierungspräsidium Darmstadt, AZ: F 126/1002 (mouse) and Regierungspräsidium Marburg AZ: V54 -19c 20 15 hours 01 MR 13/1 Nr. 78/2012 (macaque).

The human brain tissue samples were collected during neurosurgical procedures that were indicated for medical reasons and independently from this research project at the Department of Neurosurgery at the Klinikum rechts der Isar of the Technical University of Munich. They were obtained from access tissue (presumably healthy brain parenchyma that had to be removed as part of the procedure and would have been discarded otherwise) before removal of the respective target lesions, as approved by the Ethics Committee of the Technical University of Munich School of Medicine (Ethikvotum 184/16S and 273/21 S-EB). All patients had given their written informed consent.

The macaque brain tissue sample was collected at German Primate Center (DPZ) GmbH, Göttingen.

Tissue extraction and preparation

Mouse tissue was processed and imaged as described previously (*49*, *55*, *84*). These consisted of four 3D-EM datasets from layers 2/3 of mouse S1, V2, PPC and ACC sized between 72 by 93 by 141 µm and 88 by 56 by 213 µm (Fig. 1) acquired at a voxel size of 11.24 to 12 × 11.24 to 12 by 28 to 30 nm the present study densely reconstructed and reanalyzed these published datasets from mouse, together with a newly acquired dataset from mouse A2 cortex. The published mouse dataset "PPC-2" (*55*) covering layers 1 to 5 was also used for expert reconstructions.

The human and macaque samples were processed as follows. All tissue specimen were fixed by either immersion (human) or transcardial perfusion (macaque) using an EM fixative composed of 2.5% paraformaldehyde (Sigma), 1.25% glutaraldehyde (Serva) and 2 mM calcium chloride (Sigma) in 80 mM cacodylate buffer adjusted to pH 7.4 with an osmolarity ranging from 700 to 800 mOsmol/kg (*85*).

The human "H5" tissue was obtained from the right superior temporal gyrus of an 80 year old male patient during resection of a temporal mass lesion (final diagnosis: glioblastoma multiforme). After removal the sample was immersed in cold (13°C) EM fixative and transported to a nearby laboratory (transport time

about 6 min). Partially submerged in cold fixative the sample was manually trimmed along the pia-WM axis and mounted on a vibratome stage. Then, submerged in cold EM fixative, the sample was cut into 500 µm thick slices with a vibratome (Leica VT 1200S). The slices were then transported in 8 to 12°C cooled fixative and stored over night at 4°C. The next day, samples spanning the entire cortical depth and about 1.5 to 1.7 mm in width were cut out and prepared for electron microscopy as described in (*55*) with the modification that the sample was embedded in Epon Hard for sectioning in the ATUM. For this, samples were infiltrated through a graded series (3:1 for 4 hours, 1:1 12h/overnight, 1:3 for 4 hours) of aceton and Epon resin (Epon hard mixture: 5.9 g Epoxy, 2.25 g DDSA, 3.7 g NMA, 205 µl DMP; Sigma-Aldrich). Samples were then incubated in pure resin for 4 hours at room temperature, 12 hours/overnight at 4°, and another 4 to 5 hours at room temperature. Samples were directly embedded in pure resin on aluminum pins and kept in a pre-warmed oven (60°) for 2 to 3 days. Fresh resin was prepared for each incubation step.

The cured sample was trimmed into a hexagonal shape (size 3.1 by 1.8 mm) with a diamond milling head using an EM trimmer (Leica EM TRIM2, Leica Microsystems, Wetzlar, Germany). Next, the sample was cut into 35- to 40-nm-thick slices at 0.3 mm s^{-1} cutting speed using a 4 mm ultra35° knife (DiATOME, Nidau, Switzerland). The ultrathin sections were collected on plasma-treated, carbon coated Kapton tape (custom-processed) with a customized ATUM-tome (RMC Boeckeler, Tucson, USA) (*51*, *52*). The tape was then mounted on silicon wafers using double-sided adhesive carbon tape (P77819-25, Science Services GmbH, Munich, Germany). In total, 7009 slices (corresponding to an extent of 270.25 µm) were cut.

The human "H6" sample was obtained from inferior frontal gyrus from a 69 year old female patient during surgical removal of frontal mass lesion (final diagnosis: glioblastoma multiforme). Following surgical removal, tissue was directly collected in fix solution kept at 4°C. The tissue was immediately sliced into 500 µm thin slices in cold fixative using vibratome. Slices were kept at 4°C overnight. Samples were then collected using 1mm circular medical biopsy punch (covering L2/3) and prepared for SBEM as described in (*55*).

The macaque sample was acquired from a 15.6 year old right handed male animal. Transcardial perfusion was performed under anesthesia (Ketamin, Xylazin) after an additional lethal dose of Barbituate (90 to 120 mg/kg intravenously). After flushing with ~2 l of 0.15 M cacodylate buffer (pH 7.4, duration 15 min) the perfusion solution was changed to ~2 l of EM fixative. The brain was removed and the anterior half of the brain was sectioned into 5-mm-thick coronal sections. The sections were subsequently immersed in cold EM fixative, transported to the research laboratory and kept at 4°C over night. Within 48 hours the tissue was further dissected (with storage in 4°C EM fixative between processing steps). A ~5 mm wide tissue block centered on the anatomically defined arm/finger region of the left postcentral gyrus (Area 3a/b) medially adjacent to the rostral end of the intraparietal sulcus was dissected. The block was placed in 0.15 M cacodylate buffer and cut along the medio-lateral axis into 600 µm thick slices using a vibratome (Microm HM650V, Thermo Scientific). Using a medical biopsy punch (KAI medicals, USA), a 1.5-mm-wide sample spanning almost the entire thickness of the cortex was cut and subsequently prepared according to the methods described in (*55*) using an automated tissue processor (Leica EM AMW).

3D EM imaging and image alignment

SBEM datasets of the human (H5, H6), macaque and mouse (A2) samples were acquired using a custom-built SBEM microtome [(*50*) courtesy of W. Denk] mounted inside the chamber of a scanning electron microscope (FEI Verios, Thermo Fisher Scientific, USA). The image acquisition and SBEM microtome were controlled using custom written software (*55*). Focus and stigmation were adjusted using custom written auto-correction routines. Imaging parameters were as follows: 4 by 6 (macaque, H5, H6) or 3 by 4 (mouse A2) mosaic tiles of images sized 4096 by 3536 voxels with an in-plane voxel size of $(11.24\ nm)^2$ and 30 nm nominal cutting thickness at 400 ns (macaque) or 700 ns (H5, H6, A2) dwell time with a nominal beam current of 800 pA (macaque) or 400 pA (H5, H6, A2).

Two multi-SEM datasets of sample H5 were acquired as follows. In one experiment, 767 slices (476 at 35 nm, 291 at 40 nm) were imaged with a 61-beam multiSEM (multiSEM 505, Carl Zeiss Microscopy GmbH, Oberkochen, Germany) at a landing energy of 1.5 kV, a pixel size of 4 nm and a pixel dwell time of 50 ns with an FOV per slice of 1.7 by 2.1 mm. In a second experiment (H5_ext), 1342 additional slices (thickness 35 to 40 nm, corresponding to about 54 µm extent) were imaged with a smaller FOV per slice (1.1 by 1.1 mm) and aligned together with the previous experiment resulting in a total of 82 µm depth.

Image alignment for SBEM datasets was performed using global 3D relaxation of shift vectors calculated by SURF feature detection as in (*86*) followed by subimage alignment as in (*49*). The aligned image volume was then saved in the webKnossos (*87*) three-dimensional image format. Human H5 (STG), H6 (IFG), macaque (STG), and mouse (A2) SBEM datasets were aligned by scalable minds, Postdam (supplementary materials, materials and methods) (*88*).

Image alignment for the two multiSEM datasets from sample H5 were performed similar to (*55*), following the alignment routines in (*89*) and https://github.com/billkarsh/Alignment_Projects, with modifications.

Cell type classification

For analyses in Fig. 2, A and B, cell bodies were manually annotated by an expert annotator in webKnossos (*87*). All cell bodies were identified and classified into pyramidal cell, IN and glia. Pyramidal cells were identified based on presence of an apical dendrite directed toward the pia, an AIS directed toward the white matter, and spiny dendrites. INs were identified based on their large somata, which contained large numbers of mitochondria, an axon often exiting from a dendrite, lack of a clear apical dendrite or, if presence of an apical-like dendrite, lack of basal dendrites and WM-directed AIS, and smooth dendrites. Non-neuronal cells were distinguished primarily by their smaller cell bodies and different nuclear shapes.

For distinction of MP versus BP INs (Fig. 5, A to D), the dendrites of all identified INs were reconstructed and inspected in the coronal and tangential plane (*90*–*98*). INs with short dendritic lengths were excluded from this classification.

Dendrite reconstructions

Pyramidal and IN dendrites (Fig. 2, D to G) were reconstructed by an expert annotator by following their trajectory throughout the dataset volume and placing nodes as described previously in (*55*, *84*). From these dendrites for which the identity of the originating cell body had been determined, a distal stretch of 3 to 49 µm length (mouse) and 7 to 44 µm length (macaque and human) was used for annotation of all input synapses.

Annotation of input synapses on dendrite, soma, AISs

Analyses reported in Fig. 2, D to G, and fig. S1 were conducted as follows. For a given postsynaptic target class (dendrites, somata, or AIS), all input synapses were identified based on the presence of a presynaptic vesicle cloud and postsynaptic density [as described in (*55*, *67*, *84*); see following section]. These synapses were labeled as single spine when only one presynaptic bouton was found for a dendritic spine; double spine when two input synapses were found for which a clear distinction into primary (excitatory) and secondary (inhibitory) was not possible; primary spine and secondary spine when two input synapses were found for which this distinction was possible; neck for spine neck innervations, stubby spine synapses when a short dendritic protrusion of larger diameter than a spine neck and without clear diameter change at the end (no clear spine head) was synaptically innervated, and shaft

synapses when the synapse was clearly placed on the main dendritic process without noticeable protrusion. Distances of synapses from the soma were measured using minimum spanning tree on the annotated nodes of the dendrites (see "skeleton" class in code repository).

Soma based axon reconstructions

Analyses reported in Figs. 3, A to C and G to I, and 5, E and F, were conducted as follows. The axons of identified pyramidal cell and INs were reconstructed by first identifying the exit at the AIS. Then the trajectory of the axon was followed throughout the dataset and comments were added at the outgoing synapses. The post-target of each synapse was further classified into excitatory or inhibitory class based on their spine-targeting synapse density (when target was dendritic) or cell body type (when target was a soma or an AIS). Axons with fewer than 10 synapses were excluded to allow higher signal-to-noise ratio per axon.

Synapse-seeded axon reconstructions

Analyses reported in Fig. 4, A and B, were conducted as follows. For a given postsynaptic target with identified input synapses, a skeleton node was placed in the presynaptic axon's vesicle cloud and commented as "seed" synapse. The presynaptic axon was then reconstructed throughout the entire dataset volume, and all of the axons' other output synapses and their corresponding postsynaptic targets were identified as described in the following section. The "seed" synapse was excluded when quantifying axonal spine target properties in Fig. 4A.

Synapse identification for reconstructed axons

For analyses reported in Figs. 3, A to C and G to I; 4, A and B; and 5, E and F, the following synapse identification was applied. For each reconstructed axon, synapses were identified manually when following the trajectory of axon. First, vesicle clouds in the axon were identified as accumulations of vesicles. Subsequently, the most likely postsynaptic target was identified by the following criteria: direct apposition with vesicle cloud; presence of a darkening and slight broadening of the synaptic membrane; vesicles at close proximity to the plasma membrane at the site of potential synaptic contact. Synapses were marked as uncertain whenever the signs of darkened postsynaptic density could not be clearly identified. All analysis in this study were conducted only on synapses that had been classified as certain. For each axon, at every subsequent synapse location, a node was placed in the vesicle cloud and commented with the corresponding postsynaptic target's identity (for example, soma, AIS, dendritic shaft, dendritic single spine, dendritic double-innervated spine, spine neck, stub, somatic spines or filopodium). All synapses were annotated by an expert annotator; for unclear cases, these were re-annotated for expert consensus between two or three experts.

Volumetric model of synapse and axon types, inference of synapse and axon types, automated reconstruction, and error analysis

These methods are reported in the supplementary materials.

Estimates of total dendritic path length of human pyramidal cells

For the estimation of the complete synaptic input onto a L2/3 pyramidal cell (Fig. 6, A and B, and fig. S4), all dendrites of 10 L2/3 pyramidal cells were reconstructed until their end in dataset H5, which was either the actual ending of the dendrite in the neuropil, or the end of the dendrite at the dataset boundary (Fig. 6A). In addition, nine pyramidal cells in the extended dataset H5_ext were analyzed with a larger fraction of in-dataset dendrite endings. Results from both datasets yielded similar ranges for total dendritic path length estimates, as detailed in the following.

For the following calculations, only dendrites with actual in-dataset endings were used (one exception were apical tuft dendrites in which some dendrites without in-dataset endings had more branch points and were therefore included in the estimate). This approach could correspond to an underestimation of dendrite length; therefore, in addition to the length measurements described here, we also used length reports from light-microscopically imaged human L3 pyramidal neurons, which provided similar path length measurements (ranges in Fig. 6B) (*45*).

For each pyramidal cell, the following dendritic compartments were distinguished: (i) apical dendrite trunk (measured from the exit at the cell body toward pia along the cortical axis, until the main bifurcation), (ii) apical tuft (measured from the main bifurcation point of the apical dendrite), (iii) oblique dendrites (measured from the exit at the apical dendritic trunk), and (iv) basal dendrites (all dendrites exiting at the cell body except the apical dendrite) (*55*).

For the basal, oblique and apical tuft compartment n = 226, 211, and 167 dendrites were reconstructed of which n = 25, 28, and 32 dendrites with in-dataset endings were found (n = 21 cells).

For the estimation of the average number of branch points for the apical tuft compartment dendrites without in-dataset endings were included to avoid an underestimation of dendrite length as some of them had more branch points than the dendrites with in-dataset endings. For this, all apical tuft dendrites beyond a 300-μm threshold were included (n = 64 dendrites, with n = 32 in-dataset endings, n = 21 cells).

The estimations for the lengths of the above compartments were done as follows. The length for apical dendrite trunk compartment was estimated by averaging the lengths over all the samples from their soma exit until the main bifurcation. For each of the remaining compartments (i.e., apical tufts, oblique and basal dendrites), the path lengths (PLs) between consecutive branch points (BPs) were measured until a true end was reached. The order of a branch point ("order") was defined as the number of edges along the unique path between it and the cell body node (for example, order is 0 at the cell body node, 1 at first branch point, 2 at second branch point and so on). Then for the entire compartment (apical tuft, oblique or a basal dendrite), the total length was calculated by summing the path lengths of each segment times 2 to the power order of the corresponding branch point: $\sum_{\text{order}=0}^{BP} PL_{\text{seg}_o\text{rder}} * 2^{\text{order}}$ where order = order of a branch point, BP = mean number of branch points for each compartment and $PL_{\text{seg\_order}}$ = path length of the succeeding segment at the corresponding branch point order. Then, the resulting total length of each compartment was multiplied with the average number of basal, oblique or apical tuft exits to estimate lengths for all basal, oblique and apical tufts respectively. The model of the branching as binary to the final branch point order would likely overestimate total dendritic path length, thus not affecting the conclusion about low total input synapse numbers in human.

To estimate the number of input synapses for each of the dendritic compartments, we proceeded as follows. For each compartment (apical trunk, apical tuft, oblique or basal dendrite), dendritic segments of path length 10 to 50 μm were sampled and all input synapses were annotated. For computing the spine-targeting synapse density per dendritic shaft path length, both single spine head and primary spine head synapses were included. For shaft synapse density, only the shaft synapses were included. For total synapse density, additionally secondary spine head, spine neck and dendrite "stub" synapses were included (total fraction of these additional synapses: less than 10%). The average spine, shaft and total synapse densities were then multiplied with the total path lengths of the corresponding compartments. The apical dendrite trunk was additionally distinguished into a proximal segment (up to 50 μm from cell body) and a distal segment (50 μm from soma until the main bifurcation). Similarly, the basal dendrites were additionally distinguished into a proximal segment which showed high shaft synapse density (up to 30 μm from cell body for mouse, 80 μm for human).

For the basal compartment we measured spine-targeting and shaft synapse densities of 0.66 ± 0.23 per micrometer and 0.23 ±

0.09 per micrometer (n = 13), for the oblique compartment 0.67 ± 0.23 per micrometer and 0.19 ± 0.06 per micrometer (n = 14), for the apical tuft compartment 0.73 ± 0.34 per micrometer (n = 12) and 0.25 ± 0.07 per micrometer (n = 12), respectively.

For the proximal segment of the apical trunk we measured spine-targeting and shaft synapse densities of 0.06 ± 0.14 per micrometer and 0.44 ± 0.16 per micrometer (n = 9), for the distal segment until the main bifurcation 0.45 ± 0.39 per micrometer and 0.20 ± 0.08 per micrometer (n = 7), respectively.

Statistical tests

All statistical tests were KS for fractions and ratios, and Wilcoxon ranksum tests, otherwise. Data are shown as box or violin plots with median, 25th and 75th percentile and outliers (using MATLAB, Mathworks). All tests are documented in the code repository (next section).

The test of significance for increase in IN fraction from mouse to macaque/human was done by pooling all identified excitatory neurons and INs counts per species group and generating 1000 bootstrap samples (*99*) with replacement. IN fraction (IN_{frac}) was computed for each bootstrap sample. The P-value was calculated as the fraction of bootstrap samples in mouse for which IN_{frac} was greater than or equal to the lowest IN_{frac} in macaque/human bootstrap group (Fig. 2B, error bars 10th to 90th percentiles). The same bootstrap approach was used as a test of significance for increase in non- MP IN fraction from mouse to macaque/human by pooling all MP and non- MP INs in volume per species group (Fig. 5D, error bars 10th-90th percentiles) and similarly for testing increase of IN-targeting fraction of non-MP IN axons from mouse to macaque/human by pooling all output synapses of non-MP IN axons per species group (Fig. 5, F and G).

Data availability

All electron microscopy datasets are publicly available for browsing at webknossos.org: macaque L2/3 (S1), https://wklink.org/1186; macaque L2/3 (STG), https://wklink.org/1319; human (H5) L23 (STG), https://wklink.org/7861; human (H5) L23 (STG) multiSEM, https://wklink.org/5364; human (H5) L1-6 (STG) multiSEM, https://wklink.org/1742; human (H6) L2/3 (IFG), https://wklink.org/7299; mouse L2/3 (S1), https://wklink.org/9045; mouse L2/3 (PPC), https://wklink.org/2581; mouse L2 (ACC), https://wklink.org/7415; mouse L2/3 (V2), https://wklink.org/3592; mouse L2/3 (A2), https://wklink.org/7193; mouse L1-5 (PPC-2), https://wklink.org/4814.

REFERENCES AND NOTES

1. S. Herculano-Houzel, The human brain in numbers: A linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3**, 31 (2009). doi: 10.3389/neuro.09.031.2009; pmid: 19915731
2. C. S. von Bartheld, J. Bahney, S. Herculano-Houzel, The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* **524**, 3865–3895 (2016). doi: 10.1002/cne.24040; pmid: 27187682
3. R. D. Hodge *et al.*, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019). doi: 10.1038/s41586-019-1506-7; pmid: 31435019
4. F. M. Krienen *et al.*, Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020). doi: 10.1038/s41586-020-2781-z; pmid: 32999462
5. K. Letinic, R. Zoncu, P. Rakic, Origin of GABAergic neurons in the human neocortex. *Nature* **417**, 645–649 (2002). doi: 10.1038/nature00779; pmid: 12050665
6. S. R. y. Cajal, Estudios sobre la corteza cerebral humana. II. Estructura de la corteza motriz del hombre y mamíferos superiores. *Revista Trimestral Micrográfica* **4**, 117–200 (1899).
7. R. N. Kooijmans, W. Sierhuis, M. W. Self, P. R. Roelfsema, A quantitative comparison of inhibitory interneuron size and distribution between mouse and macaque V1, using calcium-binding proteins. *Cereb. Cortex Commun.* **1**, tgaa068 (2020). doi: 10.1093/texcom/tgaa068; pmid: 34296129
8. J. Berg *et al.*, Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **598**, 151–158 (2021). doi: 10.1038/s41586-021-03813-8; pmid: 34616067
9. T. E. Bakken *et al.*, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021). doi: 10.1038/s41586-021-03465-8; pmid: 34616062
10. A. Shapson-Coe *et al.*, A connectomic study of a petascale fragment of human cerebral cortex. *bioRxiv*, 446289 [Preprint] (2021): .doi: 10.1101/2021.05.29.446289
11. M. Field *et al.*, Tonic $GABA_A$ receptor mediated currents of human cortical GABAergic interneurons vary amongst cell types. *J. Neurosci.* **41**, 9702–9719 (2021). doi: 10.1523/JNEUROSCI.0175-21.2021; pmid: 34667071
12. V. Szegedi *et al.*, Robust perisomatic GABAergic self-innervation inhibits basket cells in the human and mouse supragranular neocortex. *eLife* **9**, e51691 (2020). doi: 10.7554/eLife.51691; pmid: 31916939
13. A. M. M. Sousa *et al.*, Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017). doi: 10.1126/science.aan3456; pmid: 29170230
14. E. Boldog *et al.*, Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat. Neurosci.* **21**, 1185–1195 (2018). doi: 10.1038/s41593-018-0205-2; pmid: 30150662
15. J. DeFelipe, L. Alonso-Nanclares, J. I. Arellano, Microstructure of the neocortex: Comparative aspects. *J. Neurocytol.* **31**, 299–316 (2002). doi: 10.1023/A:1024130211265; pmid: 12815249
16. A. Rollenhagen *et al.*, Synaptic organization of the human temporal lobe neocortex as revealed by high-resolution transmission, focused ion beam scanning, and electron microscopic tomography. *Int. J. Mol. Sci.* **21**, 5558 (2020). doi: 10.3390/ijms21155558; pmid: 32756507
17. M. Domínguez-Álvaro, M. Montero-Crespo, L. Blazquez-Llorca, J. DeFelipe, L. Alonso-Nanclares, 3D electron microscopy study of synaptic organization of the normal human transentorhinal cortex and its possible alterations in Alzheimer's disease. *eNeuro* **6**, ENEURO.0140-19.2019 (2019). doi: 10.1523/ENEURO.0140-19.2019; pmid: 31217195
18. P. Marco, J. DeFelipe, Altered synaptic circuitry in the human temporal neocortex removed from epileptic patients. *Exp. Brain Res.* **114**, 1–10 (1997). doi: 10.1007/PL00005608; pmid: 9125446
19. B. Wang *et al.*, A subtype of inhibitory interneuron with intrinsic persistent activity in human and monkey neocortex. *Cell Rep.* **10**, 1450–1458 (2015). doi: 10.1016/j.celrep.2015.02.018; pmid: 25753411
20. G. Molnár *et al.*, Complex events initiated by individual spikes in the human cerebral cortex. *PLOS Biol.* **6**, e222 (2008). doi: 10.1371/journal.pbio.0060222; pmid: 18767905
21. D. S. Melchitzky, S. R. Sesack, D. A. Lewis, Parvalbumin-immunoreactive axon terminals in macaque monkey and human prefrontal cortex: Laminar, regional, and target specificity of type I and type II synapses. *J. Comp. Neurol.* **408**, 11–22 (1999). doi: 10.1002/(SICI)1096-9861(19990524)408:1<11::AID-CNE2>3.0.CO;2-T; pmid: 10331577
22. M. Domínguez-Álvaro, M. Montero-Crespo, L. Blazquez-Llorca, J. DeFelipe, L. Alonso-Nanclares, 3D ultrastructural study of synapses in the human entorhinal cortex. *Cereb. Cortex* **1**, 410–435 (2020).
23. D. S. Melchitzky, D. A. Lewis, Dendritic-targeting GABA neurons in monkey prefrontal cortex: Comparison of somatostatin- and calretinin-immunoreactive axon terminals. *Synapse* **62**, 456–465 (2008). doi: 10.1002/syn.20514; pmid: 18361442
24. N. Cano-Astorga, J. DeFelipe, L. Alonso-Nanclares, Three-dimensional synaptic organization of layer III of the human temporal neocortex, three-dimensional synaptic organization of layer III of the human temporal neocortex. *Cereb. Cortex* **31**, 4742–4764 (2021). doi: 10.1093/cercor/bhab120; pmid: 33999122
25. B. G. Cragg, Ultrastructural features of human cerebral cortex. *J. Anat.* **121**, 331–362 (1976). pmid: 179969
26. I. B. Yáñez *et al.*, Double bouquet cell in the human cerebral cortex and a comparison with other mammals. *J. Comp. Neurol.* **486**, 344–360 (2005). doi: 10.1002/cne.20533; pmid: 15846784
27. E. A. Nimchinsky *et al.*, A neuronal morphologic type unique to humans and great apes. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5268–5273 (1999). doi: 10.1073/pnas.96.9.5268; pmid: 10220455
28. M. R. del Río, J. DeFelipe, Double bouquet cell axons in the human temporal neocortex: Relationship to bundles of myelinated axons and colocalization of calretinin and calbindin D-28k immunoreactivities. *J. Chem. Neuroanat.* **13**, 243–251 (1997). doi: 10.1016/S0891-0618(97)00050-1; pmid: 9412906
29. R. Benavides-Piccione, I. Ballesteros-Yáñez, J. DeFelipe, R. Yuste, Cortical area and species differences in dendritic spine morphology. *J. Neurocytol.* **31**, 337–346 (2002). doi: 10.1023/A:1024134312173; pmid: 12815251
30. V. Szegedi *et al.*, Plasticity in Single Axon Glutamatergic Connection to GABAergic Interneurons Regulates Complex Events in the Human Neocortex. *PLOS Biol.* **14**, e2000237 (2016). doi: 10.1371/journal.pbio.2000237; pmid: 27828957
31. K. D. Micheva, C. Beaulieu, Postnatal development of GABA neurons in the rat somatosensory barrel cortex: A quantitative study. *Eur. J. Neurosci.* **7**, 419–430 (1995). doi: 10.1111/j.1460-9568.1995.tb00338.x; pmid: 7773439
32. H. S. Meyer *et al.*, Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16807–16812 (2011). doi: 10.1073/pnas.1113648108; pmid: 21949377
33. D. Džaja, A. Hladnik, I. Bičanić, M. Baković, Z. Petanjek, Neocortical calretinin neurons in primates: Increase in proportion and microcircuitry structure. *Front. Neuroanat.* **8**, 103–103 (2014). pmid: 25309344
34. B. Tasic *et al.*, Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018). doi: 10.1038/s41586-018-0654-5; pmid: 30382198
35. J. Q. Ren, Y. Aika, C. W. Heizmann, T. Kosaka, Quantitative analysis of neurons and glial cells in the rat somatosensory cortex, with special reference to GABAergic neurons and parvalbumin-containing neurons. *Exp. Brain Res.* **92**, 1–14 (1992). doi: 10.1007/BF00230378; pmid: 1486945
36. C. Beaulieu, G. Campistron, C. Crevier, Quantitative aspects of the GABA circuitry in the primary visual cortex of the adult rat. *J. Comp. Neurol.* **339**, 559–572 (1994). doi: 10.1002/cne.903390407; pmid: 8144746
37. S. Lefort, C. Tomm, J. C. Floyd Sarria, C. C. H. Petersen, The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* **61**, 301–316 (2009). doi: 10.1016/j.neuron.2008.12.020; pmid: 19186171
38. C. Beaulieu, Z. Kisvarday, P. Somogyi, M. Cynader, A. Cowey, Quantitative distribution of GABA-immunopositive and -immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cereb. Cortex* **2**, 295–309 (1992). doi: 10.1093/cercor/2.4.295; pmid: 1330121
39. P. L. A. Gabbott, S. J. Bacon, Local circuit neurons in the medial prefrontal cortex (areas 24a,b,c, 25 and 32) in the monkey: I. Cell morphology and morphometrics. *J. Comp. Neurol.* **364**, 567–608 (1996). doi: 10.1002/(SICI)1096-9861(19960122)364:4<567::AID-CNE1>3.0.CO;2-1; pmid: 8821449
40. J.-P. Hornung, N. De Tribolet, Distribution of GABA-containing neurons in human frontal cortex: A quantitative immunocytochemical study. *Anat. Embryol.* **189**, 139–145 (1994). doi: 10.1007/BF00185772; pmid: 8010412
41. M. R. del Río, J. DeFelipe, Colocalization of calbindin D-28k, calretinin, and GABA immunoreactivities in neurons of the human temporal cortex. *J. Comp. Neurol.* **369**, 472–482 (1996). doi: 10.1002/(SICI)1096-9861(19960603)369:3<472::AID-CNE11>3.0.CO;2-K; pmid: 8743426
42. E. G. Jones, G. W. Huntley, D. L. Benson, Alpha calcium/calmodulin-dependent protein kinase II selectively expressed in a subpopulation of excitatory neurons in monkey sensory-

motor cortex: Comparison with GAD-67 expression. *J. Neurosci.* **14**, 611–629 (1994). doi: 10.1523/JNEUROSCI.14-02-00611.1994; pmid: 8301355
43. G. Eyal *et al*., Human cortical pyramidal neurons: From spines to spikes via models. *Front. Cell. Neurosci.* **12**, 181 (2018). doi: 10.3389/fncel.2018.00181; pmid: 30008663
44. R. Benavides-Piccione, I. Fernaud-Espinosa, V. Robles, R. Yuste, J. DeFelipe, Age-based comparison of human dendritic spine structure using complete three-dimensional reconstructions. *Cereb. Cortex* **23**, 1798–1810 (2013). doi: 10.1093/cercor/bhs154; pmid: 22710613
45. H. Mohan *et al*., Dendritic and axonal architecture of individual pyramidal neurons across layers of adult human neocortex. *Cereb. Cortex* **25**, 4839–4853 (2015). doi: 10.1093/cercor/bhv188; pmid: 26318661
46. G. N. Elston, R. Benavides-Piccione, J. DeFelipe, The pyramidal cell in cognition: A comparative study in human and monkey. *J. Neurosci.* **21**, RC163–RC163 (2001). doi: 10.1523/JNEUROSCI.21-17-j0002.2001; pmid: 11511694
47. M. B. Ramocki, H. Y. Zoghbi, Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* **455**, 912–918 (2008). doi: 10.1038/nature07457; pmid: 18923513
48. J. L. R. Rubenstein, M. M. Merzenich, Model of autism: Increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003). doi: 10.1034/j.1601-183X.2003.00037.x; pmid: 14606691
49. A. Motta *et al*., Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science* **366**, eaay3134 (2019). doi: 10.1126/science.aay3134; pmid: 31649140
50. W. Denk, H. Horstmann, Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLOS Biol.* **2**, e329 (2004). doi: 10.1371/journal.pbio.0020329; pmid: 15514700
51. K. J. Hayworth *et al*., Imaging ATUM ultrathin section libraries with WaferMapper: A multi-scale approach to EM reconstruction of neural circuits. *Front. Neural Circuits* **8**, 68 (2014). doi: 10.3389/fncir.2014.00068; pmid: 25018701
52. K. J. Hayworth, N. Kasthuri, R. Schalek, J. W. Lichtman, Automating the collection of ultrathin serial sections for large volume TEM reconstructions. *Microsc. Microanal.* **12** (S02), 86–87 (2006). doi: 10.1017/S1431927606066268
53. A. L. Eberle *et al*., High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *J. Microsc.* **259**, 114–120 (2015). doi: 10.1111/jmi.12224; pmid: 25627873
54. K. Lee, J. Zung, P. Li, V. Jain, H. S. Seung, Superhuman accuracy on the SNEMI3D connectomics challenge. arXiv:1706.00120v1 [cs.CV] (2017).
55. A. Karimi, J. Odenthal, F. Drawitsch, K. M. Boergens, M. Helmstaedter, Cell-type specific innervation of cortical pyramidal cells at their apical dendrites. *eLife* **9**, e46876 (2020). doi: 10.7554/eLife.46876; pmid: 32108571
56. V. Braitenberg, A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer Science & Business Media, 1998).
57. T. Kwon *et al*., Ultrastructural, molecular and functional mapping of GABAergic synapses on dendritic spines and shafts of neocortical pyramidal neurons. *Cereb. Cortex* **29**, 2771–2781 (2019). doi: 10.1093/cercor/bhy143; pmid: 30113619
58. Y. Kubota, F. Karube, M. Nomura, Y. Kawaguchi, The diversity of cortical inhibitory synapses. *Front. Neural Circuits* **10**, 27–27 (2016). doi: 10.3389/fncir.2016.00027; pmid: 27199670
59. G. A. Wildenberg *et al*., Primate neuronal connections are sparse in cortex as compared to mouse. *Cell Rep.* **36**, 109709 (2021). doi: 10.1016/j.celrep.2021.109709; pmid: 34525373
60. A. T. Kuan *et al*., Dense neuronal reconstruction through x-ray holographic nano-tomography. *Nat. Neurosci.* **23**, 1637–1643 (2020). doi: 10.1038/s41593-020-0704-9; pmid: 32929244
61. Y. Kubota, S. Hatada, S. Kondo, F. Karube, Y. Kawaguchi, Neocortical inhibitory terminals innervate dendritic spines targeted by thalamocortical afferents. *J. Neurosci.* **27**, 1139–1150 (2007). doi: 10.1523/JNEUROSCI.3846-06.2007; pmid: 17267569
62. M. R. del Río, J. DeFelipe, A light and electron microscopic study of calbindin D-28k immunoreactive double bouquet cells in the human temporal cortex. *Brain Res.* **690**, 133–140 (1995). doi: 10.1016/0006-8993(95)00641-3; pmid: 7496800
63. I. P. Lukacs *et al*., Differential effects of group III metabotropic glutamate receptors on spontaneous inhibitory synaptic currents in spine-innervating double bouquet and parvalbumin-expressing dendrite-targeting GABAergic interneurons in human neocortex. *bioRxiv*, 483105 [Preprint] (2022). doi: 10.1101/2022.03.05.483105
64. C. Xu *et al*., Structure and plasticity of silent synapses in developing hippocampal neurons visualized by super-resolution imaging. *Cell Discov.* **6**, 8 (2020). doi: 10.1038/s41421-019-0139-1; pmid: 32133151
65. G. M. Durand, Y. Kovalchuk, A. Konnerth, Long-term potentiation and functional synapse induction in developing hippocampus. *Nature* **381**, 71–75 (1996). doi: 10.1038/381071a0; pmid: 8609991
66. M. Lenz *et al*., All-trans retinoic acid induces synaptic plasticity in human cortical neurons. *eLife* **10**, e63026 (2021). doi: 10.7554/eLife.63026; pmid: 33781382
67. H. Schmidt *et al*., Axonal synapse sorting in medial entorhinal cortex. *Nature* **549**, 469–475 (2017). doi: 10.1038/nature24005; pmid: 28959971
68. J. Kornfeld *et al*., EM connectomics reveals axonal target variation in a sequence-generating network. *eLife* **6**, e24364 (2017). doi: 10.7554/eLife.24364; pmid: 28346140
69. C. E. Carr, M. Konishi, A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.* **10**, 3227–3246 (1990). doi: 10.1523/JNEUROSCI.10-10-03227.1990; pmid: 2213141
70. K. D. Micheva, C. Beaulieu, Quantitative aspects of synaptogenesis in the rat barrel field cortex with special reference to GABA circuitry. *J. Comp. Neurol.* **373**, 340–354 (1996). doi: 10.1002/(SICI)1096-9861(19960923)373:3<340::AID-CNE3>3.0.CO;2-2; pmid: 8889932
71. C. Calì *et al*., The effects of aging on neuropil structure in mouse somatosensory cortex-A 3D electron microscopy analysis of layer 1. *PLOS ONE* **13**, e0198131 (2018). doi: 10.1371/journal.pone.0198131; pmid: 29966021
72. J. P. Bourgeois, P. Rakic, Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. *J. Neurosci.* **13**, 2801–2820 (1993). doi: 10.1523/JNEUROSCI.13-07-02801.1993; pmid: 8331373
73. S. A. Kirov, L. J. Petrak, J. C. Fiala, K. M. Harris, Dendritic spines disappear with chilling but proliferate excessively upon rewarming of mature hippocampus. *Neuroscience* **127**, 69–80 (2004). doi: 10.1016/j.neuroscience.2004.04.053; pmid: 15219670
74. L. A. Glantz, D. A. Lewis, Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *Arch. Gen. Psychiatry* **57**, 65–73 (2000). doi: 10.1001/archpsyc.57.1.65; pmid: 10632234
75. M. Medalla, J. P. Gilman, J.-Y. Wang, J. I. Luebke, Strength and diversity of inhibitory signaling differentiates primate anterior cingulate from lateral prefrontal cortex. *J. Neurosci.* **37**, 4717–4734 (2017). doi: 10.1523/JNEUROSCI.3757-16.2017; pmid: 28381592
76. A. Hsu, J. I. Luebke, M. Medalla, Comparative ultrastructural features of excitatory synapses in the visual and frontal cortices of the adult mouse and monkey. *J. Comp. Neurol.* **525**, 2175–2191 (2017). doi: 10.1002/cne.24196; pmid: 28256708
77. J. P. Gilman, M. Medalla, J. I. Luebke, Area-specific features of pyramidal neurons-a comparative study in mouse and rhesus monkey. *Cereb. Cortex* **27**, 2078–2094 (2017). pmid: 26965903
78. A. Gidon *et al*., Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* **367**, 83–87 (2020). doi: 10.1126/science.aax6239; pmid: 31896716
79. G. Eyal *et al*., Unique membrane properties and enhanced signal processing in human neocortical neurons. *eLife* **5**, e16553 (2016). doi: 10.7554/eLife.16553; pmid: 27710767
80. B. Scholl, C. I. Thomas, M. A. Ryan, N. Kamasawa, D. Fitzpatrick, Cortical response selectivity derives from strength in numbers of synapses. *Nature* **590**, 111–114 (2021). doi: 10.1038/s41586-020-03044-3; pmid: 33328635
81. J. J. Letzkus *et al*., A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* **480**, 331–335 (2011). doi: 10.1038/nature10674; pmid: 22158104
82. J. J. Letzkus, S. B. E. Wolff, A. Lüthi, Disinhibition, a circuit mechanism for associative learning and memory. *Neuron* **88**, 264–276 (2015). doi: 10.1016/j.neuron.2015.09.024; pmid: 26494276
83. R. Kim, T. J. Sejnowski, Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *Nat. Neurosci.* **24**, 129–139 (2021). doi: 10.1038/s41593-020-00753-w; pmid: 33288909
84. A. Gour *et al*., Postnatal connectomic development of inhibition in mouse barrel cortex. *Science* **371**, eabb4534 (2021). doi: 10.1126/science.abb4534; pmid: 33273061
85. Y. Hua, P. Laserstein, M. Helmstaedter, Large-volume en-bloc staining for electron microscopy-based connectomics. *Nat. Commun.* **6**, 7923 (2015). doi: 10.1038/ncomms8923; pmid: 26235643
86. F. Drawitsch, A. Karimi, K. M. Boergens, M. Helmstaedter, FluoEM, virtual labeling of axons in three-dimensional electron microscopy data for long-range connectomics. *eLife* **7**, e38976 (2018). doi: 10.7554/eLife.38976; pmid: 30106377
87. K. M. Boergens *et al*., webKnossos: Efficient online 3D data annotation for connectomics. *Nat. Methods* **14**, 691–694 (2017). doi: 10.1038/nmeth.4331; pmid: 28604722
88. M. A. Fischler, R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**, 381–395 (1981). doi: 10.1145/358669.358692
89. L. K. Scheffer, B. Karsh, S. Vitaladevun, Automated alignment of imperfect EM images for neural reconstruction. arXiv:1304.6034v1 [q-bio.QM] (2013).
90. A. Peters, Cellular components of the cerebral cortex, in *Cerebral Cortex* (Univ. Chicago Press, 1984), vol. 1, pp. 107–200.
91. P. Somogyi, A. Cowey, Combined Golgi and electron microscopic study on the synapses formed by double bouquet cells in the visual cortex of the cat and monkey. *J. Comp. Neurol.* **195**, 547–566 (1981). doi: 10.1002/cne.901950402; pmid: 7462443
92. A. D. de Lima, J. H. Morrison, Ultrastructural analysis of somatostatin-immunoreactive neurons and synapses in the temporal and occipital cortex of the macaque monkey. *J. Comp. Neurol.* **283**, 212–227 (1989). doi: 10.1002/cne.902830205; pmid: 2567743
93. J. DeFelipe, S. H. Hendry, E. G. Jones, Synapses of double bouquet cells in monkey cerebral cortex visualized by calbindin immunoreactivity. *Brain Res.* **503**, 49–54 (1989). doi: 10.1016/0006-8993(89)91702-2; pmid: 2611658
94. J. DeFelipe, S. H. Hendry, T. Hashikawa, M. Molinari, E. G. Jones, A microcolumnar structure of monkey cerebral cortex revealed by immunocytochemical studies of double bouquet cell axons. *Neuroscience* **37**, 655–673 (1990). doi: 10.1016/0306-4522(90)90097-N; pmid: 1701039
95. P. Somogyi, T. F. Freund, A. Cowey, The axo-axonic interneuron in the cerebral cortex of the rat, cat and monkey. *Neuroscience* **7**, 2577–2607 (1982). doi: 10.1016/0306-4522(82)90086-0; pmid: 7155343
96. B. Cauli *et al*., Molecular and physiological diversity of cortical nonpyramidal cells. *J. Neurosci.* **17**, 3894–3906 (1997). doi: 10.1523/JNEUROSCI.17-10-03894.1997; pmid: 9133407
97. Y. Kawaguchi, Y. Kubota, Physiological and morphological identification of somatostatin- or vasoactive intestinal polypeptide-containing cells among GABAergic cell subtypes in rat frontal cortex. *J. Neurosci.* **16**, 2701–2715 (1996). doi: 10.1523/JNEUROSCI.16-08-02701.1996; pmid: 8786446
98. C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, M. Scanziani, Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013). doi: 10.1038/nn.3446; pmid: 23817549
99. B. Efron, Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979). doi: 10.1214/aos/1176344552
100. S. B. Hedges, S. Kumar, Genomics. Vertebrate genomes compared. *Science* **297**, 1283–1285 (2002). doi: 10.1126/science.1076231; pmid: 12193771
101. M. A. Hofman, Evolution of the human brain: When bigger is better. *Front. Neuroanat.* **8**, 15 (2014). doi: 10.3389/fnana.2014.00015; pmid: 24723857

ACKNOWLEDGMENTS

We thank K. Mätz-Rensing at the German Primate Center and F. Bremmer and A. Kaminiarz at the University of Marburg for providing tissue of macaque brain; anonymous individuals H5 and H6 for consent to provide tissue from neurosurgery interventions; H. Wissler for support with visualizations and for management of the neuron reconstruction team; I. Wolf and J. Kim for staining support; A. Karimi, F. Drawitsch, and K. M. Boergens for providing custom written image acquisition and alignment routines; S. Babl, L. Bezzenberger, R. Jakoby, R. Kneisl, and M. Kronawitter for annotator training and task management; D. Werner, V. Pinkau, G. Wiese, F. Meinel, J. Striebel, P. Otto, T. Herold, and N. Rzepka at scalable minds, Potsdam, Germany for collaboration on automated data analysis; T. Garbowski and D. Lechner (Zeiss) for help with setting up multiSEM; M. Fahland and N. Prager (Fraunhofer FEP, Dresden) for collaboration on Kapton tape; R. Schalek and J. Lichtman for initial advice on setting up ATUM; G. Tushev, A. Kotowicz, T. Zeller, and C. Guggenberger for IT support; S. Soworka, I. Wolf, S. Horn and L Dadashev for excellent technical support; and three anonymous reviewers for helpful advice. Additional methodological acknowledgments are provided in the supplementary materials.

Funding: Funding was provided by the Max-Planck Society. S.L. was partly funded by Neuronex2 grant (DFG, German Research Foundation–HE 7321/1-1). **Author contributions:** M.H. conceived, initiated and supervised the study. J.S., N.H., A.M.K., V.G., and N.J. carried out experiments. H.S.M. and J.G. provided human tissue. M.S. provided ATUM-multiSEM methodology. A.M. developed and applied analysis methodology. S.L., J.S., N.H., A.M.K., V.G., and M.H. provided reconstructions and analyzed data. S.L. and M.H. wrote the manuscript, with contributions from all authors. **Competing interests:** The authors declare no competing financial interests.

Data and materials availability: All image data and reconstructions are publicly available through https://www.webknossos.org under the links at the end of the Materials and methods section. All software used for analysis is available at the following address and have been made publicly available under the MIT license: https://gitlab.mpcdf.mpg.de/connectomics/human_primate.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0924
Materials and Methods
Supplementary Text
Figs. S1 to S4
MDAR Reproducibility Checklist

Submitted 13 January 2022; accepted 30 May 2022
Published online 23 June 2022
10.1126/science.abo0924

RESEARCH ARTICLE

ORGANIC CHEMISTRY

Digitization and validation of a chemical synthesis literature database in the ChemPU

Simon Rohrbach†, Mindaugas Šiaučiulis†, Greig Chisholm†, Petrisor-Alin Pirvan, Michael Saleeb, S. Hessam M. Mehr, Ekaterina Trushina, Artem I. Leonov, Graham Keenan, Aamir Khan, Alexander Hammer, Leroy Cronin*

Despite huge potential, automation of synthetic chemistry has only made incremental progress over the past few decades. We present an automatically executable chemical reaction database of 100 molecules representative of the range of reactions found in contemporary organic synthesis. These reactions include transition metal–catalyzed coupling reactions, heterocycle formations, functional group interconversions, and multicomponent reactions. The chemical reaction codes or χDLs for the reactions have been stored in a database for version control, validation, collaboration, and data mining. Of these syntheses, more than 50 entries from the database have been downloaded and robotically run in seven modular ChemPU's with yields and purities comparable to those achieved by an expert chemist. We also demonstrate the automatic purification of a range of compounds using a chromatography module seamlessly coupled to the platform and programmed with the same language.

To replicate a known chemical reaction the protocol must be obtained from either the literature or a database so that it can be run manually in the laboratory (*1*). However, not all literature or database entries can be easily reproduced (*2*). This is a barrier not only to the synthesis of new molecules but also to accumulation of high-quality data for machine learning (*3*–*6*) and is exacerbated by the fact that there is also no open standard for coding the procedures or a way to widely report and correct failed experiments (*7*, *8*). An approach that unambiguously captures and codes a chemical synthesis protocol for use by an automated system (*9*–*16*) with capacity to be versioned similar to software and record failed experiments would transform the field. Organic synthesis currently requires intensive, highly skilled labor (*17*) and a typical synthesis can require multiple complex unit operations that are difficult to explicitly encode. This is because the tacit knowledge required is often context-dependent, resulting in ambiguities in the published literature that limit reproducibility, automation, or data mining (*18*). These limits have been overcome in some specific areas such as oligopeptide (*19*), oligosaccharide (*20*), and oligonucleotide (*21*) chemistry, and in recent years much progress has been made in automating chemical reactions more broadly (*22*–*30*). However, most automated synthetic chemistry platforms remain task-specific or represent islands of automation (*10*) in an otherwise manual workflow, but even these have bespoke instruction sets with no simple semantic link among them or to the literature. To fully exploit the potential of automation in chemical synthesis and ensure reproducibility of procedures, progress is needed on two fronts (*31*, *32*). First, a truly universal automation platform is required that can perform all unit operations (*14*, *16*); second, a standardized and precise syntax to describe these chemical processes is essential to reliably capture all the critical details of a given chemical process (*15*). Such a code must also be independent of the type of hardware employed for automation and thus be compiled to work flawlessly on any compatible hardware system.

We present the design, construction, and validation of a workflow that allows us to capture the chemical synthesis literature from manual operation to a fully described and universal Chemical Description Language (χDL) (*15*, *33*) to be run automatically in the Chemical Processing Unit or ChemPU. The process of running the χDL on the ChemPU we call chemputation (similar to computation) and is the reliable conversion of code and reagents into products. We not only show that the χDL can be compiled to run on many different ChemPU configurations but also demonstrate the capacity of the χDL language to encode a wide range of synthetic procedures, which are representative of the organic chemistry toolbox. Overall, 103 different reactions of highly varied chemistry have been translated from the literature to reliable χDL codes, and 53 of these programs have been validated on the hardware with yields and purities comparable to that in the literature. This increased synthesis throughput would not have been possible with earlier versions of the ChemPU, which could not use χDL (*14*–*16*). It also signifies a massive step up in the number of validated χDL procedures compared with the seminal paper on χDL (*15*) and is testimony to the increased reliability of the hardware employed in this paper. We designed and built a χDL database called Chemify (*34*) for our current 103 entries and anticipate this will rapidly expand; the database will be available for anyone to run and validate on suitable hardware. Not only could these χDL entries be implemented on other automated synthesis platforms and material generated on demand, but statistics could be gathered and new versions suggested if required. In addition to directly repeating the validated procedures, the substrate scope for each χDL can be gradually expanded by changing the substrates and adjusting key parameters—such as temperature or time—of the reaction while keeping the rest of the process unchanged. Because we have selected reactions based on popularity, the resulting set of validated χDLs covers a substantial range of common reactions and constitutes an entry point to automation of the entire organic synthesis toolbox. Further, through performing 53 procedures of highly diverse chemistry, the hardware and software of the ChemPU has been pushed to the limit and a path to full universality demonstrated. To do this, key advances have been made by incorporating a χDL-enabled flash column chromatography system in the hardware library. This means that the ChemPU can perform not only the reaction, work up, and concentration, but also the chromatographic separation of the product to directly deliver the purified compound on demand. To achieve this, we show that the platform can react in a dynamic manner, responding to detection of the product to collect the appropriate fractions.

The workflow starting from a literature procedure to a validated entry in our Chemify database is illustrated in Fig. 1. By contrast to earlier work on χDL, the focus was not on an exact translation of the original procedure text to χDL but rather on the implementation of a chemical process providing the target molecule. Following this approach allowed us not only to reproduce the literature but also to improve the processes in several instances. Chemical reactions can be captured in χDL, which represents synthetic steps as sequences of physical processes such as Add, Dissolve, Evaporate, and more. There are currently 44 steps within the χDL framework with each step having a fully customizable set of parameters. All often-used tasks in organic syntheses have a boilerplate χDL step to represent them, such as EvacuateAndRefill to establish an inert atmosphere or Separate to perform a liquid-liquid separation and extraction. The

School of Chemistry, the University of Glasgow, University Avenue, Glasgow G12 8QQ, UK.
*Corresponding author. Email: lee.cronin@glasgow.ac.uk
†These authors contributed equally to this work.

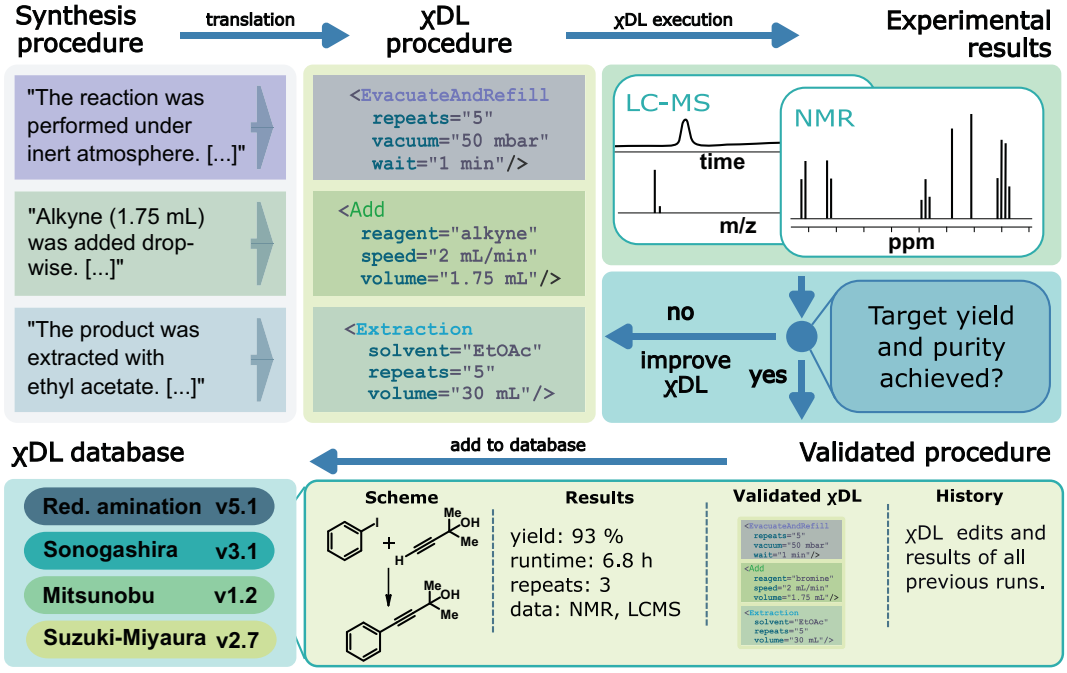


Fig. 1. Schematic of χDL protocol optimization. Chemical reactions can be captured with χDL, which represents synthetic steps as sequences of physical processes such as Add, Dissolve, and Evaporate. The initially established χDL protocol is then executed on a ChemPU and the purity and yield of the product are determined. The χDL protocol can be improved until the process meets the expectation of product purity and yield. At that stage the protocol can be added to the database as validated, backed up by the full characterization of the target product and the process development history.

χDL steps help enforce precise descriptions of the process and eliminate any ambiguity such as the number of cycles of evacuation and inert gas refill or process-critical addition speeds. To achieve this, we used our web-based Chemistry Development Environment (ChemIDE) (*33*) which aids the quick generation of χDL procedures by providing a text-to-χDL translation tool. This works by using a template library of all available χDL steps and an editor in which individual χDL steps are represented as graphical elements, which can be edited and arranged as needed (*33*). ChemIDE was used in the generation of all χDL procedures detailed in this work.

Expression of a chemical procedure in χDL does not immediately solve the problem of missing information or ambiguity present in the original prose instructions but it does provide an unambiguous path to close it. To do this, some process development and iteration may still be required to maximize yields and purity. After appropriate analysis [Nuclear magnetic resonance spectroscopy (NMR), liquid chromatography mass spectrometry, or gas chromatography mass spectrometry] of the target compound from the ChemPU execution of the χDL code, an assessment of the quality and purity of the product is made. If necessary the χDL is improved to increase the yield and purity and then executed again. The key advantage of χDL is that once a successful process has been encoded, all subsequent users who execute the code on compatible hardware can expect identical results, with no further requirements for process development. All critical knowledge needed to execute the process on qualified hardware both tangible and intangible is now captured in the χDL. At this stage the protocol can be added to the database as a validated process, backed up by the full characterization of the target product and process development history. Inclusion of process development history is a distinguishing feature of the Chemify database; by showing the results of less successful experiments and contrasting them with the final successful run, critical aspects of the process are highlighted and can be quantified.

The Chemify database persistently stores information for χDL procedures, experimental results, and relevant analyses. It is a locally hosted PostgreSQL database server containing all validated χDL scripts as described above, which can be accessed through ChemIDE (the web-based χDL development environment) or by using a Python 3–based API for automated database querying. Moreover, for end user experience, ChemIDE is equipped to display characterization parameters of each experiment such as product scale, yield, status (translated, validated, failed) and process duration. Users can submit, search, download, and reproduce trusted syntheses. The database contains final validated synthesis scripts as well as previous developmental versions, which may work to a varying degree, affording the desired products in lower yields, insufficient purities, or leading to process failures (for example, causing blockages or formation of emulsions during liquid-liquid separations) as a result of insufficient or incorrect description of the necessary process parameters for automation. Comparing failed or lower-yield experiments to successful attempts of a given specific reaction or reaction class can unveil critical aspects of the process. Further, the database also contains χDL entries that have been translated but not yet executed on a suitable automation platform. Users interested in unvalidated χDL files can access these and have the option to validate them. The χDL procedures reported here have been validated on a ChemPU, a chemistry automation platform that emulates the manual operations of a bench chemist. Although operationally simple and intuitive, the rigorous implementation means that the platform operates as a finite state machine (Fig. 2). It can be in one of a finite number of states and transitions from one state to the next based on well-defined operations. These operations are defined by the program—the χDL synthesis protocol—as well as the sensor feedback [e.g., temperature, conductivity, pressure, or ultraviolet (UV) absorbance]. The direct mapping of the χDL synthesis instructions to state transitions or "unit operations" highlights the rigorous abstraction of synthesis processes in χDL. Moreover, the clear definition of state transitions as defined in the χDL procedure is critical to ensuring the reproducibility of the χDL synthesis, including on different layouts of the ChemPU and potentially entirely different qualified hardware setups.

The ChemPU state machine consists of three logical parts: the physical input or output (I/O), digital I/O, and the processing unit. The processing unit can transition through several states based on either the initial conditions of the ChemPU or a combination of the physical and digital I/O, that is, the current conditions as defined by the sensors, the process variables, and the χDL step being executed. The execution of the χDL step according to the

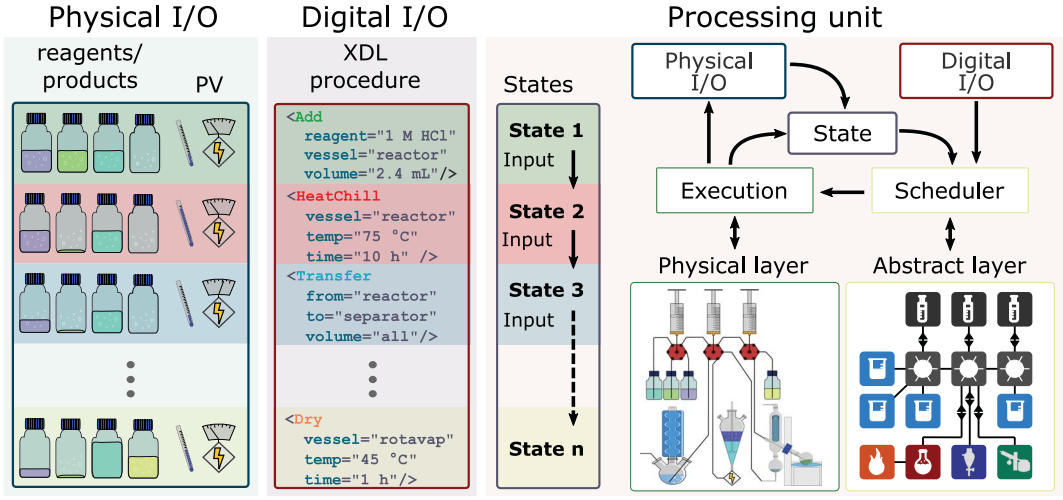


Fig. 2. The ChemPU can be regarded as a chemical state machine. The ChemPU processes the physical input (i.e., reagents, solvents) based on the digital input (i.e., the synthesis script). Each unit operation defines a route for the system to progress from the current state to the next. The state of the system is always well defined, and every detail of the synthesis is known and can be documented. PV, process variables.

scheduler gives rise to a new state to be acted upon in later steps and results in a physical change to the physical I/O, e.g., a change of location of reagents, a change in temperature, the phase boundary in a liquid-liquid separation, or the peak elution during chromatography. The scheduler resorts to a graph representation of the hardware (abstract layer) to interpret the χDL script and orchestrate the hardware for concerted tasks (e.g., moving liquid through the liquid-handling backbone). The abstract layer defines the locations and connections of the hardware devices as nodes and contains specific information on each node such as the IP address and temperature limits of the device in question. The graph file together with the χDL file can be compiled into an execution file (executable χDL or xdlexe), which is platform-specific. The strict separation of the description of the chemical process into the χDL file and the hardware platform description into the graph file ensures that the χDL file remains platform-independent. It also allows for flexibility in how the platform is designed and what its exact physical layout is. This means that each χDL can be versioned and compiled to run on any suitable platform and that the ChemPU system is highly modular, flexible, and extensible (Fig. 3).

By mirroring the unit operations of batch synthetic chemistry, the ChemPU represents a universal, programmable hardware platform for execution of synthetic chemistry as previously demonstrated (*14*–*16*). The platform can be readily expanded as a result of its modular nature, with individual modules being connected through the liquid-handling backbone, analogous to the bus of a conventional computer. Connection to the liquid-handling backbone (consisting of pumps and valves) is through a single piece of flexible tubing, which allows modules to be easily removed for maintenance or rearranged to optimize operations (e.g., by segregating aqueous and water-sensitive parts of the process). The liquid-handling backbone consists of a series of syringe pumps and valves. A typical backbone consists of six of each; however, the backbone is readily contracted or expanded to accommodate the requirements of the desired chemical process. The valves have six positions and seven ports each. Each valve in the liquid-handling backbone is connected to a pump, its nearest neighboring valves, and a waste container, and can connect to three to four different reagents, solvents, or hardware modules. The connectivity of the modules to the backbone is represented in an abstract manner by a graph as described above. Cleaning of the backbone is carried out through an automated cleaning routine that can be defined by the user to account for different types of contamination present after different procedures. In addition to the liquid-handling backbone, the ChemPU systems used to execute the syntheses reported here incorporated a reaction module consisting of a standard hotplate controlled through an Ethernet-to-serial convertor, a separator for liquid-liquid extractions equipped with an overhead stirrer for agitation, as well as a conductivity sensor for phase boundary detection; it also includes a jacketed filter for precipitation and recrystallization of products, a number of reagent flasks, a rotary evaporator, and an optional chromatography system.

Validation of literature procedures on the ChemPU

With the abstraction of chemputation, the χDL language, and the ChemPU platform, we set out to translate and automate the typical reactions from the organic chemistry toolbox. Organic chemistry encompasses an immense diversity of transformations. Despite a large degree of variety, most reactions can be classified succinctly with fewer than ten categories. Several studies have analyzed the reaction frequencies in different fields, e.g., medicinal chemistry, process chemistry, and total synthesis (*35*–*38*). There are some notable differences in the distribution of reaction classes used in synthesis, depending on the primary goal; for example, medicinal chemistry researchers may prefer transition metal–catalyzed C-C bond-forming reactions which allow for convenient generation of large numbers of related compounds for biological assays, whereas modern total synthesis relies more heavily on elaborate ring-forming reactions for assembly of complex molecular scaffolds in the fewest number of steps possible (*39*).

Additionally, although protecting-group chemistry is the cornerstone of some synthesis fields such as peptide synthesis (*40*) or carbohydrate chemistry (*41*), researchers working on total syntheses often prefer more elegant protecting-group–free approaches (*42*). Despite the minor variations, these categories embody the varied toolbox of modern organic chemistry. To represent these categories with the examples from all types of reactions we chose to translate the χDLs of these procedures and validate them with the ChemPU (Fig. 4). The carbon-carbon bond-forming reaction class was further separated into transition metal–catalyzed and transition metal–free reactions. Furthermore, a separate multicomponent reaction class was introduced as these reactions generally accomplish multiple chemical transformations in one synthetic operation. The initial reactions were chosen from the most cited papers in the journal *Organic Syntheses* (*43*). This journal is notable in the organic chemistry field in that it publishes practical methods for either synthesis of notable compounds or execution of important synthetic methods, and the submitted procedures have been repeated at least once by expert chemists independent of those who submitted the original synthesis. Although the procedures from this journal generally have a high level of detail there was still a need for some process

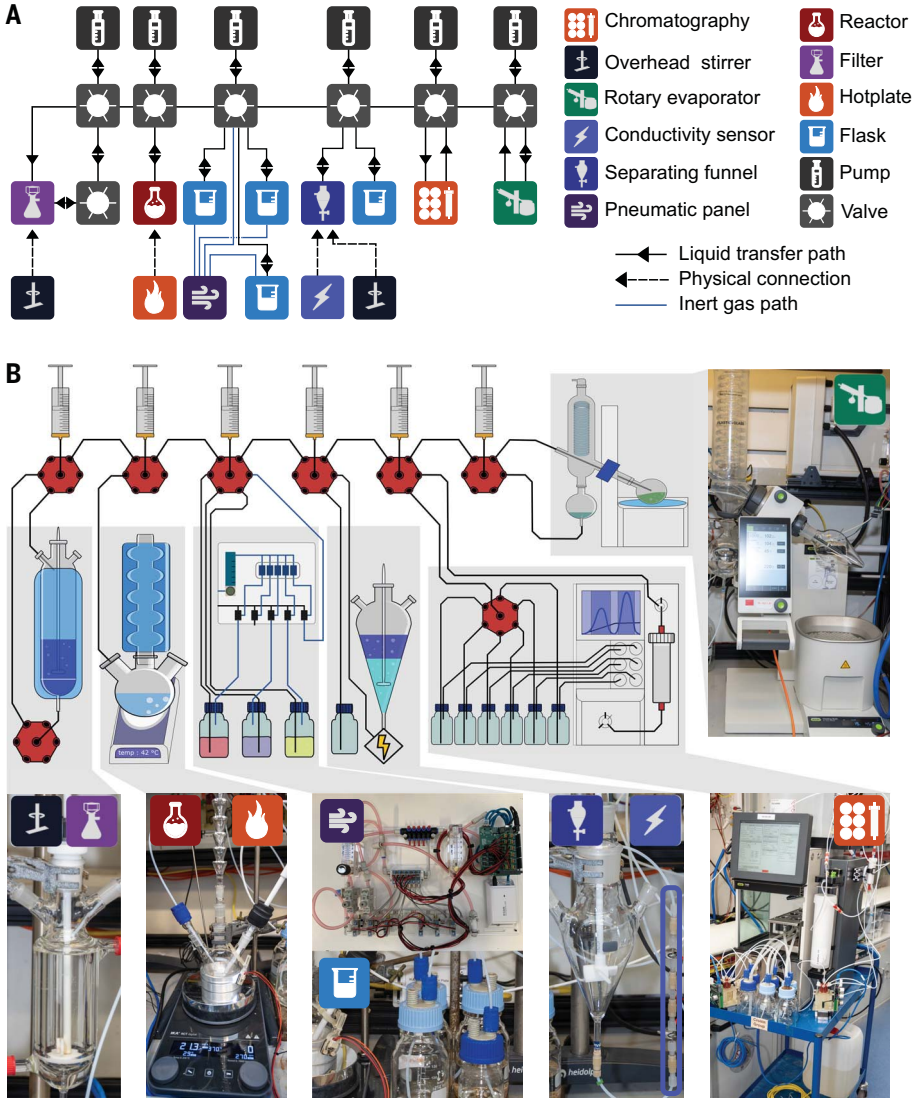


Fig. 3. The physical layout of the ChemPU and the available hardware library can be represented as a graph. (**A**) The exact state of the ChemPU for each individual synthesis is represented as a graph. The nodes indicate the modules and the edges define the tubing connections for liquid transfers and physical connections. (**B**) A schematic representation of the ChemPU with images of the individual modules. The ChemPU emulates the manual batch chemistry workflow and uses much of the typical laboratory hardware. The latest key addition to the ChemPU hardware library, as described in this paper, is a flash chromatography system which allows for fully automated purification of the reaction products. The liquid-handling backbone consists of an array of pumps and valves which transfer reagents, solvents, and solutions of intermediates between the different units of the system. Reactions are either carried out in a round bottom flask reactor or a filter. The work up is performed in a separator that is agitated with an overhead stirrer. The phase boundary is detected with a conductivity sensor. Solutions are concentrated in a rotary evaporator. Column chromatography is performed on a flash column chromatography machine.

development, highlighting the difficulty of capturing all necessary information in an unstructured prose text format as opposed to χDL. Selecting these highly cited papers from *Organic Syntheses* covered the top reaction classes but provided an uneven distribution. Hence, further examples were manually selected from notable literature sources to achieve a more balanced representation of the organic chemistry toolbox with our dataset.

The reactions chosen for each of the categories include well-established classical reactions and important contemporary reactions, as well as some more unconventional synthetic transformations (Fig. 4); for a comprehensive list of all translated reactions, see supplementary materials fig. S143. The selected transition metal–catalyzed carbon-carbon bond forming reactions included commonly used Suzuki, Heck, and Sonogashira couplings, as well as a stereoselective Carroll rearrangement.

The transition metal–free carbon-carbon bond-forming reaction class encompasses such classical reactions as the Wittig reaction, Friedel-Crafts alkylation, and the Aldol and Claisen condensations. Different types of heteroatom alkylations are represented by palladium-catalyzed Buchwald-Hartwig coupling, copper-catalyzed alkylation, S_NAr reaction of heteroarenes, and reductive amination reactions. Functional group interconversions include a Mitsunobu reaction, nitrile formation, and esterifications, among others. Manipulations of protecting groups include common boc, benzyl, and tosyl groups. Ring and heterocycle formations include both classical syntheses such as the Fischer indole synthesis and a more exotic formation of a trisubstituted pyrylium salt. Reduction and oxidation reactions span conventional hydride reduction, Jones oxidation, and a palladium-catalyzed hydrogen transfer reaction. Finally, the multicomponent reactions include the well-known Ugi reaction as well as other more unusual cascade reactions and one-pot multistep manifolds. This diverse set of reactions (for a comprehensive list of all validated and translated reactions, see supplementary materials figs. S140 and S143, respectively) covers the standard organic chemistry toolbox. Crucially, automating further reactions simply requires translation of the original synthesis procedure to χDL.

The average procedure consists of 20 discrete, high-level instructions such as Add, Separate, and Evaporate with some procedures having up to 40 such instructions (Fig. 4C). Unpacking these high-level χDL steps into the corresponding unit operation—e.g., StartStir, WaitForTemp, ApplyVacuum—gives an average of 266 operations that have been executed. The successful executions of all χDL scripts took >1000 hours of chemputation across seven different systems. This figure only includes the operations from the final iteration of each χDL protocol and includes the reaction time but does not account for asynchronous steps, i.e., steps in which two processes are running in parallel on the same ChemPU hardware, such as a cleaning step for a rotary

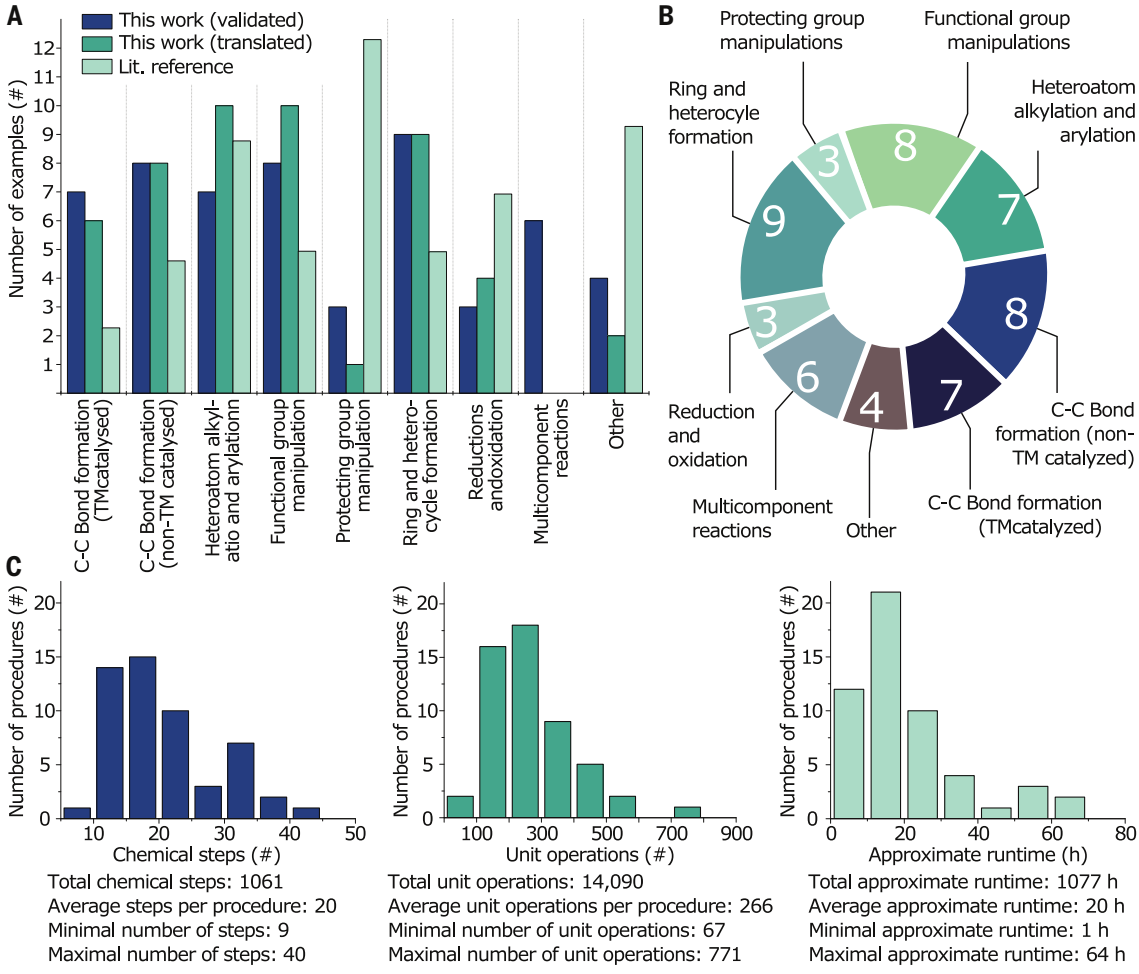


Fig. 4. A representative selection of most-used reaction classes have been translated to χDL and validated on a ChemPU. (**A**) Number of examples per reaction class. Once a literature procedure has been captured by χDL it is marked as "translated." When a translated procedure is successfully executed on a ChemPU it is moved to the "validated" class of χDL scripts. For reference, the average frequency of reactions over the fields of medicinal chemistry, process chemistry, and total syntheses are shown (*35*–*38*). (**B**) The distribution of validated reactions and a specific example for illustration is shown. (**C**) Chemical operations, unit operations, and total runtime per procedure.

evaporator running at the same time as a reaction. The yields of the reactions performed on the ChemPU were in general comparable to that of the literature yields after a period of process development. This could be required to fill the gaps in the original protocol and is common to all synthetic development whether manual or automated, or to adapt elements of the protocol not amenable to automation, such as unexpected formation of precipitates that lead to blocked lines. A selection of reactions is shown in Fig. 5 to illustrate the performance of the platforms and give specific examples to show the breadth of chemistry that has been performed.

Automation of a diverse set of reactions on the ChemPU

The system is tolerant of moisture-sensitive or highly reactive reagents such as potassium bis (trimethylsilyl)amide (KHMDS) used in a copper-mediated alkynylation of a carbamate to afford **3**, boron trifluoride used in a Friedel-Crafts alkylation of a steroid estrone to afford derivative **4**, or Eaton's reagent (10% phosphorus pentoxide solution in methanesulfonic acid) used in a Fischer indole synthesis of **6**. Additionally, reactions requiring inert atmosphere were successfully executed on the platform including a palladium-catalyzed enantioselective Carroll rearrangement to give **5**. Procedures of up to 90 mmol scale were efficiently executed on our ChemPU platform. Conveniently, once a χDL script is produced, a particular reaction can be scaled up or down within the constraints of the available vessel sizes and the chemical process, such as safety considerations or heat and mass transfer (see supplementary materials section 3, fig. S141, for the full distribution of reaction scales). The χDL procedures for the generation of more complex products arising from multicomponent and cascade reactions were also successfully executed on the platform. For example, a Petasi/Diels-Alder cyclization cascade has been used for rapid generation of a scaffold containing multiple stereogenic centers **7**, with potential for further derivatization in a library synthesis. Similarly, a copper(I)-catalyzed three-component coupling/palladium(0)-catalyzed annulation cascade was also successfully applied, affording product **8** which contains the indenoisoquinoline scaffold.

Expanding the substrate scope

The substrate scope of validated χDL procedures can be expanded by generating a compound library with the ChemPU. One particularly attractive prospect is the use of validated χDL procedures for the construction of large libraries of compounds for biological screening. Such libraries could conveniently be accessed simply by changing the starting materials without major modifications to the synthesis scripts; i.e., once a process has been established it can be applied to many different substrates as a general procedure by only

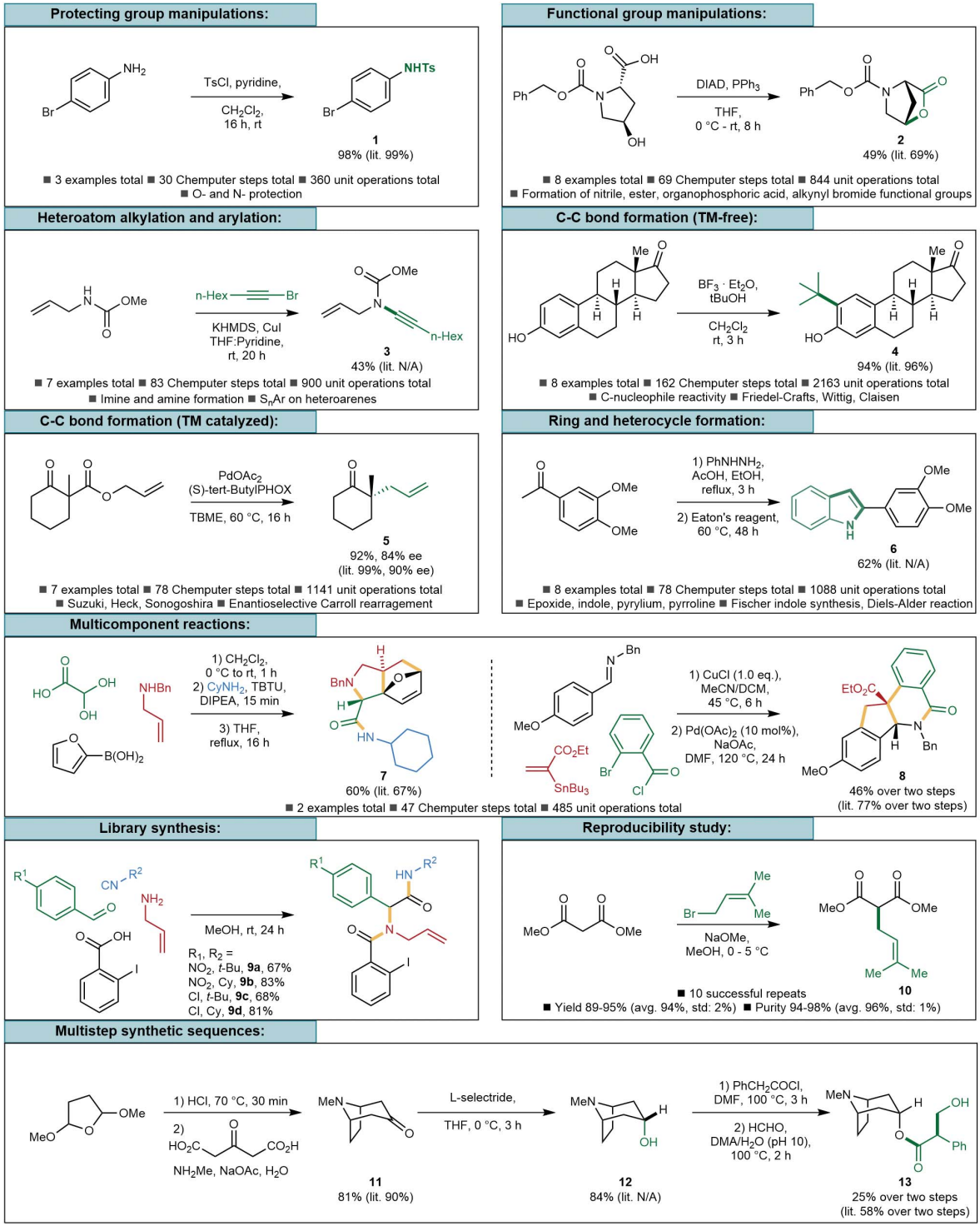


Fig. 5. Representative examples of χDL procedures validated on the ChemPU. Transformations from all the main reaction classes afforded products in yields comparable to those reported in the literature for manual synthesis. Additionally, multicomponent reactions were performed and a small library of compounds was prepared by varying the starting materials of one of the multicomponent reactions. Finally, the reproducibility of a validated χDL procedure has been examined and a χDL procedure of a multistep synthetic sequence has been validated. The counts for total steps and total unit operations include all examples in a given category.

varying key parameters such as the substrates, reaction solvent, and reaction time. To showcase such an approach, a small library of α-acylamino amides **9a to 9d** was synthesized through a multicomponent Ugi reaction. To do this, we conducted simultaneous execution of multiple or "multithreaded" reactions in parallel on the ChemPU by using reactant combinations from two different isocyanide and two aldehyde starting materials affording four structurally related α-acylamino amide products. Further expansion of the set of reactants

used would rapidly expand the number of products generated and allow for swift generation of larger libraries.

Reproducibility of the ChemPU synthesis

To examine the consistency and reliability of executing the curated χDL procedures, we set out to repeat the same reaction protocol multiple times on the ChemPU platform. An alkylation of malonate ester (affording **10**) was chosen as a suitable reaction for the reproducibility study, as accurate temperature control and rate of addition are key to the success of the process. After the initial process development, a validated χDL procedure script was obtained and the reaction protocol was successfully replicated 10 times in 12 attempted runs. The two failures were caused by incorrect phase boundary determination during liquid-liquid separations; product could have been recovered through manually restarting the system, but that was not done here. Crucially, execution of the curated χDL procedure reliably afforded the product in consistent yields (avg. 94%, min 89%, std: 2%) and purities (avg. 96%, min 94%, std: 1%). Together with the ability to generate libraries of compounds, the ChemPU can be used to automate the highly repetitive work of generating multiple batches of the same material or repeating the same reaction with different substrates once the initial protocol has been set up.

Multistep synthesis

The versatility of the platform is further demonstrated by the ability to execute multistep synthetic sequences. Atropine **13**, an anticholinergic medication used in treatment of nerve agent poisoning, was synthesized in four steps from simple commercially available starting materials. Synthetic protocols for individual steps from multiple sources—as well as a reduction protocol that was previously reported for related substrates but not for the synthesis

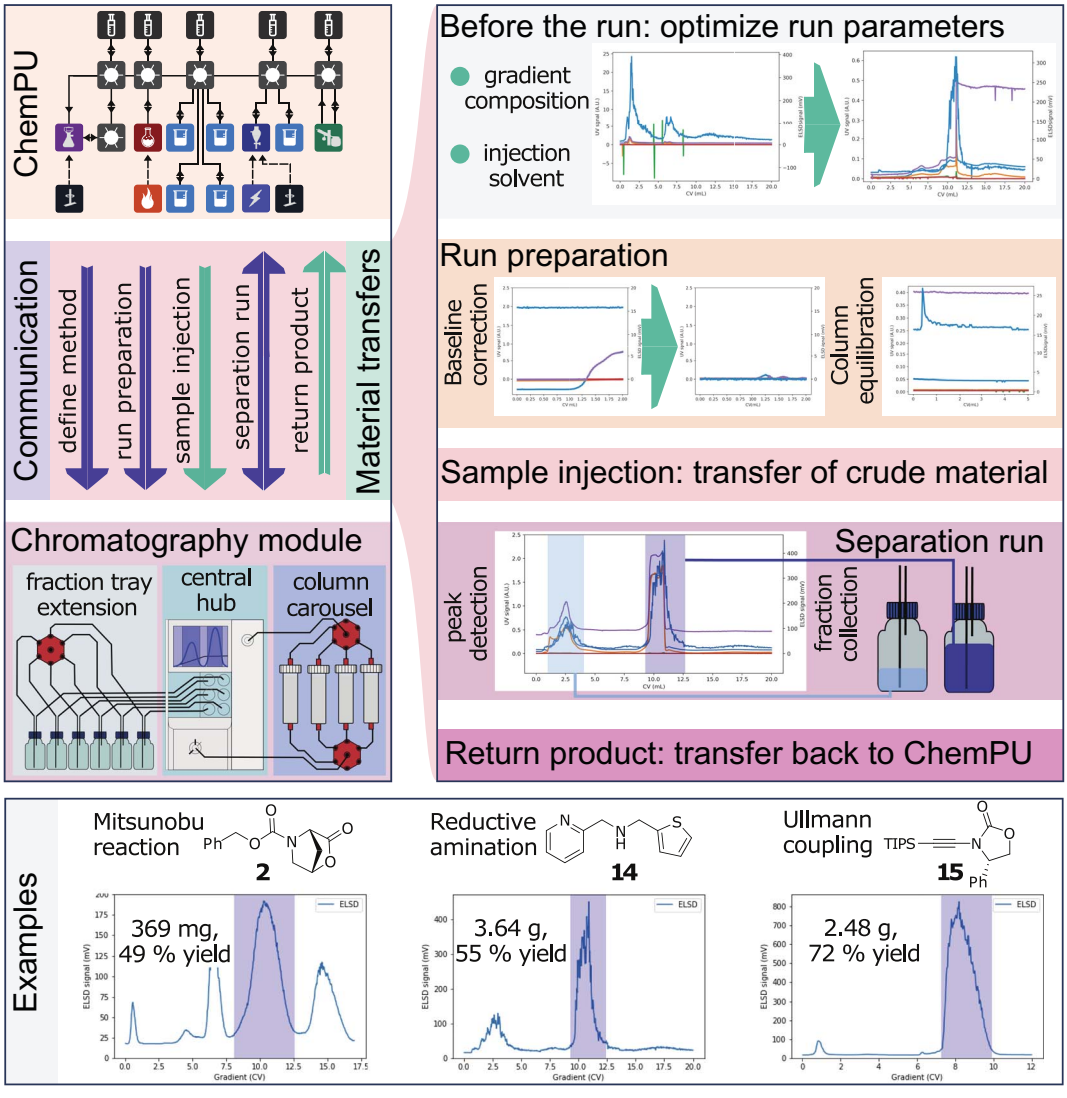


Fig. 6. The chromatography module. This component stands out from the ChemPU hardware library in terms of the complexity of the information and material flow between the module and the controller and the other hardware. The chromatography method consists of two crucial parts: the sample injection protocol and the solvent gradient. Once these parameters have been optimized the separation run can be initiated. The ChemPU defines the run parameters on the commercial chromatography unit. Then the run preparations (baseline correction over the gradient, and column equilibration) are performed. Next the sample is injected onto the column. During the gradient run the chromatography machine sends the detector signals to the ChemPU controller in real time. The controller performs the peak detection and triggers the fraction collection mechanism of the chromatography machine. When the separation run is complete the product peak is identified and transferred to the next module (usually the rotary evaporator).

of **12**—were successfully converted to χDL procedures. The ability to efficiently execute multistep reaction protocols combined with the reliability offered by reproducible execution of a well-defined synthesis script reaffirms the universality of the platform toward the breadth of synthetic organic chemistry.

Fully automated purification on the ChemPU

Chromatographic separation of the product compound from a reaction is the go-to method of purification for small- and medium-scale organic syntheses. Many commercially available chromatography systems exist for assisting lab-based chemists in chromatographic separations. However, these systems still require a substantial amount of user interaction. For example, the crude material must be manually loaded onto the column and the product fractions must be manually identified, washed out of the fraction vials, and combined. Further, these commercial systems require user interactions at several different stages, thus tying the chemist to the lab even if it is only for a trivial task such as loading the sample onto the column. To integrate the Buchi Pure C-815 chromatography system with the ChemPU, two auxiliary hardware units were built: a column carousel that allows preinstallation of different columns on the system and an extension to the fraction tray. The latter allows for recovery of the product fraction by the ChemPU. The first operation that is challenging to automate is the sample loading onto the column. The laboratory-based chemist usually chooses between dry-loading and liquid injection of the sample. We aimed to implement the liquid injection method which ties in nicely with the liquid-handling backbone of the ChemPU; further, the liquid injection sample loading method entailed little process development, requiring only the identification of a suitable solvent mixture and volume to dissolve the crude material. The second challenge to full automation of normal-phase chromatography is to reliably select the product peak. Usually, chemists need to analyze individual fractions by thin-layer chromatography, mass spectrometry, or NMR after chromatographic separation. For the ChemPU integration of the module several alternative options were considered. We found that considering the UV/visible response or the signal from the elastic light scattering detector of the eluting fractions and choosing the peak with the largest area under the curve for a specified signal trace gave the best trade-off between reliability and flexibility; for a given well-performing reaction the product peak can correctly be identified independent of the exact retention time. Moreover, this method does not rely on more elaborate product identification such as mass spectrometry or NMR.

Once the method is developed and coded in χDL it can be executed on the ChemPU or equivalent automation system as shown in Fig. 6. The platform controller starts the chromatography process by defining the run parameters on the commercial chromatography unit (central hub), such as flowrate and detector settings. The actual run preparations, such as baseline corrections and the equilibration of the column, are then executed. Next, the sample of crude material is dissolved, transferred to the chromatography machine, and injected onto the column. The sample injection process also includes a rinsing sequence to minimize loss of material during the sample dissolution and transfer. Once the sample loading is complete the gradient run is commenced. During the gradient run the chromatography machine continuously reads the detector signals and sends them to the ChemPU controller software in real time.

The ChemPU controller then performs the peak detection and triggers the fraction collection mechanism of the chromatography machine. The controller also keeps track of fraction vial filling levels and various run parameters such as back pressure buildup, solvent vapor levels, and solvent levels of the gradient solvents and the solvent waste drum. If any of these parameters exceed the specified threshold, an appropriate error-handling routine is initiated that pauses the chromatographic separation in a controlled way. When the separation run is complete the product peak is identified and transferred to the next module (usually the rotary evaporator). The crude material is typically transferred from the rotary evaporator to the chromatography module followed by transfer of the purified product back from the chromatography module to the rotary evaporator, and as such the rotary evaporator flask needs to be cleaned in between. Hence an optional cleaning routine for the target vessel of the purified product has been implemented and can be performed during the chromatographic separation. The integrated chromatographic separation was used for three reactions. The process of these chromatographic separations has been captured by χDL, specifying every minute and critical detail in a concise, easy-to-understand way. Hence, reproduction of the chromatographic separations on another ChemPU or equivalent system or even manually with a commercially available chromatography machine is readily possible.

Outlook

We have shown how the chemical synthesis literature can be easily converted to a universal chemical code that can run on any robot capable of chemputation; the only requirements for this are a batch reactor, a separator, evaporator, and purification system. This means that potentially many different robotic approaches will be able to use identical χDL codes to produce identical results. The use of a χDL Chemify database will not only facilitate reproduction of published procedures but also provide the community with a rich source of validated data amenable to state-of-the-art machine learning for reaction optimization, route planning, increased safety, and reduced environmental impact of synthesis while substantially reducing labor for bench chemists repeating well-known procedures.

REFERENCES AND NOTES

1. W. A. Warr, *Mol. Inform.* **33**, 469–476 (2014).
2. M. Baker, *Nature* **533**, 452–454 (2016).
3. H. Gelernter, J. R. Rose, C. H. Chen, *J. Chem. Inf. Comput. Sci.* **30**, 492–504 (1990).
4. M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* **555**, 604–610 (2018).
5. O. Engkvist *et al*., *Drug Discov. Today* **23**, 1203–1218 (2018).
6. B. A. Grzybowski, K. J. Bishop, B. Kowalczyk, C. E. Wilmer, *Nat. Chem.* **1**, 31–36 (2009).
7. I. W. Davies, *Nature* **570**, 175–181 (2019).
8. N. Matosin, E. Frank, M. Engel, J. S. Lum, K. A. Newell, *Dis. Model. Mech.* **7**, 171–173 (2014).
9. M. Trobe, M. D. Burke, *Angew. Chem. Int. Ed.* **57**, 4192–4214 (2018).
10. J. Li *et al*., *Science* **347**, 1221–1226 (2015).
11. C. W. Coley *et al*., *Science* **365**, eaax1566 (2019).
12. T. Jiang *et al*., *Chem. Sci.* **12**, 6977–6982 (2021).
13. A. C. Bédard *et al*., *Science* **361**, 1220–1225 (2018).
14. S. Steiner *et al*., *Science* **363**, eaav2211 (2019).
15. S. H. M. Mehr, M. Craven, A. I. Leonov, G. Keenan, L. Cronin, *Science* **370**, 101–108 (2020).
16. D. Angelone *et al*., *Nat. Chem.* **13**, 63–69 (2021).
17. P. G. Nantermet, *Chem* **1**, 335–336 (2016).
18. Z. Wang, W. Zhao, G. F. Hao, B. A. Song, *Drug Discov. Today* **25**, 2006–2011 (2020).
19. R. B. Merrifield, *Science* **150**, 178–185 (1965).
20. O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **291**, 1523–1527 (2001).
21. G. Alvarado-Urbina *et al*., *Science* **214**, 270–274 (1981).
22. M. Legrand, P. Bolla, *J. Automat. Chem.* **7**, 31–37 (1985).
23. S. B. Boga *et al*., *React. Chem. Eng.* **2**, 446–450 (2017).
24. B. Burger *et al*., *Nature* **583**, 237–241 (2020).
25. A. G. Godfrey, T. Masquelin, H. Hemmerle, *Drug Discov. Today* **18**, 795–802 (2013).
26. B. P. MacLeod *et al*., *Sci. Adv.* **6**, eaaz8867 (2020).
27. H. Okamoto, K. Deuchi, *Lab. Robot. Autom.* **12**, 2–11 (2000).
28. A. Orita, Y. Yasui, J. Otera, *Org. Process Res. Dev.* **4**, 333–336 (2000).
29. Y. Tanaka, S. Fuse, H. Tanaka, T. Doi, T. Takahashi, *Org. Process Res. Dev.* **13**, 1111–1121 (2009).
30. S. Chatterjee, M. Guidi, P. H. Seeberger, K. Gilmore, *Nature* **579**, 379–384 (2020).
31. A. J. S. Hammer, A. I. Leonov, N. L. Bell, L. Cronin, Chemputation and the Standardization of Chemical Informatics. *JACS Au* **1**, 1572–1587 (2021).
32. L. Wilbraham, S. H. M. Mehr, L. Cronin, *Acc. Chem. Res.* **54**, 253–262 (2021).
33. M. Craven, G. Keenan, A. Khan, M. Lee, L. Wilbraham, ChemIDE. https://croningroup.gitlab.io/chemputer/xdlapp/ (2021).
34. Cronin Group, Chemify Database: https://croningroup.gitlab.io/chemputer/chemify-database/.
35. S. D. Roughley, A. M. Jordan, *J. Med. Chem.* **54**, 3451–3479 (2011).
36. N. I. Vasilevich, R. V. Kombarov, D. V. Genis, M. A. Kirpichenok, *J. Med. Chem.* **55**, 7003–7009 (2012).
37. N. Schneider, D. M. Lowe, R. A. Sayle, M. A. Tarselli, G. A. Landrum, *J. Med. Chem.* **59**, 4385–4402 (2016).
38. J. S. Carey, D. Laffan, C. Thomson, M. T. Williams, *Org. Biomol. Chem.* **4**, 2337–2347 (2006).
39. P. S. Baran, *J. Am. Chem. Soc.* **140**, 4751–4755 (2018).
40. A. Isidro-Llobet, M. Alvarez, F. Albericio, *Chem. Rev.* **109**, 2455–2504 (2009).
41. A. G. Volbeda, G. A. Marel, J. D. C. Codée, in *Protecting Groups*. S. Vidal, Ed. (Wiley, 2019), pp. 1–27.

42. R. A. Fernandes, P. Kumar, P. Choudhary, *Chem. Commun.* **56**, 8569–8590 (2020).
43. W. R. Roush *et al*., Organic Syntheses (2021); http://www.orgsyn.org/.
44. S. Rohrbach synthesis literature database in the Chemputer, Zenodo (2022); doi:10.5281/zenodo.6534009.

ACKNOWLEDGMENTS

We thank BUCHI for supplying us with a pure C-815 chromatography system and API to interface it with the Chemputer software package. The authors gratefully acknowledge the assistance of D. Doran and V. Sandoval in the preparation of and interfacing with the χDL Chemify database. **Funding:** We gratefully acknowledge financial support from the EPSRC (EP/L023652/1, EP/R020914/1, EP/S030603/1, EP/R01308X/1, EP/S017046/1, and EP/S019472/1), the ERC (670467 SMART-POM), the EC (766975 MADONNA), and DARPA (W911NF-18- 2-0036, W911NF-17-1-0316, and HR001119S0003). **Author contributions:** L.C. conceived the concept, architecture, and programming approach. S.R., M.Š., A.P., M.S., H.M.M., E.T., A.I.L., and A.H. configured the robots, ran the synthetic protocols, and characterized the products. G.K. and A.K. helped with the development of the database and integration with the ChemIDE. L.C. wrote the paper together with S.R., M.Š., and G.C. with help from all authors. **Competing interests:** L.C. is the founder of Chemify Ltd. L.C. is listed as an inventor on the UK patent GB 2209476.7., which describes this system. **Data and materials availability:** Supplementary materials include full details to reproduce this work, including instructions for how to build and run the platform. Additional details of the electronic and mechanical components of the platform, videos of the platform working, and the software to produce and run the χDL files, and the raw analytical data for all experiments are available at Zenodo (*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0058
Materials and Methods
Figs. S1 to S143
Tables S1 to S34
References (*45*–*101*)

Submitted 6 January 2022; accepted 26 May 2022
10.1126/science.abo0058

BIOENGINEERING

Recreating the heart's helical structure-function relationship with focused rotary jet spinning

Huibin Chang[1]†, Qihan Liu[1,2]†, John F. Zimmerman[1]†, Keel Yong Lee[1], Qianru Jin[1], Michael M. Peters[1], Michael Rosnach[1], Suji Choi[1], Sean L. Kim[1], Herdeline Ann M. Ardoña[1,3], Luke A. MacQueen[1], Christophe O. Chantre[1], Sarah E. Motta[1,4], Elizabeth M. Cordoves[1], Kevin Kit Parker[1]*

Helical alignments within the heart's musculature have been speculated to be important in achieving physiological pumping efficiencies. Testing this possibility is difficult, however, because it is challenging to reproduce the fine spatial features and complex structures of the heart's musculature using current techniques. Here we report focused rotary jet spinning (FRJS), an additive manufacturing approach that enables rapid fabrication of micro/nanofiber scaffolds with programmable alignments in three-dimensional geometries. Seeding these scaffolds with cardiomyocytes enabled the biofabrication of tissue-engineered ventricles, with helically aligned models displaying more uniform deformations, greater apical shortening, and increased ejection fractions compared with circumferential alignments. The ability of FRJS to control fiber arrangements in three dimensions offers a streamlined approach to fabricating tissues and organs, with this work demonstrating how helical architectures contribute to cardiac performance.

The heart's musculature is organized in a helical fashion, with cardiomyocytes in the left ventricle smoothly transitioning transmurally from a left- to right-handed helix (*1*). This helical alignment results in a "wringing motion" that was first described in 1669 by Lower (*2*). Over the past half-century, it has been argued that this helical arrangement represents a fundamental structural design critical to achieving large ejection fractions (EFs) (*1*, *3*–*6*). Although in vivo studies have proven to be a powerful tool, both in defining these structural features (*7*–*9*) and in correlating misalignments with cardiac disease and reduced ventricle function (*5*, *9*, *10*), they are often characterized by concomitant changes in protein expression and metabolism (*11*) and are limited by their ability to control cardiac alignments (*12*, *13*). This makes it difficult to distinguish between biomolecular and biomechanical contributions to cardiac dysfunction. However, understanding how helical structures contribute to cardiac function is important, as some cardiomyopathies can exhibit maladaptive tissue remodeling (*14*, *15*), which may result in more circumferential alignments of the musculature (*12*, *13*, *15*).

To study the role of helical muscular alignment on cardiac function, it is necessary to recreate the multiscale architectures of the heart, with both controlled alignments and three-dimensional (3D) geometries. In the heart, helical alignments are supported by extracellular matrix (ECM) proteins such as collagen fibrils (*8*, *16*), the diameters of which are on the order of a single micrometer. Approaches that use 3D extrusion printing have demonstrated important milestones toward replicating such structures, including the production of microphysiological devices (*17*, *18*), microvasculature systems (*19*), and spontaneously beating heart models (*20*, *21*). However, reproducing the fine spatial features required to potentiate muscular alignment while retaining practical production rates is difficult. This is the result of a fundamental limitation in 3D extrusion printing, whereby throughput declines rapidly with respect to feature size. Consider a full-size human heart. Printing the ECM components at current resolutions (~250 μm feature sizes) takes hours to days (*22*), but at native feature sizes (1 μm) could take hundreds of years if current scaling trends are followed (eq. S1). Because fiber-spinning techniques can reproduce these fine spatial features with higher throughputs, they offer a potential solution and have been used previously to engineer tissue scaffolds, such as heart valves (*23*) and ventricle models (*24*). However, unlike 3D printing, fiber-spinning approaches often fail to recreate complex 3D geometries while maintaining controlled alignments.

We developed focused rotary jet spinning (FRJS), an additive manufacturing method that uses centrifugal jet spinning to rapidly form polymeric micro/nanofibers, which are then focused and spatially patterned by means of a controlled airstream. This approach allows for the rapid manufacturing of fiber constructs with programmable fiber alignments in three dimensions. Because fibers can be used to direct tissue formation, it is possible to recreate complex anatomies that were hitherto unobtainable via current biofabrication techniques. Basing our study on Sallin's analytical model of the heart (*3*), we used FRJS to manufacture both helically aligned (HA) and circumferentially aligned (CA) tissue-engineered models of the left ventricle, showing that the biomechanics of these single-layer ventricle models is consistent with theoretical predictions from the past 50 years (*3*, *4*, *25*). Collectively, this work reveals how a FRJS-based approach to biofabrication can be used to rapidly create robust model organs, with increased geometric complexity.

Fiber manufacture using focused rotary jet spinning

Methods such as electrospinning (*26*), melt blowing (*27*), pull spinning (*28*), and traditional centrifugal spinning (*29*) can form micro/nanofibers. When producing single-micrometer features, these techniques can offer orders of magnitude greater throughputs than those achieved by 3D extrusion printing (*30*). However, fiber-spinning methods are often less precise, failing to accurately recreate complex 3D geometries and alignments. This lack of precision is partly because fiber formation and patterning are traditionally interrelated during fiber production. In FRJS, decoupling the fiber formation and patterning processes is realized by creating a focused stream of preformed fibers (Fig. 1, A to C; fig. S1, and movie S1). Rotary jet spinning produces fibers by centrifugal force, pushing polymer solutions through a small orifice in the spinneret. Subsequent jet elongation results in free-floating single-micrometer fibers (*29*). This allows for a formation period that is independent of fiber patterning, generating a cloud of fibers that surrounds the spinneret. Next, in a phenomenon known as entrainment (*31*), fibers are pulled into a jet stream blown from the center of the spinneret. Because air speeds are orders of magnitude slower outside of the jet (fig. S2), this entrainment process minimally perturbs fiber formation (fig. S3). The jet stream then allows for fibers to become aligned and confined in a small region, focusing them for patterning.

Focusing the fibers in this manner enables them to be conformally deposited, meaning that fibers will adhere well to convex structures or features with curvature radii that exceed the focal spot size (fig. S4). To demonstrate this focusing effect in FRJS, polycaprolactone fibers were collected at regular intervals from within the airstream to measure deposition profiles. This showed that, at the narrowest point, 95% of fiber deposition occurred within a spot size of 5.0 ± 0.3 cm (2σ of Gaussian distribution) (Fig. 1B, ii, and fig. S4). To demonstrate conformal deposition, fibers were then spun onto both a rotating collection mandrel (movie S2) and a 1-cm ventricle model

[1]Disease Biophysics Group, John A. Paulson School of Engineering and Applied Science, Harvard University, Boston, MA 02134, USA. [2]Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA 15261, USA. [3]Department of Chemical and Biomolecular Engineering, Samueli School of Engineering, University of California, Irvine, CA 92697, USA. [4]Institute for Regenerative Medicine, University of Zurich, Zurich, Switzerland.
*Corresponding author. Email: kkparker@seas.harvard.edu
†These authors contributed equally to this work.

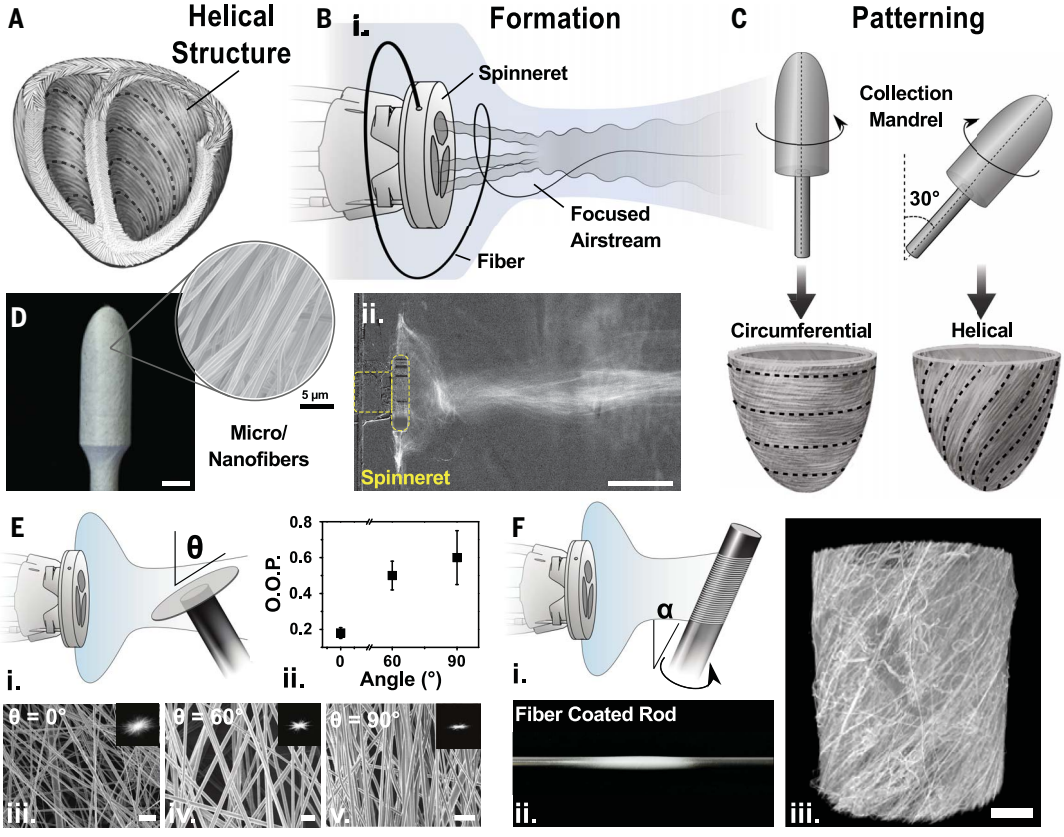


Fig. 1. Focused rotary jet spinning for producing helical structures. (**A**) Schematic diagram of the helical alignment of a human heart. (**B** and **C**) FRJS uses focused air to separate fiber manufacture into formation (B, i) and patterning (C) phases, allowing for controlled alignments during deposition. (B, ii) Differential contrast projection of the fiber stream (maximal projection; scale bar, 5 cm). (**D**) Image of polycaprolactone spun onto a mandrel (scale bar, 5 mm), with corresponding scanning electron microscopy (SEM) image, showing aligned fiber formation (mean fiber diameter, ~900 nm; scale bar, 5 μm). (**E**) Schematic diagram showing that collection angle (θ) dictates fiber deposition alignment (i), with OOP indicating the relative average alignment (error bars, mean ± SD) (ii). SEM micrographs (iii to v; 0°, 60°, and 90°, respectively) with corresponding 2D Fourier transforms inset, indicating that the degree of alignment is based on collection angle (scale bars, 5 μm). (**F**) Schematic diagram showing HA fiber manufacture based on angle (i), α, with a representative fiber-coated rod (ii) and reconstructed micro–computed tomograph of a HA scaffold (iii; scale bar, 200 μm).

(Fig. 1D). In each case, targets were coated in ~5 min, forming standalone structures that could be removed and manipulated. Additionally, we showed that this approach could be used with a variety of material compositions, such as nylon, polyurethane, and gelatin, while maintaining single-micrometer fiber diameters (fig. S5).

Controlled distribution of fiber alignment

Envisioning that micro/nanofibers could mimic the structural features of ECM proteins, we examined how FRJS could be used to control the anisotropic distribution of fiber alignments. We hypothesized that the alignment of fibers in the air stream could enable controlled deposition, in which tangential collection (θ = 0°) should minimally perturb airflow, while head-on deposition (θ = 90°) would enable divergent patterning. To test this hypothesis, the collector angle relative to the fiber stream was modulated during deposition (Fig. 1E), with the orientation order parameter (OOP) used as a metric of subsequent fiber organization (*32*). The resulting fibers displayed angle-dependent anisotropy, with tangential, intermediate, and perpendicular collection leading to highly anisotropic (θ = 0°, OOP = 0.60 ± 0.15), intermediate (θ = 60°, OOP = 0.50 ± 0.08), and randomly aligned (θ = 90°, OOP = 0.18 ± 0.03) fiber distributions, respectively.

This ability to control fiber orientation suggested that more complex patterning could be achieved by moving the target relative to the stream. To test whether FRJS could recreate key geometric features of the heart, such as the helical alignments and laminar tissue structures identified in a rodent model (fig. S6, A and B), fibers were collected on an inclined rotating cylinder to generate helical alignments (Fig. 1F and movie S3) and on an incrementally rotating disk to generate multilayered fiber sheets (fig. S6C and movie S4). We observed that FRJS was able to accurately reproduce these structures, as confirmed by x-ray micro–computed tomography (μCT). This suggested that FRJS could be effectively employed in hierarchical biofabrication, in which the focused air stream provides gross structural morphology (centimeter scale), whereas fibers provide fine structural (single-micrometer scale) cues to promote tissue morphogenesis.

Building helically aligned models of the left ventricle

We reasoned that the ability of FRJS to control fiber alignment could be used to test Sallin's hypothesis (*3*) and the advantages of helically versus circumferentially aligned ventricular geometries. Basing our experimental approach Sallin's analytical model of the heart (*3*), we then used FRJS to produce both CA (α = 0°) and HA (α = 30°) (60° with respect to the ventricle's long axis) single-layer models of the left ventricle (Fig. 2A). HA fiber orientations were selected on the basis of energy minimization constraints, with angles between 20° and 45° predicted to be the most energy efficient for pumping fluids (*4*). Fiber scaffolds were designed to have mechanical properties consistent with those of human heart tissues (fig. S7 and table S2).

After fiber manufacture, gelatin fiber scaffolds were then seeded with either primary neonatal rat ventricular myocytes (NRVMs) or human induced pluripotent stem cell–derived cardiomyocytes (hiPSC-CMs), resulting in confluent tissues with depths between one to three layers of cardiomyocytes (fig. S8). hiPSC-CMs were differentiated in vitro for 15 days total before seeding, forming visible spontaneous contractions after 7 days and staining positive for sarcomeric α-actinin (fig. S9). To ensure that FRJS fiber spinning could be used to direct tissue alignment, we first examined NRVMs seeded onto laminar

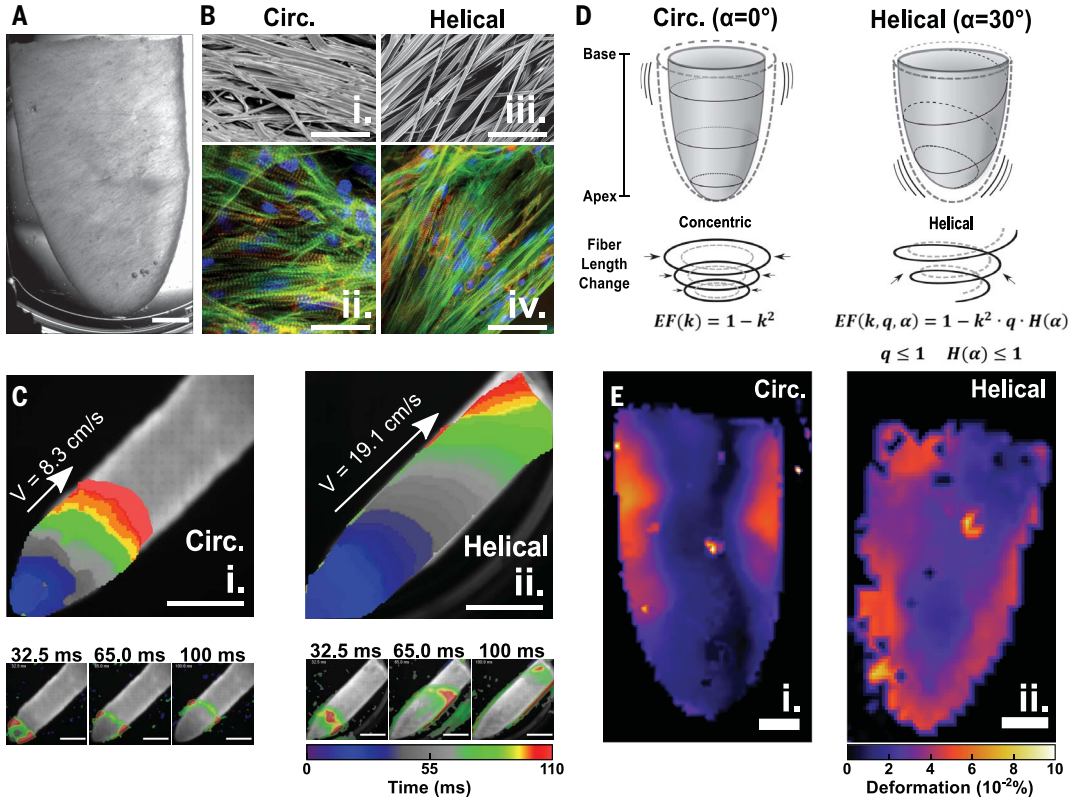


Fig. 2. Tissue scaffolds with controlled helical alignments. (**A**) Bright-field micrograph of a HA ventricle model (gelatin fibers seeded with NRVMs; scale bar, 2 mm). (**B**) SEM micrograph of fibers from CA [(i), α = 0°] and HA [(iii), α = 45°] cylinders, with corresponding immunofluorescent staining of cardiomyocytes (ii and iv) [NRVM; blue, DAPI (4′,6-diamidino-2-phenylindole); green, f-actin; red, sarcomeres], showing that fibers help direct tissue alignment (scale bars, 50 μm). Circ., circumferential. (**C**) Isochrones (top) with corresponding still frames (bottom), indicating calcium transience along an extended ventricle surface, showing increased transverse wave propagation for HA scaffolds. Tissues were point-stimulated apically [(i), CA; (ii), HA] (scale bars, 5 mm). (**D**) Schematic diagram of a CA (left) and HA (right) ventricle, illustrating differences in wall displacement during contraction. CA contracts as concentric rings, whereas HA follows a wringing motion, resulting in different predicted EFs [k, strain, q, apical shortening, $H(\alpha)$, radial shortening]. (**E**) Deformation maps generated by CA (left) and HA (right) ventricle models during contraction (scale bar, 2 mm).

tissue constructs. This resulted in tissues with a highly anisotropic distribution and cell alignment that conformed to the underlying fiber orientation (Fig. 2B and fig. S10). Next, after seeding 3D scaffolds with either NRVMs (movie S5) or hiPSC-CMs (movie S6), spontaneous contractions were observed after 3 to 5 days. These data demonstrate that FRJS could be used to form 3D contractile cardiac models.

Calcium wave propagation

To demonstrate syncytium formation, we examined the ability of the model ventricles to sustain uniform calcium wave propagation. In healthy tissues, action potentials propagate faster in the direction of cell alignment, with myocytes showing slower conduction velocities (CVs) in the transverse direction (*33*). Given this insight, we reasoned that electrical signal propagation should vary depending on the ventricle's fiber alignment and could be used to confirm long-range tissue formation and directional alignment in these model systems. To measure directional signal conduction in tissue scaffolds, laminar tissues were electrically stimulated (corner point stimulation), and the resulting calcium propagation was measured with an optical mapping system (fig. S11). We observed higher CVs in the direction of fiber alignment [longitudinal CVs (LCVs)] with respect to transverse CVs (TCVs), with LCVs and TCVs of 14.9 ± 5.8 cm/s and 9.0 ± 3.7 cm/s, respectively (n = 9 samples) at ratios of ~1.5 to 2.0. These values were consistent with previous reports of in vitro tissues composed of immature cardiomyocytes, where factors such as geometry, calcium handling, and gap junction expression can influence CV (*24*, *34*, *35*). Turning to 3D models, ventricles with extended basal regions, allowing for increased travel distances, were apically stimulated. Isochrones of the calcium propagation revealed that CA scaffolds displayed only modest CVs (8.3 cm/s) along the ventricle's long axis, whereas HA ventricles showed increased CVs (19.1 cm/s) with respect to CA ventricles (Fig. 2C and movie S7). Together, these findings demonstrated that our tissues maintained their alignment and confluence over the length of the ventricle (~1 cm) and further suggested the importance of cardiomyocyte alignment in regulating the spatiotemporal control of excitation-contraction coupling.

Ventricular deformation

It has been hypothesized that HA ventricles would display increased apical shortening (*3*, *36*) and reduced basal displacement (*9*, *36*), owing to differences in myofiber lengths. Because cardiomyocytes contract along their long axis during systole, while connecting transversely as a bulk material (*37*), they generate forces primarily along the length of the fibers. This means that longer fibers should result in greater total displacements. For CA ventricles, which are made up of concentric rings, fiber length is greatest at the base, where the radius is largest, and smallest at the apex (Fig. 2D). This should result in nonuniform deformations, with greater displacements in the basal region and negligible displacements at the apex. For HA ventricles, however, fiber length should vary minimally, as fibers extend uniformly from the apex to the base of the ventricle. Overall, this should result in more homogenous deformations for HA ventricles, with increased apical shortening.

Examining how CA and HA models deformed during contraction, we visualized displacement by submerging these ventricles into a solution containing nonspecifically adherent fluorescent beads. Using digital image cross-correlation, we performed deformation mapping across the ventricle's surface during 1-Hz field stimulation (movie S8). As predicted, CA ventricles displayed greater deformations near the base of the ventricle, whereas HA scaffolds showed more-uniform deformations (Fig. 2E). Probing further, we measured changes in the boundary shape using an elliptical fit, indicating substantial differences in deformation

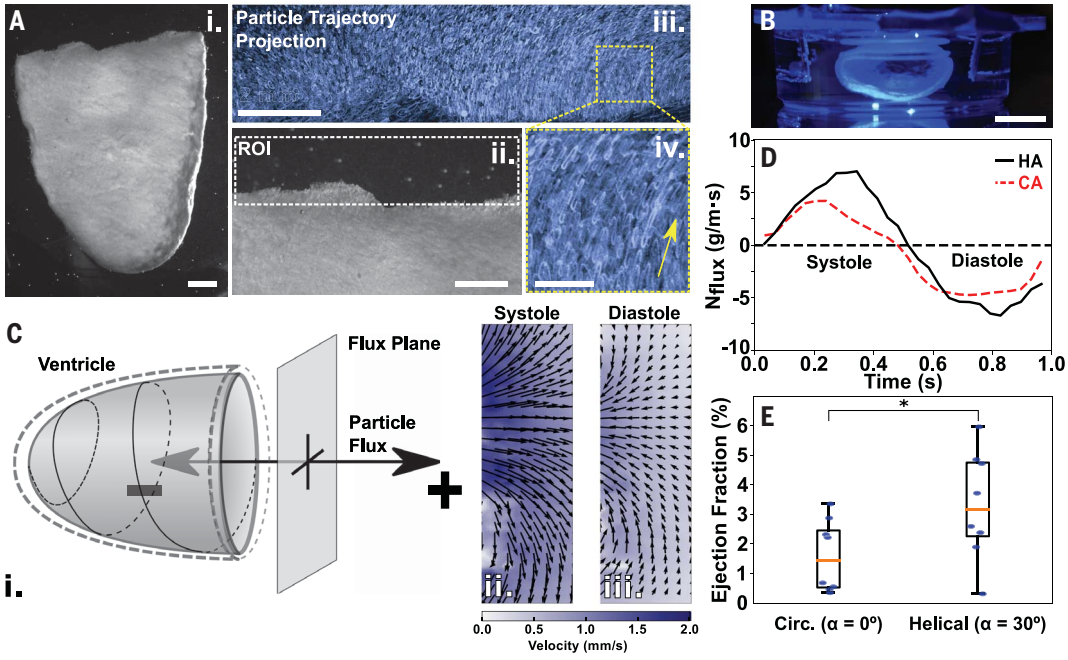


Fig. 3. Alignment dictates ventricular ejection fractions. (**A**) (i) Bright-field micrograph of a tissue-engineered ventricle, with (ii) a magnified view of the ventricular basal region. ROI, region of interest. (iii) Maximum intensity projection of fluorescent beads taken over a single contraction cycle, showing particle displacement. (iv) High-magnification image of the boxed region in (iii); the arrow indicates the direction of fluid displacement [scale bars in (i) to (iii), 2 mm; in (iv), 0.5 mm]. (**B**) Side view of the ventricle (scale bar, 10 mm). (**C**) (i) Schematic diagram of PIV measurement, with velocity fields [(ii) and (iii)] taken from the base of a HA ventricle scaffold during peak systole [(ii), t = 0.3 s] and diastole [(iii), t = 0.7 s]. (**D**) Representative measurements of the instantaneous mass flux (N_{flux}) in the region of interest as function of contraction time and ventricle angle. (**E**) Ensemble measurements of EF for CA and HA ventricle scaffolds (n = 8 ventricles for each angle; *P < 0.05 by Student's t test; box plot given in quartiles).

between each alignment. CA samples showed greater basal and minimal longitudinal constriction, whereas HA ventricles displayed significant apical shortening with reduced basal deformation ($n \geq 5$ ventricles for each condition) (fig. S12). Overall, these results were consistent with the projections of previous analytical and computational models (*3*, *25*, *36*) in that they showed more-homogeneous deformations for HA ventricles.

These changes in ventricular contraction were also suggestive of further functional differences between myofiber alignments. For instance, apical shortening indicated the potential for a wringing motion or ventricle twist to occur in HA ventricles. Additionally, this suggested the possibility for variable EFs on the basis of fiber alignment, as these differences in deformation could result in distinct volumetric displacements. It has been predicted that EFs should scale as $1 - k^2$ for CA ventricles (eq. S5) and $1 - k^2 \cdot q \cdot H(\alpha)$ for HA ventricles (eq. S15) (Fig. 2D) (*3*), with k being the strain along the fiber axis, q being the apical shortening, and $H(\alpha)$ being the radial shortening based on the fiber's helical angle (eqs. S5 to S15). q and $H(\alpha)$ are defined as ≤1, which suggests that HA ventricles could produce larger EFs.

Ventricular twist

To determine whether ventricular twist was preserved in our model system, we measured rotational displacement at the apex of suspended ventricle scaffolds (fig. S13, A to B). In a healthy heart, kinetic energy is stored during contraction in the sarcomeric protein titin and is subsequently released during relaxation (*38*), giving rise to rotational displacements, or ventricular twist (*1*, *38*). HA and CA scaffolds were sutured at the base to a fixed support, allowing the apex to move freely, and were monitored from below during field stimulation. Using edge features to detect rotation (fig. S13, C to E), we observed that CA scaffolds showed minimal twist (1.35 ± 1.1°, n = 7 ventricles). Conversely, HA scaffolds displayed approximately four times as much twist (5.4 ± 3.6°, n = 7 ventricles) (movie S9), modeling physiological ventricular twist in an in vitro system.

Cardiac output and ejection fractions

To determine whether these structural changes in alignment and deformation lead to altered function, we then used cardiac output (CO) and EF as quantitative metrics of cardiac performance. To measure these values, we first used catheterization to monitor pressure-volume changes in the ventricle scaffolds, observing the formation of complete oblate loops (fig. S14). However, submerged gelatin fibers are a poor dielectric, making it difficult to obtain consistent results in our synthetic scaffolds by means of conductance catheterization. Consequently, particle imaging velocimetry (PIV) was used to further quantify cardiac performance.

Model ventricles were suspended in a bath containing neutrally buoyant fluorescent beads, and bead displacement was tracked. This allowed for the construction of 2D velocity fields surrounding the basal opening, as shown for a HA ventricle model (Fig. 3, A and B). Using PIV, we then evaluated the instantaneous mass flux resulting from ventricle contraction (fig. S15 and movie S10) and observed cyclic outputs, with fluid being expelled during systole and refill occurring during diastole (Fig. 3, C and D). This was performed for ventricles with both varying fiber alignments and fibroblast compositions (fig. S16). Summation of the fluid displacement over systole yields the total CO, with values of 11.7 ± 8.9 μl/s and 24.3 ± 13.5 μl/s (n = 8 samples each) for CA and HA ventricle scaffolds, respectively (fig. S15G). This observation represented a significant (P < 0.05) CO increase based purely on ventricular tissue alignment. Normalizing the CO by diastolic ventricle volume and fluid density, we calculated the resulting EFs. This yielded average EFs of 1.6 ± 1.1% and 3.3 ± 1.7% in the CA and HA case, respectively (n = 8 samples each), with a maximum EF of 5.9% observed in the helical case (Fig. 3E). This indicated that helical alignments confer a relative increase in ventricle output. To compare our findings with Sallin's analytical model of ventricle contraction (*3*), we then normalized our results on the basis of relative contractile strain (eqs. S5 to S15). Sallin's model predicted an EF increase of 54 to 64% for HA ventricles, which was consistent with our experimental observations (within the standard error of the mean; Fig. 3E). This indicated that our model system was capable of preserving these fundamental scaling laws, signifying that helical

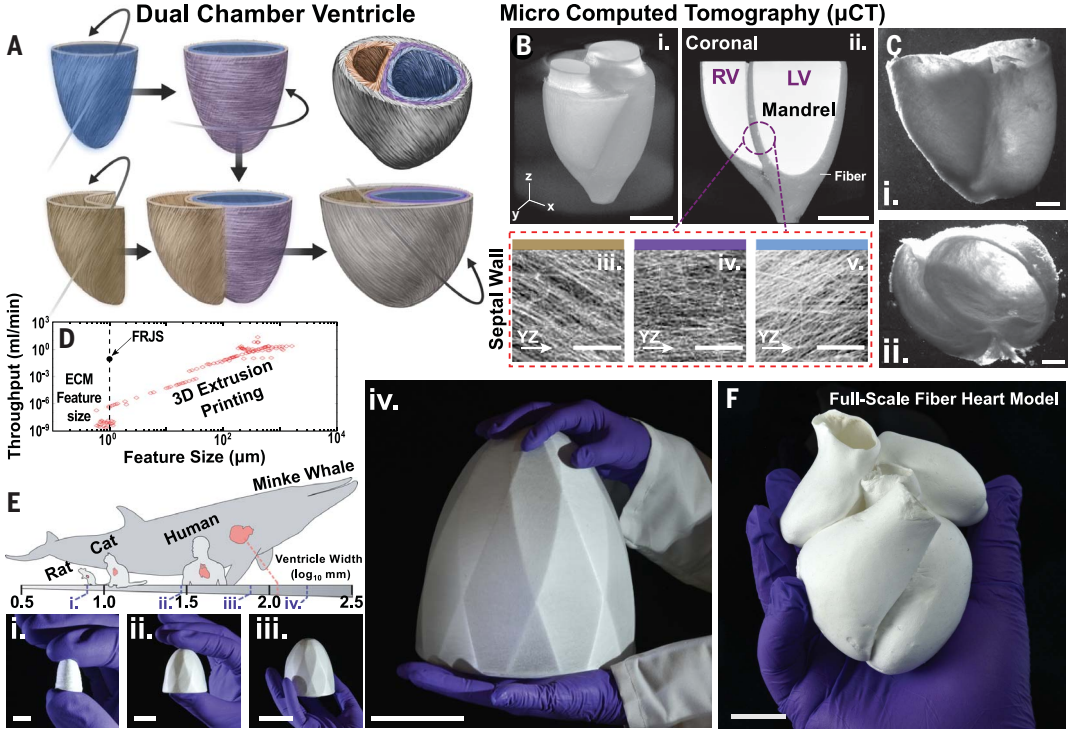


Fig. 4. Multiscale heart models. (**A**) Simplified design of a trilayered DCV that mimics the native ECM alignment of the heart, highlighting the four-step manufacturing process. (**B**) (i) μCT imaging of the DCV suspended on the collection mandrel, with corresponding coronal cross sections (ii) (scale bars, 5 mm). (iii to v) High-magnification μCT images taken from the intraventricular septum (highlighted in purple), showing trilayer alignments (scale bars, 25 μm). RV, right ventricle; LV, left ventricle. (**C**) DCV in culture seeded with NRVMs, as viewed from the side [(i), coronal] and looking into the two chambers [(ii), transverse]. (**D**) 3D extrusion printing scales as a power law with respect to feature size, showing that the throughput of FRJS is ~10^6 times greater than that of 3D extrusion printing for single-micrometer features. (**E**) Left ventricle width for different species. (i to iv) Single-layer ventricles of different sizes, which can be rapidly manufactured owing to increased fiber production rates, while maintaining a single-micrometer feature scale [scale bars (left to right), 5 mm; 15 mm; 4 cm; and 8 cm). (**F**) Full-scale four-chambered human heart model composed of single-micrometer fibers (scale bar, 2 cm).

alignments of the ventricular myocardium result in increased EFs.

Dual-chambered ventricles and full-scale heart models

Scaling Sallin's model to physiological relevant strains (~15 to 20%) (*3*), we noted that single-angle ventricles could achieve EFs of only up to ~43% ($\alpha = 30°$) (see supplementary materials for details). This EF is comparable to values experienced during borderline heart failure and underscores the importance of multiple helical alignments in maintaining healthy cardiac function. To examine whether we could mimic these architectures using FRJS, we manufactured both a dual-chambered ventricle (DCV) with multiple helical angles and a full-scale human heart model (Fig. 4). DCVs were fabricated using a multistage process, creating inner and outer helical layers, with an intermediate circumferential sheet reminiscent of the native myocardium (fig. S17, Fig. 4, A and B, and movie S11). These competing helical structures were confirmed with μCT, indicating three distinct transmural layers along the septal wall (Fig. 4B and movies S12 and S13). DCVs were then seeded with either hiPSC-CMs or NRVMs, forming contractile tissue constructs (Fig. 4C, fig. S18A, and movies S14 to S15). Examining excised segments of the left ventricle, cultured with hiPSC-CMs, we observed that cells adhered to the fibers, forming aligned tissues (fig. S18B). Additionally, calcium imaging revealed sustained wave propagation across the ventricle surface, primarily in the direction of fiber alignment, indicating the formation of a continuous cardiac syncytium ($n \geq 3$ samples for each condition) (fig. S18, C and D). Overall, these observations indicated that fiber scaffolds can also support human stem cell–derived tissues in complex geometries.

We then spun fibers onto ventricle-shaped targets of varying size, ranging from ~1 to 20 cm in diameter (fig. S19 and movie S16). In each case, targets were coated in <35 min with micro/nanofibers, generating conformal coatings on the exterior. Throughout this process, fibers were produced with a total throughput rate of 0.1 g/min, or 0.03 ghm (grams per hole per minute). This throughput is comparable to those of melt-blowing processes (*27*) but is orders of magnitude greater (~10^6) than that of 3D extrusion printing, the speed of which declines rapidly at single-micrometer features (as a power law, with a scaling factor of ~2.8) (Fig. 4, D and E, and eq. S1). Collectively, these observations demonstrated that FRJS is amenable to rapidly manufacturing fiber scaffolds spanning multiple length scales and can be adapted to different geometries.

To demonstrate the ability of FRJS to enable hierarchical biofabrication, a full-size model of the human heart musculature was constructed. This model was assembled by individually patterning micro/nanofibers onto dissolvable collectors in the shape of each of the heart's four chambers (fig. S20). Individual chambers were then connected by chemical annealing before the interior supports were dissolved, resulting in standalone fiber scaffolds (Fig. 4F). Although full-size anatomical models have previously been produced with thermoplastics and hydrogels (*22*), these systems typically lack the micrometer-scale features needed to direct myocyte alignment. With our fiber-based approach, these local structures can be preserved across entire tissue volumes, allowing for the hierarchical assembly of tissues. These proof-of-concept 3D organ models demonstrate that FRJS scaffolds support human-derived tissues and allow for rapid assembly of full-size models of the musculature. These key features mark fiber-based manufacturing as a promising approach for achieving whole-organ biofabrication, which can be used as an alternative to or in conjunction with emerging biomanufacturing platforms such as 3D extrusion printing.

Discussion

Biofabrication using FRJS allowed for the rapid assembly of functional 3D ventricle models capable of recapitulating emergent phenomena in vitro, including ventricular twist, strain displacement, and myocardial angle-dependent EFs. Although the metrics presented

here (e.g., EF) were greater (by a factor of ~3) than for recently reported in vitro ventricle models (*24*, *39*), they were still substantially lower than in vivo values (tables S2 and S3). This difference may be the result of limited cardiomyocyte penetration into the fiber scaffolds (figs. S7 and S8). Consequently, further work is needed to achieve full-scale de novo organ fabrication; this includes improved cardiac maturation, vascularization, and incorporating consistent back-pressures (Frank-Starling mechanism). There is also a need to generate the large numbers of cardiomyocytes and other diverse cell populations (e.g., neurons, endothelial cells, fibroblasts) required to support organ biofabrication. However, the work presented here provides an initial pathway toward achieving hierarchical patterning while maintaining 3D cell alignment.

In addition to biofabrication, FRJS may serve an important role in other additive manufacturing applications, as it provides production rates comparable to those of current industrial processes while enabling micro/nanoscale feature sizes and controlled 3D alignments. This includes applications in which a material's properties are determined by its microstructure and alignment. The high surface area–to–volume ratio of micro- and nanofibers makes them ideal candidates for the controlled absorbance and release of chemical species. This suggests that FRJS may be an important technique for industrial manufacturing process that use hierarchical design principles, for which control over several orders of spatial magnitude is needed.

REFERENCES AND NOTES

1. P. P. Sengupta, A. J. Tajik, K. Chandrasekaran, B. K. Khandheria, *JACC Cardiovasc. Imaging* **1**, 366–376 (2008).
2. R. Lower, *Tractatus de Corde. Item de Motu & Colore Sanguinis et Chyli in eum Transitu* (Allestry, 1669); available at https://www.woodlibrarymuseum.org/rare-book/lower-r-tractatus-de-corde-item-de-motu-colore-sanguinis-et-chyli-in-eum-transitu-1669/.
3. E. A. Sallin, *Biophys. J.* **9**, 954–964 (1969).
4. A. Grosberg, M. Gharib, A. Kheradvar, *Bull. Math. Biol.* **71**, 1580–1598 (2009).
5. J. S. Davis *et al*., *Cell* **107**, 631–641 (2001).
6. R. S. Stephenson *et al*., *Clin. Anat.* **29**, 316–332 (2016).
7. P. Agger, R. S. Stephenson, *J. Cardiovasc. Dev. Dis.* **7**, 47 (2020).
8. K. D. Costa, Y. Takayama, A. D. McCulloch, J. W. Covell, *Am. J. Physiol.* **276**, H595–H607 (1999).
9. E. D. Carruth, A. D. McCulloch, J. H. Omens, *Prog. Biophys. Mol. Biol.* **122**, 215–226 (2016).
10. S. Puwanant *et al*., *Circulation* **121**, 259–266 (2010).
11. J. Zhang, *Clin. Exp. Pharmacol. Physiol.* **29**, 351–359 (2002).
12. P. Agger *et al*., *J. Cardiovasc. Magn. Reson.* **19**, 93 (2017).
13. E. D. Carruth *et al*., *J. Cardiovasc. Magn. Reson.* **22**, 21 (2020).
14. M. L. McCain, S. P. Sheehy, A. Grosberg, J. A. Goss, K. K. Parker, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9770–9775 (2013).
15. C. L. Hung *et al*., *J. Am. Coll. Cardiol.* **56**, 1812–1822 (2010).
16. J. B. Caulfield, T. K. Borg, *Lab. Invest.* **40**, 364–372 (1979).
17. J. U. Lind *et al*., *Nat. Mater.* **16**, 303–308 (2017).
18. F. B. Coulter *et al*., *Matter* **1**, 266–279 (2019).
19. M. A. Skylar-Scott *et al*., *Sci. Adv.* **5**, eaaw2459 (2019).
20. A. Lee *et al*., *Science* **365**, 482–487 (2019).
21. N. Noor *et al*., *Adv. Sci.* **6**, 1900344 (2019).
22. E. Mirdamadi, J. W. Tashman, D. J. Shiwarski, R. N. Palchesko, A. W. Feinberg, *ACS Biomater. Sci. Eng.* **6**, 6453–6459 (2020).
23. A. K. Capulli *et al*., *Biomaterials* **133**, 229–241 (2017).
24. L. A. MacQueen *et al*., *Nat. Biomed. Eng.* **2**, 930–941 (2018).
25. B. Baillargeon, N. Rebelo, D. D. Fox, R. L. Taylor, E. Kuhl, *Eur. J. Mech. A* **48**, 38–47 (2014).
26. D. Li, Y. Xia, *Adv. Mater.* **16**, 1151–1170 (2004).
27. Y. Kara, K. Molnár, *J. Ind. Text.* 10.1177/15280837211019488 (2021).
28. L. F. Deravi *et al*., *Macromol. Mater. Eng.* **302**, 1600404 (2017).
29. M. R. Badrossamay, H. A. McIlwee, J. A. Goss, K. K. Parker, *Nano Lett.* **10**, 2257–2261 (2010).
30. J. Go, S. N. Schiffres, A. G. Stevens, A. J. Hart, *Addit. Manuf.* **16**, 1–11 (2017).
31. P. Davidson, *Turbulence: An Introduction for Scientists and Engineers* (Oxford Univ. Press, 2015).
32. F. S. Pasqualini, S. P. Sheehy, A. Agarwal, Y. Aratyn-Schaus, K. K. Parker, *Stem Cell Rep.* **4**, 340–347 (2015).
33. M. Valderrábano, *Prog. Biophys. Mol. Biol.* **94**, 144–168 (2007).
34. S. Rohr, J. P. Kucera, A. G. Kléber, *Circ. Res.* **83**, 781–794 (1998).
35. A. G. Kléber, Q. Jin, *Biophys. Rev.* **2**, 031301 (2021).
36. N. B. Ingels Jr., *Technol. Health Care* **5**, 45–52 (1997).
37. R. H. Anderson, P. F. Niederer, D. Sanchez-Quintana, R. S. Stephenson, P. Agger, *J. Anat.* **235**, 697–705 (2019).
38. A. M. S. Omar, S. Vallabhajosyula, P. P. Sengupta, *Circ. Cardiovasc. Imaging* **8**, e003029 (2015).
39. M. E. Kupfer *et al*., *Circ. Res.* **127**, 207–224 (2020).
40. SeasDBG, SeasDBG/FRJS: Recreating the Heart's Helical Structure-Function Relationship with Focused Rotary Jet Spinning- Data Analysis, *Zenodo* (2022); https://doi.org/10.5281/zenodo.6547775.

ACKNOWLEDGMENTS

The authors thank A. G. Kleber for discussions regarding cardiac physiology, P. Campbell for rodent heart isolation, W. T. Pu for providing hiPSC cell lines, and H.-Y. G. Lin for assistance with micro-CT. **Funding:** H.A.M.A. thanks the American Chemical Society for support through the Irving S. Sigal Postdoctoral Fellowship. This work was sponsored by the John A. Paulson School of Engineering and Applied Sciences at Harvard University, the Wyss Institute for Biologically Inspired Engineering at Harvard University, the Harvard Materials Research Science and Engineering Center (DMR-1420570 and DMR-2011754), and the National Institutes of Health with the Center for Nanoscale Systems (S10OD023519) and National Center for Advancing Translational Sciences (UH3TR000522 and 1-UG3-HL-141798-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. **Author contributions:** K.K.P. supervised the research. K.K.P., H.C., Q.L., J.F.Z., and L.A.M. designed the study. Q.L. and H.C. designed and produced the fiber-spinning platform. H.C. and J.F.Z. manufactured, cultured, and performed the ventricle experiments. K.Y.L. performed optical mapping experiments. H.C., Q.L., J.F.Z., and K.Y.L. analyzed the data. Q.L. performed simulations. L.A.M., C.O.C., S.E.M., G.T., and E.M.C. performed additional supporting experiments. S.L.K., H.A.M.A., S.C., and Q.J. helped with animal protocols and stem cell culture to obtain cardiomyocytes. M.M.P. and H.C. performed SEM experiments and analysis. M.R., H.C., J.F.Z., and C.O.C. produced the full-size heart model. All authors discussed the results and contributed to the writing of the final manuscript. **Competing interests:** Harvard University filed for intellectual property relevant to this manuscript, listing J.F.Z., Q.L., H.C., and K.K.P. as inventors (US Patent Application 17/421,047 and US Provisional Patent Application 63/234,287). **Data and materials availability:** All data are available in the manuscript, the supplementary materials, or Zenodo (*40*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl6395
Materials and Methods
Figs. S1 to S20
Tables S1 to S3
References (*41*–*54*)
MDAR Reproducibility Checklist
Movies S1 to S16

Submitted 26 July 2021; resubmitted 1 April 2022
Accepted 16 May 2022
10.1126/science.abl6395

MEMBRANES

A two-phase model that unifies and extends the classical models of membrane transport

Varun H. Hegde, Michael F. Doherty, Todd M. Squires*

Two models describe solvent transport through swollen, nonporous membranes. The pore-flow model, based on fluid mechanics, works for porous membranes, whereas the solution-diffusion model invokes molecular diffusion to treat nonporous membranes. Both approaches make valid arguments for swollen polymer membranes, but they disagree in their predictions of intramembrane pressure and concentration profiles. Using a fluid-solid model that treats the solvent and membrane matrix as separate phases, we show both classical models to be valid, to represent complementary approaches to the same phenomenon, and to make identical predictions. The fluid-solid model clarifies recent reverse osmosis measurements; provides a predictive and mechanistic basis for empirical high-pressure limiting flux phenomena, in quantitative agreement with classic measurements; and gives a framework to treat nonporous but mechanically heterogeneous membrane materials.

Polymer membranes play an essential role in both classic and emerging technologies, including water treatment, gas separations, fuel cells, and flow batteries (*1*, *2*). Membranes generally fall into two broad classes, on the basis of morphological structure and mechanism of action. Porous membranes (e.g., Fig. 1, A to C) are permeated by fixed pores that permit the passage of some species but not others, typically by means of size exclusion and/or electrostatics (*3*). Examples include the micro-, nano-, and ultrafiltration of contaminated water (*3*). By contrast, nonporous membranes (e.g., Fig. 1D) are structurally homogeneous, with permeability and selectivity determined by solubility and diffusivity within the membrane. Examples include gas separation membranes (*3*), organic solvent nanofiltration (*4*, *5*), and reverse osmosis desalination (*6*).

Two types of models are used to describe transport through these two classes of membranes. Porous membranes are treated using pore-flow (PF) models, in which pressure gradients force fluid to flow within the membrane pores. Nonporous membranes are described using solution-diffusion (SD) models, wherein solute and solvent dissolve into the membrane and diffuse down concentration gradients. Each approach is built on a distinct transport mechanism: PF involves the mechanical forcing of fluid through porous membranes by means of a momentum balance, whereas SD describes the thermodynamic forcing down concentration gradients, as described by a mass balance (*3*).

However, some membranes exhibit features common to both SD and PF mechanisms. Swollen polymer membranes consist of statistically homogeneous polymer meshes, in persistent diffusive motion relative to the permeating fluid, consistent with the SD mechanism (*7*–*9*). At the same time, solvent within the permeating fluid networks may traverse such membranes without ever encountering polymer, and so it would be described by fluid dynamics, consistent with the PF model (*10*–*12*). Decades ago, a spirited debate emerged over which approach is appropriate for swollen membranes (*7*–*12*), which persists to this day. Variants of both models continue to be discussed and criticized, for example, in treating reverse osmosis membranes (*13*, *14*), with ongoing reframing of the classical models (*15*, *16*). For example, recent neutron scattering studies of water transport in polyamide reverse osmosis (PARO) membranes suggested SD to be invalid and motivated a proposed change in the strategic design of future PARO membranes, toward reducing tortuosity rather than enhancing selective solubility and diffusivity (*17*).

Here, we introduce a two-phase fluid-solid (FS) model for solvent transport through swollen, homogeneous membranes that reconciles the PF and SD mechanisms, revealing them to be entirely consistent with one another, meaning that they make quantitatively identical predictions. Moreover, the FS model predicts the PF and SD parameters to be connected through the longitudinal osmotic modulus of the membrane, which can be independently measured. In fact, FS model predictions are consistent with neutron scattering measurements of PARO membranes (*17*), indicating that neither PF nor SD is inappropriate for these membranes.

Background

The PF model treats solvent flux through a membrane as a pressure-driven flow through a porous network with fixed geometry. In principle, the Navier-Stokes equations could be solved to relate the solvent flux to the pressure gradient, if the particular morphology of the membrane were known (*1*) (e.g., isoporous or track-etched membranes in Fig. 1, A and B). More often, however, the pore-scale structure is known, at best, only statistically (e.g., Fig. 1C), instead requiring a Darcy-flow approach, $v_s = -\nabla p/\xi$, wherein a local pressure gradient ∇p drives a local solvent velocity v_s, where ξ is a local solvent/matrix friction coefficient. Because solvent only occupies a volume fraction ϕ_s within the membrane, the volumetric solvent flux J_s is a factor ϕ_s smaller than v_s. The macroscopic solvent flux through a membrane of thickness L is then given by

$$J_s = \frac{\phi_s}{\xi}\frac{\Delta p}{L} \equiv \kappa_{PF}\frac{\Delta p}{L} \quad (1)$$

In practice, the friction coefficient ξ or permeability κ_{PF} is not determined independently but rather is typically extracted from macroscopic permeance measurements. Explicit hydrodynamic calculations for simplified geometries [see the supplementary materials (SM)] relate ξ to pore size R_p and solvent viscosity η, typically via $\xi \sim \eta/R_p^2$ (*18*, *19*).

The SD model describes solvent transport through nonporous membranes, whose uniform structure and composition lack clear, continuous pores. This model assumes solvent to diffuse down concentration gradients within the matrix, with flux

$$J_s = \frac{D}{1-\phi_s}\frac{\Delta\phi_s}{L} \quad (2)$$

where D is the binary diffusivity between solvent and polymer, and $\Delta\phi_s$ is the drop in solvent volume fraction between the feed and permeate sides of the membrane (see the SM). Equation 2 is difficult to apply directly, because the solvent volume fraction profile ϕ_s is generally not known within the membrane. Instead, $\Delta\phi_s$ is related to the transmembrane pressure drop Δp. Imposing contact equilibrium for solvent molecules at the feed and permeate boundaries of the membrane (described in the SM) gives

$$J_s = \kappa_{SD}\frac{\Delta p}{L} \equiv \frac{D\phi_s^0}{1-\phi_s^0}\left(\frac{\overline{V}_s}{RT}\frac{\partial \ln\phi_s}{\partial \ln a_s}\bigg|_{\phi=\phi_0}\right)\frac{\Delta p}{L} \quad (3)$$

where $\overline{V}$s and a_s are the molar volume and activity of the solvent, respectively (*2*, *8*).

Both SD and PF approaches have been used to treat solvent transport through highly swollen membranes, and reasonable arguments can be made to support either. The fundamentally diffusive nature of even equilibrium swollen membranes, along with the lack of any clear pores, suggests the stochastic, diffusive approach of SD. However, polymer-free "tubes" of solvent permeate highly swollen membranes,

Department of Chemical Engineering, University of California, Santa Barbara, CA, USA.
*Corresponding author. Email: tsquires@ucsb.edu

along which the hydrodynamic pressure must decrease continually as fluid traverses the membrane—no other element enters the momentum balance within the tube. From this standpoint, the pressure-driven PF approach is warranted.

Both SD and PF models predict flux relations of the same macroscopic form, $J_s = \kappa \Delta p / L$, allowing experimentally measured permeabilities (κ) to be interpreted through either model to determine R_p and ξ for PF, and D and ϕ_s for SD. Without independent measurements of those quantities, macroscopic permeability measurements cannot differentiate between the predictions of the two approaches. Support for SD came from the classic experiments of Rosenbaum and Cotton (*20*), which measured concentration gradients across a multilayer stack of swollen gel membranes. Additionally, Ham and colleagues (*21*) demonstrated that the solvent concentration ϕ_s^0 at the feed-side boundary remains constant as pressure increases. Likewise, Gehman (*22*) measured ϕ_s to decrease on the permeate side, as the membrane was compressed against the support layer. A number of apparent discrepancies arise between the PF and SD treatments of swollen membranes. First, the internal pressure and concentration profiles predicted by these two models appear to be at odds with one another (Fig. 2, C and D). SD predicts the pressure p to be constant within the membrane but a steady gradient in solvent concentration ϕ_s (Fig. 2D). By contrast, PF predicts a pressure gradient ∇p within the membrane but a constant concentration ϕ_s (Fig. 2C). Second, the two models describe solvent transport through fundamentally distinct physical mechanisms. The PF approach is mechanical and deterministic in nature and assumes purely convective flow, whereas SD is stochastic and thermodynamic, driven by mutual diffusion between polymer and solvent. Because the binary diffusive flux adds to the volume-averaged reference velocity in Eq. 2, SD transport contains both diffusive and convective components. However, the Fick's law basis of SD underscores the fundamental role of diffusion. Finally, the flux relations predicted by the SD and PF models depend on physically different, and seemingly unrelated, parameters.

The fluid-solid model for gel dynamics

We now describe the FS model, building on the poroelastic theory developed by Biot and Willis to describe soil compaction (*23*–*25*) and extended to gel dynamics by Tanaka and others (*26*–*29*). This approach has been used to probe gel swelling and deswelling, drying, crack propagation, and water permeation (*29*–*31*) and the dynamic swelling, buckling, wrinkling and folding of thin polymer gel films (*32*). The FS model treats the solvent and polymer matrix individually, as distinct

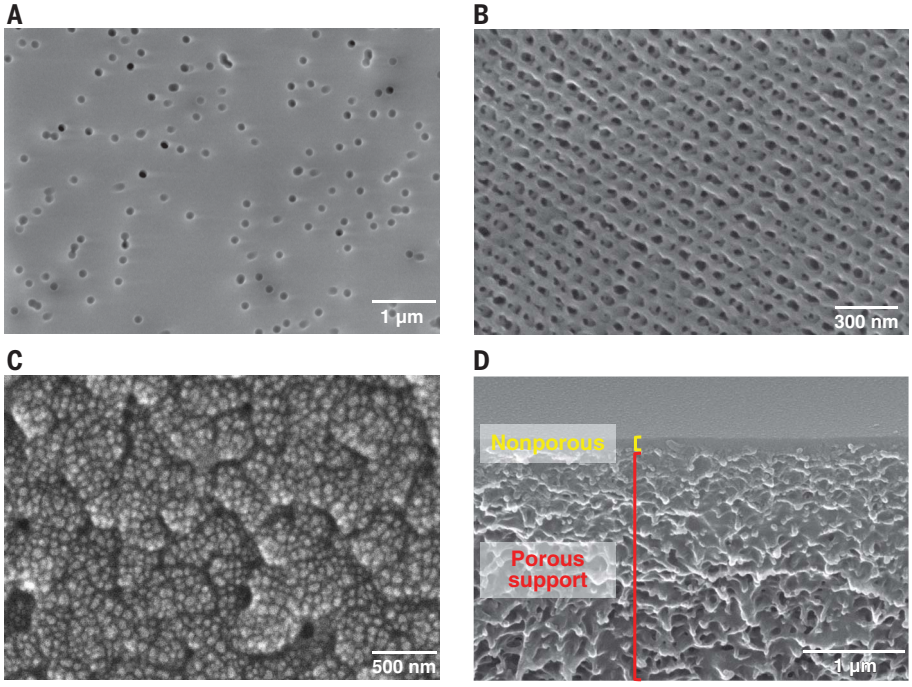


Fig. 1. SEM images of membrane surfaces. (**A**) Scanning electron microscopy (SEM) image of a track-etched polycarbonate membrane with 0.1-μm pores. (**B**) SEM image of an isoporous PS-P4VP [polystyrene-poly(4-vinylpyridine)] membrane where the skin layer is composed of uniformly spaced nanoscale nodules. (**C**) SEM image of a commercial polystyrene-20 ultrafiltration membrane where the skin layer is composed of randomly spaced nanoscale nodules. (**D**) SEM image of a P1-8 asymmetric gas separation membrane with a dense nonporous skin layer bound to a porous support.

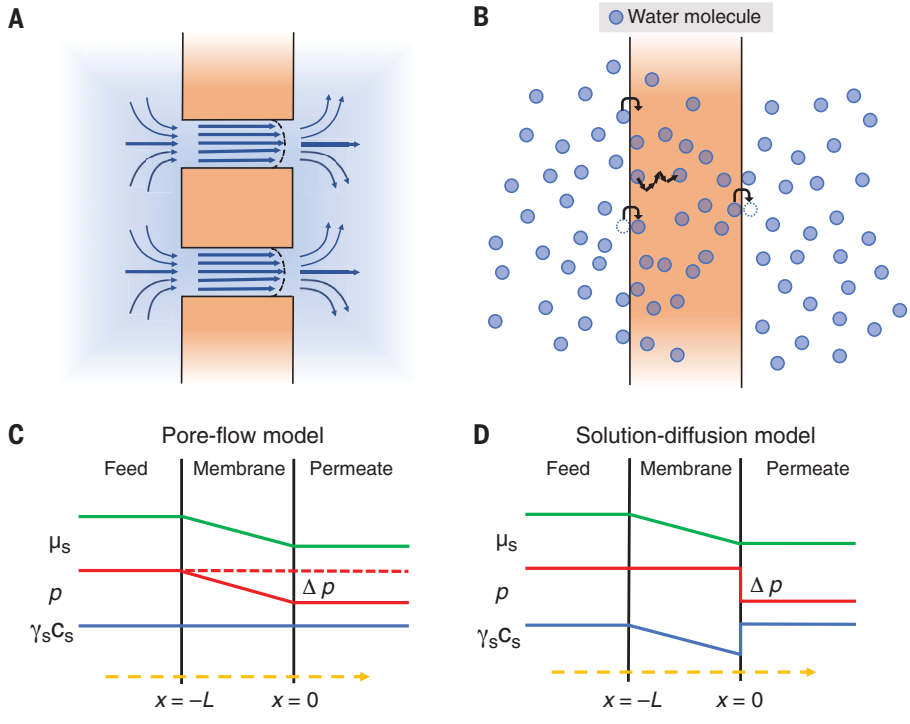


Fig. 2. Classical membrane transport models. (**A** and **B**) Pictorial representation of the pore-flow and solution-diffusion models, respectively. (**C** and **D**) Solvent chemical potential μ_s, pressure p, and solvent activity a_s profiles across a pore-flow and a solution-diffusion membrane, respectively.

IMAGES: MOSTAFA NASSER (FIG. 1A); ALEXANDER BRIDGE (FIG. 1, B TO D)

but interpenetrating phases. The solvent phase is a Newtonian liquid with local velocity field $\mathbf{v}_s(r,t)$ and pressure p and occupies volume fraction ϕ_s. The polymer phase, which occupies volume fraction ϕ_p, is treated as a homogeneous elastic network with displacement field $\mathbf{u}_p(r,t)$ that describes local deformations from the equilibrium state. Together, the solvent and matrix phases fill space—meaning $\phi_s + \phi_p = 1$. Even though the physical solvent and matrix materials are themselves assumed to be incompressible, the matrix network may itself be compressed or dilated, because solvent flows into or out of the network to satisfy the space-filling requirement. The volumetric flux of the FS composite is given by the local volume-averaged velocity, $\mathbf{v} = \phi_p\mathbf{v}_p + \phi_s\mathbf{v}_s$, where $\dot{\mathbf{u}}_p = \partial\mathbf{u}_p/\partial t$ is the local polymer velocity field (*33*).

The governing equations of the FS model

$$\nabla\cdot\mathbf{v} = \nabla\cdot\left[\phi_p\mathbf{u}_p + (1-\phi_p)\mathbf{v}_s\right] = 0 \quad (4)$$

$$\nabla\cdot\boldsymbol{\sigma}_p = \left(K + \frac{G}{3}\right)\nabla(\nabla\cdot\mathbf{u}_p) + G\nabla^2\mathbf{u}_p = -\xi(\mathbf{v}_s - \dot{\mathbf{u}}_p) \quad (5)$$

$$\nabla\cdot\boldsymbol{\sigma}_s = -\nabla p = \xi(\mathbf{v}_s - \dot{\mathbf{u}}_p) \quad (6)$$

enforce conservation of volume, polymer momentum, and solvent momentum, respectively. Polymer and fluid inertia are neglected in both Eq. 5 and Eq. 6, owing to the small length and long time scales of membrane permeation.

The first two terms in the polymer momentum balance (Eq. 5) represent the elastic forces exerted by a deformed, isotropic elastic solid with shear and compressive moduli G and K, respectively (*34*). In the gel context, K is called the osmotic modulus, rather than the bulk modulus, because it is the swollen network that is compressed, rather than the polymer itself, with solvent flowing in or out as needed to conserve volume. The final term represents the drag force (per unit volume) exerted on the network by the solvent, with friction coefficient ξ. The network exerts an equal and opposite force on the fluid phase, which is treated as Darcy flow driven by a hydrodynamic pressure p.

Two important relations emerge from the FS model that play a key role in the PF–SD bridge. The first relates the mutual diffusivity D to the Darcy friction coefficient ξ, and the second relates the thermodynamic properties of a swollen membrane to its mechanical properties.

Solving the polymer momentum equation (Eq. 5) for v_s, inserting into Eq. 4, and relating $\nabla\cdot u_p$ to changes in ϕ_p gives a diffusion-like equation for compressive deformations of the swollen gel

$$\frac{\partial\phi_p}{\partial t} = \frac{M\phi_s}{\xi}\nabla^2\phi_p \quad (7)$$

where $M = K + 4G/3$ is the longitudinal osmotic modulus, associated with free-draining uniaxial compression. Concentration fluctuations of polymer and solvent diffuse relative to one another with mutual diffusivity

$$D = \frac{M\phi_s}{\xi} \equiv M\kappa_{PF} \quad (8)$$

which is called the collective diffusivity by the gel dynamics community (*27*–*29*, *35*). The relationship between the collective diffusivity, which follows from deterministic arguments, and the stochastic solvent-polymer mutual diffusivity, is a consequence of the fluctuation-dissipation theorem, analogous to the Stokes-Einstein relation. Indeed, Paul and Ebra-Lima suggested a relation of this sort should exist (*5*).

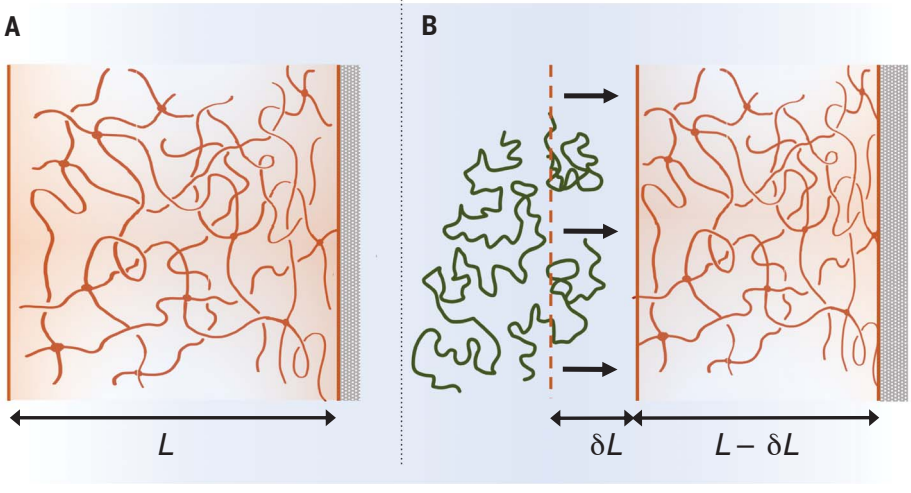


Fig. 3. Hydrogel deswelling. (**A**) A gel initially swollen to thickness L in pure solvent (**B**) de-swells by δL when an osmolyte lowers the osmotic pressure by $\delta\pi$.

Tanaka and co-workers verified Eq. 8 experimentally, using dynamic light scattering to measure D within swollen gels, macroscopic permeation measurements to determine ξ, mechanical experiments to measure elastic moduli (K and G, and therefore M), and swelling experiments to measure ϕ_s (*28*).

The second key relation connects the mechanical response properties (moduli) of the matrix to the equilibrium thermodynamics of swollen membranes (see the SM). Figure 3A shows a gel initially swollen to thickness L in pure solvent, which deswells when an osmolyte (e.g., a polymer excluded from the swollen gel) reduces the solvent activity a_s outside the membrane (Fig. 3B). The osmotic pressure difference exerts a stress $\delta\pi$ on the membrane, compressing the matrix until it is balanced by the elastic stress $-M\nabla\cdot u$. Solving this stress balance gives the longitudinal osmotic modulus M

$$M = \frac{RT}{\overline{V}_s}\frac{\phi_p^0}{\phi_s^0}\left.\frac{\partial\ln a_s}{\partial\ln\phi_s}\right|_{T;\phi_s^0} \quad (9)$$

in terms of the thermodynamic relation between solvent swelling (ϕ_s) and solvent activity (a_s).

Even without explicitly solving the FS equations (Eqs. 4 to 6), Eqs. 8 and 9 reveal three distinct expressions for the permeability

$$\kappa_{SD} = \frac{D}{1-\phi_s}\left(\frac{\phi_s\overline{V}_s}{RT}\left.\frac{\partial\ln\phi_s}{\partial\ln a_s}\right|_{T;\phi_s^0}\right) = \frac{D}{M} = \frac{\phi_s}{\xi} = \kappa_{PF} \quad (10)$$

to be quantitatively identical and to all follow from the FS model. The equality between the first two relates the thermodynamic factor in the general solution-diffusive flux (Eq. 3) to the compressional stiffness of the matrix using Eq. 9. The second equality relates the binary diffusivity to the mechanical membrane properties M and ξ (Eq. 8) to reproduce the PF permeability (Eq. 1). Although these permeability expressions appear very different, they simply reflect different ways of expressing the same, fundamental driving force—in purely thermodynamic terms (Eq. 3), in purely mechanical terms (Eq. 1), or in hybrid terms (Eq. 10).

Figure 4 shows FS model predictions (solutions to Eqs. 4 to 6) for a swollen membrane of equilibrium thickness L, through which a transmembrane pressure drop Δp drives a solvent flux J_s that is gentle enough to maintain a linear pressure-flux relationship. The polymer displacement u_p is highest $[L\Delta p/(2M)]$ on the upstream side and vanishes at the rigid support on the permeate side. The corresponding solvent concentration profile ϕ_s decreases linearly by $(1-\phi_s^0)\Delta p/M$ over the membrane thickness L, consistent with SD. Equal and opposite

gradients in ϕ_p, not shown, are required by space filling. The fluid-phase pressure p decreases continuously within the membrane, with constant gradient $-\Delta p/L$, while the compressive normal stress σ_p^{xx} on the polymer phase grows linearly from zero on the (uncompressed) feed side, to a maximum at the permeate side, with gradient equal and opposite to the fluid pressure.

Resolving discrepancies

The normal stress profiles in Fig. 4 clarify the different pressure profiles predicted by PF and SD—the key distinction ultimately reflects differences in how the two models define pressure. PF focuses exclusively on the mechanics of the solvent phase and uses p to refer to the hydrodynamic pressure within the solvent phase. SD, by contrast, combines solvent and polymer as one composite phase and uses p to mean the composite stress $\sigma_T = \sigma_s + \sigma_p$ of both fluid and solid components. Given that no other forces are exerted within the membrane, mechanical equilibrium requires σ_T to be constant—indeed, adding the polymer and solvent momentum equations (Eqs. 5 and 6) gives

$$\frac{\partial \sigma_T}{\partial x} = \frac{\partial \sigma_s}{\partial x} + \frac{\partial \sigma_p}{\partial x} = \xi v_s - \xi v_s = 0 \qquad (11)$$

Equation 11 holds irrespective of the specific form of the constitutive relations σ_s and σ_p or the solvent-polymer friction ξ. Treating the solvent and polymer phases separately, as in FS, gives a nonzero solvent-phase pressure gradient that is balanced by the elastic stress of the compressed network.

The FS profiles in Fig. 4 resolve the apparent discrepancies between PF and SD predictions of concentration and pressure fields. In particular, nonzero gradients arise for both concentration ϕ_s and fluid pressure p. Concentration gradients are required by SD but are assumed not to exist in PF. Likewise, the pressure gradient is required by PF but assumed to be zero in SD. In this regard, SD and PF are both correct in treating their respective gradient but incorrect in assuming the other gradient to be zero. Notably, the PF model does not require ϕ_s to be constant; it simply does not rely on a nonzero $\nabla\phi_s$. Similarly, no pressure gradient is needed to drive a SD flux; instead, mechanical equilibrium requires a constant normal stress.

Three small quantities arise in the linear-response limit of solvent permeation of a swollen gel: the solvent flux J_s, the solvent pressure gradient ∇p, and the solvent (or polymer) concentration gradient $\nabla\phi_s$. Each quantity is proportional to one of the others while neglecting the third. Solving the solvent equation alone relates J_s to ∇p, while ignoring $\nabla\phi_s$, giving the PF flux (Eq. 1). Enforcing polymer momentum alone relates J_s to $\nabla\phi_s$ and ignores ∇p; using Eq. 8 to eliminate M and ξ reproduces the general SD flux (Eq. 2). Finally, combining the fluid and solid momentum equations relates $\nabla\phi_p$ directly to ∇p, without explicit reference to J_s.

Consequences

In addition to bridging the two distinct models, the FS model brings qualitative and quantitative insight to experimental results that affect the research workflows in next-generation membranes. Recently, for example, the structural and dynamic properties of dry and swollen PARO membranes were studied using neutron scattering, with the explicit goal of identifying whether SD or PF correctly treated water permeation (*17*). Small-angle neutron scattering (SANS) found structural length scales to not change appreciably when swelling, and quasi-elastic neutron scattering (QENS) suggested water diffusivity $D \sim 2 \times 10^{-9}$ m^2/s, comparable to its self-diffusivity in bulk. Both results seemed inconsistent with SD, which requires coupled solvent/matrix diffusion driven by concentration gradients. Moreover, recent advances in microscopy revealed nanoscale density variations in PARO membranes (*36*), which could violate the homogeneity assumed by SD. These results motivated an argument that future PARO membranes should be developed to reduce tortuosity—as appropriate for PF—rather than working to enhance selective solubility.

Reassessing these results using the FS model, however, reveals no inconsistency. Namely, Eq. 8 connects the collective diffusivity D to hydraulic permeability κ through the longitudinal osmotic modulus M_0, which can be measured independently, for example, using a swelling pressure apparatus (*22*) or nanoindentation (*37*, *38*). Shin *et al.* measured both hydraulic permeance $P_w \sim 1.6$ liters per m^2·hour·bar and reduced Young's moduli ($E_r \sim 1.6$ GPa) of water-swollen PA membranes of thickness $L \sim 330 \pm 75$ nm (*38*). Using these measurements as characteristic of PARO films, the FS model predicts diffusivities $D = MP_wL \sim 2 \times 10^{-9}$ to 4×10^{-9} m^2/s. While the relatively large experimental uncertainties may be responsible for the aphysically high end of this range, the diffusivities measured by QENS has the magnitude expected for a membrane this stiff, with quantitative values consistent with FS predictions. Moreover, Shin *et al.* conditioned PARO membranes with a series of solvents, finding higher permeance P_w with softer moduli, as expected from Eq. 8, owing to the higher volume fraction ϕ_s and larger pore sizes (lower ξ) (*38*).

The FS model highlights the fundamentally interrelated nature of the mechanical (hydrodynamic) and thermodynamic (diffusive) mechanisms of solvent transport in osmotically active membranes. Collective diffusivity in PARO membranes is high because the swollen membrane is very stiff, owing to its high cross-link diffusivity, low swelling ratio, and therefore large thermodynamic factor. High cross-link densities give small mesh spacing (akin to pore size R_p), increasing the friction factor ξ and decreasing permeability κ.

Any case in which the FS connection fails to be made directly signals a breakdown in the universality of the mechanism governing solvent transport in the membrane. For example, dense polymer membranes for gas separation do not swell appreciably with added gas; rather, gas molecules dissolve into the free volume of the polymer network. Moreover, the pressure-driven flow of bulk gasses looks nothing like the molecular hopping of dissolved gas molecules. The elastic moduli of such membranes reflect

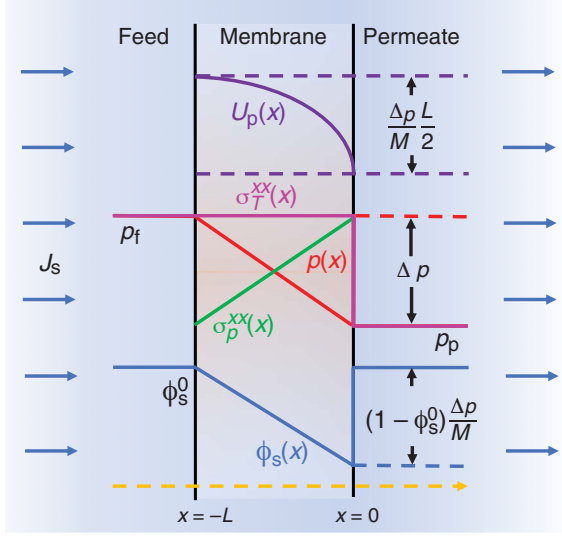


Fig. 4. Steady-state FS profiles across a solvent-swollen nonporous membrane bound to a permeable support at the permeate boundary. Polymer network displacement, purple (SM, eq. S73); normal solvent stress, red (eq. S63); normal polymer stress, green (eq. S75); normal composite stress, pink (Eq. 11); and solvent volume fraction, blue (eq. S76). A constant solvent flux J_s compresses the polymer network. Compression of the polymer network is dictated by the imposed pressure drop Δp and the modulus of uniaxial compression M. The sum of the solvent and polymer stress, that is, the normal composite stress, is constant throughout the membrane, with a discontinuity at the permeate boundary.

the compressibility of the polymer itself, with negligible influence of the gas phase. Likewise, the moduli of rigid, phase-separated, porous membranes (e.g., zeolites, macroporous polymers, or track-etched membranes) are not set by balancing the elastic stresses of the network against the mixing and solvating stresses of the solvent but instead reflect the stiffness of its elements and the morphology of the network. In both cases, the osmotic (free-draining) moduli of the networks are little affected by the presence (or absence) of solvent, breaking the interrelation between hydraulic and diffusive transport captured by the FS model. Whether the nanoscale heterogeneities observed in PARO membranes (*36*) violate the homogeneity assumed by SD and FS depends on whether they are osmotically active when hydrated and can therefore be treated as a swellable elastic network that is statistically homogeneous on length scales that are larger than the heterogeneities but smaller than membrane dimensions.

Another process involves high-pressure membrane performance under extreme pressures, as found in organic solvent nanofiltration (OSN) (*4*, *5*) and high-pressure reverse osmosis, for example, as is used for brine treatment (*39*–*41*). In these and other cases, membrane permeance decreases at high pressures, which has been hypothesized to result from membrane compaction (*4*, *39*, *41*).

Our analysis thus far has focused on lower-pressure permeation, where flux grows linearly with applied pressure. In this linear-response limit, the physical properties of the membrane change very little and are well approximated by their equilibrium values. Because the local friction coefficient ξ, longitudinal osmotic modulus M, and solvent velocity $v_s = j_s/\phi_s$ all depend on ϕ_s, concentration gradients can affect the flux expressions beyond their linear-response forms (Eq. 10).

The FS model provides a natural framework to understand, analyze, and predict nonlinear pressure-flux relations that are expected to arise in high-pressure reverse osmosis brine treatment (*39*–*41*) and OSN (*4*, *5*). Namely, the permeation flux stops increasing linearly at sufficiently high pressures, asymptotically approaching a limiting flux, as identified experimentally by Paul and Ebra-Lima (*5*) and in the OSN context (*4*). Hints of a limiting flux already appear in the leading-order FS solutions (Fig. 4), because $u_p \to L$ once $\Delta p \sim M$. The weak compression limit $\left(\delta\phi_p \ll \phi_p^0\right)$ is therefore only valid when Δp is much smaller than the longitudinal osmotic modulus M, which emerges as a natural scale for the transmembrane pressure. When $\Delta p \ll M$, polymer and solvent concentrations are perturbed only slightly, and the solvent flux varies linearly with Δp. When Δp approaches or exceeds M, however, the network compresses more substantially, the flux develops a nonlinear dependence on Δp, and material properties (e.g., M, ξ, ϕ_p, ϕ_s, and D) deviate substantially from their equilibrium values. A systematic perturbation solution to the unidirectional permeation problem, described in the SM, reveals the general first correction to the flux relation to be

$$J_s = \frac{\frac{D\Delta p}{M_0 L}}{1 + \frac{\Delta p}{2M_0}\left[-\phi_p^0(\ln\xi_0)' + \frac{\phi_p^0}{\phi_s^0} - 1\right]} \qquad (12)$$

where $(\ln\xi_0)' = d\ln\xi/d\phi_s|_{\phi_s^0} < 0$ captures the decrease in friction with solvent swelling. The second term in brackets reflects changes in flow velocity due to increased (compressed) polymer fraction ϕ_p, and the third reflects the decreased thickness of the compressed membrane.

The FS limiting flux relation (Eq. 12) justifies the empirical expression $a\Delta p/(1 + b\Delta p)$ found by Paul and Ebra-Lima to fit nonlinear flux measurements (*5*). Additionally, Eq. 12 predicts the physical form of the fitting parameters a and b and thus provides design guidance for future membranes. A key feature is that the longitudinal osmotic modulus M_0 sets the pressure scale above which the permeation flux starts to saturate. Although M_0 was not measured alongside the flux measurements, FS predicts the limiting (ceiling) flux at large pressure

$$J_s^{\text{ceiling}} \sim \frac{2D}{L}\left[-\phi_p^0(\ln\xi_0)' + \frac{\phi_p^0}{\phi_s^0} - 1\right]^{-1} \qquad (13)$$

to be independent of M_0, which can be compared to the measured ratio a/b. Paul and Ebra-Lima tabulated a, b, D, L, ϕ_p, and ϕ_s for 12 organic solvents in cross-linked rubber membranes (*5*); only $(\ln\xi_0)'$ was not determined. Basing ξ on Zick and Homsy's calculation of fluid flow through periodic arrays of rigid spheres (*19*) reveals Eq. 13 to predict measurements to within a factor of four for all 12 solvents (SM).

Furthermore, the fluid-solid model lays out a path to treat membranes for which neither the PF nor the SD approach is appropriate. For example, hydrogel pore-filled composite membranes (*42*) immobilize a homogeneous, nonporous polymer matrix within the pores of a rigid membrane support. The rigid pore walls constrain the deformations of the nonporous gel in a way that would be difficult to capture with SD, yet straightforward with FS. More generally, the FS approach provides a natural framework to treat mechanically heterogeneous membranes with rigid structural elements that pin or restrict the deformation of nonporous networks. Double network hydrogels (*43*, *44*), composed of an expanded rigid polymer mesh interpenetrated by an elastic swollen polymer network, have both high water content and high mechanical strength and toughness. Combining these two networks provides orthogonal control over permeability and mechanical stiffness and thus greater control over the limiting flux.

The FS model is primed to determine the dynamics of solutes and solvents in hydrogel systems placed under mechanical stresses or pressures—for example, hydrogels injected into the human body for drug delivery (*45*, *46*), tissue scaffolds (*47*), and chemomechanic sensors (*48*, *49*). The FS description of these bioengineered hydrogels connects system performance with material chemistry. Although we have specifically focused on binary solvent-polymer systems here, it is natural to expand the FS formulation to account for additional solute or solvent concentration fields, which occur both in membranes and more generally.

More broadly, the FS model lays out a framework for the design of membrane materials on the basis of ex situ characterizations. The degree of swelling $\left(\phi_s^0\right)$ and mechanical moduli (G, K, and M) may be measured macroscopically for a swollen gel of interest, and the collective diffusivity may be measured using dynamic light scattering, following Tanaka *et al.* (*28*), to give all the needed parameters for FS description. Moreover, performing such experiments while varying the osmotic pressure of the swelling solution could elucidate the dependence of M, ξ, and ϕ_s on osmotic or transmembrane pressure, providing a complete set of parameters for nonlinear flux relations such as Eq. 12. Finally, measuring these parameters as functions of the composition of the swelling solution (solvent mixtures, solute concentrations) would provide the parameters needed to predict multicomponent transport through membranes.

Here, we have shown that the pore-flow and solution-diffusion approaches both provide correct and consistent descriptions of solvent transport through nonporous, swollen membranes, despite the apparent discrepancies in their predictions of pressure and concentration profiles. These two approaches represent complementary descriptions of the same process and ultimately yield identical predictions. Neither approach is incorrect; instead, the predictions of one model are in perfect quantitative agreement with those of the other. After all, momentum and mass must both be conserved, mechanical and thermodynamic balances must both be respected. A subtle distinction between the two approaches lies in the phases they treat: Pore flow focuses exclusively on the liquid phase, whereas solution diffusion treats the polymer and solvent as a single, inseparable continuum phase. The fluid-solid model is general enough to enable transport modeling for membrane materials that neither of the classical models can handle. Given the central role of the longitudinal osmotic modulus M_0

in the fluid-solid model, it is imperative that it be measured as part of the standard suite of characterization experiments for swollen membranes (*50*).

REFERENCES AND NOTES

1. G. M. Geise *et al.*, *J. Polym. Sci. B Polym. Phys.* **48**, 1685–1718 (2010).
2. G. M. Geise, D. R. Paul, B. D. Freeman, *Prog. Polym. Sci.* **39**, 1–42 (2014).
3. J. G. Wijmans, R. W. Baker, *J. Membr. Sci.* **107**, 1–21 (1995).
4. H. Y. Jang *et al.*, *AIChE J.* **65**, e16757 (2019).
5. D. R. Paul, O. M. Ebra-Lima, *J. Appl. Polym. Sci.* **14**, 2201–2224 (1970).
6. L. F. Greenlee, D. F. Lawler, B. D. Freeman, B. Marrot, P. Moulin, *Water Res.* **43**, 2317–2348 (2009).
7. D. R. Paul, *J. Polym. Sci. Polym. Phys. Ed.* **11**, 289–296 (1973).
8. D. R. Paul, *J. Polym. Sci. Polym. Phys. Ed.* **12**, 1221–1230 (1974).
9. D. R. Paul, *Separ. Purif. Methods* **5**, 33–50 (1976).
10. M. L. White, *J. Phys. Chem.* **64**, 1563–1565 (1960).
11. H. Yasuda, C. E. Lamaze, A. Peterlin, *J. Polym. Sci. A-2 Polym. Phys.* **9**, 1117–1131 (1971).
12. A. Peterlin, H. Yasuda, *J. Polym. Sci. Polym. Phys. Ed.* **12**, 1215–1220 (1974).
13. M. Qasim, M. Badrelzaman, N. N. Darwish, N. A. Darwish, N. Hilal, *Desalination* **459**, 59–104 (2019).
14. P. M. Biesheuvel, S. Porada, M. Elimelech, J. E. Dykstra, *J. Membr. Sci.* **647**, 120221 (2022).
15. L. Song, M. Heiranian, M. Elimelech, *Desalination* **520**, 115360 (2021).
16. W. K. Kim, S. Milster, R. Roa, M. Kanduč, J. Dzubiella, *Macromolecules* 10.1021/acs.macromol.2c00605 (2022).
17. E. P. Chan *et al.*, *Macromolecules* **53**, 1443–1450 (2020).
18. E. M. Furst, T. M. Squires, *Microrheology* (Oxford Univ. Press, 2018).
19. A. A. Zick, G. M. Homsy, *J. Fluid Mech.* **115**, 13–26 (1982).
20. S. Rosenbaum, O. Cotton, *J. Polym. Sci. A-1 Polym. Chem.* **7**, 101–109 (1969).
21. J. S. Ham, M. C. Bolen, J. K. Hughes, *J. Polym. Sci.* **57**, 25–40 (1962).
22. S. D. Gehman, *Rubber Chem. Technol.* **40**, 532–543 (1967).
23. M. A. Biot, *J. Appl. Phys.* **12**, 155–164 (1941).
24. M. A. Biot, *J. Appl. Mech.* **23**, 91–96 (1956).
25. M. A. Biot, D. G. Willis, *J. Appl. Mech.* **24**, 594–601 (1957).
26. M. Tokita, T. Tanaka, *J. Chem. Phys.* **95**, 4613–4619 (1991).
27. T. Tanaka, D. J. Fillmore, *J. Chem. Phys.* **70**, 1214–1218 (1979).
28. T. Tanaka, L. O. Hocker, G. B. Benedek, *J. Chem. Phys.* **59**, 5151–5159 (1973).
29. M. Doi, *J. Phys. Soc. Jpn.* **78**, 052001 (2009).
30. C. Y. Hui, V. Muralidharan, *J. Chem. Phys.* **123**, 154905 (2005).
31. J. Yoon, S. C. Cai, Z. Suo, R. C. Hayward, *Soft Matter* **6**, 6004–6012 (2010).
32. D. Chen, J. Yoon, D. Chandra, A. J. Crosby, R. C. Hayward, *J. Polym. Sci. B Polym. Phys.* **52**, 1441–1461 (2014).
33. R. B. Bird, W. E. Stewart, E. N. Lightfoot, *Transport Phenomena, Revised 2nd Edition* (John Wiley & Sons, 2006).
34. L. D. Landau, E. M. Lifshitz, *Theory of Elasticity*, vol. 7 of *Course of Theoretical Physics*, J. B. Sykes, W. H. Reid, Trans. (Addison-Wesley, 1959).
35. T. Tanaka, S. Ishiwata, C. Ishimoto, *Phys. Rev. Lett.* **38**, 771–774 (1977).
36. T. E. Culp *et al.*, *Science* **371**, 72–75 (2021).
37. Y. Hu, X. Zhao, J. J. Vlassak, Z. Suo, *Appl. Phys. Lett.* **96**, 121904 (2010).
38. M. G. Shin *et al.*, *J. Membr. Sci.* **578**, 220–229 (2019).
39. M. T. M. Pendergast, J. M. Nygaard, A. K. Ghosh, E. M. Hoek, *Desalination* **261**, 255–263 (2010).
40. D. M. Davenport *et al.*, *J. Membr. Sci.* **610**, 118268 (2020).
41. D. M. Davenport, A. Deshmukh, J. R. Werber, M. Elimelech, *Environ. Sci. Technol. Lett.* **5**, 467–475 (2018).
42. N. Adrus, M. Ulbricht, *J. Mater. Chem.* **22**, 3088–3098 (2012).
43. J. P. Gong, *Soft Matter* **6**, 2583–2590 (2010).
44. S. Ahmed, T. Nakajima, T. Kurokawa, M. Anamul Haque, J. P. Gong, *Polymer* **55**, 914–923 (2014).
45. T. R. Hoare, D. S. Kohane, *Polymer* **49**, 1993–2007 (2008).
46. J. Li, D. J. Mooney, *Nat. Rev. Mater.* **1**, 16071 (2016).
47. J. M. Cloyd *et al.*, *Eur. Spine J.* **16**, 1892–1898 (2007).
48. H. Ding *et al.*, *Carbohydr. Polym.* **248**, 116797 (2020).
49. G. Lin, S. Chang, C. H. Kuo, J. Magda, F. Solzbacher, *Sens. Actuators B Chem.* **136**, 186–195 (2009).
50. J. M. Dennison, X. Xie, C. J. Murphy, D. G. Cahill, *ACS Appl. Nano Mater.* **1**, 5008–5018 (2018).

ACKNOWLEDGMENTS

Funding: This work was supported as part of the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under award DE-SC0019272. **Author contributions:** V.H.H.: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, and Writing – review & editing. M.F.D.: Conceptualization, Writing – review & editing, Supervision, Project administration, and Funding acquisition. T.M.S.: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision, Project administration, and Funding acquisition. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm7192
Supplementary Text
Table S1
Nomenclature
References

Submitted 6 October 2021; accepted 17 May 2022
10.1126/science.abm7192

GENDER-BASED VIOLENCE

Policing in patriarchy: An experimental evaluation of reforms to improve police responsiveness to women in India

Sandip Sukhtankar[1,2]*, Gabriele Kruks-Wisner[3], Akshay Mangla[4]

Gender-targeted police reforms are frequently proposed to tackle the global problem of rising yet under-reported gender-based violence (GBV)—but with mixed and often disappointing results. We explore this issue in India, a country with alarming rates of GBV and limited police capacity, by studying the impact of Women's Help Desks (WHDs): dedicated spaces for women in local police stations, staffed by trained officers. Drawing on the largest randomized controlled trial of a police reform to date (180 police stations serving 23.4 million people), we find that officers in stations with WHDs are more likely to register cases of GBV, particularly where female officers run the desks. This suggests that even in resource-constrained and patriarchal environments, police responsiveness can be improved by focusing and mainstreaming attention to women's cases and by greater gender representation within the police.

Countries worldwide grapple with the dual problems of rising gender-based violence (GBV) and weak law enforcement. The ability to register crimes with the police is one essential step in accessing the justice system and promoting women's security. Yet social barriers, including stigma, can deter women from reporting crimes, as can low trust in the police (*1*, *2*). The police, moreover, are often unresponsive to women's concerns, leading to both under-reporting and under-recording of crime and violence (*3*). The result is a pronounced gap between the incidence of crime and the rates at which such crime is formally recognized.

Gender-targeted police reforms, which seek to increase the accessibility and accountability of the police to women, are regularly proposed to tackle these problems. One set of reforms have worked to increase women's representation within police forces. These initiatives build on theories of representative bureaucracy, which hold that the presence of members of marginalized groups within a public agency improves performance with respect to those groups (*4*, *5*). Evidence on the efficacy of these reforms, however, remains mixed. Some studies find that reporting and arrests for sexual assault increase with the presence of female officers (*6*, *7*). Other research suggests that gender differentials in police behavior are diminished as female officers operate within male-dominated policing cultures (*8*, *9*), noting a lack of increased gender sensitivity among female officers (*10*, *11*) who, like their male counterparts, often blame victims or dismiss their claims (*12*, *13*).

A second set of reforms involve gender-segregated services such as all-women police stations, established in countries as diverse as Brazil, Ghana, India, Kosovo, and the Philippines. These women-only stations, like other gender-segregated spaces such as women's train carriages or girls' schools (*14*, *15*), operate on the theory that women may be more comfortable in the absence of men. The assumption is that female victims of crime will be more likely to report cases in women-only police stations, and that female officers in those stations will be less constrained by patriarchal policing cultures (*16*). However, recent research suggests that separating women's cases from other police work makes it less likely that officers in regular, mixed-gender police stations will file cases related to GBV, creating barriers to access that further marginalize women (*11*, *13*, *17*).

In this article, we examine a large-scale gender-targeted police intervention that seeks to mainstream rather than segregate attention to women's cases within routine police work, while also testing whether increased representation of female officers better meets women's needs. Our setting is India, a country plagued by some of the world's highest rates of violence against women. Home to entrenched patriarchal norms and chronically weak state capacity (*18*), India is both a critical case and a representative one, given the prevalence of such conditions in settings throughout the Global South. Theoretical predictions about whether an intervention of this kind ought to have an impact are ambiguous, given the mixed evidence of existing research, in addition to the difficulty of implementing such reforms. Moreover, efforts to mainstream rather than segregate attention to women's cases have not been rigorously studied.

To fill these gaps, we employ what is to our knowledge the largest randomized controlled trial of a police reform to date, implemented across 180 stations serving a population of 23.4 million. The trial evaluates the impact of introducing Women's Help Desks (WHDs): dedicated spaces mandated to respond to women's cases, located within regular (mixed gender) police stations, staffed by officers who are trained in assisting women, and supported through routine monitoring, coupled with community outreach. We find that officers in stations with WHDs are more likely to register cases of GBV and other complaints filed by women, particularly where female officers are (randomly) assigned to run the help desks. This suggests that, even in resource-constrained and patriarchal environments, police responsiveness can be improved by mainstreaming and focusing attention on women's cases and by greater gender representation within the police. We explore the conditions under which these changes occurred, as well as the limits of the reform—which influenced officer behaviors and interactions with complainants but had little observable effect on the overall rates at which women turn to the police.

Reforming police under patriarchy

India, which was recently labeled the "world's most dangerous country for women" owing to high rates of sexual violence (*19*), is an instructive place to study gender-targeted police reforms. India ranks 140 of 156 countries on the Global Gender Gap Index—an international measure of gender inequality. This inequality is visible in India's skewed sex ratio at birth (929 girls/1000 boys), low female labor force participation rate (22.3%) (*20*), and high rates of GBV. An estimated 4 in 10 Indian women report experiencing domestic violence in their lifetimes (*21*).

GBV is a deeply rooted problem reflecting a nexus of social, economic, and political factors, and state-led interventions—including policing—are only one slice of the necessary responses. Policing is a particularly crucial but fraught arena in which to consider efforts to address GBV. The police are the primary institution for public security, as well as the most visible arm of the state's coercive apparatus. The police are charged with protecting women at the same time that law enforcement itself can play a role in perpetrating violence—including violence against women (*22*). The police are also critical gatekeepers to the broader justice system, yet where embedded in patriarchal norms can serve to deter or block women's legal claims. We focus on police registration of cases of GBV as a preliminary step in a long chain to justice, including case investigation, arrests, judicial action, and referrals to social services. Although case registration is just one step, it is a critical

[1]Department of Economics, University of Virginia, Charlottesville, VA 22904, USA. [2]Abdul Lateef Jameel Poverty Action Lab (J-PAL), Cambridge, MA, USA. [3]Department of Politics, University of Virginia, Charlottesville, VA 22904, USA. [4]Saïd Business School, University of Oxford, Oxford OX1 1HP, UK.
*Corresponding author. Email: srs8yk@virginia.edu

one: If police cases are not filed, further legal action cannot occur.

In India, registration takes the form of a First Information Report (FIR). Lodged at a police station, the FIR records information about a "cognizable" offense (classified in the Indian Penal Code as offenses of a serious nature for which a police officer has authority to make an arrest without a warrant and to start investigation without permission from a court) (*23*). The FIR initiates case investigation and criminal proceedings, as stipulated under the Indian Penal Code. Domestic violence can also be registered in a Domestic Incident Report (DIR), a complaint mechanism created under the Domestic Violence Act of 2005. Filing a DIR initiates civil proceedings and referrals to social services, allows for protection orders and economic support, and may also lead to criminal proceedings. Unlike FIRs, which are filed at police stations, the DIR is filed with the local magistrate (judge). In the latter instance, the police can serve as designated Protection Officers who compile the DIR and bring it before the magistrate. More details on FIRs and DIRs are in Appendix S1.

Women face major obstacles to case registration, reflecting barriers to both demand (that inhibit the reporting of crime) and supply (that inhibit officers from recording cases) (*24*–*26*). Even when a woman overcomes social and familial pressures to report a case, officers often resist officially recording it—despite their legal obligation to do so (*27*). The police's hesitancy to file cases is driven in part by acute resource and capacity constraints, which push officers to lessen their caseloads (*28*), as well as by political pressure to show lower official crime rates (*29*). Patriarchal policing norms also push against case registration: Officers are encouraged to "protect families" by promoting reconciliation rather than the legal rights of women (*27*, *30*–*32*) and often blame victims of sexual assault or question the validity of their claims. A recent report, for example, found that 39% of officers believe that complaints of GBV are unfounded (*33*). The same report also highlighted patriarchy within India's police stations. Women make up just 7% of the force nationally and face heavy work burdens, as well as workplace discrimination. Female officers, operating in these highly masculine settings, may feel pressure to act as "one of the boys," replicating and expressing patriarchal norms (*34*). These factors push against case registration: Indeed, there were fewer than four GBV-related FIRs registered per police station (serving 130,000 people on average) per month at baseline in our data.

Intervention and randomization

With these barriers to case registration in mind, we conducted an experiment to test the impact of a bundle of gender-targeted reforms designed to make the police both more accessible and accountable to female complainants (*35*). Our study setting is Madhya Pradesh, a large (population 81 million) and ethnically diverse state in north-central India. Madhya Pradesh is representative, in socioeconomic indicators and gender norms, of much of northern India. It is also an illustrative setting in which to explore the problem of under-reported GBV. A report from the state's four largest cities, for example, found that only 1% of women who had experienced violence had reported it to the police (*36*).

India's federal architecture assigns state governments the responsibility for policing, as well as social programming for women's health and security. The WHD intervention was designed by the Madhya Pradesh Police (MPP) Research and Training Department, in consultation with our research team, lawyers, and GBV experts from civil society, with the goal of overcoming barriers to both the reporting of crimes by women and the recording of such cases by the police.

Undertaking research on GBV and policing is a fraught endeavor, which requires careful ethical consideration. We recognize that our roles as researchers are not and never can be fully neutral and that there is the potential, through research, to introduce risk, activate trauma, and create harm. We discuss our efforts to minimize such risks in Appendix S2. We are guided by recommendations developed specifically for research on GBV (*37*, *38*), as well as by the insights of our local partners, which enabled us to ground global best practices in our study context.

The WHD intervention consists of four components: (i) private spaces (such as a room or cubicle) for female complaints within police stations; (ii) standard operating procedures on how to register cases and assist women visitors, along with officer training on those procedures and routine implementation monitoring; (iii) outreach to local women's and community safety networks; and (iv) assignment of female officers (at the rank of Assistant Sub-Inspector or higher) to run the WHDs.

Police stations—the unit of randomization—assigned to the first treatment arm ("regular" WHDs) received the first three components, whereas those assigned to the second arm (woman-run WHDs) received all four components. WHDs in the second treatment arm were directed to have designated female officers (with 90% compliance). Although there was no prohibition on assigning female officers to WHDs in the first treatment arm, most (72%) were run by male officers. The control group continued with business as usual.

For the study, the MPP purposively selected 12 districts (out of 51) across the state, automatically including those home to the state's four largest cities and selecting the remainder to be representative of geography, demographics, and socioeconomic conditions across the state (*39*). We excluded specialized police stations (state or district headquarters, cyber cells, or all-women stations), as well as outposts (smaller than a station). MPP also excluded stations that were remote, servicing entirely rural populations, because they did not consider traffic to those stations high enough to justify a continuously staffed desk. The resulting sample of 180 police stations should thus be viewed as being representative of primarily (although not exclusively) urban and large police stations. Within each district, we stratified these 180 stations by geography (those that were fully urban, and those that also included rural areas) and by the first principal component of a vector of police station size characteristics (*35*). Within each of these strata, we then randomly assigned the stations into three groups: 61 to the first treatment arm (regular WHDs), 59 to the second (woman-run WHDs), and 60 to the control arm, which received no help desk intervention.

We consider May 2019 as the start date of the full intervention. Training began as early as July 2018, although state elections in December 2018 meant that training did not fully ramp up until March 2019. Our field teams monitored implementation in December 2019, by which point all treatment stations had at least some training, 94% had set aside requisite space, 87% had conducted a community outreach event, and 90% of women-run WHDs had a female officer assigned. However, only 67% had full training at the station level, and community outreach was limited overall given the huge jurisdictions these stations served. Appendix S4 details implementation efforts and timelines.

Data sources

Our aim was to evaluate whether the establishment of the WHDs, as well as assignment of female personnel to the WHDs, improved the responsiveness of police officers to women (*40*). To that end, we gathered data from five main sources.

1) Administrative data on crimes registered by the police at study police stations from May 2018 through March 2020. Our data are aggregated with no individual or identifiable case details, but FIR categories indicate (i) the number of "crimes against women" (CAW) cases, which include officially designated cases of GBV such as sexual assault, rape, dowry, and other cases (details in Appendix S1); and (ii) whether the case was filed by a woman, the latter also incorporating nonviolent crimes.

2) CCTV data from the video feeds of cameras, present in all police stations, focused on the station entrance. MPP provided us with a week's worth of data from each study station, for the hours of 10 a.m. to 10 p.m. each day, at

Table 1. Primary outcomes from administrative data. Each observation represents data at the police station–month level. DIRs are Domestic Incident Reports representing civil complaints of domestic violence (columns 1 and 5). FIRs are First Information Reports, either in cases of Crimes Against Women (CAW) filed by anyone (columns 2 and 6) or in all criminal cases filed by women (columns 3 and 7). Arrests correspond to arrests in CAW cases in a given month (columns 4 and 8). Regular WHD refers to Women's Help Desks without an assigned female officer; woman-run WHDs include an assigned female officer. Strata FE refers to fixed effects for district × urban/rural strata. Standard errors clustered by police station are in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DIRs filed	FIRs filed in CAW cases	FIRs filed by women	Arrests	DIRs filed	FIRs filed in CAW cases	FIRs filed by women	Arrests
Treatment	1.452***	0.542*	0.269	0.265				
	(0.180)	(0.308)	(0.168)	(0.433)				
Regular WHD					1.492***	0.095	0.073	0.126
					(0.250)	(0.339)	(0.188)	(0.457)
Woman-run WHD					1.410***	1.014**	0.476**	0.412
					(0.205)	(0.400)	(0.207)	(0.506)
Baseline control	0.192	0.409***	0.344***	0.210***	0.191	0.408***	0.344***	0.209***
	(0.268)	(0.041)	(0.036)	(0.041)	(0.265)	(0.040)	(0.035)	(0.040)
Observed	1980	1980	1980	1980	1980	1980	1980	1980
Adjusted R^2	0.179	0.270	0.254	0.164	0.179	0.279	0.259	0.164
Control mean	0.047	3.847	2.577	3.433	0.047	3.847	2.577	3.433
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
***p* value: T1 = T2**					0.772	0.030	0.058	0.498

both baseline (n = 12,537, January to March 2019) and endline (n = 9757, February to March 2020). This footage enables us to measure the number of men and women entering or exiting the police station.

3) A user survey of members of the public who had visited study police stations, asking about their satisfaction with their visit. This was conducted at endline only (n = 3251, February to March 2020), drawing from a random selection of visitors within the span of 1 week.

4) A police survey, carried out at baseline (n =1950, September to October 2018) and endline (n = 1961, February to March 2020) of personnel in different roles and ranks in study stations. The survey—a representative individual panel with equivalent rank replacements for any transferred officers—captured police perceptions and attitudes on crimes against women.

5) A survey of citizens, carried out at baseline (n = 5648 women, 871 men, November 2018 and March to April 2019) and endline (n = 3376, July to December 2020, phone survey). The survey—an individual panel sampled from all adult residents living in study police station jurisdictions—asked about perceptions of safety, opinions of and contact with the police, and experiences of crimes.

Concurrent to this data collection, we also carried out qualitative research—observation and interviews with officers of various ranks—in eight police stations, selected to represent each arm of the study in two purposively selected districts (*41*).

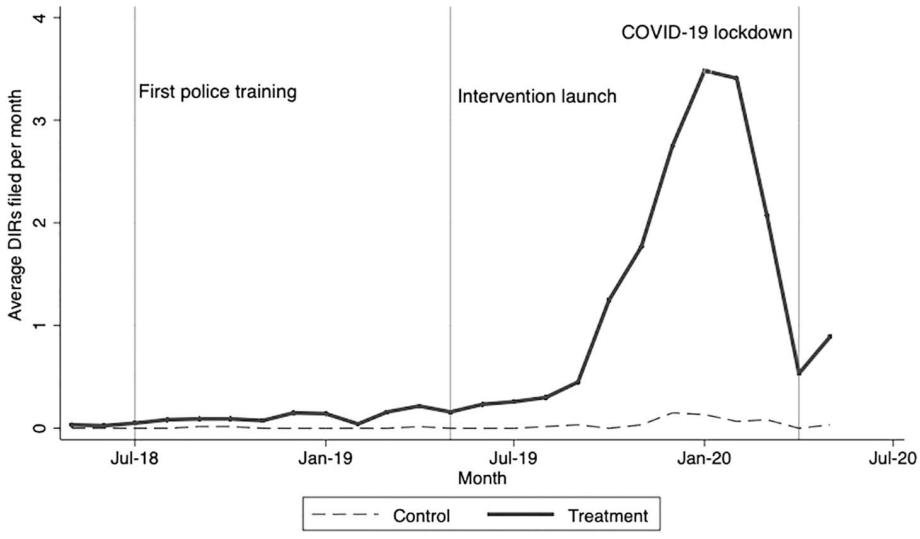


Fig. 1. Domestic incident registration increased sharply. Source: Raw official administrative data.

Statistical methods

We report Intent to Treat estimates comparing average outcomes in treated stations (with either regular or woman-run WHDs) to outcomes in the control stations, as well as outcomes between the two treatment groups, using the most disaggregated measure available. All regressions include district-geographic stratum fixed effects (the level at which treatment probabilities are equal), with a control for the first principal component of variables used for further stratification. Sampled observations are weighted using inverse sampling probabilities to make the outcomes representative at the police station level. Where available, we include the baseline value of the outcome as a control. We cluster standard errors at the police station, the level of randomization. Our preferred specifications are

(1) Y_ips = alpha + beta*treatment_ps + gamma*Y0_ips + delta_s + pc_ps + epsilon_ips

(2) Y_ips = alpha + beta1*treat1_ps + beta2*treat2_ps + gamma*Y0_ips + delta_s + pc_ps + epsilon_ipswhere i is the individual, p is the police station, s is the stratum, Y is endline and Y0 is the baseline outcome. Using

baseline outcomes, we find no more imbalances across treatment groups than would be expected by chance (tables S1 to S3) (*42*).

Results

The WHD intervention led to increased registration of women's cases (Table 1). First, we observe an increase in the number of DIRs filed, from practically zero in the control group (where officers had little to no knowledge of the DIR) to 1.5 monthly cases in the treatment group ($p < 0.01$). Figure 1 shows the sharp increase in DIRs after the official launch of the program in May 2019, although some cases were registered soon after training started in treatment stations as early as July 2018. Second, we also observe sizable increases in the number of FIRs filed in CAW cases (14.1%, $p = 0.08$), as well as FIRs filed by women (10.4%, $p = 0.1$). These results reflect an additional 1905 DIRs and 3360 FIRs registered in treatment stations over the 11 months of the intervention. There were no significant changes in the number of arrests in CAW cases; the coefficient is positive but standard errors are large [95% confidence interval (CI) −0.58, 1.11] (*43*).

Separating the analysis by treatment group, we see that officers in stations with both regular and woman-run WHDs filed increased numbers of DIRs, with the coefficients statistically indistinguishable from each other. However, the increase in FIRs is entirely driven by the woman-run WHDs (Fig. 2). In regular WHDs, the treatment effects are small and statistically indistinguishable from zero. In the stations with woman-run WHDs, by contrast, FIRs in CAW cases increased by 26.4%, whereas FIRs filed by women rose by 18.5%; these treatment effects are statistically distinguishable from those in regular WHDs ($p = 0.03, 0.06$; Table 1). Meanwhile, implementation quality and officer training both have a strong and significant impact on the number of cases registered for all outcomes (table S4). Finally, increases in registration of CAW cases do not come at the cost of reductions in other cases, with no discernible spillover effect on the overall number of FIRs or on other kinds of police reports (table S5).

These results are driven by changes in police behavior, reflected in a greater likelihood of registering a case once a woman has reported it, but not by any observable changes in the rates at which women report cases or approach the police. There is no impact on the overall CAW rates reported by women in our citizen survey (table S17). Neither is there an observable impact on the rates at which women visited police stations, based on analysis of CCTV

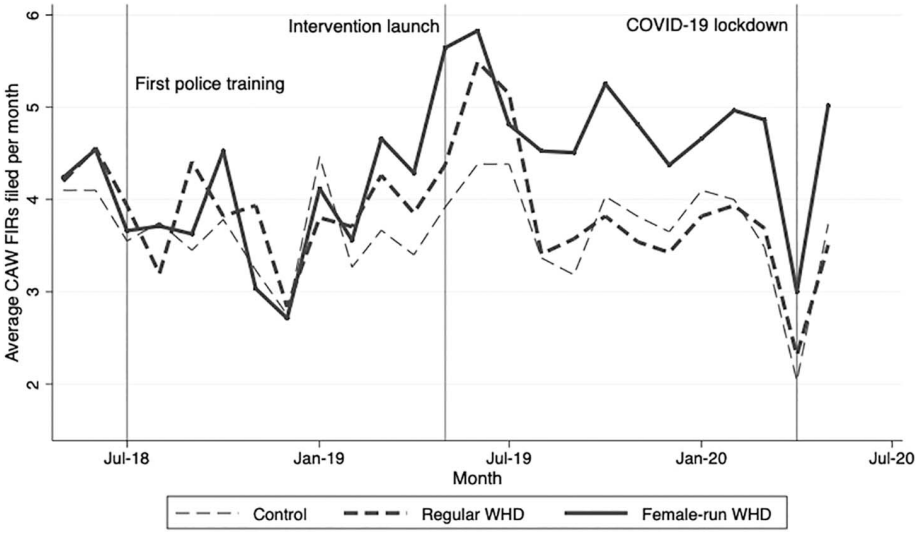


Fig. 2. Registration of crimes against women increased in women-run WHDs. Source: Raw official administrative data.

Table 2. Primary outcomes from CCTV data. Observations are at time window–day–police station level, as described in Appendix S1. The number of female visitors is the per-hour count of all women who entered a police station at any time between 10 a.m. and 10 p.m. (columns 1 and 3). The proportion of female visitors is the number of female visitors divided by the number of all visitors captured by CCTV camera in the same hourly durations (columns 2 and 4). Regular WHD refers to Women's Help Desks without an assigned female officer; woman-run WHDs include an assigned female officer. We use district FE instead of strata FE owing to the unavailability of data from 41 police stations, which results in an unbalanced distribution of treatment and control police stations across some strata. Though not shown in the table, the regressions control for the number of female officers in each police station at endline; fixed effects for time window of day and day of week; the average number of frames per second in the video; and the average number of pixels in the video.

	(1)	(2)	(3)	(4)
	No. of female visitors	**Percent female visitors**	**No. of female visitors**	**Percent female visitors**
Treatment	−0.590	0.005		
	(1.656)	(0.011)		
Regular WHD			−0.811	0.009
			(1.894)	(0.011)
Woman-run WHD			−0.402	0.001
			(2.097)	(0.012)
Baseline control	0.226***	0.359***	0.227***	0.359***
	(0.052)	(0.062)	(0.052)	(0.061)
Observed	1832	1831	1832	1831
Adjusted R^2	0.162	0.168	0.162	0.169
Control mean	13.706	0.092	13.706	0.092
District FE	Yes	Yes	Yes	Yes
***p* value: T1 = T2**			0.857	0.403

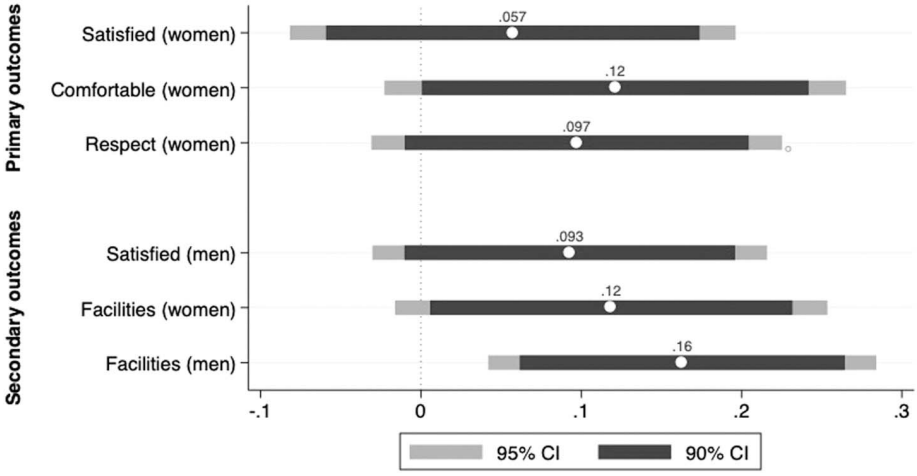


Fig. 3. Women visiting treatment stations were slightly more satisfied. Source: User Survey; coefficients and notes in table S7.

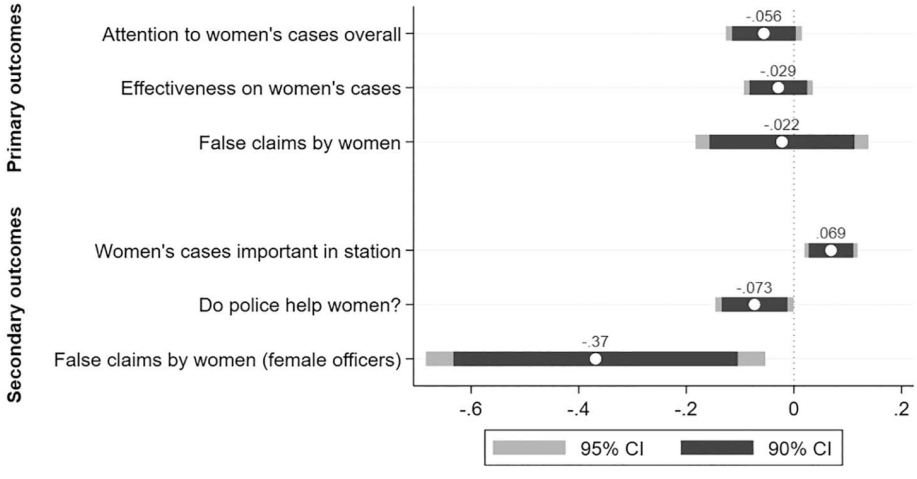


Fig. 4. Police attitudes and awareness. Source: Police survey; coefficients and notes in tables S13 and S14.

data (Table 2), although these latter results must be viewed with caution given missing data issues (*44*). These non-impacts are not a result of these variables being uninformative or unpredictive of outcomes in general: Survey CAW numbers are predictive of FIRs registered, baseline CCTV counts are predictive of endline CCTV counts, and the number of female staff (from administrative data) is predictive of the number of female visitors in the CCTV data. We do not observe any broad changes in citizen behavior, although it is possible that there were changes in complainant behavior within the police stations, as women decide whether or not to pursue cases once at the station and in interaction with officers. Our research design and ethical parameters precluded us from observing any such interactions, but we posit that changes in complainant behavior are likely supported by changes in police behavior that we do observe. Among complainants, we do find minor changes in the satisfaction of female visitors to police stations, expressed in exit interviews (Fig. 3). In control stations, respondents generally expressed high levels of satisfaction with their visits (3.16 on a 4-point scale), agreed they were treated respectfully (3.31), and that they felt comfortable discussing their concerns (3.3). The high levels of reported satisfaction may reflect desirability bias among respondents hesitant to critique the police at the station, and given top-coded values, there may be limited scope to observe any improvement due to the WHD intervention. It follows that there are few statistically distinguishable differences in satisfaction between treatment and control stations; only the "comfort" variable is significantly improved given the presence of a WHD (p = 0.07, table S7). Notably, however, all three measures of user satisfaction were significantly higher, comparing among treatment stations, in those with better implementation and more training (table S8). In addition, both female and male visitors to treatment stations were significantly more likely to profess satisfaction with the physical infrastructure and conditions of the station, compared to visitors to control stations. Overall, these results suggest modest improvements in citizen perceptions of the police given the presence of a WHD. Meanwhile, deep-seated police attitudes about gender did not appear to shift overall; there was no significant change in the rates at which officers reported the belief that women file so-called "false cases" against men (an indicator of whether they are inclined to believe women or dismiss their claims) (Fig. 4) (*45*). Notably, however, female officers in both treatment arms were less likely than those in control stations to ascribe to the narrative of false cases, though there was no such effect on male officers; this suggests that police attitudes shifted more for women. There is some evidence of a shift in officers' awareness of the general inadequacies of policing with regard to women's cases (Fig. 4). Both male and female officers in treatment stations were significantly less likely to describe the police as helpful to victims of crimes against women and were also less likely to believe that the police pay sufficient attention to women's cases relative to other law and order issues (this last result is borderline significant; p = 0.12, table S13). We do not interpret these findings as evidence of a shift in gender attitudes, but rather of increased cognizance of the gap between how women's cases should be handled and how they are handled in practice—a likely consequence of WHD training. Both male and female officers in treatment stations were also more likely to state (unprompted) that cases of crimes against women are among the top priorities within their stations—a likely effect of the attention-focusing presence of the WHDs.

Discussion

Our findings suggest that even in resource-constrained and patriarchal environments, efforts that focus attention on women's cases can have a meaningful impact on police behavior, making officers more responsive to women's security concerns. This is visible in the higher registration of both FIRs and DIRs (respectively, criminal and civil complaints). The rise in DIRs across both treatment arms is pronounced, reflecting the adoption of a relatively new practice (since 2005) that remains largely unknown to the police without WHD training, and so is all but absent at baseline and in control stations. Police officers are required to assist women in creating the report and must also ensure that the form is lodged with a local magistrate and that social services (such as shelter homes) are accessible—activities that extend beyond conventional police work. Through

WHD training, officers gained knowledge of the DIR and learned to coordinate with other state and civil society agencies.

The increase in FIRs for women's cases is also notable, as is the fact that it is driven almost entirely by woman-run WHDs. This is not simply a function of increased personnel assigned to the woman-run WHDs: Both woman-run WHDs and regular WHDs received additional high-ranked officers to operate the help desk (table S9). It appears, then, that the presence of additional female officers is critical for overcoming barriers to FIR registration. This gender-differentiated effect for FIRs but not DIRs, we suggest, reflects the higher costs to officers of filing an FIR. Unlike the DIR, the FIR automatically initiates a criminal case, requiring substantial investments of police time for investigation and in court proceedings. Moreover, to file an FIR, officers must push against strong norms within the police—articulated in our qualitative research—that prioritize "protecting families" by avoiding legal proceedings, in addition to dismissive narratives about "false cases." The DIR, though certainly not trivial, requires less from officers, in part because it initiates a civil rather than criminal case, and in part because it passes some of the burden to other government agencies. It is also a relatively new practice that does not have the same norms pushing against it. Filing an FIR, in sum, requires higher levels of officer commitment, which we see primarily within woman-run WHDs.

Questions remain about the sources of commitment and the agency of female officers. More research is required to explore the mechanisms through which female officers affected change through the WHDs, as well as on how the WHD bundle affected female officers. However, our qualitative research suggests two avenues that, together, appear to empower female officers and increase their responsiveness to women. First, the intervention worked to build station-level capacity for action on women's cases, channeling resources to treatment stations, while also introducing manuals to clarify complex legal procedures, alongside intensive training. These investments in officer capacity may have been most acutely felt and utilized by female officers, who tend to be more overburdened—given their small numbers and the many tasks assigned to them—relative to male officers (*33*, *34*). Female officers appear to have been particularly responsive to WHD training, which emphasized the legal requirement to file FIRs in the case of cognizable offenses, and which also contained gender sensitization modules that urged officers to listen to women's claims and not dismiss them out of hand. The impact of this training is visible in the gender-differentiated rates of filing FIRs (more likely in woman-run stations) and shifts in beliefs concerning "false cases" (significant for female but not male officers).

Second, the WHD intervention worked to mainstream gender-responsive policing practices. By locating the help desks in regular police stations, and by training both male and female officers on the mission and operation of the desks, the WHDs helped to give visibility and ascribe value to work on women's cases, rather than casting such work as peripheral to and therefore of lesser importance than other crime prevention tasks. Further study is required to examine the effects of this mainstreaming, which—in creating an enabling environment in which to focus on women's cases—may help us understand why female officers were particularly receptive to WHD training and protocols.

These observations carry important insights for debates over representative bureaucracy and the question of whether the presence of members of marginalized groups within a public agency, such as the police, improves performance (*4*, *5*, *30*). Our findings suggest that descriptive representation does matter; female officers played a critical role in shaping the impact of the help desks. This, however, was not simply a matter of "feminizing" a police station by adding female officers. Far from it the agency of female officers must be understood as part of the full WHD bundle, including training, infrastructure, and higher-level supports that enabled those officers to work—both as women and for women. Efforts to enhance bureaucratic representation, this suggests, may hinge on institutional supports that activate the agency of underrepresented public personnel. In the case of the WHDs, the mainstreamed nature of the intervention (housed in regular, mixed-gender police stations) may have played a critical role. Although we are not able to directly compare the WHDs to all-women stations, recent research suggests that the act of "segregating" women's cases may have the perverse effect of marginalizing female complainants (creating barriers and displacing cases) as well as female officers (who are isolated and sidelined from broader policing structures) (*13*, *30*, *46*). Locating women's help desks in regular stations, by contrast, appears to have worked to increase attention to women's security within "normal" police work. Future research is required to systematically explore such mainstreaming attempts, which may prove critical to broader efforts to build and support more representative bureaucracies.

Our study also highlights the limitations of police-centered reforms. First, gender attitudes among the police are difficult to move (although perhaps less so for female officers). Second, even as the WHD program increased the likelihood of the police recording crimes against women, barriers to women reporting such crimes remain. This may reflect limited community outreach implementation, with only 10% of women in our citizen survey aware of WHDs (table S20). Citizens' attitudes toward the police, moreover, did not shift, although this may reflect the short duration of the intervention; beliefs about the police may shift over time if the intervention becomes institutionalized and more visible to citizens. This, however, will depend not only on the ease with which women are able to register cases, but also on the functioning of the broader criminal justice system. Last, the WHD intervention did little to address deep-seated social and economic structures that both drive violence against women and inhibit women's access to justice. Efforts to address GBV necessarily require broader, multipronged approaches that extend beyond police reforms.

Conclusion

The urgent challenge of women's security has prompted police reform proposals worldwide, calling for increased female representation among officers, gender-sensitization training, and specialized services for women. To date, however, the limited evidence available on the efficacy of these reforms has been mixed, if not pessimistic, especially within patriarchal settings. Moreover, efforts aimed at mainstreaming (rather than segregating) these reforms have not been rigorously studied. Our experiment suggests that cautious optimism may be warranted, providing evidence that gender-targeted reforms can, under certain conditions, improve police responsiveness to women.

In addition to furthering our understanding of police reform and representative bureaucracy, our results also carry policy implications as governments in India and elsewhere consider gender-targeted reforms. Indeed, informed in part by these results, MPP has begun scaling up the WHD intervention to 700 police stations across the state. Studying this scale-up may help us understand whether the observed changes in police behavior can be sustained, as well as whether and how attitudinal change among both police and citizens can take hold.

REFERENCES AND NOTES

1. G. Blair *et al*., *Science* **374**, eabd3446 (2021).
2. Human Rights Watch, *"Everyone Blames Me": Barriers to Justice and Support Systems for Sexual Assault Survivors in India* (2017).
3. R. Manjoo, "Report of the Special Rapporteur on violence against women, its causes and consequences" (United Nations Office of the High Commissioner for Human Rights, 2012).
4. B. A. Ba, D. Knox, J. Mummolo, R. Rivera, *Science* **371**, 696–702 (2021).
5. L. Keiser, V. M. Wilkins, K. J. Meier, C. A. Holland, *Am. Polit. Sci. Rev.* **96**, 553–564 (2002).
6. K. Meier, J. Nicholson-Crotty, *Public Adm. Rev.* **66**, 850–860 (2006).
7. A. Miller, C. Segal, *Rev. Econ. Stud.* **86**, 2220–2247 (2019).
8. M. Morash, R. Haarr, *Fem. Criminol.* **7**, 3–23 (2012).
9. S. Karim, *Am. Polit. Sci. Rev.* **114**, 536–551 (2020).
10. S. Karim, M. J. Gilligan, R. Blair, K. Beardsley, *Int. Stud. Q.* **62**, 618–631 (2018).
11. N. Wagner, M. Rieger, A. Bedi, W. Hout, *J. Afr. Econ.* **26**, 492–515 (2017).
12. C. Santos, *Women's Police Stations: Gender, Violence, and Justice in São Paulo, Brazil* (Palgrave Macmillan, 2005).
13. N. Jassal, *Am. Polit. Sci. Rev.* **114**, 1035–1054 (2020).
14. C. Jackson, *J. Public Econ.* **96**, 173–187 (2012).

15. V. Ceccato, A. Loukaitou-Sideris, *Transit Crime and Sexual Violence in Cities: International Evidence and Prevention* (Routledge, 2020).
16. E. Perova, S. A. Reynolds, *Soc. Sci. Med.* **174**, 188–196 (2017).
17. S. Amaral, S. Bhalotra, N. Prakash, "Gender, crime and punishment: Evidence from women police stations in India" Working paper, 2021.
18. D. Kapur, *J. Econ. Perspect.* **34**, 31–54 (2020).
19. Thomas Reuters Foundation, "The World's Most Dangerous Countries for Women – 2018"; https://www.reuters.com/article/us-women-dangerous-poll-exclusive/exclusive-india-most-dangerous-country-for-women-with-sexual-violence-rife-global-poll-idUSKBN1JM01X.
20. *The Global Gender Gap Report 2021* (World Economic Forum, 2021).
21. A. Kalokhe *et al*., *Glob. Public Health* **12**, 498–513 (2017).
22. M. Berry, M. Lake, *To End Violence Against Women, We Need to Imagine New Futures* (IPI Global Observatory, 2021).
23. Noncognizable offences, by contrast, require police officers to obtain warrants before arrests and court orders before initiating investigation.
24. P. Roychowdhury, *Capable Women, Incapable States: Negotiating Violence and Rights in India* (Oxford Univ. Press. 2020).
25. A. Banerjee, R. Chattopadhyay, E. Duflo, D. Keniston, N. Singh, *Am. Econ. J. Econ. Policy* **13**, 36–66 (2021).
26. L. Iyer, A. Mani, P. Mishra, P. Topalova, *Am. Econ. J. Appl. Econ.* **4**, 165–193 (2012).
27. Commonwealth Human Rights Initiative, *Barriers in Accessing Justice* (CHRI, 2020).
28. Nationally, there are just 1.2 officers per 1000 people in India. The United States, by comparison, has 2.3 police officers per 1000 people, according to the UN Office on Drugs and Crime.
29. On politicization of the police in India, see (*47*).
30. N. Jassal, *J. Asian Stud.* **80**, 631–661 (2021).
31. CSDS, *Status of Policing in India Report 2018: A Study of Performance and Perceptions* (Centre for the Study of Development Societies, New Delhi, 2018).
32. The practice of reconciliation builds on Indian legislation, including the 1984 Family Courts Act, and various Supreme Court directives, which advise officers to provide "counseling" to preserve families.
33. Common Cause *et al*., *Status of Policing in India 2019: Police Adequacy and Working Conditions* (Common Cause and Lokniti, 2019).
34. Commonwealth Human Rights Initiative, *Rough Roads to Equality: Women Police in South Asia* (CHRI, 2015).
35. Our Pre-Analysis Plans were registered with the American Economic Association's RCT Registry at https://www.socialscienceregistry.org/trials/3357.
36. S. Neville *et al*., "Evaluation of the Madhya Pradesh Safe Cities Initiative. Baseline Report" (DFID, London, 2015).
37. C. Bradbury-Jones *et al*., *Best Practice Guidelines in Relation to Gender-Based Violence Research in Low- and Middle-Income Countries* (Univ. of Birmingham/Institute for Global Innovation, 2020).
38. World Health Organization, *Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women* (WHO, 2003).
39. Bhopal, Gwalior, Indore, and Jabalpur districts contain the (eponymous) four largest cities in MP; the other districts are Baitul, Balaghat, Morena, Panna, Ratlam, Rewa, Seoni, and Vidisha.
40. The registration of crime must be distinguished from the actual incidence of crime which, by nature, is not observable through official data given the many deterrents to both reporting and recording cases (*1*).
41. Interviews took place over repeated station visits during a 10-month period and were carried out by two Research Assistants (RAs) trained in qualitative methods. RAs visited each station approximately twice per month. Each visit lasted for a full day, yielding an average of 21.25 day-long visits per station. RAs interviewed four to six officers of different ranks in each station (including station chiefs, assigned help desk officers, and female officers where present). Interviews lasted for 1 hour on average, and most officers were interviewed multiple times. The qualitative research design is described in our Pre-Analysis Plan, Section 8 (*35*).
42. DIRs filed are significantly higher if data until April 2019 are considered, which is not surprising because many treatment stations had received training by that time. There is no difference in DIRs when including only data before July 2018 (prior to training). Excluding DIRs, out of 45 coefficients only one is significant at the 5% and one at the 10% level.
43. This lack of movement on arrests is unsurprising given the complex mix of factors that shape whether an arrest is appropriate (given the nature of a case), as well as the fact that some complainants, facing family pressure, may not want perpetrators arrested. Lack of increased arrests, therefore, should not be interpreted solely as an indicator of police inaction.
44. Because of data collection issues (COVID-related and otherwise) with the CCTV data, nearly one-third (55) of our study stations are not included in our endline analysis of these outcomes, so these results must be interpreted with caution.
45. The relatively tight standard error bounds visible in the figure suggest that these are precise null effects, which is unsurprising given deep-rooted attitudes.
46. Our aim and empirical strategy did not include a direct comparison of the WHD intervention and all-women stations, which are entirely different models of policing. Moreover, there were only six all-women police stations in our study districts in MP during our study period.
47. S. K. Mitra, "Policing and strategic reform" in *The puzzle of India's Governance: Culture, Context and Comparative Theory* (Routledge, 2007).
48. S. Sukhtankar, G. Kruks-Wisner, A. Mangla, 'Policing in patriarchy: An experimental evaluation of reforms to improve police responsiveness to women in India,' Harvard Dataverse, Version 2 (2022). https://doi.org/10.7910/DVN/R75XVZ.

ACKNOWLEDGMENTS

This study would not have been possible without the cooperation of the Madhya Pradesh Police department and its officials, in particular V. Kapoor, A. Shankar, R. Kumar Shukla, S. Rana, and R. Tiwari; as well as the enormous efforts of the JPAL-SA research team, in particular S. Bhattacharya, S. Suganya, S. Chandrasekhar, M. Tikku, M. Walia, P. Kumar, P. Baruah, K. Ravishankar, L. Diaz-Martin, S. Ramesh, and A. Bhargava; and the analysis team of J. Le, E. Robertson, and A. Kumar at UVA. We thank J. Doleac, O. Dube, S. Karim, S. Lakhtakia, A. Miller, P. Niehaus, S. Ostermann, N. Prakash, S. Sekhri, M. Williams, four anonymous referees, and participants in various seminars and conferences for comments. **IRB:** The study received IRB approval from the University of Virginia (#2548), the Institute for Financial Management and Research in India (#7107), and Oxford University (SSH_SBS_C1A_18_065). **Registration:** The trial is registered (with Pre-Analysis Plans) in the American Economic Association's randomized control trial registry under ID AEARCTR-0003357. **Funding:** This study was funded primarily by the Crime and Violence Initiative (CVI) of the Jameel Poverty Action Lab (JPAL), as well as the World Bank's Sexual Violence Research Initiative (SVRI), with supplemental funding provided by the University of Virginia's Center for Global Inquiry and Innovation (CGII) and the University of Oxford. **Author contributions:** Authors contributed equally to conceptualization, design, and writing. S.S. was responsible for randomization and data analysis. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** A Pre-Analysis Plan was registered with the American Economic Association's RCT Registry at: https://www.socialscienceregistry.org/trials/3357. All code, and survey and administrative data used in the analysis are available (with PII and police station names extracted to preserve anonymity as required by IRB) on the JPAL dataverse hosted by Harvard University (*48*). The raw CCTV videos will not be available given privacy issues, but the final data used for analysis are available. All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm7387
Appendix S1: Clarifications and further background
Appendix S2: Ethics statement
Appendix S3: Tables S1 to S24
Appendix S4: Implementation timeline and details, table S25, and figs. S1 to S6
References (*49*–*65*)

Submitted 7 October 2021; accepted 4 May 2022
10.1126/science.abm7387

NEUROSCIENCE

Sound induces analgesia through corticothalamic circuits

Wenjie Zhou[1]†, Chonghuan Ye[1]†, Haitao Wang[2,3]†, Yu Mao[1,4]†, Weijia Zhang[1], An Liu[5], Chen-Ling Yang[5], Tianming Li[6], Lauren Hayashi[6], Wan Zhao[7], Lin Chen[2], Yuanyuan Liu[6]*‡, Wenjuan Tao[5]*‡, Zhi Zhang[1]*‡

Sound—including music and noise—can relieve pain in humans, but the underlying neural mechanisms remain unknown. We discovered that analgesic effects of sound depended on a low (5-decibel) signal-to-noise ratio (SNR) relative to ambient noise in mice. Viral tracing, microendoscopic calcium imaging, and multitetrode recordings in freely moving mice showed that low-SNR sounds inhibited glutamatergic inputs from the auditory cortex (ACx^{Glu}) to the thalamic posterior (PO) and ventral posterior (VP) nuclei. Optogenetic or chemogenetic inhibition of the ACx^{Glu}→PO and ACx^{Glu}→VP circuits mimicked the low-SNR sound–induced analgesia in inflamed hindpaws and forepaws, respectively. Artificial activation of these two circuits abolished the sound-induced analgesia. Our study reveals the corticothalamic circuits underlying sound-promoted analgesia by deciphering the role of the auditory system in pain processing.

As early as 1960, there were accounts from dental operations showing that music and noise can induce analgesic effects (*1*). Music delivered as an intervention can alleviate postoperative and procedural pain (*2*–*4*) and even refractory pain in the clinic (*5*, *6*). Because diverse genres of music and even nature sounds can relieve pain to an equal extent (*7*), the inherent characteristics of music or contextual factors—that is, not only to music per se—have been hypothesized to drive these analgesic effects (*8*). However, it is still unknown how it works.

Functional magnetic resonance imaging (fMRI) studies have implicated changes in the activity of multiple brain areas mediating pain processing in humans exposed to music (*9*–*12*). To date, the neural substrates underpinning cross-modal audio-somatosensory interactions remain unclear. The thalamus relays multimodal sensory information (including auditory and somatosensory) (*13*–*15*) and reciprocally connects with pain-associated regions to process sensory discrimination and affective-motivation of pain through diverse subregions (*16*–*19*). Notably, the activity of the somatosensory thalamus can be affected by music and innocuous salient auditory stimuli (*15*, *20*). Thus, the thalamus might function as a bridge for audio-somatosensory processing. However, the precise cell type–specific organization and the function(s) of the thalamic circuits mediating sound-induced analgesia remain largely unknown.

[1]Department of Anesthesiology and Pain Medicine, The First Affiliated Hospital of USTC, Hefei National Laboratory for Physical Sciences at the Microscale, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, PR China. [2]Auditory Research Laboratory, Department of Neurobiology and Biophysics, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, PR China. [3]School of Integrated Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei, PR China. [4]Department of Anesthesiology and Pain Management, The First Affiliated Hospital of Anhui Medical University, Hefei, PR China. [5]Department of Physiology, School of Basic Medical Sciences, Anhui Medical University, Hefei, PR China. [6]Somatosensation and Pain Unit, National Institute of Dental and Craniofacial Research (NIDCR), National Center for Complementary and Integrative Health (NCCIH), National Institutes of Health (NIH), Bethesda, MD, USA. [7]Department of Otolaryngology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, PR China.
*Corresponding author. Email: zhizhang@ustc.edu.cn (Z.Z.); wjtao01@ahmu.edu.cn (W.T.); yuanyuan.liu@nih.gov (Y.L.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

Sound-induced analgesia depends on a low signal-to-noise ratio (SNR) relative to ambient noise in mice

Inspired by findings in humans, we investigated whether consonant sound (e.g., pleasant music for humans) may elicit analgesic effects in mice with inflammatory pain induced by hindpaw injection of complete Freund's adjuvant (CFA) (Fig. 1A and fig. S1A). Because loud noises (above ~75 dB) induce escape behavior in mice (*21*), we delivered consonant sound to CFA mice at a sound pressure level (SPL) of ~50 or ~60 dB through a closely positioned audio speaker (ambient noise at 45-dB SPL). Using von Frey filaments to assess mechanical sensitivity, we found that delivery of consonant sound at a SPL of ~50 dB, but not of ~60 dB, elevated the inflamed hindpaw nociceptive threshold compared with CFA mice exposed only to ambient noise (Fig. 1B). Notably, this effect lasted for at least 2 days after repeated exposure to ~50-dB SPL consonant sound for 3 days (20 min per day). Notably, no difference was detected in the nociceptive thresholds among CFA mice exposed to ~50-dB SPL consonant sound, dissonant sound, or white noise (Fig. 1, C and D, and fig. S1B).

Music-induced analgesia in humans may be attributable to the treatment environment (*8*). To test whether 50-dB SPL or the intensity difference relative to ambient noise (i.e., the SNR) increased the nociceptive threshold, we delivered 62-, 67-, 72-, or 77-dB SPL white noise to CFA mice under conditions of 57-dB SPL ambient noise. The mechanical nociceptive threshold was elevated only after exposure to 62-dB (5-dB SNR) SPL white noise (Fig. 1E). In addition, 35-dB (5-dB SNR) SPL white noise, but not 40-, 45-, or 50-dB SPL, produced similar effects in an environment with 30-dB SPL ambient noise (fig. S1C). We also found that 5-dB SNR sound elevated the thermal nociceptive threshold in the Hargreaves test (Fig. 1F and fig. S1D). Furthermore, the effects of 5-dB SNR sound on pain hypersensitivity were observed in mice with neuropathic pain induced by spared nerve injury (SNI) and tonic pain induced by capsaicin in both males and females (figs. S2 to S4). On the basis of these findings, we refer to 5-dB SNR as low SNR; henceforth, the ambient noise level used was 45-dB SPL, unless otherwise stated.

We next conducted conditioned place aversion (CPA) and conditioned place preference (CPP) tests to assay the potential effects of sound on the affective component of pain. Five-dB, but not 15-dB, SNR white noise abolished the subthreshold von Frey (0.04 g) stimuli-induced place aversion in the CPA test and evoked a preference for the sound-delivery side in the CPP test (Fig. 1, G to J, and fig. S5). Taken together, these findings support that low-SNR sound induces analgesia.

Pain perception is affected by emotion and stress (*22*, *23*). Behavioral tests, including open field, light-dark box, and elevated plus maze, showed that neither 5-dB nor 15-dB SNR white noise evoked anxiety-like behaviors in mice under acute pain conditions (CFA 3 days or SNI 7 days) (fig. S6, A to F), and neither sound reduced anxiety under chronic pain conditions (CFA 21 days or SNI 42 days) (fig. S6, G to L). Moreover, no stress effects were involved in the sound-induced analgesia: (i) The serum corticosterone level was unaffected by 5-dB SNR sound exposure (fig. S7), and (ii) intrathecal application of the μ opioid receptor antagonist naloxone had no effect on 5-dB SNR sound–induced increase in nociceptive thresholds (fig. S8).

Glutamatergic neurons in the auditory cortex (ACx) are functionally connected to the thalamic posterior (PO) and ventral posterior (VP) nuclei

We next investigated the neural circuit(s) underlying the observed low-SNR sound–induced analgesia. In the brain, sound signals are gated by the medial geniculate body (MGB) before reaching the ACx, which serves as the convergence point for detailed analysis of sound (*24*, *25*). Because both of these nuclei are primarily composed of glutamatergic neurons (*26*), we examined the presynaptic outputs from the

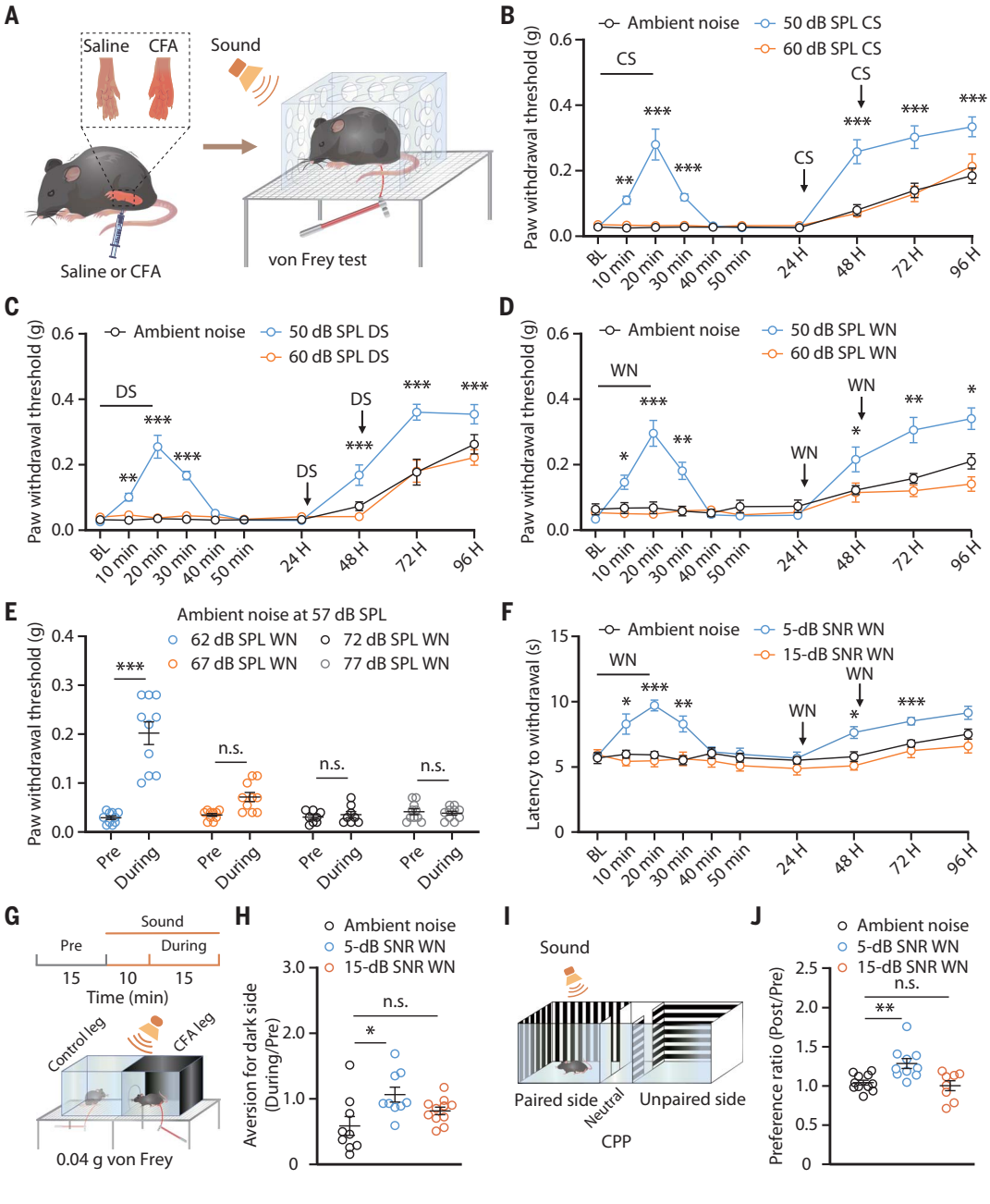


Fig. 1. Low-intensity sound relative to ambient noise induces analgesia in mice. (**A**) Schematic for inducing inflammatory pain and the von Frey test to assess the mechanical nociceptive threshold. Sound refers to delivered consonant sound (CS), dissonant sound (DS), or white noise (WN) given at the indicated SPLs. (**B** to **D**) The mechanical nociceptive threshold in CFA mice treated with or without consonant sound (ambient noise, n = 10 mice; 50-dB SPL, n = 10 mice; 60-dB SPL, n = 8 mice; $P < 0.0001$) (B), dissonant sound (n = 10 mice each group; $P < 0.0001$) (C), and white noise (ambient noise, n = 10 mice; 50-dB SPL, n = 10 mice; 60-dB SPL, n = 8 mice; $P < 0.0001$) (D) in an environment with an ambient noise at 45-dB SPL. BL, baseline. (**E**) The mechanical nociceptive threshold of CFA mice exposed to white noise at different intensities in an environment with ambient noise at 57-dB SPL (62-dB SPL, n = 10 mice; 67-dB SPL, n = 10 mice; 72-dB SPL, n = 8 mice; 77-dB SPL, n = 10 mice; $P < 0.0001$). (**F**) The thermal nociceptive threshold assessed by the Hargreaves test in CFA mice exposed to different SNR white noise (n = 10 mice each group; $P < 0.0001$). (**G**) Schematic for the CPA test. (**H**) Summarized data for the von Frey filament stimulus-induced place aversion in CFA mice treated with or without white noise (ambient noise, n = 9 mice; 5-dB SNR, n = 9 mice; 15-dB SNR, n = 11 mice; P = 0.0165). (**I**) Schematic for the CPP test. (**J**) Summarized data for preference ratio for the sound-delivery side in CFA mice from the indicated group (ambient noise, n = 11 mice; 5-dB SNR, n = 10 mice; 15-dB SNR, n = 8 mice; P = 0.0015). The data are expressed as the means ± SEMs. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; n.s., not significant. Details of the statistical analyses are presented in table S1.

ACx^{Glu} and MGB^{Glu} neurons into the brain regions tightly linked to pain processing (*27*–*29*). Specifically, we injected an adeno-associated viral vector (AAV)–DIO–membrane-bound green fluorescent protein (mGFP)–Synaptophysin-mRuby into these nuclei in *Ca^{2+}/calmodulin-dependent protein kinase II* (*CaMKII*)–*Cre* mice (fig. S9, A and B, and fig. S10, A and B).

$mRuby^{+}$ terminals originating from ACx^{Glu} neurons were observed in the insular cortex (ICx) but were very scarce in the medial prefrontal cortex, the anterior and posterior cingulate cortices, the primary and secondary somatosensory cortices, the nucleus accumbens, the ventrolateral periaqueductal gray, the dorsal raphe nucleus, the rostroventral medulla, the parabrachial nucleus, and the basolateral and central amygdala, and no signals were detected in the spinal cord (fig. S9). We detected abundant mRuby signals in thalamic nuclei, including the PO and VP, but not in the mediodorsal and central medial nuclei (fig. S9, M to P and R). Notably, we detected no mRuby signals originating from MGB^{Glu} neurons in any of these nuclei (fig. S10). Because the PO and VP receive prominent inputs from the ACx and are known to relay nociceptive information (*30*–*32*), we subsequently focused on the role of ACx→PO and ACx→VP circuits in sound-induced analgesia.

In vivo multitetrode recordings in freely moving CFA mice showed that the spontaneous neuronal activity in the ACx was decreased in response to white noise at 5-dB, but not 15-dB, SNR (Fig. 2, A to C). Five-dB SNR white noise–induced neuronal inhibition was mimicked on

the basis of (i) ACx infusion of *Cre*-dependent chemogenetic inhibitory hM4Di-mCherry viruses (AAV-DIO-hM4Di-mCherry) and (ii) intraperitoneal injection with the hM4Di-agent clozapine-*N*-oxide (CNO) in *CaMKII-Cre* mice. The nociceptive thresholds were elevated upon selective inactivation of ACx^{Glu} neurons; conditioned aversion was abolished upon such inhibition (Fig. 2, D and E). Conversely, bilateral optical activation of ACx^{Glu} neurons expressing AAV-DIO-ChR2-mCherry abolished the 5-dB SNR white noise–induced analgesia in *CaMKII-Cre* mice (fig. S11).

We then characterized both ACx→PO and ACx→VP circuits in greater detail. We performed anterograde monosynaptic tracing by ACx injection with AAV1-Cre–enhanced green fluorescent protein (EGFP) virus along with ipsilateral PO and VP injection of Cre-dependent DIO-EGFP, which revealed numerous $EGFP^+$ neurons in the PO and VP (Fig. 2F). These PO and VP $EGFP^+$ neurons colocalized with glutamatergic neurons but not with γ-aminobutyric acid–releasing (GABAergic) neurons (Fig. 2, G and H, and fig. S12, A to C).

Retrograde tracing by PO injection with a retroAAV expressing tdTomato (retroAAV-tdTomato) and VP injection with a retroAAV expressing EGFP (retroAAV-EGFP) showed that retrogradely labeled $tdTomato^+$ neurons were abundant in ACx layer 6 (L6), whereas $EGFP^+$ neurons were abundant in ACx L5 (Fig. 2, I and J). Both the PO- and VP-projecting ACx neurons colocalized with glutamatergic rather than GABAergic neurons (Fig. 2, K and L, and fig. S12, D to F). Whole-cell recordings combined with optogenetics in brain slices showed that brief light stimulation of ChR2-containing ACx^{Glu} terminals in the PO or VP reliably elicited glutamate-mediated excitatory postsynaptic currents (EPSCs) (Fig. 2, M to P).

Inhibition of the ACx^{Glu}→PO circuit mediates sound-induced hindpaw analgesia

We conducted in vivo multitetrode and fiber photometry recordings in freely moving CFA mice. PO rather than VP neurons were rapidly activated by punctate mechanical stimulation of inflamed hindpaws (Fig. 3, A to C; fig. S13, A to D; and movies S1 and S2). PO neuronal activity was significantly enhanced in CFA mice compared with saline-treated controls (fig. S13, E and F), which was attenuated by 5-dB, but not 15-dB, SNR white noise (Fig. 3, D and E, and fig. S13G). Similar sound-induced inhibitory effects on PO neurons were observed after optical inhibition of the ACx^{Glu} terminals in the PO of *CaMKII-Cre* mice treated with ACx infusion of a Cre-dependent AAV carrying eNpHR3.0–enhanced yellow fluorescent protein (EYFP) (AAV-DIO–eNpHR3.0–EYFP) (Fig. 3F and fig. S13, H and I).

We then performed bilateral optical inhibition of the ACx^{Glu}→PO circuit and conducted von Frey, Hargraves, CPA, and CPP tests. This selective inhibition mimicked the effects from the 5-dB SNR white noise in both male and female mice (Fig. 3, G to J; fig. S13J; and fig. S14). By contrast, optical activation of the ACx^{Glu}→PO circuit abolished both neuronal inhibition in the PO and analgesia induced by 5-dB SNR sound (fig. S15).

To selectively monitor the response of PO-projecting ACx neurons to 5-dB SNR sound at single-neuron resolution, we infused a retroAAV-Cre virus into the PO and an AAV expressing Cre-dependent fluorescent Ca^{2+} indicator GCaMP6m (AAV-DIO-GCaMP6m) into the ipsilateral ACx, accompanied with the mounting of a microendoscopic gradient index (GRIN) lens at the top of the ACx (fig. S16A). The Ca^{2+} transient frequency of these ACx neurons was decreased after exposure to 5-dB, but not 15-dB, SNR white noise in freely moving CFA mice (fig. S16, B and C). The activity of PO neurons receiving ACx projections (PO^{ACx}) was measured in mice with ACx infusion of AAV1-Cre virus and ipsilateral PO infusion of AAV-DIO-GCaMP6m (Fig. 3, K and L). The Ca^{2+} transient frequency of PO^{ACx} neurons was rapidly increased by punctate mechanical stimuli (movies S3 and S4). The spontaneous Ca^{2+} transient frequency of these neurons was elevated in CFA mice compared with saline mice (fig. S16, D to F), and the increased frequency was attenuated during exposure to 5-dB, but not 15-dB, SNR white noise (Fig. 3, M and N).

The PO^{ACx} neurons were then selectively manipulated to experimentally validate their participation in sound-induced analgesia. We injected AAV1-Cre virus into the ACx while concurrently injecting a viral vector expressing either Cre-dependent hM3Dq-mCherry or hM4Di-mCherry into the PO (fig. S16G). We found bilateral chemogenetic inhibition of PO^{ACx} neurons recapitulated 5-dB SNR sound–induced analgesia (fig. S16H). By contrast, chemogenetic activation of these neurons blocked the 5-dB SNR sound–induced analgesia (fig. S16I).

Sound-induced forepaw analgesia mediated by inhibition of the ACx^{Glu}→VP circuit

The VP mediates somatosensation of the upper limbs (*33*). Given our findings of a functional ACx^{Glu}→VP connection, we further investigated whether this circuit participates in 5-dB SNR sound–induced analgesia. We observed a significant increase in nociceptive thresholds of the CFA-injected forepaw along with a reduction in place aversion upon 5-dB, but not 15-dB, SNR white noise (fig. S17, A to E). Both in vivo multitetrode and fiber photometry recordings in freely moving mice revealed that neuronal activity in the VP—but not the PO—was rapidly enhanced by punctate mechanical stimulation of inflamed forepaws (Fig. 4, A to C; fig. S17, F to H; and movies S5 and S6). Compared with saline mice, the VP neuronal activity was significantly increased in CFA mice (fig. S17, I and J), and this elevation in neuronal activity was attenuated by exposure to 5-dB SNR white noise, but not 15-dB SNR (Fig. 4, D and E, and fig. S17K).

In both male and female mice, bilateral optical inhibition of the eNpHR3.0-EYFP–containing ACx^{Glu} terminals in the VP mimicked 5-dB SNR white noise–induced VP neuronal inhibition (Fig. 4, F and G) and analgesia (Fig. 4, H to J; fig. S17, L and M; and fig. S18). Optical activation of the ACx^{Glu}→VP circuit abolished both elevation of nociceptive thresholds (fig. S19, A to C) and neuronal inhibition in the VP induced by 5-dB SNR sound (fig. S19, D to F).

Microendoscopic calcium imaging showed that the spontaneous Ca^{2+} transient frequency of VP-projecting ACx neurons was attenuated during exposure to 5-dB, but not 15-dB, SNR white noise (fig. S20, A to D). In contrast to that of PO^{ACx} neurons, the Ca^{2+} transient frequency of VP neurons receiving ACx projections (VP^{ACx}) was rapidly increased by punctate mechanical stimulation of inflamed forepaws, but not hindpaws (movies S7 and S8). The spontaneous Ca^{2+} transient frequency of these neurons was enhanced in CFA mice compared with saline mice (fig. S20, E and F), which was attenuated by exposure to 5-dB, but not 15-dB, SNR white noise (Fig. 4, K to M).

The 5-dB SNR sound–induced elevation of the forepaw nociceptive threshold was mimicked upon chemogenetic inhibition of VP^{ACx} neurons and was blocked upon activation of these neurons (fig. S20, G to I). Notably, the hindpaw nociceptive threshold was not affected by optical activation or inhibition of the ACx^{Glu}→VP circuit, and neither optical activation nor inhibition of the ACx^{Glu}→PO circuit affected the forepaw nociceptive threshold (fig. S21). Comparable results were obtained upon chemogenetic manipulations of VP^{ACx} or PO^{ACx} neurons (fig. S22).

Given that the ICx, a higher-order cortex influenced by music in humans (*9*–*11*), receives ACx projections (fig. S9C), we examined the potential role of the ACx^{Glu}→ICx circuit in sound-induced analgesia. Optical activation of ACx^{Glu} terminals in the ICx did not affect 5-dB SNR sound–induced elevation of nociceptive thresholds of inflamed paws, whereas optical inhibition of this circuit in the absence of sound had no effects on pain hypersensitivity (fig. S23). Taken together, our results demonstrate that the observed analgesia from 5-dB SNR sound is mediated by the ACx→PO circuit for hindpaws and by the ACx→VP circuit for forepaws (fig. S24).

Discussion

The neural circuit responsible for processing sound-induced analgesia has long remained elusive. Here, we describe the analgesic effects of sound at low SNR through inhibition of audio-somatosensory corticothalamic circuits.

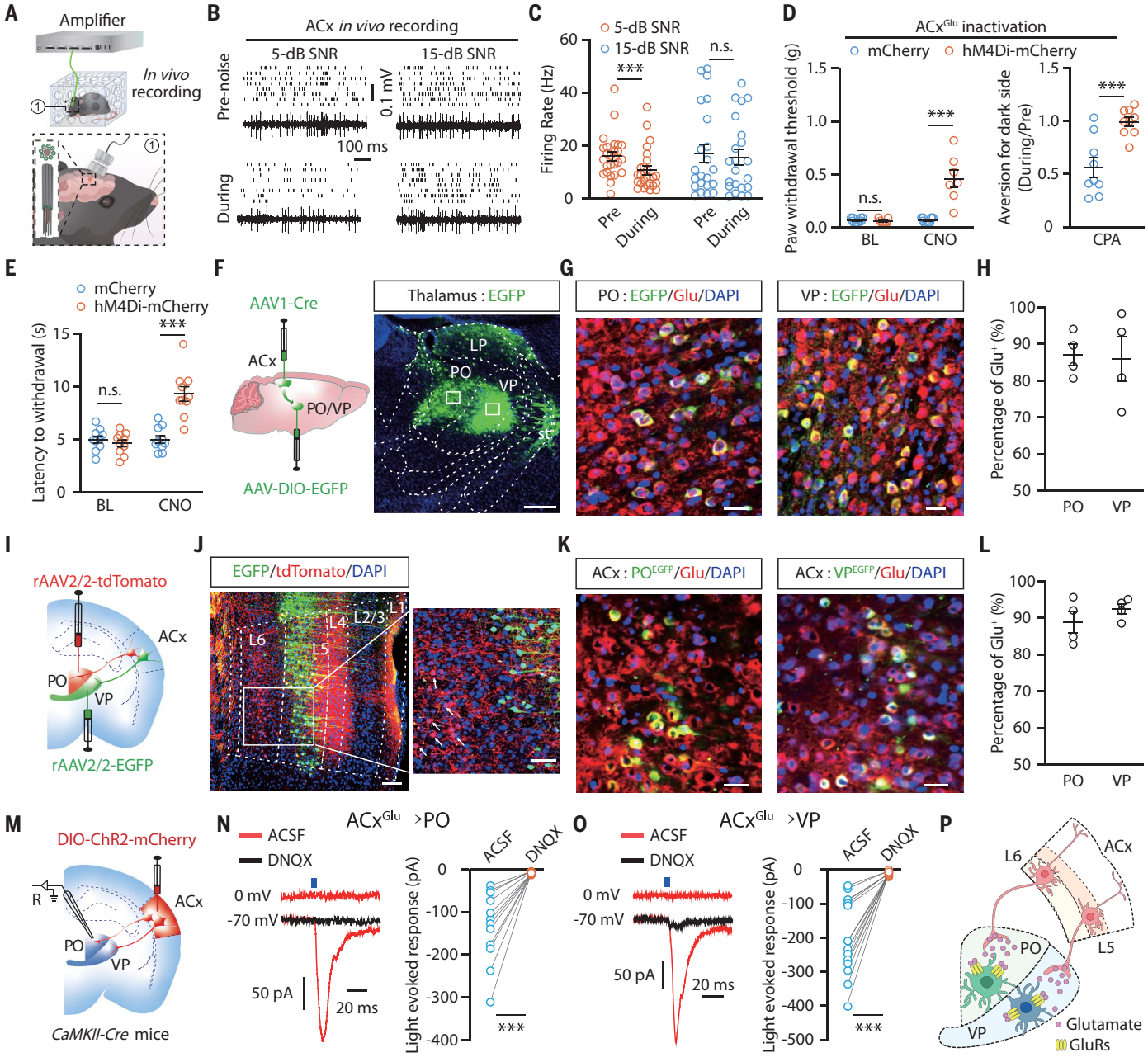


Fig. 2. ACx^{Glu} neurons project to VP^{Glu} and PO^{Glu} neurons. (**A**) Schematic for multitetrode recording in freely moving mice. (**B** and **C**) Raster plots and voltage traces of the spontaneous firings recorded in the ACx (B) and summarized data (5-dB SNR, n = 25 cells from four mice; 15-dB SNR, n = 22 cells from four mice; P = 0.0053) (C). (**D** and **E**) Summarized data for the mechanical nociceptive threshold (mCherry, n = 10 mice; hM4Di-mCherry, n = 8 mice; BL, P = 0.3816; CNO, P < 0.0001) [(D), left], place aversion (n = 9 mice each group; P = 0.0006) [(D), right], and thermal nociceptive threshold (n = 10 mice each group; P < 0.0001) (E) in CFA mice upon chemogenetic inhibition of ACx^{Glu} neurons. (**F**) Schematic for anterograde tracing and representative image of EGFP-expressing neurons in the PO and VP. Scale bar, 500 μm. LP, lateral posterior thalamic nucleus; st, stria terminalis. (**G** and **H**) Representative images showing the colocalization of EGFP-labeled neurons with glutamate (Glu) immunofluorescence (G) and summarized data (n = 4 slices) (H). Scale bars, 50 μm. DAPI, 4′,6-diamidino-2-phenylindole. (**I**) Schematic for retrograde tracing. (**J**) Representative images showing EGFP$^+$ and tdTomato$^+$ neurons in the ACx. Scale bars, 100 μm. (**K** and **L**) Representative images of the colocalization of EGFP-labeled PO- and VP-projecting ACx neurons with glutamate immunofluorescence (K) and summarized data (n = 4 slices) (L). Scale bars, 50 μm. (**M**) Schematic for viral injection and whole-cell recordings. R, recording. (**N** and **O**) Representative traces and summarized data for light-evoked postsynaptic currents recorded in PO neurons (n = 12 cells from four mice; P = 0.0002) (N) and VP neurons (n = 14 cells from four mice; P < 0.0001) (O). ACSF, artificial cerebrospinal fluid; DNQX, 6,7-dinitroquinoxaline-2,3-dione. (**P**) A model of ACx^{Glu}→PO and ACx^{Glu}→VP circuits. GluRs, glutamate receptors. The data are expressed as the means ± SEMs. ***P < 0.001; n.s., not significant. Details of the statistical analyses are presented in table S1.

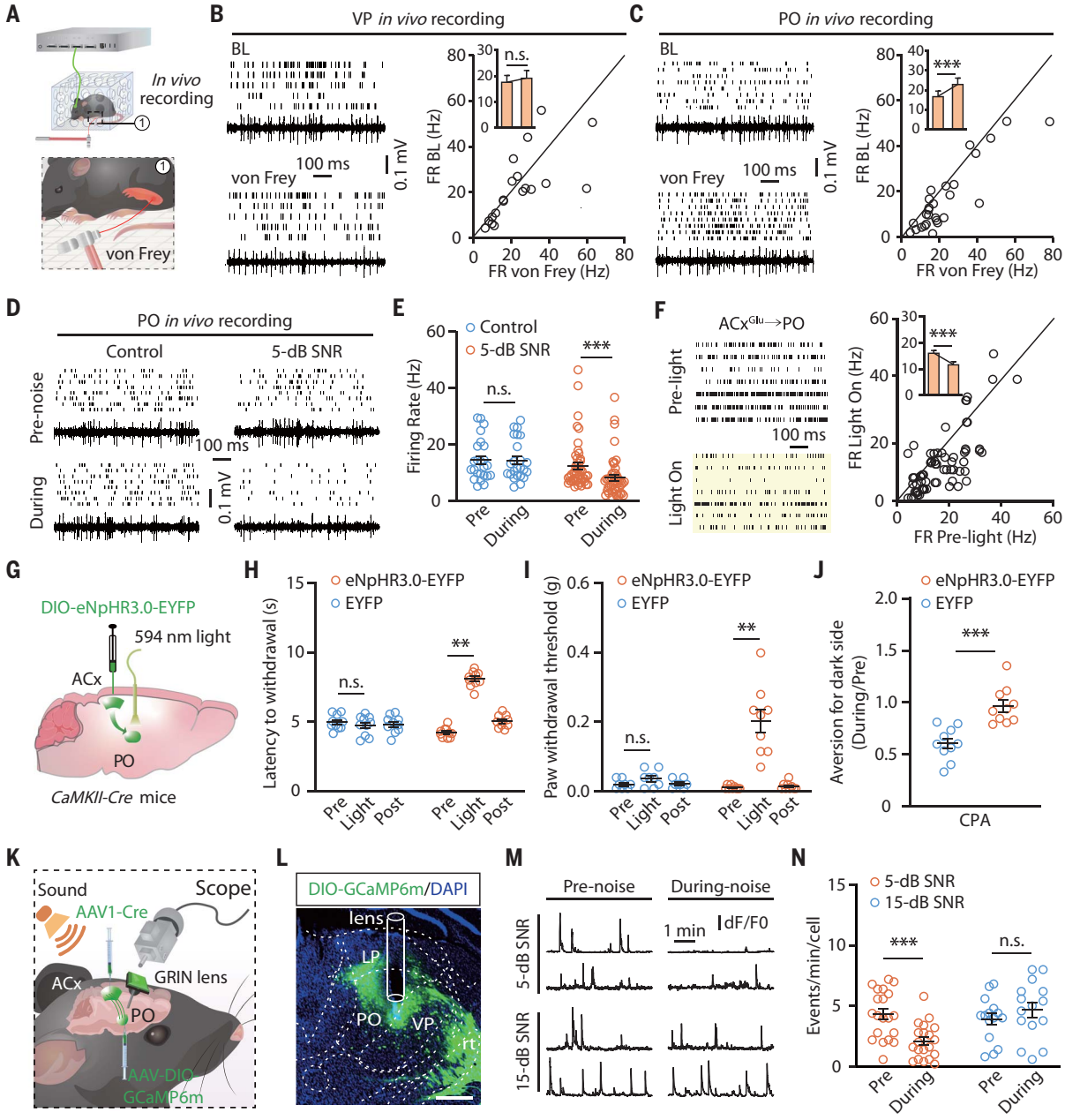


Fig. 3. Low-SNR sound inhibits the ACx^{Glu}→PO circuit to induce analgesia on hindpaws. (**A**) Schematic for multitetrode recording in freely moving mice with punctate mechanical stimulation (von Frey filament, 0.04 g). (**B** and **C**) Raster plots, voltage traces, and summarized data for the spontaneous firings recorded in VP neurons (n = 19 cells from four mice; P = 0.5079) (B) and in PO neurons (n = 27 cells from five mice; P = 0.0003) (C) before and during punctate mechanical stimulation of CFA-injected hindpaws. FR BL, firing rate baseline. (**D** and **E**) Raster plots and voltage traces of the spontaneous firings recorded in PO neurons in CFA mice with or without 5-dB SNR white noise exposure (D) and summarized data (control, n = 24 cells from four mice; 5-dB SNR, n = 47 cells from eight mice; P < 0.0001) (E). (**F**) Raster plots of the spontaneous firings recorded in PO neurons before and during optical inhibition of the ACx^{Glu}→PO circuit (left) and summarized data (n = 71 cells from seven mice; P < 0.0001) (right). (**G**) Schematic for optogenetic inhibition of the ACx^{Glu}→PO circuit. (**H** to **J**) Summarized data for the thermal nociceptive threshold (EYFP, n = 10 mice; eNpHR3.0-EYFP, n = 9 mice; P < 0.0001) (H), mechanical nociceptive threshold (EYFP, n = 8 mice; eNpHR3.0-EYFP, n = 9 mice; P < 0.0001) (I), and place aversion (n = 10 mice each group; P = 0.0001) (J) after optical inhibition of the ACx^{Glu}→PO circuit in CFA mice. (**K**) Schematic for vial injection and microendoscopic calcium imaging. (**L**) A typical image showing the GCaMP6m fluorescence and track of lens in the PO. Scale bar, 200 μm. rt, reticular thalamic nucleus. (**M** and **N**) Representative traces of spontaneous Ca^{2+} signal transient recorded in PO neurons receiving ACx projections (M) and summarized data (5-dB SNR, n = 20 cells from four mice; 15-dB SNR, n = 15 cells from four mice; P < 0.0001) (N). dF/F0, the change in fluorescence (dF) over the baseline fluorescence (F0) of calcium spikes. The data are expressed as the means ± SEMs. **P < 0.01; ***P < 0.001; n.s., not significant. Details of the statistical analyses are presented in table S1.

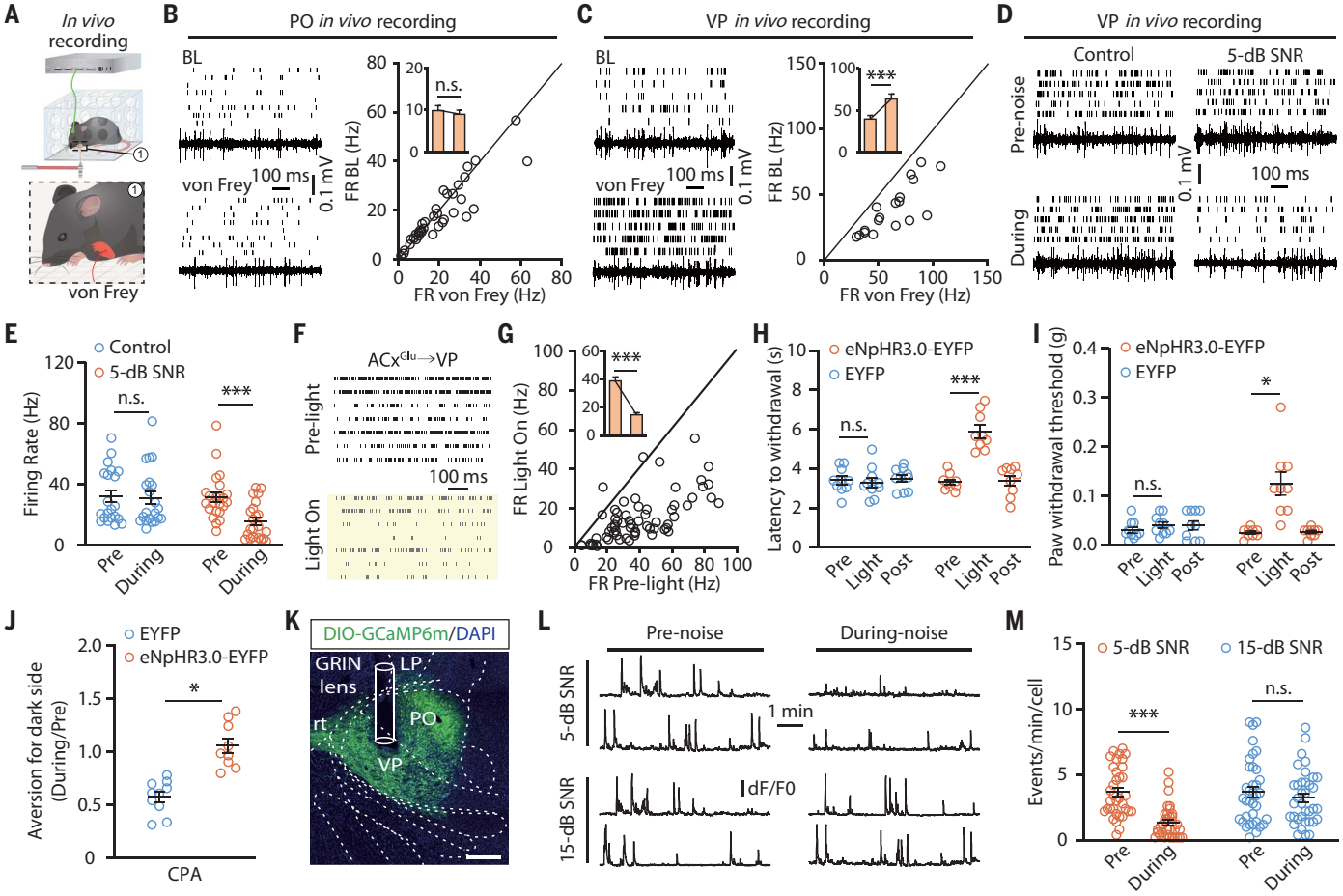


Fig. 4. Inhibition of the ACx^{Glu}→VP circuit mediates low-SNR sound–induced analgesia on forepaws. (**A**) Schematic for multitetrode recording in the VP or PO of freely moving mice. (**B** and **C**) Raster plots, voltage traces, and summarized data for the spontaneous firings recorded in PO neurons (n = 36 cells from four mice; t_{35} = 1.749; P = 0.089) (B) and VP neurons (n = 18 cells from four mice; t_{17} = 7.373; P < 0.0001) (C) before and during punctate mechanical stimulation (von Frey filament, 0.02 g) of inflamed forepaws. (**D** and **E**) Raster plots and voltage traces of the spontaneous firings recorded in VP neurons from CFA mice with or without 5-dB SNR white noise exposure (D) and summarized data (control, n = 21 cells from four mice; 5-dB SNR, n = 23 cells from four mice; $F_{1,42}$ = 24.18; P < 0.0001) (E). (**F** and **G**) Raster plots of the spontaneous activity recorded in VP neurons before and during optical inhibition of the ACx^{Glu}→VP circuit in CFA-treated mice (F) and summarized data (n = 67 cells from seven mice; t_{66} = 12.14; P < 0.0001) (G). (**H** to **J**) Summarized data for the thermal (EYFP, n = 10 mice; eNpHR3.0-EYFP, n = 9 mice; $F_{2,34}$ = 20.98; P < 0.0001) (H) and mechanical (EYFP, n = 10 mice; eNpHR3.0-EYFP, n = 9 mice; $F_{2,34}$ = 13.25; P < 0.0001) (I) nociceptive thresholds of CFA-injected forepaws and place aversion (EYFP, n = 10 mice; eNpHR3.0-EYFP, n = 9 mice; t_{17} = 5.648; P < 0.0001) (J) upon optical inhibition of the ACx^{Glu}→VP circuit. (**K**) A typical image of GCaMP6m fluorescence and track of the lens in the VP. Scale bar, 200 μm. (**L** and **M**) Representative traces (L) of spontaneous Ca^{2+} signals recorded in VP neurons receiving ACx projections and summarized data (5-dB SNR, n = 35 cells from four mice; 15-dB SNR, n = 36 cells from four mice; $F_{1,69}$ = 24.24; P < 0.0001) (M). The data are expressed as the means ± SEMs. *P < 0.05; ***P < 0.001; n.s., not significant. Details of the statistical analyses are presented in table S1.

Specifically, we reveal that the distinct roles of the ACx^{Glu}→PO and ACx^{Glu}→VP circuits in sound-induced analgesia depend on the physical location of the pain.

In mice, we found that sound-induced analgesia depended on its low SNR rather than harmony, which is supported by a previous hypothesis that music-induced analgesia is attributable to contextual factors of the treatment, not only to the music per se (*8*). It bears mention that the observed low-SNR sound–induced analgesia is unlikely to result from some reduction in anxiety or stress, and it probably does not directly involve attention-distraction in affecting pain perception (*33*), given that the analgesic effects persisted for at least 2 days after sound withdrawal.

The neural mechanisms underlying music-induced analgesia in humans are doubtlessly more complicated than those revealed in mice (*34*). In humans, multiple areas that are involved in pain processing, including the ICx, striatum, and midbrain dopamine system, do respond to music (*35*–*38*). Therefore, whether other brain regions receiving ACx projections (except the PO and VP) or how downstream projections of the PO and VP account for music-induced analgesia deserve further investigation.

This study suggests a complement for the classic pain sensation pathway that is implicated in the effect of sound on pain processing and could expedite the study of music-induced analgesia. In the future, these findings could spur the development of alternative interventions for treating pain.

REFERENCES AND NOTES

1. W. J. Gardner, J. C. Licklider, A. Z. Weisz, *Science* **132**, 32–33 (1960).
2. A. Keenan, J. K. Keithley, *Oncol. Nurs. Forum* **42**, E368–E375 (2015).
3. T. N. Nguyen, S. Nilsson, A.-L. Hellström, A. Bengtson, *J. Pediatr. Oncol. Nurs.* **27**, 146–155 (2010).
4. L. Hartling *et al.*, *JAMA Pediatr.* **167**, 826–835 (2013).
5. C. Boyd-Brewer, R. McCaffrey, *Holist. Nurs. Pract.* **18**, 111–118 (2004).

6. M. P. Jensen, S. Hakimian, L. H. Sherlin, F. Fregni, *J. Pain* **9**, 193–199 (2008).
7. E. A. Garza Villarreal, E. Brattico, L. Vase, L. Østergaard, P. Vuust, *PLOS ONE* **7**, e29397 (2012).
8. S. J. Lunde, P. Vuust, E. A. Garza-Villarreal, L. Vase, *Pain* **160**, 989–993 (2019).
9. C. E. Dobek, M. E. Beynon, R. L. Bosma, P. W. Stroman, *J. Pain* **15**, 1057–1068 (2014).
10. M. E. Sachs, A. Habibi, A. Damasio, J. T. Kaplan, *Neuroimage* **218**, 116512 (2020).
11. C. Usui *et al*., *Pain Med.* **21**, 1546–1552 (2020).
12. L. Bonetti *et al*., *Neuroimage* **245**, 118735 (2021).
13. E. G. Jones, *The Thalamus* (Cambridge Univ. Press, ed. 2, 2007).
14. S. M. Sherman, R. W. Guillery, *Phil. Trans. R. Soc. Lond. B* **357**, 1695–1708 (2002).
15. A. Mouraux, A. Diukova, M. C. Lee, R. G. Wise, G. D. Iannetti, *Neuroimage* **54**, 2237–2249 (2011).
16. B. A. Vogt, D. L. Rosene, D. N. Pandya, *Science* **204**, 205–207 (1979).
17. A. N. Viaene, I. Petrof, S. M. Sherman, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18156–18161 (2011).
18. K. S. Meda *et al*., *Neuron* **102**, 944–959.e3 (2019).
19. X. Zhu *et al*., *Nat. Neurosci.* **24**, 542–553 (2021).
20. G. Wagner, M. Koschke, T. Leuf, R. Schlösser, K.-J. Bär, *Neuropsychologia* **47**, 980–987 (2009).
21. X. R. Xiong *et al*., *Nat. Commun.* **6**, 7224 (2015).
22. G. Gilam, J. J. Gross, T. D. Wager, F. J. Keefe, S. C. Mackey, *Neuron* **107**, 17–21 (2020).
23. J. W. Grau, R. L. Hyson, S. F. Maier, J. Madden4th, J. D. Barchas, *Science* **213**, 1409–1411 (1981).
24. S. Koelsch, *Nat. Rev. Neurosci.* **15**, 170–180 (2014).
25. H. L. Read, J. A. Winer, C. E. Schreiner, *Curr. Opin. Neurobiol.* **12**, 433–440 (2002).
26. J. A. Winer, in *The Mammalian Auditory Pathway: Neuroanatomy*, D. B. Webster, A. N. Popper, R. R. Fay, Eds. (Springer, 1992), pp. 222–409.
27. T. V. Salomons, G. D. Iannetti, M. Liang, J. N. Wood, *JAMA Neurol.* **73**, 755–756 (2016).
28. A. V. Apkarian, M. C. Bushnell, R. D. Treede, J. K. Zubieta, *Eur. J. Pain* **9**, 463–484 (2005).
29. W. Zhou *et al*., *Nat. Neurosci.* **22**, 1649–1658 (2019).
30. D. D. Price, *Mol. Interv.* **2**, 392–403 (2002).
31. H. Head, G. Holmes, *Brain* **34**, 102–254 (1911).
32. M. E. Diamond, M. Armstrong-James, M. J. Budway, F. F. Ebner, *J. Comp. Neurol.* **319**, 66–84 (1992).
33. K. Wiech, M. Ploner, I. Tracey, *Trends Cogn. Sci.* **12**, 306–313 (2008).
34. S. Leknes, I. Tracey, *Nat. Rev. Neurosci.* **9**, 314–320 (2008).
35. W. Ren *et al*., *Nat. Neurosci.* **19**, 220–222 (2016).
36. M. L. Smith, N. Asada, R. C. Malenka, *Science* **371**, 153–159 (2021).
37. V. N. Salimpoor, M. Benovoy, K. Larcher, A. Dagher, R. J. Zatorre, *Nat. Neurosci.* **14**, 257–262 (2011).
38. M. N. Baliki *et al*., *Nat. Neurosci.* **15**, 1117–1119 (2012).

ACKNOWLEDGMENTS

We thank Y. Yang for providing Ai9 (RCL-tdT) mice. We thank F. Xu and T. Xue for technical support, C. Weisz, and K. Suthakar for critical reading. **Funding:** This work was supported by the National Key Research and Development Program of China Brain Science and Brain-Like Intelligence Technology (2021ZD0203100), the National Natural Science Foundation of China (32025017, 32000717, and 82101301), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (32121002), the CAS Project for Young Scientists in Basic Research (YSBR-013), the Natural Science Foundation of Anhui Province (KJ2020A0138, 2008085QC115, and 2008085QC114), the University of Science and Technology of China Research Funds of the Double First-Class Initiative (YD3460002001), the NIH NIDCR intramural research program (ZIA DE000757), and the Innovative Research Team of High-level Local Universities in Shanghai. **Author contributions:** W.Zho., C.Y., Y.L., W.T., and Z.Z. initiated and designed the research. W.Zho., H.W., Y.L., W.T., and Z.Z. wrote the manuscript. W.Zho., C.Y., H.W., Y.M., W.Zhan., C.-L.Y., T.L., L.H., A.L., and W.Zhao performed all experiments and analyzed and interpreted the results. L.C. contributed to the discussion of the results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn4663
Materials and Methods
Figs. S1 to S24
Tables S1 and S2
References (*39*, *40*)
MDAR Reproducibility Checklist
Movies S1 to S8

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

Submitted 30 November 2021; accepted 18 May 2022
10.1126/science.abn4663

REVIEW SUMMARY

PHYSICS

Stranger than metals

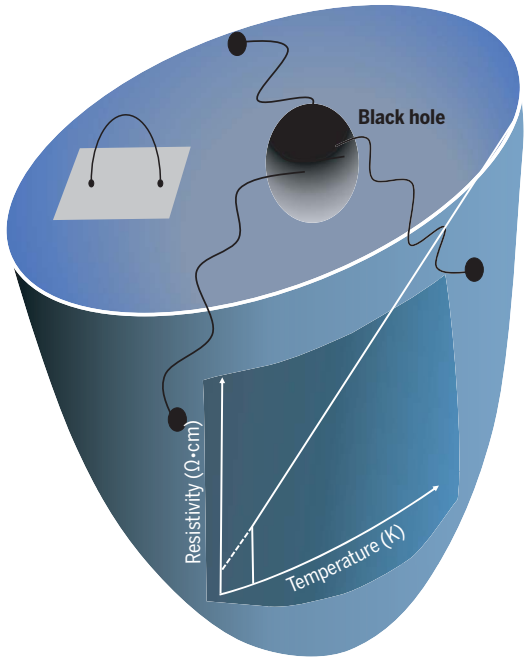
Philip W. Phillips*, Nigel E. Hussey, Peter Abbamonte

BACKGROUND: Landau's Fermi liquid theory provides the bedrock on which our understanding of metals has developed over the past 65 years. Its basic premise is that the electrons transporting a current can be treated as "quasiparticles"—electron-like particles whose effective mass has been modified, typically through interactions with the atomic lattice and/or other electrons. For a long time, it seemed as though Landau's theory could account for all the many-body interactions that exist inside a metal, even in the so-called heavy fermion systems whose quasiparticle mass can be up to three orders of magnitude heavier than the electron's mass. Fermi liquid theory also lay the foundation for the first successful microscopic theory of superconductivity.

In the past few decades, a number of new metallic systems have been discovered that violate this paradigm. The violation is most evident in the way that the electrical resistivity changes with temperature or magnetic field. In normal metals in which electrons are the charge carriers, the resistivity increases with increasing temperature but saturates, both at low temperatures (because the quantized lattice vibrations are frozen out) and at high temperatures (because the electron mean free path dips below the smallest scattering pathway defined by the lattice spacing). In "strange metals," by contrast, no saturation occurs, implying that the quasiparticle description breaks down and electrons are no longer the primary charge carriers. When the particle picture breaks down, no local entity carries the current.

ADVANCES: A new classification of metallicity is not a purely academic exercise, however, as strange metals tend to be the high-temperature phase of some of the best superconductors available. Understanding high-temperature superconductivity stands as a grand challenge because its resolution is fundamentally rooted in the physics of strong interactions, a regime where electrons no longer move independently. Precisely what new emergent phenomena one obtains from the interactions that drive the electron dynamics above the temperature where they superconduct is one of the most urgent problems in physics, attracting the attention of condensed matter physicists as well as string theorists. One thing is clear in this regime: The particle picture breaks down. As particles and locality are typically related, the strange metal raises the distinct possibility that its resolution must abandon the basic building blocks of quantum theory.

We review the experimental and theoretical studies that have shaped our current understanding of the emergent strongly interacting physics realized in a host of strange metals, with a special focus on their poster-child: the copper oxide high-temperature superconductors. Experiments are highlighted that attempt to link the phenomenon of nonsaturating resistivity to parameter-free universal physics. A key experimental observation in such materials is that removing a single electron affects the spectrum at all energy scales, not just the low-energy sector as in a Fermi liquid. It is observations of this sort that reinforce the breakdown of the single-particle concept. On the theoretical side, the modern accounts that borrow from the conjecture that strongly interacting physics is really about gravity are discussed extensively, as they have been the most successful thus far in describing the range of physics displayed by strange metals. The foray into gravity models is not just a pipe dream because in such constructions, no particle interpretation is given to the charge density. As the breakdown of the independent-particle picture is central to the strange metal, the gravity constructions are a natural tool to make progress on this problem. Possible experimental tests of this conjecture are also outlined.



Curved spacetime with a black hole in its interior and the strange metal arising on the boundary. This picture is based on the string theory gauge-gravity duality conjecture by J. Maldacena, which states that some strongly interacting quantum mechanical systems can be studied by replacing them with classical gravity in a spacetime in one higher dimension. The conjecture was made possible by thinking about some of the fundamental components of string theory, namely D-branes (the horseshoe-shaped object terminating on a flat surface in the interior of the spacetime). A key surprise of this conjecture is that aspects of condensed matter systems in which the electrons interact strongly—such as strange metals—can be studied using gravity.

OUTLOOK: As more strange metals emerge and their physical properties come under the scrutiny of the vast array of experimental probes now at our disposal, their mysteries will be revealed and their commonalities and differences cataloged. In so doing, we should be able to understand the universality of strange metal physics. At the same time, the anomalous nature of their superconducting state will become apparent, offering us hope that a new paradigm of pairing of nonquasiparticles will also be formalized. The correlation between the strength of the linear-in-temperature resistivity in cuprate strange metals and their corresponding superfluid density, as revealed here, certainly hints at a fundamental link between the nature of strange metallicity and superconductivity in the cuprates. And as the gravity-inspired theories mature and overcome the challenge of projecting their powerful mathematical machinery onto the appropriate crystallographic lattice, so too will we hope to build with confidence a complete theory of strange metals as they emerge from the horizon of a black hole. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dimer@illinois.edu
Cite this article as P. W. Phillips *et al.*, *Science* **377**, eabh4273 (2022). DOI: 10.1126/science.abh4273

REVIEW

PHYSICS

Stranger than metals

Philip W. Phillips[1]*, Nigel E. Hussey[2,3], Peter Abbamonte[4]

In traditional metals, the temperature (*T*) dependence of electrical resistivity vanishes at low or high *T*, albeit for different reasons. Here, we review a class of materials, known as "strange" metals, that can violate both of these principles. In strange metals, the change in slope of the resistivity as the mean free path drops below the lattice constant, or as $T \to 0$, can be imperceptible, suggesting continuity between the charge carriers at low and high *T*. We focus on transport and spectroscopic data on candidate strange metals in an effort to isolate and identify a unifying physical principle. Special attention is paid to quantum criticality, Planckian dissipation, Mottness, and whether a new gauge principle is needed to account for the nonlocal transport seen in these materials.

To understand the essential tension between quantum mechanics and gravity, simply imagine two electrons impinging on the event horizon of a black hole. Whereas classical gravity predicts that they meet at the center, quantum mechanics forbids this if the electrons have the same spin. In essence, classical gravity has no way of preserving Pauli exclusion. Replacing classical general relativity with a quantum theory of gravity at small enough scales resolves the problem, but what is this scale?

In 1899, Planck formulated a universal length now regarded as the scale below which a quantum theory of gravity supplants its classical counterpart. The Planck scale,

$$\ell_P = \sqrt{\frac{\hbar G}{c^3}} \tag{1}$$

obtains by pure dimensional analysis on three fundamental constants: the speed of light, c, Newton's gravitational constant, G, and the quantum of uncertainty, $\hbar$ (Planck's constant, h, divided by 2π). This leads naturally to a Planck time as the ratio of the Planck length to the speed of light, ℓ_P/c. Such a Planckian analysis can be extended to many-body systems in contact with a heat bath. All that is necessary is to include the temperature T. A similar dimensional analysis then leads to

$$\tau_P = \frac{\hbar}{k_B T} \tag{2}$$

as the shortest time for heat loss in a many-body system obeying quantum mechanics, where k_B is Boltzmann's constant. Because no system parameters enter τ_P, this quantity occupies a similar fundamental role in analogy to the Planck length and is referred to as the Planckian dissipation time. Equation 2 has had previous incarnations (*1*, *2*); in the realm of charge transport, relevant to this article, it defines the time scale for scale-invariant or Planckian dissipation (*3*). Scale invariance follows because there is no scale other than temperature appearing in τ_P. Achieving such scale invariance necessitates a highly entangled many-body state. Such a state would lead to the breakdown of a local single-particle framework and the advent of new collective nonlocal entities as the charge carriers. Identifying the new propagating degrees of freedom constitutes the key mystery of strange metals.

Whereas the Planck scale ℓ_P requires high-energy accelerators much beyond anything now in use, such is not the case with physics at the Planckian dissipation limit. Early tabletop experiments on cuprate superconductors, for example, revealed a "strange metal" regime defined by a robust *T*-linear resistivity extending to the highest temperatures measured (*4*–*6*) (Fig. 1), a possible harbinger of Planckian dissipation. Recall that in a Fermi liquid, the conductivity can be well described by a Drude formula,

$$\sigma = \frac{n_e e^2}{m}\tau_{tr} \tag{3}$$

where n_e is the charge carrier density, e is the charge of an electron, m is its mass, and τ_{tr} is the transport lifetime, defined as

$$\tau_{tr} = \frac{\hbar E_F}{(k_B T)^2} = \frac{E_F}{k_B T}\tau_P \tag{4}$$

which contains the Fermi energy E_F of the quasiparticles. No such energy scale appears in Eq. 2. If the scattering rate in cuprates is directly proportional to the resistivity, as it is in simple metals, *T*-linear resistivity is equivalent to scale-invariant Planckian dissipation only if $\tau_{tr} = \alpha_1\tau_P$ with α_1 (the coefficient of *T*-linear resistivity) ~ 1. Although this state of affairs seems to be realized in a host of correlated metals, including the cuprates (*7*–*10*), there is no consensus concerning how accurately α_1 can be known and the assumptions that go into its determination. Regardless of the possible relationship with Planckian dissipation, what makes *T*-linear resistivity in the cuprates truly novel is its persistence—from millikelvin temperatures (in both the electron- and hole-doped cuprates) (*11*, *12*) up to 1000 K (in the hole-doped cuprates) (*4*, *6*)—and its omnipresence, the strange metal regime dominating large swathes of the temperature versus doping phase diagram (*13*). In normal metals (*14*, *15*) as well as some heavy fermion materials (*16*), the resistivity asymptotically approaches a saturation value at which the mean free path ℓ becomes comparable with the interatomic spacing a or more generally the Fermi wavelength λ_F—the minimum length over which a Bloch wave and its associated Fermi velocity and wave vector can be defined. In many correlated metals, collectively referred to as "bad metals," $\ell < a$ at high *T*, which violates the so-called Mott-Ioffe-Regel (MIR) limit (*5*, *6*, *1*, *16*–*18*). Remarkably, no saturation occurs in these bad metals across the MIR threshold, implying that the whole notion of a Fermi velocity of quasiparticles breaks down at high *T*. In certain cases, an example of which is shown in Fig. 1, there is no discernible change in slope as the MIR limit is exceeded. Although this circumstance occurs only in a narrow doping window (in cuprates) (*18*), such continuity does suggest that, even at low *T*, quasiparticles (*19*) cannot be the

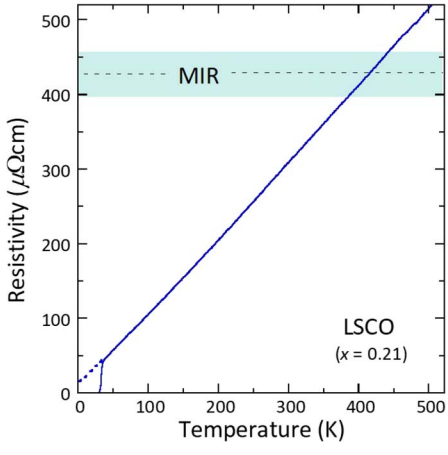


Fig. 1. *T*-linear resistivity in strange metals. Shown is the in-plane resistivity of $La_{2-x}Sr_xCuO_4$ (x = 0.21). The dotted points are extrapolated from high-field magnetoresistance data (*8*). The shaded area shows the Mott-Ioffe-Regel (MIR) boundary defined here as when the mean free path becomes comparable to the Fermi wavelength λ_F. [Adapted from (*8*, *18*)]

[1]Department of Physics and Institute for Condensed Matter Theory, University of Illinois, Urbana, IL 61801, USA. [2]H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK. [3]High Field Magnet Laboratory (HFML-EMFL) and Institute for Molecules and Materials, Radboud University, 6525 ED Nijmegen, Netherlands. [4]Department of Physics, University of Illinois, Urbana, IL 61801, USA.
*Corresponding author. Email: dimer@illinois.edu

effective propagating degrees of freedom. Evidently, in strongly correlated electron matter, the current-carrying degrees of freedom in the infrared (IR) need not have a particle interpretation.

Over time, the label "strange metal" has seemingly become ubiquitous, used to describe any metallic system whose transport properties display behavior that is irreconcilable with conventional Fermi liquid or Boltzmann transport theory. This catch-all phraseology, however, is unhelpful, as it fails to differentiate between the various types of non–Fermi liquid behavior observed, some of which deserve special deliberation on their own. We attempt to bring strange metal phenomenology into sharper focus by addressing a number of pertinent questions. Does the term refer to the resistive behavior of correlated electron systems at high or low temperatures, or both? Does it describe any T-linear resistivity associated with the Planckian time scale, or something unique? Does it describe the physics of a doped Mott insulator or the physics associated with quantum criticality (whose underlying origins may or may not include Mottness as a key ingredient)? Finally, does anything local carry the current, and if not, does explicating the propagating degrees of freedom in the strange metal require a theory as novel as quantum gravity?

Is strange metallicity ubiquitous?

A survey of the DC transport properties of several strange metal candidates is presented in Table 1 (*4*, *6*, *8*, *10*, *20*–*61*). In addressing the above question, we must first acknowledge the many definitions of strange metallic behavior that exist, the simplest being a material hosting a metallic-like resistivity in the absence of quasiparticles. A more precise, if empirical, definition centers on the T-linear resistivity, specifically one that is distinguishable from the resistivity manifested by simple metals (which is attributed to electron-phonon scattering). For a metal to be classified as strange, the T-linearity must extend far beyond the typical bounds associated with phonon-mediated resistivity. At low T, this is typically one-third of the Debye temperature, whereas at high T, it is the temperature at which the magnitude of the resistivity is roughly half the value corresponding to the MIR limit. A subset of correlated metals, such as $SrRuO_3$ (*62*) and Sr_2RuO_4 (*36*), exhibit T-linear resistivity at high T with a magnitude that clearly violates the MIR limit, but as the system cools down, conventional Fermi-liquid behavior is restored (*37*, *63*). Hence, although they are bona fide bad metals—exhibiting metallic resistivity beyond the MIR limit—they do not classify as strange (*16*, *64*).

Another subset, identified here as quantum critical metals, exhibit T-linear resistivity down to the lowest temperatures studied, but only at a singular quantum critical point (QCP) in their T versus tuning parameter g phase diagram at which a continuous quantum phase transition to a symmetry-broken phase is suppressed to $T = 0$. Here, g can be pressure, magnetic field, doping level, or even strain. In most cases, the phase transition in question is associated with finite-momentum antiferromagnetism [as in pure $YbRh_2Si_2$ (*49*), $CeCoIn_5$ (*55*), and $BaFe_2(As_{1-x}P_x)_2$ (*45*)], although similar behavior has recently been reported in systems exhibiting zero-momentum order, such as nematic $FeSe_{1-x}S_x$ (*41*) or ferromagnetic $CeRh_6Ge_4$ (*57*). Away from the QCP, the low-T resistivity recovers the canonical T^2 Fermi liquid form, albeit with a coefficient that is enhanced as the

Table 1. Summary of the DC transport properties of various strange metal candidates. First column: Candidate compound or family of compounds. For the hole-doped cuprates, underdoped (UD), optimally doped (OP) and overdoped (OD) compounds are treated separately; the transport properties of individual compounds within each subset are generic. For the electron-doped cuprates, only $La_{2-x}Ce_xCuO_4$ is listed; Pr- and Nd-based compounds show similar behavior. Second column: Bad metallic behavior; a check mark indicates T-linear resistivity beyond the Mott-Ioffe-Regel (MIR) limit. A cross indicates either a tendency toward saturation or a marked reduction in slope near the MIR limit. Third column: A check mark identifies systems that at any point in their respective phase diagram(s) exhibit T-linear resistivity down to the lowest temperatures studied. Fourth column: "Extended criticality" refers to systems where a predominant T-linear resistivity at low T extends over a finite region of the phase diagram. Fifth column: T^2 dependence of the inverse Hall angle $\cot\Theta_H$ in the same temperature range where $\rho(T)$ is T-linear. Sixth column: Compounds satisfying the "Modified Kohler's" label have a low-field magnetoresistance (MR), defined as $[\rho(H, T) - \rho(0, T)]/\rho(0, T)$, that exhibits a T dependence similar to that of $\tan^2\Theta_H$. Seventh and eighth columns: High-field MR behavior of strange metal candidates. The observation of an H-linear MR at high fields does not imply that the MR exhibits quadrature scaling over all fields and temperatures. *$FeSe_{1-x}S_x$ H-linear/quadrature MR seen in this family coexists with a more conventional MR contribution, indicating the presence of both strange metal and Fermi liquid–like components in the DC transport. **In $YbBAl_4$, although T-linear resistivity is observed over a wide pressure range, its limiting low-T dependence is $T^{1.5}$. A dash indicates no reports confirming or disproving the considered behavior.

	$\rho \propto T$ as $T \to \infty$	$\rho \propto T$ as $T \to 0$	Extended criticality	$\cot\Theta_H \propto T^2$ (at low H)	Modified Kohler's (at low H)	H-linear MR (at high H)	Quadrature MR
UD p-cuprates	√ (*6*)	× (*20*)	× (*21*)	√ (*22*)	√ (*23*)	–	–
OP p-cuprates	√ (*4*)	–	–	√ (*24*)	√ (*25*)	√ (*26*)	× (*27*)
OD p-cuprates	√ (*6*)	√ (*8*)	√ (*8*)	√ (*28*)	× (*29*)	√ (*29*)	√ (*29*)
$La_{2-x}Ce_xCuO_4$	× (*30*)	√ (*31*, *32*)	√ (*31*, *32*)	× (*33*)	× (*34*)	√ (*35*)	× (*35*)
Sr_2RuO_4	√ (*36*)	× (*37*)	× (*38*)	× (*39*)	× (*37*)	× (*37*)	× (*37*)
$Sr_3Ru_2O_7$	√ (*10*)	√ (*10*)	× (*10*)	×	–	–	–
$FeSe_{1-x}S_x$	× (*40*)	√ (*41*)	× (*41*)	√ (*42*)	√ (*42*)	√* (*43*)	√* (*43*)
$BaFe_2(As_{1-x}P_x)_2$	× (*44*)	√ (*45*)	× (*45*)	–	√ (*46*)	√ (*47*)	√ (*47*)
$Ba(Fe_{1/3}Co_{1/3}Ni_{1/3})_2As_2$	–	√ (*48*)	× (*48*)	–	–	√ (*48*)	√ (*48*)
$YbRh_2Si_2$	× (*49*)	√ (*50*)	√ (*51*)	√ (*52*)	–	–	–
$YbBAl_4$	× (*53*)	√** (*53*)	√** (*53*)	–	–	–	–
$CeCoIn_5$	× (*54*)	√ (*55*, *56*)	× (*55*, *56*)	√ (*54*)	√ (*54*)	–	–
$CeRh_6Ge_4$	× (*57*)	√ (*57*)	× (*57*)	–	–	–	–
$(TMTSF)_2PF_6$	–	√ (*58*)	√ (*58*)	–	–	–	–
MATBG	√ (*59*)	√ (*60*)	√ (*60*)	√ (*61*)	–	–	–

QCP is approached and the order parameter fluctuations soften.

By contrast, in overdoped cuprates [both hole- (*8*, *9*) and electron-doped (*31*, *32*)], Ge-doped $YbRh_2Si_2$ (*51*), $YbBAl_4$ (*53*), and the organic Bechgaard salts (*58*), $\rho(T)$ is predominantly T-linear down to low T not at a singular point in these materials' respective phase diagrams, but over an extended range of the relevant tuning parameter. At first sight, this "extended criticality" is difficult to reconcile with current theories of quantum criticality, which predict a crossover to a purely T^2 resistivity and thus a recovery of Fermi liquid behavior at low T everywhere except at the (singular) QCP. Arguably, it is this feature—incompatibility with both standard Fermi liquid and quantum critical scenarios—that distinguishes a genuine strange metal. Intriguingly, in many of these systems, α_1 is found to scale with the superconducting transition temperature T_c. Moreover, for $La_{2-x}Ce_xCuO_4$ (*31*, *32*) and bis-(tetramethyltetraselenafulvalene) hexafluorophosphate $[(TMTSF)_2PF_6]$ (*58*), extended criticality emerges once the spin density wave transition has been fully suppressed, suggesting an intimate link between the strange metal transport, superconductivity, and the presence of critical or long-wavelength spin fluctuations. In hole-doped cuprates, however, the strange metal regime looks different, in the sense that the extended criticality emerges beyond the end of the pseudogap regime that does not coincide with a magnetic quantum phase transition (*65*). Furthermore, although the pseudogap plays host to a multitude of broken-symmetry states, the jury is still out as to whether any of these are responsible for pseudogap formation or are merely instabilities of it.

Besides T-linear resistivity, strange metals also exhibit anomalous behavior in their magnetotransport, including (i) a quadratic temperature dependence of the inverse Hall angle $\cot\Theta_H = \sigma_{xy}/\sigma_{xx}$ (*24*), (ii) a transverse magnetoresistance (MR) that at low field exhibits modified Kohler's scaling [$\Delta\rho/\rho(0) \propto \tan^2\Theta_H \propto (1/T^2)^2$ or $1/(A + BT^2)^2$ (*25*)], and/or (iii) an H-linear MR at high fields that may or may not follow quadrature scaling [whereby $\Delta\rho/T \propto \sqrt{1+\gamma(H/T)^2}$] (*43*, *47*). The combination of a modified Kohler's rule and T^2 Hall angle has been interpreted to indicate the presence of distinct relaxation times, either for different loci in momentum space (*22*) or for relaxation processes normal and tangential to the underlying Fermi surface (*24*). The H-linear MR, on the other hand, is inextricably tied to the T-linear zero-field resistivity via its H/T scaling relation, a relation that can also extend over a broad range of the relevant tuning parameter (*29*). In some cases, this link can be obscured, either because $\rho(T)$ itself is not strictly T-linear (*29*) or because the quadrature-scaling MR coexists with a more conventional orbital MR (*43*). Both sets of behavior highlight once again the possible coexistence of two relaxation times or two distinct charge-carrying sectors in real materials. Curiously, in single-band materials, quadrature scaling breaks down inside the pseudogap regime (*26*, *27*), whereas modified Kohler's scaling is recovered (*23*, *25*), suggesting that the two phenomena may be mutually exclusive in such materials. In multiband materials such as $FeSe_{1-x}S_x$, on the other hand, these different manifestations of strange metallic transport appear side by side (*42*, *43*). Irrespective of these caveats and complexities, what is striking about the quadrature-scaling MR is that it occurs in systems with varied Fermi surface topologies, dominant interactions, and energy scales, hinting at some universal but as yet unidentified organizing principle.

Restricting the strange metal moniker, as done here, to materials that exhibit low-T T-linear resistivity over an extended region of phase space likewise restricts strange metallicity to a select "club." The following sections explore various possible attributes that they have in common.

Is it quantum criticality?

Scale-free T-linear resistivity is highly suggestive of some form of underlying quantum criticality in which the only relevant scale is the temperature governing collisions between excitations of the order parameter (*66*). In fact, following the advent of marginal Fermi liquid (MFL) phenomenology and its associated (T, ω)-linear self energies (*67*), the common interpretation of such T-linear resistivity was and still remains the nucleus of ideas centered on quantum criticality. The strict definition of quantum criticality requires the divergence of a thermodynamic quantity. In heavy fermion metals, the electronic heat capacity ratio C_{el}/T indeed grows as $\ln(1/T)$ as the antiferromagnetic correlations diverge (*49*, *55*, *68*). In certain hole-doped cuprates, C_{el}/T also scales as $\ln(1/T)$ at doping levels close to the end of the pseudogap regime (*69*), although here, evidence for a divergent length scale of an associated order parameter is currently lacking (*70*). Moreover, photoemission suggests that at a T-independent critical doping $p_c \approx 0.19$, all signatures of incoherent spectral features that define the strange metal cease, giving way to a more conventional coherent response (*71*). The abruptness of the transition suggests that it is first-order, posing a challenge to interpretations based solely on criticality.

As mentioned above, another major hurdle for the standard criticality scenario is that the T-linear resistivity persists over a wide range of the relevant tuneable parameter, whether doping [as is the case for cuprates (*8*, *9*, *31*, *32*, *72*) and magic-angle twisted bilayer graphene (MATBG) (*60*)], pressure [for $YbBAl_4$ (*53*) and the organics (*58*)], or magnetic field [for Ge-doped $YbRh_2Si_2$ (*51*)]. If quantum criticality is the cause, then it is difficult to fathom how a thermodynamic quantity can be fashioned to diverge over an entire phase.

Despite these difficulties, it is worth exploring the connection T-linear resistivity has with continuous quantum critical phenomena, which for the sake of argument we presume to be tied to a singular point in the phase diagram. Regardless of the origin of the QCP, universality allows us to answer a simple question: What constraints does quantum criticality place on the T dependence of the resistivity? The answer to this question should be governed only by the fundamental length scale for the correlations. The simplest formulation of quantum criticality is single-parameter scaling in which the spatial and temporal correlations are governed by the same diverging length (see Fig. 2). Making the additional assumption that the relevant charge carriers are formed from the quantum critical fluctuations, we find that a simple scaling analysis on the singular part of the free energy results in the scaling law

$$\sigma(\omega = 0, T) = \frac{q^2}{\hbar} f(\omega = 0)\left(\frac{k_B T}{\hbar c}\right)^{(d-2)/z} \quad (5)$$

(*73*) for the T dependence of the conductivity, where $f(\omega = 0)$ is a nonzero constant, q is the

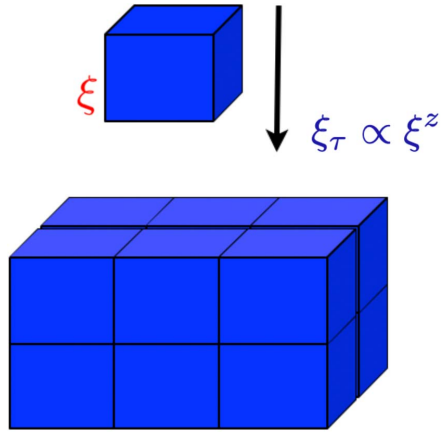


Fig. 2. Single-parameter scaling hypothesis. Depicted here is the collective scaling of a physical system near a critical point. The essential idea is that the correlations within each of the blocks shown are independent of one another. That is, spatial correlations in a volume smaller than the correlation volume, ξ^d, and temporal correlations on a time scale shorter than ξ_τ are small, and space-time regions of size $\xi^d\xi_\tau$ behave as independent blocks. At the critical point, the correlation length diverges. The single-scaling parameter hypothesis assumes that temporal correlations diverge also as a simple power of the spatial correlation length, namely $\xi_\tau \propto \xi^z$, where z is known as the dynamical critical exponent and by causality must exceed unity.

charge, and z is the dynamical exponent, which from causality must obey the inequality $z \geq 1$. Absent from this expression is any dependence on an ancillary energy scale—for example, E_F or the plasma frequency ω_p—as the only assumption is scale-invariant transport irrespective of the details of the system. The analogous expression for the optical conductivity is

$$\sigma(\omega, T = 0) \propto \omega^{(d-2)/z} \quad (6)$$

(*74*). In pure $YbRh_2Si_2$, for example, $\sigma^{-1}(\omega)$ follows an ω-linear dependence at low frequencies in the same region of the (T, H) phase diagram—the quantum critical "fan"—where $\rho(T)$ is also linear, consistent with this notion of single-parameter scaling (*75*). In cuprates, on the other hand, the situation is more nuanced. At intermediate frequencies—sometimes referred to as the mid-IR response—$\sigma(\omega)$ exhibits a ubiquitous $\omega^{-2/3}$ dependence (*7*). Although this feature in $\sigma(\omega)$ has been interpreted in terms of quantum critical scaling (*7*), it is inconsistent with the single-parameter scaling described above. At any doping level, $\sigma(\omega)$ in the cuprates exhibits a minimum at roughly the charge transfer scale of 1 eV. This is traditionally (*76*, *77*) used as the energy scale demarcating the separation between intraband and interband transitions and hence serves to separate the low-energy from the high-energy continua. It has long been debated whether the broad sub-eV $\sigma(\omega)$ response in cuprates is best analyzed in terms of one or two components (*77*, *78*). In the former case, the $\omega^{-2/3}$ tail is simply a consequence of the strong ω-linear dependence in $1/\tau_{tr}(\omega)$—as in MFL—whereas in the latter, it forms part of an incoherent response that is distinct from the coherent Drude weight centered at ω = 0, which itself is described with either a constant or ω-dependent scattering rate.

Returning to the DC resistivity, we find that in cuprates, where $d = 3$, an exponent $z = -1$ is required to account for the T-linear dependence; this value is strictly forbidden by causality (*73*). For $d = 2$, as in the case of MATBG, the T dependence vanishes. This can be fixed with the replacement of $d \rightarrow d^* = 1$ for both materials. Although d^* can be construed as the number of dimensions (*79*) transverse to the Fermi surface, it is difficult to justify such a procedure here, as the persistence of T-linearity with no change in slope above and below the MIR requires a theory that does not rely on FL concepts such as a Fermi velocity or energy. Furthermore, it is well known that introducing d^* yields a power law for the heat capacity, $C \propto T^{3/2}$, which is not seen experimentally (*80*). On dimensional grounds, the $z = -1$ result in the context of the Drude formula is a consequence of compensating the square power of the plasma frequency with powers of T so that the scaling form of Eq. 5 is maintained. A distinct possibility is that perhaps some other form of quantum criticality beyond single-parameter scaling, such as a noncritical form of the entropy suggested recently (*81*), is at work here (see below).

Another critical feature of the conductivity is its behavior at finite wave vector k, which may be quantified by the dynamic charge susceptibility,

$$\chi''(k, \omega) = -\frac{k^2}{\omega e^2} \Re\sigma(k, \omega) \quad (7)$$

determined from electron energy-loss spectroscopy (EELS). A restriction on EELS is that it measures the longitudinal charge response, whereas optics yields the transverse. At vanishing momentum they are expected to be equal. Because optics has no momentum resolution, comparison with EELS can only be made as $k \rightarrow 0$. The primary charge excitation in strange metals is a 1-eV plasmon that was long believed to exhibit the same behavior as in a normal Fermi liquid (*82*, *83*). Recent high-resolution momentum-resolved EELS (M-EELS) measurements have called this belief into question, showing that the large-k response is dominated by a continuum that remains flat to high energies, roughly 2 eV (*84*–*86*). Such behavior is reminiscent of the MFL (*67*) scenario except in that picture, the continuum persists up to a cutoff scale determined by the temperature, and not the Mott scale of 2 eV. In addition, the continuum exhibits scale-invariant features but with a dynamical critical exponent, $z \sim \infty$, not possible from a simple QCP.

We conclude, then, that no form of traditional quantum criticality can account easily for the power laws seen in strange metallic transport (although we recognize that T-linear resistivity is observed above what appear to be genuine singular QCPs). The photoemission experiments (*71*) indicating a first-order transition pose an additional problem exacerbated by the possibility that the criticality might be relevant to a whole region (*8*, *9*, *51*, *53*, *58*, *60*, *65*, *87*, *88*) rather than a point. Such criticality over an extended region is reminiscent of critical charged matter (*89*, *90*) arising from dilatonic models in gauge-gravity duality. These ideas have been the most successful thus far in reproducing the various characteristics of strange metal physics (see below and Table 2).

Is it Planckian dissipation?

Whereas the electrical resistivity in metals can be measured directly, the scattering rate is entirely an inferred quantity. Herein lies the catch with Planckian dissipation. Angle-resolved photoemission (ARPES) experiments on cuprates as early as 1999 reported that the width of the momentum distribution curves (MDCs) at optimal doping along the nodal direction [(0, 0) to (π, π)] scale as a linear function of temperature and $a_0 + 0.75\omega$ for frequencies that exceed $2.5k_BT$ (*91*). The momentum linewidth, which in photoemission enters as Im Σ—the imaginary part of the self energy—can be used to define a lifetime through

$$\hbar v_k \Delta k = \text{Im}\,\Sigma(k, \omega) = 2\frac{\hbar}{\tau} \quad (8)$$

where v_k is the group velocity for momentum state k. Extracting the slope from the data in Fig. 2 of (*92*) and using the experimentally reported Fermi velocity v_F = 1.1 eV/Å, we find that the single-particle scattering rate $\hbar/\tau \sim 1.7k_BT$ (i.e., on the order of the Planckian limit). Similar results were obtained in subsequent ARPES studies (*92*–*94*) with a key extension introduced in (*88*) whereby the width of nodal states was observed to obey a quadrature form described by the expression $[(\hbar\omega)^2 + (\beta k_BT)^2]^\lambda$, where λ is a doping-dependent exponent equal to ½ at optimal doping.

This extraction of the scattering rate from ARPES, however, is not entirely problem-free, as v_F is hard to define in ARPES experiments at energies close to the Fermi level and where, for the most part, the width of the state exceeds its energy. Indeed, the integral of the density of states using as input the v_F extracted from APRES measurements is found to account for only half of the as-measured electronic specific heat coefficient (*95*). Furthermore, this reliance on Fermiology (fermion phenomenology), present also in (*10*), leaves open the precise meaning of figure 2 of (*10*), in which the magnitude of the T-linear transport scattering rate extracted from the DC resistivity is found to scale as $1/v_F$ for a series of materials that violate the MIR limit at intermediate to high temperatures. Despite this, a similar extraction in (*9*), again using Fermiology but focusing on the low-T resistivity, also found a transport scattering rate close to the Planckian bound. This consistency between the two analyses reflects the curious fact that the T-linear slope of the DC resistivity does not vary markedly as the MIR threshold is crossed. It does not, however, necessarily justify either approach in validating T-linear scattering at the Planckian limit. Finally, although T-linearity and Planckian dissipation appear synonymous in the cuprates, this is not universally the case. In $YbRh_2Si_2$ (*75*), for example, the T-linear scattering rate is found to deviate strongly from the Planckian limit with $\tau_{tr} \sim 0.1\tau_P$ (*52*), and in the electron-doped cuprates, the notion of a Planckian limit to the scattering rate has recently been challenged (*96*). This certainly adds to the intrigue regarding quantum criticality as the underlying cause of Planckian dissipation, for which several alternative proposals have recently emerged (*97*, *98*).

In principle, the optical conductivity permits an extraction of τ without recourse to Fermiology. Within a Drude model, the optical conductivity

$$\sigma(\omega) = \frac{1}{4\pi}\frac{\omega_p^2\tau_{tr}}{1 + i\omega\tau_{tr}} \quad (9)$$

Table 2. Snapshot of current theoretical modeling of the strange metal regime. Indicated are consistency with *T*-linear resistivity, $\omega^{-2/3}$ scaling of the mid-IR optical conductivity, quadrature-scaling magnetoresistance, extended quantum criticality, and what predictions are made in terms of experimental observables. Scenarios: MFL, marginal Fermi liquid; EFL, ersatz Fermi liquid; SYK, Sachdev-Ye-Kitaev; AdS/CFT, anti–de Sitter space/conformal field theory conjecture; AD/EMD, anomalous dimensions/Einstein-Maxwell-dilaton; HM, Hubbard model; QMC, quantum Monte Carlo; ED, exact diagonalization; CA, cold atoms; DMFT/EDMFT, dynamical mean-field theory/embedded dynamical mean-field theory; A-B, Aharonov-Bohm effect; ECFL, extremely correlated Fermi liquid; QCP, quantum critical point. *T-linear resistivity is an input. **A slope change occurs through the MIR. ***Quadrature scaling obtained only for a bivalued random resistor model (*121*) with equal weights (*27*). ****Although this scaling was thought to arise in pure AdS with an inhomogeneous charge density (*123*), later studies (*124*, *125*) found otherwise.

	$\rho \propto T$ as $T \to 0$	$\rho \propto T$ as $T \to \infty$	$\sigma \propto \omega^{-2/3}$	Quadrature MR	Extended criticality	Experimental prediction
			Phenomenological			
MFL	√ (*67*)	× (*67*)	×	×	×	Loop currents (*107*)
EFL	— *	—	—	×	×	Loop currents (*108*)
			Numerical			
ECFL	×	(*109*)	—	—	×	×
HM (QMC/ED/CA)	— (*110*)	√ (*110*–*114*)	×	—	—	—
DMFT/EDMFT	√ (*115*)	√ (*116*, *117*)	×	—	√ (*117*)	—
QCP	(*118*)	—	—	—	×	—
			Gravity-based			
SYK	√ (*119*, *120*)	√** (*120*)	×	√*** (*121*)	—	×
AdS/CFT	√ (*122*)	√ (*122*)	√**** (*90*, *126*)	×	×	×
AD/EMD	√ (*127*–*129*)	√ (*90*, *126*, *127*, *129*, *130*)	√ (*90*, *126*, *130*)	×	√ (*126*)	Fractional A-B (*129*)

contains only τ_{tr} and $\omega_p = \sqrt{4\pi n_e e^2/m}$. At zero frequency, the Drude formula naturally yields the DC conductivity σ_{DC} and an estimate for the relaxation rate can be extracted from the width at half maximum of the full Drude response. However, there is an important caveat: τ_{tr} is frequency-dependent in the cuprates, a condition that is consistent with various physical models including both the Fermi liquid and MFL scenarios as well as the large body (*88*, *91*) of MDC analysis performed on the cuprates. Although this prevents a clean separation of the conductivity into coherent and incoherent parts, it was shown in (*7*) that in the low-frequency limit, $\omega < 1.5k_BT/\hbar$, $\tau_{tr} \sim 0.8\tau_P$, in agreement with the DC analysis of (*9*).

A second key issue remains: How can such Drude analysis be justified for those strange metals in which the MIR limit is violated and the Drude peak shifts to finite frequencies (*16*)? Indeed, in the high-*T* limit, "bad metallicity" can be ascribed to a transfer of spectral weight from low to high ω, rather than to an ever-increasing scattering rate (that within a Drude picture results in a continuous broadening of the Lorentzian fixed at zero frequency). Given the marked crossover in the form of σ(ω) at low frequencies, it is indeed mysterious that the slope of the *T*-linear resistivity continues unabated with no discernible change.

Is it Mottness?

Table 1 encompasses a series of ground states from which *T*-linear resistivity emerges. In some of these materials, such as the heavy fermions, the high- and low-energy features of the spectrum are relatively distinct in the sense that spectral weight transfer from the UV to the IR is absent. On the other hand, hole or electron doping of the parent cuprate induces a marked transfer of spectral weight of roughly 1 to 2 eV. As a result, the low-energy spectral weight grows (*77*, *99*–*102*) at the expense of the degrees of freedom at high energy, a trend that persists (*76*) even inside the superconducting state. This is an intrinsic feature of Mott systems, namely that the number of low-energy degrees of freedom is derived from the high-energy spectral weight. As this physics is distinct from that of a Fermi liquid and intrinsic to Mott physics, it is termed "Mottness" (*102*). Notably, the mid-IR response with its characteristic $\omega^{-2/3}$ scaling is absent from the parent Mott insulating state. Hence, it must reflect the doping-induced spectral weight transfer across the Mott gap. It is perhaps not a surprise, then, that no low-T_c material exhibits such a mid-IR feature. In fact, some theories of cuprate superconductivity (*103*) credit its origin to the mid-IR scaling. We can quantify the total number of low-energy degrees of freedom that arise from the UV-IR mixing across the Mott gap by integrating the optical conductivity,

$$N_{\mathrm{eff}}(\Omega) = \frac{2mV_{\mathrm{cell}}}{\pi e^2}\int_0^{\Omega}\sigma(\omega)d\omega \qquad (10)$$

up to the optical gap $\Omega \approx 1.2$ eV, where V_{cell} is the unit cell volume. The energy scale of 1.2 eV corresponds to the minimum of the optical conductivity, as mentioned above. In a rigid-band semiconductor model in which such spectral weight transfer is absent, $N_{eff} = x$, where x is the number of holes. In the cuprates, however, N_{eff} exceeds x (Fig. 3). This is the defining feature of Mottness (*99*–*102*), as it is ubiquitous in Mott systems and strictly absent in weakly correlated metals. Even in many of the strange or quantum critical metals described in Table 1, there is little or no evidence that Mottness is playing a notable role. Such a distinction may thus offer a hint to the source of the uniqueness of the cuprate strange metal. In bad metals, on the other hand, a gradual transfer of low-frequency spectral weight out to energies on the order of the Mott scale is almost universally observed with increasing temperature (*16*), suggesting that Mottness is one of the key components of bad metallic transport.

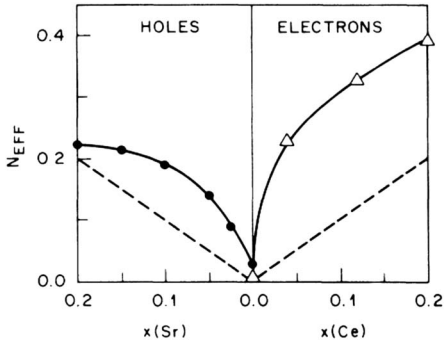


Fig. 3. Integrated optical conductivity N_{eff} exceeds dopant concentration. Shown is N_{eff} for electron-doped $Pr_{2-x}Ce_xCuO_{4-\delta}$ (triangles) and hole-doped $La_{2-x}Sr_xCuO_{4-\delta}$ (circles) The dashed line indicates what is expected for doping a semiconductor. The expectation is that each Ce or Sr atom contributes just a single charge carrier. [Reprinted with permission from (*77*)]

The optical response in cuprates tells us that there are degrees of freedom that couple to electromagnetism that have no interpretation in terms of doped holes. That is, they are not local entities, as they arise from the mixing of UV and IR degrees of freedom. It is such mixing that could account for the lack of any distinctive energy scale (*102*)—that is, scale invariance—underlying the strange metal. Additionally, optical conductivity studies showed (*104*) that throughout the underdoped regime of the cuprate phase diagram, the effective mass remains constant. As a result, the Mott transition proceeds by a vanishing of the carrier number rather than the mass divergence of the Brinkman-Rice scenario (*105*). [Note that although quantum oscillation experiments on underdoped cuprates show evidence for mass enhancement (*106*), this is thought to be tied to the charge order around 1/8 doping.] Such dynamical mixing between the UV and IR scales in Mott systems is well known to give rise to spectral weight in the lower Hubbard band (*100*–*102*) that exceeds the number of electrons, strictly $1 + x$, that the band can hold. Consequently, part of the electromagnetic response of strange metals at low energies has no interpretation in terms of electron quasiparticles, as it arises strictly from UV-IR mixing. Precisely how such mixing leads to scale-invariant T-linear resistivity remains unresolved.

Is it about gravity?

To frame the theoretical modeling of strange metallicity, we group the work into three principal categories: phenomenological, numerical, and gravity-related. Table 2 lists some representative contributions (*67*, *90*, *107*–*130*); because of space limitations, it is not possible to cite all relevant work. Both phenomenological models considered here require (such as the ersatz Fermi liquid or EFL) or predict (MFL) loop currents, but they do so for fundamentally different reasons. For EFL (*108*), such current order is needed to obtain a finite resistivity in the absence of momentum relaxation (certainly not a natural choice given the Drude fit to the optical conductivity discussed previously), whereas in MFL, loop currents (*107*) are thought to underpin the local fluctuation spectrum (*67*). The extremely correlated Fermi liquid theory (ECFL) (*109*) predicts a resistivity that interpolates between Fermi liquid–like T^2 at low T to T-linear for $T >> T_{FL}$. Quantum Monte Carlo (QMC) simulations (*110*, *111*, *113*, *114*) as well as cold atom experiments (*112*) on the Hubbard model have established that at high temperatures, the resistivity is indeed T-linear. The fermion-sign problem, however, prevents any definitive statement about the low-T behavior in the presence of Mott physics. Non–Fermi liquid transport in Sachdev-Ye-Kitaev (SYK)–based models (*131*–*133*) is achieved by an all-to-all random interaction. Although such interactions might seem initially unphysical, SYK models are nonetheless natural candidates to destroy Fermi liquids, which, by their nature, permit a purely local description in momentum space. As a result, Fermi liquids are impervious to generic repulsive local-in-space interactions (*134*). Coupling a Fermi liquid to an array of disordered SYK islands, however, leads (*120*, *121*) to a nontrivial change in the electron Green function across the MIR, and hence a change in slope of the resistivity is unavoidable (*121*), although it can be minimized through fine tuning (*120*).

An added feature of these disordered models is that in certain limits, they have a gravity dual (*132*, *133*, *135*, *136*). This state of affairs arises because the basic propagator (*132*, *133*, *135*) in the SYK model in imaginary time describes the motion of fermions, with appropriate boundary conditions, between two points of the asymptotic boundary of a hyperbolic plane. In real time, simply replacing the hyperbolic plane with the spacetime equivalent, namely two-dimensional anti–de Sitter (AdS) space (a maximally symmetric Lorentzian manifold with constant negative curvature), accurately describes all the correlators. It is from this realization that the dual description between a random spin model and gravity in AdS_2 lies (*132*, *133*, *137*). Hence, although the origins of SYK were independent of gravity, its correlators can be deduced from the asymptotics of the corresponding spacetime. At the asymptote, only the time coordinate survives and hence, ultimately, SYK dynamics is ultralocal in space with correlations diverging only in time, an instantiation of local quantum criticality.

Such local quantum criticality is not a new concept in condensed matter systems and indeed lies at the heart of MFL phenomenology (*67*) as well as dynamical mean field theory (DMFT) (*116*), and is consistent with the momentum-independent continuum found in the M-EELS data discussed earlier (*85*). The deeper question is why gravity has anything to do with a spin problem with nonlocal interactions. The issue comes down to criticality and to the structure of general relativity. The second equivalence principle on which general relativity is based states that no local measurement can detect a uniform gravitational field. A global measurement is required. The same is true for a critical system because no local measurement can discern criticality. Observables tied to the diverging correlation length are required. Hence, at least conceptually, it is not unreasonable to expect a link between critical matter and gravity. The modern mathematical machinery that makes it possible to relate the two is the gauge-gravity duality or the AdS/CFT (anti–de Sitter/conformal field theory) conjecture. The key claim of this duality (*138*–*140*) is that some strongly interacting quantum theories, namely ones that are at least conformally invariant in d dimensions, are dual to a theory of gravity in a $d + 1$ spacetime that is asymptotically AdS. The radial direction represents the energy with the quantum theory residing at the UV boundary, whereas the IR limit is deep in the interior at the black hole horizon. Hence, intrinsic to this construction is a separation between bulk (gravitational) and boundary (quantum mechanical) degrees of freedom. That the boundary of a gravitational object has features distinct from the bulk dates back to the observations of Beckenstein (*141*) and Hawking (*142*, *143*) that the information content of a black hole scales with the area, not the volume. The requirement that the boundary theory be strongly coupled then arises by maintaining that the AdS radius of curvature exceeds the Planck length ℓ_P. More explicitly, because the AdS radius and the coupling constant of the boundary theory are proportional, the requirement $R >> \ell_P$ translates into a boundary theory that is strongly coupled.

The first incarnation (*122*, *144*, *145*) of this duality in the context of fermion correlators involved modeling fermions at finite density in 2 + 1 dimensions. From the duality, the conformally invariant vacuum of such a system corresponds to gravity in AdS_4, the extra dimension representing the radial direction along which identical copies of the boundary CFT lie (albeit with differing energy scales). Surprisingly, what was shown (*122*) is that the low-energy (IR) properties of such a system in the presence of a charge density are determined by an emergent $AdS_2 \times R^2$ (with R^2 representing a plane) spacetime at the black hole horizon. The actual symmetry includes scale invariance and is denoted by $SL(2, R)$ (a special Lie group of real 2×2 matrices with a unit determinant). Once again, the criticality of boundary fermions is determined entirely by the fluctuations in time, that is, local quantum criticality as seen in SYK. The temperature and frequency dependence of the conductivity are then determined by the same exponent (*122*) as expected from Eqs. 5 and 6 and, as a result, a simultaneous description of T-linearity and $\omega^{-2/3}$ dependence is not possible, as noted in Table 2.

This particular hurdle is overcome by the anomalous dimensions/Einstein-Maxwell-dilaton (AD/EMD) theories (*89*, *90*, *126*, *146*–*149*), which, as indicated in Table 2, have been the most successful to date in describing the range of physics observed in strange metals. What is new here is the introduction of extra fields, dilatons for example, that permit hyperscaling violation (*79*) and anomalous dimensions (*89*, *90*, *126*, *146*–*149*) for all operators. Consequently, under a scale change of the coordinates, the metric is no longer unscathed. That is, the manifold is not fixed and it is the matter fields that determine the geometry. Such

systems have a covariance, rather than scale invariance indicative of pure AdS metrics. A consequence of this covariance is that even the currents acquire anomalous dimensions. But how is this possible given that a tenet of field theory is that no amount of renormalization can change the dimension of the current (*150*) from $d - 1$? What makes this possible is that in EMD theories, the extra radial dimension allows physics beyond Maxwellian electromagnetism. For example, the standard Maxwell action, $S = \int dV_d F^2$ where $F = dA$, requires that the dimension of the gauge field be fixed to unity, $[A] = 1$ (*151*). EMD theories use instead an action of the form $S = \int dV_d dy y^a F^2$ where y is the radial coordinate of the $d + 1$ AdS spacetime. Comparing these two actions leads immediately to the conclusion that the dimension of A now acquires the value $[A] = 1 - (a/2)$. Hence, even in the bulk of the geometry, the dimension of the gauge field is not unity. Depending on the value of a, $a < 0$ at the UV conformal boundary or $a > 0$ at the IR at the black hole horizon, the equations of motion are nonstandard and obey fractional electromagnetism (*128*, *152*) consistent with a nontraditional dimension for the gauge field. In EMD theories, it is precisely the anomalous dimension (*89*, *90*, *126*, *146*–*148*) for conserved quantities that gives rise to the added freedom for extended quantum criticality to occur, enabling the simultaneous fitting (*130*) of T-linearity and $\omega^{-2/3}$ of the optical conductivity, and establishing the basis for a proposal for the strange metal based on $[A] = 5/3$ (*127*).

Within these holographic systems, a Drude-like peak in the optical conductivity can emerge both from the coherent (quasiparticle-like) sector (*153*) as well as from the incoherent ["unparticle" (*154*)] sector (*155*–*158*). Application of EMD theory has also provided fresh insights into the phenomenon of "lifetime separation" seen in the DC and Hall conductivities of hole-doped cuprates (*22*, *24*, *28*) as well as in other candidate strange metals (*52*, *61*). For a system with broken translational invariance, the finite density conductivity comprises two distinct components (*159*). The DC resistivity is dominated by the particle-hole symmetric term—whose Hall conductivity is consequently zero—whereas the T dependence of the Hall angle is set by the more conventional (umklapp) momentum relaxation that governs the response of the coherent charge density.

The success of EMD theories in the context of strange metal physics raises a philosophical question: Is all of this just a game? That is, is the construction of bulk theories with funky electromagnetism fundamental? The answer lies in Nöther's Second Theorem (NST) (*102*, *128*, *152*), a theorem far less known than her ubiquitous first theorem but ultimately of more importance as it identifies a shortcoming. To illustrate her first theorem, consider Maxwellian electromagnetism, which is invariant under the transformation $A_\mu \rightarrow A_\mu + \partial_\mu \Lambda$. This theorem states that there must be a conservation law with the same number of derivatives as in the gauge principle. Hence, the conservation law only involves a single derivative, namely $\partial_\mu J_\mu = 0$. This is Nöther's First Theorem (*160*) in practice.

What Nöther (*160*) spent the second half of her famous paper trying to rectify is that the form of the gauge transformation is not unique; hence, the conservation law is arbitrary. It is for this reason that in the second half (*160*) of her foundational paper, she retained all possible higher-order integer derivatives in the gauge principle. These higher-order derivatives both add constraints to and change the dimension of the current. Stated succinctly, NST (*160*) dictates that the full family of generators of U(1) invariance determines the dimension of the current. It is easy to see how this works. Suppose we can find a quantity $\hat{Y}$ that commutes with ∂_μ. That is, $\partial_\mu \hat{Y} = \hat{Y} \partial_\mu$. If this is so, then we can insert this into the conservation law with impunity. What this does is redefine the current: $\partial_\mu \hat{Y} J^\mu = \partial_\mu \tilde{J}^\mu$. The new current $\tilde{J}^\mu$ acquires whatever dimensions $\hat{Y}$ has, such that $[\tilde{J}^\mu] = d - 1 - d_Y$. But because of the first theorem, $\hat{Y}$ must have come from the gauge transformation and hence must ultimately be a differential operator itself. That is, there is an equally valid class of electromagnetisms with gauge transformations of the form $A_\mu \rightarrow A_\mu + \partial_\mu \hat{Y} \Lambda$. For EMD theories (*102*, *128*, *152*), $\hat{Y}$ is given by the fractional Laplacian, $\Delta^{(\gamma-1)/2}$ with $[A_\mu] = \gamma$ [with $\gamma = 1 - (a/2)$ to make contact with the EMD theories introduced earlier]. For most matter as we know it, $\gamma = 1$. The success of EMD theories raises the possibility that the strangeness of the strange metal hinges on the fact that $\gamma \neq 1$. This can be tested experimentally using the standard Aharonov-Bohm geometry (*128*, *129*) in which a hole of radius r is punched into a cuprate strange metal. Because $[A]$ is no longer unity, the integral of $A \cdot d\ell$ is no longer the dimensionless flux. For physically realizable gauges, this ultimately provides an obstruction to charge quantization. As a result, deviations (*128*, *129*) from the standard $\pi r^2 \times B$ dependence for the flux would be the key experimental feature that a nonlocal gauge principle is operative in the strange metal. An alternative would be, as Anderson (*161*) advocated, the use of fractional or unparticle propagators with the standard gauge principle. However, in the end, it all comes down to gauge invariance. The standard gauge-invariant condition prevents the power laws in unparticle stuff from influencing the algebraic fall-off of the optical conductivity (*130*, *162*), as they offer just a prefactor to the polarizations (*163*). The escape route, an anomalous dimension for the underlying gauge field, offers a viable solution, but the price is abandoning locality (*164*) of the action.

Is it important?

Given the immense difficulty in constructing a theory of strange metals, one might ask why bother? To gauge the importance of strange metals, look no further than Fig. 4. This figure shows that the coefficient α_1 of the T-linear resistivity component in the strange metal regime of overdoped hole-doped cuprates tracks the doping dependence of the $T = 0$ superfluid density $n_s(0)$. As mentioned earlier, a similar correlation exists between α_1 and T_c in electron-doped cuprates (*31*, *32*), the Bechgaard salts (*58*), and the iron pnictides (*58*), establishing a fundamental link between high-temperature superconductivity and strange metals.

For a long time, the drop in $n_s(0)$ with doping in cuprates was attributed to pair breaking, a symptom of the demise of the dominant pairing interaction within a disordered lattice. Recent mutual inductance measurements, however, have challenged this view, arguing that the limiting low-T behavior of $n_s(T)$ was incompatible with conventional pair-breaking scenarios (*165*). Certainly, the correlation between α_1 and $n_s(0)$ is unforeseen in such models. Moreover, if the strange metal regime is indeed populated with non-quasiparticle states, then Fig. 4 indicates a pivotal role for these states in the pairing condensate (*166*, *167*). On more general grounds, this result informs us that the door to unlocking cuprate superconductivity is through the strange metal regime, and any theory that divorces superconductivity from strange metals is unlikely to be a promising avenue. To conclude, solving the strange metal regime kills two birds with one stone. Perhaps there is some justice here. After all, we know from Pippard's (*168*) work, which can be reformulated (*128*, *152*) in terms of fractional

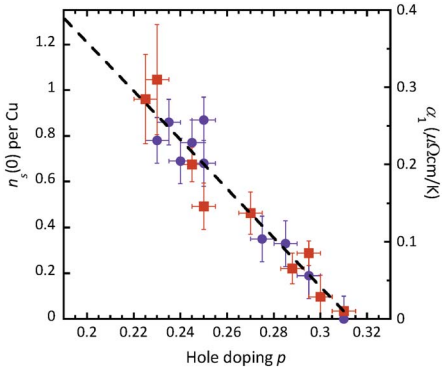


Fig. 4. Correlation between the superfluid density $n_s(0)$ and the coefficient a_1 of the T-linear resistivity. The data shown are for $Tl_2Ba_2CuO_{6+\delta}$ (Tl2201) across the strange metal regime. [Adapted from (*166*, *167*)]

Laplacians, that explaining superconductivity even in elemental metals necessitates a nonlocal relationship between the current and the gauge field. What seems to be potentially new about the cuprates is that now the normal state, as a result of the strange metallicity, also requires nonlocality.

REFERENCES AND NOTES

1. T. Matsubara, A New Approach to Quantum-Statistical Mechanics. *Prog. Theor. Phys.* **14**, 351–378 (1955). doi: 10.1143/PTP.14.351
2. S. Chakravarty, B. I. Halperin, D. R. Nelson, Two-dimensional quantum Heisenberg antiferromagnet at low temperatures. *Phys. Rev. B* **39**, 2344–2371 (1989). doi: 10.1103/PhysRevB.39.2344; pmid: 9948474
3. J. Zaanen, Superconductivity: Why the temperature is high. *Nature* **430**, 512–513 (2004). doi: 10.1038/430512a; pmid: 15282588
4. M. Gurvitch, A. T. Fiory, Resistivity of $La_{1.825}Sr_{0.175}CuO_4$ and YBa_2Cu_3O to 1100 K: Absence of saturation and its implications. *Phys. Rev. Lett.* **59**, 1337–1340 (1987). doi: 10.1103/PhysRevLett.59.1337; pmid: 10035206
5. S. Martin, A. T. Fiory, R. M. Fleming, L. F. Schneemeyer, J. V. Waszczak, Normal-state transport properties of $Bi_{2+x}Sr_{2-y}CuO_{6+\delta}$ crystals. *Phys. Rev. B* **41**, 846–849 (1990). doi: 10.1103/PhysRevB.41.846; pmid: 9992839
6. H. Takagi *et al*., Systematic evolution of temperature-dependent resistivity in $La_{2-x}Sr_xCuO_4$. *Phys. Rev. Lett.* **69**, 2975–2978 (1992). doi: 10.1103/PhysRevLett.69.2975; pmid: 10046689
7. D. van der Marel *et al*., Quantum critical behaviour in a high-T_c superconductor. *Nature* **425**, 271–274 (2003). doi: 10.1038/nature01978; pmid: 13679910
8. R. A. Cooper *et al*., Anomalous criticality in the electrical resistivity of $La_{2-x}Sr_xCuO_4$. *Science* **323**, 603–607 (2009). doi: 10.1126/science.1165015; pmid: 19074310
9. A. Legros *et al*., Universal T-linear resistivity and Planckian dissipation in overdoped cuprates. *Nat. Phys.* **15**, 142–147 (2019). doi: 10.1038/s41567-018-0334-2
10. J. A. N. Bruin, H. Sakai, R. S. Perry, A. P. Mackenzie, Similarity of scattering rates in metals showing T-linear resistivity. *Science* **339**, 804–807 (2013). doi: 10.1126/science.1227612; pmid: 23413351
11. P. Fournier *et al*., Insulator-Metal Crossover near Optimal Doping in $Pr_{2-x}Ce_xCuO_4$: Anomalous Normal-State Low Temperature Resistivity. *Phys. Rev. Lett.* **81**, 4720–4723 (1998). doi: 10.1103/PhysRevLett.81.4720
12. A. P. Mackenzie, S. R. Julian, D. C. Sinclair, C. T. Lin, Normal-state magnetotransport in superconducting $Tl_2Ba_2CuO_{6+\delta}$ to millikelvin temperatures. *Phys. Rev. B* **53**, 5848–5855 (1996). doi: 10.1103/PhysRevB.53.5848; pmid: 9984193
13. N. Nagaosa, P. A. Lee, Ginzburg-Landau theory of the spin-charge-separated system. *Phys. Rev. B* **45**, 960 (1992). doi: 10.1103/PhysRevB.45.966
14. A. F. Ioffe, A. R. Regel, Non-crystalline, amorphous, and liquid electronic semiconductors. *Prog. Semicond.* **4**, 237 (1960).
15. M. Gurvitch, Ioffe-Regel criterion and resistivity of metals. *Phys. Rev. B* **24**, 7404–7407 (1981). doi: 10.1103/PhysRevB.24.7404
16. N. E. Hussey, K. Takenaka, H. Takagi, Universality of the Mott-Ioffe-Regel limit in metals. *Philos. Mag.* **84**, 2847–2864 (2004). doi: 10.1080/14786430410001716944
17. N. F. Mott, Conduction in non-crystalline systems IX. the minimum metallic conductivity. *Philos. Mag.* **26**, 1015–1026 (1972). doi: 10.1080/14786437208226973
18. N. E. Hussey *et al*., Dichotomy in the *T*-linear resistivity in hole-doped cuprates. *Philos. Trans. R. Soc. A* **369**, 1626–1639 (2011). doi: 10.1098/rsta.2010.0196; pmid: 21422018
19. V. J. Emery, S. A. Kivelson, Superconductivity in bad metals. *Phys. Rev. Lett.* **74**, 3253–3256 (1995). doi: 10.1103/PhysRevLett.74.3253; pmid: 10058150
20. C. Proust, B. Vignolle, J. Levallois, S. Adachi, N. E. Hussey, Fermi liquid behavior of the in-plane resistivity in the pseudogap state of $YBa_2Cu_4O_8$. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13654–13659 (2016). doi: 10.1073/pnas.1602709113; pmid: 27856753
21. N. Barišic *et al*., Universal sheet resistance and revised phase diagram of the cuprate high-temperature superconductors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12235–12240 (2013). doi: 10.1073/pnas.1301989110; pmid: 23836669
22. A. Carrington, A. P. Mackenzie, C. T. Lin, J. R. Cooper, Temperature dependence of the Hall angle in single-crystal $YBa_2(Cu_{1-x}Co_x)_3O_{7-\delta}$. *Phys. Rev. Lett.* **69**, 2855–2858 (1992). doi: 10.1103/PhysRevLett.69.2855; pmid: 10046606
23. M. K. Chan *et al*., In-plane magnetoresistance obeys Kohler's rule in the pseudogap phase of cuprate superconductors. *Phys. Rev. Lett.* **113**, 177005 (2014). doi: 10.1103/PhysRevLett.113.177005; pmid: 25379934
24. T. R. Chien, Z. Z. Wang, N. P. Ong, Effect of Zn impurities on the normal-state Hall angle in single-crystal $YBa_2Cu_{3-x}Zn_xO_{7-\delta}$. *Phys. Rev. Lett.* **67**, 2088–2091 (1991). doi: 10.1103/PhysRevLett.67.2088; pmid: 10044332
25. J. M. Harris *et al*., Violation of Kohler's rule in the normal-state magnetoresistance of $YBa_2Cu_3O_{7-\delta}$ and $La_2Sr_xCuO_4$. *Phys. Rev. Lett.* **75**, 1391–1394 (1995). doi: 10.1103/PhysRevLett.75.1391; pmid: 10060281
26. P. Giraldo-Gallo *et al*., Scale-invariant magnetoresistance in a cuprate superconductor. *Science* **361**, 479–481 (2018). doi: 10.1126/science.aan3178; pmid: 30072535
27. C. Boyd, P. W. Phillips, Single-parameter scaling in the magnetoresistance of optimally doped $La_{2-x}Sr_xCuO_4$. *Phys. Rev. B* **100**, 155139 (2019). doi: 10.1103/PhysRevB.100.155139
28. T. Manako, Y. Kubo, Y. Shimakawa, Transport and structural study of $Tl_2Ba_2CuO_{6+\delta}$ single crystals prepared by the KCl flux method. *Phys. Rev. B* **46**, 11019–11024 (1992). doi: 10.1103/PhysRevB.46.11019; pmid: 10002965
29. J. Ayres *et al*., Incoherent transport across the strange-metal regime of overdoped cuprates. *Nature* **595**, 661–666 (2021). doi: 10.1038/s41586-021-03622-z; pmid: 34321672
30. N. R. Poniatowski, T. Sarkar, S. Das Sarma, R. L. Greene, Resistivity saturation in an electron-doped cuprate. *Phys. Rev. B* **103**, L020501 (2021). doi: 10.1103/PhysRevB.103.L020501
31. K. Jin, N. P. Butch, K. Kirshenbaum, J. Paglione, R. L. Greene, Link between spin fluctuations and electron pairing in copper oxide superconductors. *Nature* **476**, 73–75 (2011). doi: 10.1038/nature10308; pmid: 21814279
32. J. Yuan *et al*., Scaling of the strange-metal scattering in unconventional superconductors. *Nature* **602**, 431–436 (2022). doi: 10.1038/s41586-021-04305-5; pmid: 35173341
33. P. Li, F. F. Balakirev, R. L. Greene, High-field Hall resistivity and magnetoresistance of electron-doped $Pr_{2-x}Ce_xCuO_{4-\delta}$. *Phys. Rev. Lett.* **99**, 047003 (2007). doi: 10.1103/PhysRevLett.99.047003; pmid: 17678392
34. N. R. Poniatowski, T. Sarkar, R. L. Greene, Anomalous normal-state magnetotransport in an electron-doped cuprate. *Phys. Rev. B* **103**, 125102 (2021). doi: 10.1103/PhysRevB.103.125102
35. T. Sarkar, P. R. Mandal, N. R. Poniatowski, M. K. Chan, R. L. Greene, Correlation between scale-invariant normal-state resistivity and superconductivity in an electron-doped cuprate. *Sci. Adv.* **5**, eaav6753 (2019). doi: 10.1126/sciadv.aav6753; pmid: 31114800
36. A. W. Tyler, A. P. Mackenzie, S. NishiZaki, Y. Maeno, High-temperature resistivity of Sr_2RuO_4: Bad metallic transport in a good metal. *Phys. Rev. B* **58**, 10107 (1998). doi: 10.1103/PhysRevB.58.R10107
37. N. E. Hussey *et al*., Normal-state magnetoresistance of Sr_2RuO_4. *Phys. Rev. B* **57**, 5505–5511 (1998). doi: 10.1103/PhysRevB.57.5505
38. M. E. Barber, A. S. Gibbs, Y. Maeno, A. P. Mackenzie, C. W. Hicks, Resistivity in the Vicinity of a van Hove Singularity: Sr_2RuO_4 under Uniaxial Pressure. *Phys. Rev. Lett.* **120**, 076602 (2018). doi: 10.1103/PhysRevLett.120.076602; pmid: 29542933
39. A. P. Mackenzie *et al*., Hall effect in the two-dimensional metal Sr_2RuO_4. *Phys. Rev. B* **54**, 7425–7429 (1996). doi: 10.1103/PhysRevB.54.7425; pmid: 9984367
40. S. Kasahara *et al*., Field-induced superconducting phase of FeSe in the BCS-BEC cross-over. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16309–16313 (2014). doi: 10.1073/pnas.1413477111; pmid: 25378706
41. S. Licciardello *et al*., Electrical resistivity across a nematic quantum critical point. *Nature* **567**, 213–217 (2019). doi: 10.1038/s41586-019-0923-y; pmid: 30760921
42. W. K. Huang *et al*., Non-Fermi liquid transport in the vicinity of the nematic quantum critical point of superconducting $FeSe_{1-x}S_x$. *Phys. Rev. Res.* **2**, 033367 (2020). doi: 10.1103/PhysRevResearch.2.033367
43. S. Licciardello *et al*., Coexistence of orbital and quantum critical magnetoresistance in $FeSe_{1-x}S_x$. *Phys. Rev. Res.* **1**, 023011 (2019). doi: 10.1103/PhysRevResearch.1.023011
44. D. Hu *et al*., ω/T scaling and magnetic quantum criticality in $BaFe_2(As_{0.7}P_{0.3})_2$. arXiv 1812.11902 (2018).
45. J. G. Analytis *et al*., Transport near a quantum critical point in $BaFe_2(As_{1-x}P_x)_2$. *Nat. Phys.* **10**, 194–197 (2014). doi: 10.1038/nphys2869
46. S. Kasahara *et al*., Evolution from non-Fermi- to Fermi-liquid transport via isovalent doping in $BaFe_2(As_{1-x}P_x)_2$ superconductors. *Phys. Rev. B* **81**, 184519 (2010). doi: 10.1103/PhysRevB.81.184519
47. I. M. Hayes *et al*., Scaling between magnetic field and temperature in the high-temperature superconductor $BaFe_2(As_{1-x}P_x)_2$. *Nat. Phys.* **12**, 916–919 (2016). doi: 10.1038/nphys3773
48. Y. Nakajima *et al*., Quantum-critical scale invariance in a transition metal alloy. *Commun. Phys.* **3**, 181 (2020). doi: 10.1038/s42005-020-00448-5
49. O. Trovarelli *et al*., $YbRh_2Si_2$: Pronounced non-Fermi-liquid effects above a low-lying magnetic phase transition. *Phys. Rev. Lett.* **85**, 626–629 (2000). doi: 10.1103/PhysRevLett.85.626; pmid: 10991356
50. J. Custers *et al*., The break-up of heavy electrons at a quantum critical point. *Nature* **424**, 524–527 (2003). doi: 10.1038/nature01774; pmid: 12891349
51. J. Custers *et al*., Evidence for a non-Fermi-liquid phase in Ge-substituted $YbRh_2Si_2$. *Phys. Rev. Lett.* **104**, 186402 (2010). doi: 10.1103/PhysRevLett.104.186402; pmid: 20482193
52. S. Paschen *et al*., Hall-effect evolution across a heavy-fermion quantum critical point. *Nature* **432**, 881–885 (2004). doi: 10.1038/nature03129; pmid: 15602556
53. T. Tomita, K. Kuga, Y. Uwatoko, P. Coleman, S. Nakatsuji, Strange metal without magnetic criticality. *Science* **349**, 506–509 (2015). doi: 10.1126/science.1262054; pmid: 26228141
54. Y. Nakajima *et al*., Non-Fermi Liquid Behavior in the Magnetotransport of $CeMIn_5$ (*M*: Co and Rh): Striking Similarity between Quasi Two-Dimensional Heavy Fermion and High-T_c Cuprates. *J. Phys. Soc. Jpn.* **76**, 024703 (2007). doi: 10.1143/JPSJ.76.024703
55. A. Bianchi, R. Movshovich, I. Vekhter, P. G. Pagliuso, J. L. Sarrao, Avoided antiferromagnetic order and quantum critical point in $CeCoIn_5$. *Phys. Rev. Lett.* **91**, 257001 (2003). doi: 10.1103/PhysRevLett.91.257001; pmid: 14754138
56. J. Paglione *et al*., Field-induced quantum critical point in $CeCoIn_5$. *Phys. Rev. Lett.* **91**, 246405 (2003). doi: 10.1103/PhysRevLett.91.246405; pmid: 14683139
57. B. Shen *et al*., Strange-metal behaviour in a pure ferromagnetic Kondo lattice. *Nature* **579**, 51–55 (2020). doi: 10.1038/s41586-020-2052-z; pmid: 32132691
58. N. Doiron-Leyraud *et al*., Correlation between linear resistivity and T_c in the Bechgaard salts and the pnictide superconductor $Ba(Fe_{1-x}Co_x)_2As_2$. *Phys. Rev. B* **80**, 214531 (2009). doi: 10.1103/PhysRevB.80.214531
59. H. Polshyn *et al*., Large linear-in-temperature resistivity in twisted bilayer graphene. *Nat. Phys.* **15**, 1011–1016 (2019). doi: 10.1038/s41567-019-0596-3
60. Y. Cao *et al*., Strange Metal in Magic-Angle Graphene with near Planckian Dissipation. *Phys. Rev. Lett.* **124**, 076801 (2020). doi: 10.1103/PhysRevLett.124.076801
61. R. Lyu *et al*., Strange metal behavior of the Hall angle in twisted bilayer graphene. *Phys. Rev. B* **103**, 245424 (2021). doi: 10.1103/PhysRevB.103.245424
62. P. B. Allen *et al*., Transport properties, thermodynamic properties, and electronic structure of $SrRuO_3$. *Phys. Rev. B* **53**, 4393–4398 (1996). doi: 10.1103/PhysRevB.53.4393; pmid: 9983992
63. A. P. Mackenzie *et al*., Observation of quantum oscillations in the electrical resistivity of $SrRuO_3$. *Phys. Rev. B* **58**, 13318 (1998). doi: 10.1103/PhysRevB.58.R13318
64. O. Gunnarsson, M. Calandra, J. E. Han, Saturation of electrical resistivity. *Rev. Mod. Phys.* **75**, 1085–1099 (2003). doi: 10.1103/RevModPhys.75.1085
65. N. E. Hussey, J. Buhot, S. Licciardello, A tale of two metals: Contrasting criticalities in the pnictides and hole-doped cuprates. *Rep. Prog. Phys.* **81**, 052501 (2018). doi: 10.1088/1361-6633/aaa97c; pmid: 29353812
66. K. Damle, S. Sachdev, Nonzero-temperature transport near quantum critical points. *Phys. Rev. B* **56**, 8714–8733 (1997). doi: 10.1103/PhysRevB.56.8714
67. C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, E. Abrahams, A. E. Ruckenstein, Phenomenology of the normal state of Cu-O high-temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989). doi: 10.1103/PhysRevLett.63.1996; pmid: 10040734
68. Hv. Löhneysen *et al*., Non-Fermi-liquid behavior in a heavy-fermion alloy at a magnetic instability. *Phys. Rev. Lett.* **72**, 3262–3265 (1994). doi: 10.1103/PhysRevLett.72.3262; pmid: 10056148

69. B. Michon *et al*., Thermodynamic signatures of quantum criticality in cuprate superconductors. *Nature* **567**, 218–222 (2019). doi: 10.1038/s41586-019-0932-x; pmid: 30760922
70. J. G. Storey, J. L. Tallon, G. V. M. Williams, Pseudogap ground state in high-temperature superconductors. *Phys. Rev. B* **78**, 140506 (2008). doi: 10.1103/PhysRevB.78.140506
71. S.-D. Chen *et al*., Incoherent strange metal sharply bounded by a critical doping in Bi2212. *Science* **366**, 1099–1102 (2019). doi: 10.1126/science.aaw8850; pmid: 31780552
72. N. E. Hussey, H. Gordon-Moys, J. Kokalj, R. H. McKenzie, Generic strange-metal behaviour of overdoped cuprates. *J. Phys. Conf. Ser.* **449**, 012004 (2013). doi: 10.1088/1742-6596/449/1/012004
73. P. Phillips, C. Chamon, Breakdown of one-parameter scaling in quantum critical scenarios for high-temperature copper-oxide superconductors. *Phys. Rev. Lett.* **95**, 107002 (2005). doi: 10.1103/PhysRevLett.95.107002; pmid: 16196953
74. X.-G. Wen, Scaling theory of conserved current and universal amplitudes at anisotropic critical points. *Phys. Rev. B* **46**, 2655–2662 (1992). doi: 10.1103/PhysRevB.46.2655; pmid: 10003951
75. L. Prochaska *et al*., Singular charge fluctuations at a magnetic quantum critical point. *Science* **367**, 285–288 (2020). doi: 10.1126/science.aag1595; pmid: 31949077
76. H. J. A. Molegraaf, C. Presura, D. van Der Marel, P. H. Kes, M. Li, Superconductivity-induced transfer of in-plane spectral weight in $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Science* **295**, 2239–2241 (2002). doi: 10.1126/science.1069947; pmid: 11910103
77. S. L. Cooper *et al*., Growth of the optical conductivity in the Cu-O planes. *Phys. Rev. B* **41**, 11605–11608 (1990). doi: 10.1103/PhysRevB.41.11605; pmid: 9993591
78. M. A. Quijada *et al*., Anisotropy in the *ab*-plane optical properties of $Bi_2Sr_2CaCu_2O_8$ single-domain crystals. *Phys. Rev. B* **60**, 14917–14934 (1999). doi: 10.1103/PhysRevB.60.14917
79. S. A. Hartnoll, A. Lucas, S. Sachdev, *Holographic Quantum Matter* (MIT Press, 2018).
80. J. Loram, K. Mirza, J. Wade, J. Cooper, W. Liang, The electronic specific heat of cuprate superconductors. *Physica* **235C–240C**, 134–137 (1994). doi: 10.1016/0921-4534(94)91331-5
81. J. Zaanen, Planckian dissipation, minimal viscosity and the transport in cuprate strange metals. *SciPost Phys.* **6**, 61 (2019). doi: 10.21468/SciPostPhys.6.5.061
82. N. Nücker *et al*., Plasmons and interband transitions in $Bi_2Sr_2CaCu_2O_8$. *Phys. Rev. B* **39**, 12379–12382 (1989). doi: 10.1103/PhysRevB.39.12379; pmid: 9948098
83. N. Nücker, U. Eckern, J. Fink, P. Müller, Long-wavelength collective excitations of charge carriers in high-T_c superconductors. *Phys. Rev. B* **44**, 7155(R) (1991).
84. S. Vig *et al*., Measurement of the dynamic charge response of materials using low-energy, momentum-resolved electron energy-loss spectroscopy (M-EELS). *SciPost Phys.* **3**, 026 (2017).
85. M. Mitrano *et al*., Anomalous density fluctuations in a strange metal. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5392–5396 (2018). doi: 10.1073/pnas.1721495115; pmid: 29735712
86. A. A. Husain *et al*., Crossover of Charge Fluctuations across the Strange Metal Phase Diagram. *Phys. Rev. X* **9**, 041062 (2019). doi: 10.1103/PhysRevX.9.041062
87. R. L. Greene, P. R. Mandal, N. R. Poniatowski, T. Sarkar, The Strange Metal State of the Electron-Doped Cuprates. *Annu. Rev. Phys.* **11**, 213–229 (2020). doi: 10.1146/annurev-conmatphys-031119-050558
88. T. J. Reber *et al*., A unified form of low-energy nodal electronic interactions in hole-doped cuprate superconductors. *Nat. Commun.* **10**, 5737 (2019). doi: 10.1038/s41467-019-13497-4; pmid: 31844065
89. C. Charmousis, B. Goutéraux, B. Soo Kim, E. Kiritsis, R. Meyer, Effective holographic theories for low-temperature condensed matter systems. *J. High Energy Phys.* **2010**, 151 (2010). doi: 10.1007/JHEP11(2010)151
90. B. Goutéraux, E. Kiritsis, Quantum critical lines in holographic phases with (un)broken symmetry. *J. High Energy Phys.* **2013**, 53 (2013). doi: 10.1007/JHEP04(2013)053
91. T. Valla *et al*., Evidence for quantum critical behavior in the optimally doped cuprate $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Science* **285**, 2110–2113 (1999). doi: 10.1126/science.285.5436.2110; pmid: 10497125
92. A. Kaminski *et al*., Momentum anisotropy of the scattering rate in cuprate superconductors. *Phys. Rev. B* **71**, 014517 (2005). doi: 10.1103/PhysRevB.71.014517
93. J. M. Bok *et al*., Momentum dependence of the single-particle self-energy and fluctuation spectrum of slightly underdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$ from high-resolution laser angle-resolved photoemission. *Phys. Rev. B* **81**, 174516 (2010). doi: 10.1103/PhysRevB.81.174516
94. A. Damascelli, Z. Hussain, Z.-X. Shen, Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003). doi: 10.1103/RevModPhys.75.473
95. T. Yoshida *et al*., Low-energy electronic structure of the high-T_c cuprates $La_{2-x}Sr_xCuO_4$ studied by angle-resolved photoemission spectroscopy. *J. Phys.* **19**, 125209 (2007). doi: 10.1088/0953-8984/19/12/125209
96. N. Poniatowski, T. Sarkar, R. Lobo, S. Das Sarma, R. L. Greene, Counterexample to the conjectured Planckian bound on transport. *Phys. Rev. B* **104**, 235138 (2021). doi: 10.1103/PhysRevB.104.235138
97. S. Hartnoll, A. P. Mackenzie, Planckian Dissipation in Metals. arXiv 2107.07802 (2021).
98. Z. Nussinov, S. Chakrabarty, Exact Universal Chaos, Speed Limit, Acceleration, Planckian Transport Coefficient, "Collapse" to equilibrium, and Other Bounds in Thermal Quantum Systems. *Ann. Phys*. 10.1016/j.aop.2022.168970 (2022). doi: 10.1016/j.aop.2022.168970
99. C. T. Chen *et al*., Electronic states in $La_{2-x}Sr_xCuO_{4+\delta}$ probed by soft-x-ray absorption. *Phys. Rev. Lett.* **66**, 104–107 (1991). doi: 10.1103/PhysRevLett.66.104; pmid: 10043153
100. M. B. J. Meinders, H. Eskes, G. A. Sawatzky, Spectral-weight transfer: Breakdown of low-energy-scale sum rules in correlated systems. *Phys. Rev. B* **48**, 3916–3926 (1993). doi: 10.1103/PhysRevB.48.3916; pmid: 10008840
101. H. Eskes, A. M. Oleś, M. B. Meinders, W. Stephan, Spectral properties of the Hubbard bands. *Phys. Rev. B* **50**, 17980–18002 (1994). doi: 10.1103/PhysRevB.50.17980; pmid: 9976228
102. P. W. Phillips, Identifying the propagating charge modes in doped Mott insulators. *Rev. Mod. Phys.* **82**, 1719–1742 (2010). doi: 10.1103/RevModPhys.82.1719
103. A. J. Leggett, A "midinfrared" scenario for cuprate superconductivity. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8365–8372 (1999). doi: 10.1073/pnas.96.15.8365; pmid: 10411881
104. Y. S. Lee *et al*., Electrodynamics of the nodal metal state in weakly doped high- T_c cuprates. *Phys. Rev. B* **72**, 054529 (2005). doi: 10.1103/PhysRevB.72.054529
105. W. F. Brinkman, T. M. Rice, Application of Gutzwiller's Variational Method to the Metal-Insulator Transition. *Phys. Rev. B* **2**, 4302–4304 (1970). doi: 10.1103/PhysRevB.2.4302
106. B. J. Ramshaw *et al*., Quasiparticle mass enhancement approaching optimal doping in a high-T_c superconductor. *Science* **348**, 317–320 (2015). doi: 10.1126/science.aaa4990; pmid: 25814065
107. M. E. Simon, C. M. Varma, Detection and implications of a time-reversal breaking state in underdoped cuprates. *Phys. Rev. Lett.* **89**, 247003 (2002). doi: 10.1103/PhysRevLett.89.247003; pmid: 12484975
108. D. V. Else, T. Senthil, Strange Metals as Ersatz Fermi Liquids. *Phys. Rev. Lett.* **127**, 086601 (2021). doi: 10.1103/PhysRevLett.127.086601; pmid: 34477402
109. B. S. Shastry, P. Mai, Aspects of the normal state resistivity of cuprate superconductors. *Phys. Rev. B* **101**, 115121 (2020). doi: 10.1103/PhysRevB.101.115121
110. E. W. Huang, R. Sheppard, B. Moritz, T. P. Devereaux, Strange metallicity in the doped Hubbard model. *Science* **366**, 987–990 (2019). doi: 10.1126/science.aau7063; pmid: 31753997
111. J. Kokalj, Bad-metallic behavior of doped Mott insulators. *Phys. Rev. B* **95**, 041110 (2017). doi: 10.1103/PhysRevB.95.041110
112. P. T. Brown *et al*., Bad metallic transport in a cold atom Fermi-Hubbard system. *Science* **363**, 379–382 (2019). doi: 10.1126/science.aat4134; pmid: 30523078
113. A. Vranić *et al*., Charge transport in the Hubbard model at high temperatures: Triangular versus square lattice. *Phys. Rev. B* **102**, 115142 (2020). doi: 10.1103/PhysRevB.102.115142
114. J. Vučičević *et al*., Conductivity in the Square Lattice Hubbard Model at High Temperatures: Importance of Vertex Corrections. *Phys. Rev. Lett.* **123**, 036601 (2019). doi: 10.1103/PhysRevLett.123.036601; pmid: 31386456
115. P. Cha, N. Wentzell, O. Parcollet, A. Georges, E.-A. Kim, Linear resistivity and Sachdev-Ye-Kitaev (SYK) spin liquid behavior in a quantum critical metal with spin-1/2 fermions. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18341–18346 (2020). doi: 10.1073/pnas.2003179117; pmid: 32699148
116. X. Deng *et al*., How bad metals turn good: Spectroscopic signatures of resilient quasiparticles. *Phys. Rev. Lett.* **110**, 086401 (2013). doi: 10.1103/PhysRevLett.110.086401; pmid: 23473178
117. W. Wú, X. Wang, A. M. S. Tremblay, Non-Fermi liquid phase and linear-in-temperature scattering rate in overdoped two-dimensional Hubbard model. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115819119 (2022). doi: 10.1073/pnas.2115819119; pmid: 35320041
118. S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, Superconductivity and non-Fermi liquid behavior near a nematic quantum critical point. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4905–4910 (2017). doi: 10.1073/pnas.1620651114; pmid: 28439023
119. A. A. Patel, S. Sachdev, Theory of a Planckian Metal. *Phys. Rev. Lett.* **123**, 066601 (2019). doi: 10.1103/PhysRevLett.123.066601; pmid: 31491164
120. P. Cha, A. A. Patel, E. Gull, E.-A. Kim, Slope invariant T-linear resistivity from local self-energy. *Phys. Rev. Res.* **2**, 033434 (2020). doi: 10.1103/PhysRevResearch.2.033434
121. A. A. Patel, J. McGreevy, D. P. Arovas, S. Sachdev, Magnetotransport in a Model of a Disordered Strange Metal. *Phys. Rev. X* **8**, 021049 (2018). doi: 10.1103/PhysRevX.8.021049
122. T. Faulkner, N. Iqbal, H. Liu, J. McGreevy, D. Vegh, Strange metal transport realized by gauge/gravity duality. *Science* **329**, 1043–1047 (2010). doi: 10.1126/science.1189134; pmid: 20688983
123. G. T. Horowitz, J. E. Santos, D. Tong, Optical conductivity with holographic lattices. *J. High Energy Phys.* **2012**, 168 (2012). doi: 10.1007/JHEP07(2012)168
124. B. W. Langley, G. Vanacore, P. W. Phillips, Absence of power-law mid-infrared conductivity in gravitational crystals. *J. High Energy Phys.* **2015**, 163 (2015). doi: 10.1007/JHEP10(2015)163
125. A. Donos, J. P. Gauntlett, Holographic Q-lattices. *J. High Energy Phys.* **2014**, 40 (2014). doi: 10.1007/JHEP04(2014)040
126. E. Kiritsis, Y. Matsuo, Hyperscaling-violating Lifshitz hydrodynamics from black-holes: Part II. *J. High Energy Phys.* **2017**, 41 (2017). doi: 10.1007/JHEP03(2017)041
127. S. A. Hartnoll, A. Karch, Scaling theory of the cuprate strange metals. *Phys. Rev. B* **91**, 155126 (2015). doi: 10.1103/PhysRevB.91.155126
128. G. La Nave, K. Limtragool, P. W. Phillips, Fractional electromagnetism in quantum matter and high-energy physics. *Rev. Mod. Phys.* **91**, 021003 (2019). doi: 10.1103/RevModPhys.91.021003
129. K. Limtragool, P. W. Phillips, Anomalous dimension of the electrical current in strange metals from the fractional Aharonov-Bohm effect. *Europhys. Lett.* **121**, 27003 (2018). doi: 10.1209/0295-5075/121/27003
130. A. Karch, K. Limtragool, P. W. Phillips, Unparticles and anomalous dimensions in the cuprates. *J. High Energy Phys.* **2016**, 175 (2016). doi: 10.1007/JHEP03(2016)175
131. S. Sachdev, J. Ye, Gapless spin-fluid ground state in a random quantum Heisenberg magnet. *Phys. Rev. Lett.* **70**, 3339–3342 (1993). doi: 10.1103/PhysRevLett.70.3339; pmid: 10053843
132. A. Kitaev, "A Simple Model of Quantum Holography" [video] (2015); https://online.kitp.ucsb.edu/online/entangled15/kitaev/.
133. A. Kitaev, S. J. Suh, The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual. *J. High Energy Phys.* **2018**, 183 (2018). doi: 10.1007/JHEP05(2018)183
134. J. Polchinski, Effective Field Theory and the Fermi Surface [lectures]. arXiv 9210046v2 (1992).
135. S. Sachdev, Holographic metals and the fractionalized Fermi liquid. *Phys. Rev. Lett.* **105**, 151602 (2010). doi: 10.1103/PhysRevLett.105.151602; pmid: 21230891
136. J. Maldacena, D. Stanford, Remarks on the Sachdev-Ye-Kitaev model. *Phys. Rev. D* **94**, 106002 (2016). doi: 10.1103/PhysRevD.94.106002
137. S. Sachdev, *Quantum Phase Transitions* (Cambridge Univ. Press, ed. 2, 2011).
138. J. Maldacena, The Large-N Limit of Superconformal Field Theories and Supergravity. *Int. J. Theor. Phys.* **38**, 1113–1133 (1999). doi: 10.1023/A:1026654312961
139. E. Witten, Anti de Sitter space and holography. *Adv. Theor. Math. Phys.* **2**, 253–291 (1998). doi: 10.4310/ATMP.1998.v2.n2.a2
140. S. Gubser, I. Klebanov, A. Polyakov, Gauge theory correlators from non-critical string theory. *Phys. Lett. B* **428**, 105–114 (1998). doi: 10.1016/S0370-2693(98)00377-3
141. J. D. Bekenstein, Black Holes and Entropy. *Phys. Rev. D* **7**, 2333–2346 (1973). doi: 10.1103/PhysRevD.7.2333

142. S. W. Hawking, Breakdown of predictability in gravitational collapse. *Phys. Rev. D* **14**, 2460–2473 (1976). doi: 10.1103/PhysRevD.14.2460
143. S. W. Hawking, Particle creation by black holes. *Commun. Math. Phys.* **43**, 199–220 (1975). doi: 10.1007/BF02345020
144. S.-S. Lee, Non-Fermi liquid from a charged black hole: A critical Fermi ball. *Phys. Rev. D* **79**, 086006 (2009). doi: 10.1103/PhysRevD.79.086006
145. M. Cubrović, J. Zaanen, K. Schalm, String theory, quantum phase transitions, and the emergent Fermi liquid. *Science* **325**, 439–444 (2009). doi: 10.1126/science.1174962; pmid: 19556462
146. B. Goutéraux, Charge transport in holography with momentum dissipation. *J. High Energy Phys.* **2014**, 181 (2014). doi: 10.1007/JHEP04(2014)181
147. A. Karch, Conductivities for hyperscaling violating geometries. *J. High Energy Phys.* **2014**, 140 (2014). doi: 10.1007/JHEP06(2014)140
148. S. Cremonini, A. Hoover, L. Li, Backreacted DBI magnetotransport with momentum dissipation. *J. High Energy Phys.* **2017**, 133 (2017). doi: 10.1007/JHEP10(2017)133
149. E. Blauvelt, S. Cremonini, A. Hoover, L. Li, S. Waskie, Holographic model for the anomalous scalings of the cuprates. *Phys. Rev. D* **97**, 061901 (2018). doi: 10.1103/PhysRevD.97.061901
150. D. J. Gross, in *Les Houches 1975, Sesssion XXVIII: Methods in Field Theory* (North-Holland, 1976), pp. 181–250.
151. What is really required is that $[qA] = 1$, where q is the charge. In insisting that $[A] = 1$, we are setting $q = 1$, but still all of our statements refer to the product qA.
152. G. L. Nave, P. W. Phillips, Anomalous Dimensions for Boundary Conserved Currents in Holography via the Caffarelli–Silvestre Mechanism for p-forms. *Commun. Math. Phys.* **366**, 119–137 (2019). doi: 10.1007/s00220-019-03292-z
153. R. A. Davison, B. Goutéraux, Dissecting holographic conductivities. *J. High Energy Phys.* **2015**, 90 (2015). doi: 10.1007/JHEP09(2015)090
154. P. W. Phillips, B. W. Langley, J. A. Hutasoit, Un-Fermi liquids: Unparticles in strongly correlated electron matter. *Phys. Rev. B* **88**, 115129 (2013). doi: 10.1103/PhysRevB.88.115129
155. S. A. Hartnoll, J. Polchinski, E. Silverstein, D. Tong, Towards strange metallic holography. *J. High Energy Phys.* **2010**, 120 (2010). doi: 10.1007/JHEP04(2010)120
156. E. Kiritsis, F. Peña Benitez, Scaling of the holographic AC conductivity for non-Fermi liquids at criticality. *J. High Energy Phys.* **2015**, 177 (2015). doi: 10.1007/JHEP11(2015)177
157. C.-F. Chen, A. Lucas, Origin of the Drude peak and of of zero sound in probe brane holography. *Phys. Lett. B* **774**, 569–574 (2017). doi: 10.1016/j.physletb.2017.10.023
158. R. A. Davison, S. A. Gentle, B. Goutéraux, Slow Relaxation and Diffusion in Holographic Quantum Critical Phases. *Phys. Rev. Lett.* **123**, 141601 (2019). doi: 10.1103/PhysRevLett.123.141601; pmid: 31702178
159. M. Blake, A. Donos, Quantum critical transport and the Hall angle in holographic models. *Phys. Rev. Lett.* **114**, 021601 (2015). doi: 10.1103/PhysRevLett.114.021601; pmid: 25635540
160. E. Noether, Invariante Variationsprobleme. *Nachr. Ges. Wiss. Gottingen Math. Phys. Kl.* **1918**, 235–257 (1918).
161. P. W. Anderson, Infrared conductivity of cuprate metals: Detailed fit using Luttinger-liquid theory. *Phys. Rev. B* **55**, 11785–11788 (1997). doi: 10.1103/PhysRevB.55.11785
162. K. Limtragool, P. W. Phillips, Power-law optical conductivity from unparticles: Application to the cuprates. *Phys. Rev. B* **92**, 155128 (2015). doi: 10.1103/PhysRevB.92.155128
163. Y. Liao, Effects of unparticles on running of gauge couplings. *Eur. Phys. J. C* **55**, 483–488 (2008). doi: 10.1140/epjc/s10052-008-0601-y
164. B. Basa, G. La Nave, P. W. Phillips, Classification of nonlocal actions: Area versus volume entanglement entropy. *Phys. Rev. D* **101**, 106006 (2020). doi: 10.1103/PhysRevD.101.106006
165. M. Čulo *et al.*, Possible superconductivity from incoherent carriers in overdoped cuprates. *SciPost Phys.* **11**, 012 (2021). doi: doi:1468/SciPostPhys.11.1.012
166. C. Putzke *et al.*, Reduced Hall carrier density in the overdoped strange metal regime of cuprate superconductors. *Nat. Phys.* **17**, 826–831 (2021). doi: 10.1038/s41567-021-01197-0
167. I. Božović, X. He, J. Wu, A. T. Bollinger, Dependence of the critical temperature in overdoped copper oxides on superfluid density. *Nature* **536**, 309–311 (2016). doi: 10.1038/nature19061; pmid: 27535534
168. B. Pippard, An experimental and theoretical study of the relation between magnetic field and current in a superconductor. *Proc. R. Soc. London Ser. A* **216**, 547–568 (1953). doi: 10.1098/rspa.1953.0040

ACKNOWLEDGMENTS

Funding: Supported by the Center for Emergent Superconductivity, a DOE Energy Frontier Research Center, grant DE-SC0021238 (P.A. and P.W.P.) and by Netherlands Organisation for Scientific Research (NWO) ("Strange Metals" grant 16METL01), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant 835279-Catch-22) (N.E.H.). The work on anomalous dimensions for conserved currents was funded through NSF DMR-2111379.
Competing interests: The authors declare no competing interests.

Submitted 15 May 2021; accepted 25 May 2022
10.1126/science.abh4273

REPORT

CATALYSIS

Maximizing noble metal utilization in solid catalysts by control of nanoparticle location

Kang Cheng[1,2]†, Luc C. J. Smulders[1]†, Lars I. van der Wal[1], Jogchum Oenema[1], Johannes D. Meeldijk[1,3], Nienke L. Visser[1], Glenn Sunley[4], Tegan Roberts[4], Zhuoran Xu[5], Eric Doskocil[5], Hideto Yoshida[1,6], Yanping Zheng[2], Jovana Zečević[1], Petra E. de Jongh[1], Krijn P. de Jong[1]*

Maximizing the utilization of noble metals is crucial for applications such as catalysis. We found that the minimum loading of platinum for optimal performance in the hydroconversion of *n*-alkanes for industrially relevant bifunctional catalysts could be reduced by a factor of 10 or more through the rational arranging of functional sites at the nanoscale. Intentionally depositing traces of platinum nanoparticles on the alumina binder or the outer surface of zeolite crystals, instead of inside the zeolite crystals, enhanced isomer selectivity without compromising activity. Separation between platinum and zeolite acid sites preserved the metal and acid functions by limiting micropore blockage by metal clusters and enhancing access to metal sites. Reduced platinum nanoparticles were more active than platinum single atoms strongly bonded to the alumina binder.

Noble metals (NMs) are widely used in a variety of commercial and emerging technologies, including catalytic converters in automobiles; electrocatalysts in hydrogen fuel cells; and catalysts for petroleum, biomass, and waste conversion (*1*–*5*). Increasing demand for NMs in these applications is driving approaches that make more efficient use of NMs (*6*–*9*), including so-called single-atom catalysts (SACs) in which isolated single metal atoms or ions are stabilized by the support (*10*–*13*). However, strong metal-support interactions often lead to limited reducibility and in some cases low reactivity. Confinement of NMs inside zeolite channels or cages can benefit the adsorption of reagents and stabilization of reaction intermediates to improve catalytic activity, product selectivity, or both (*14*–*16*). Successful examples of confinement effects have been demonstrated for the conversion of small molecules, including carbon monoxide (CO) oxidation, methane (CH_4) oxidation, and the water-gas shift reaction (*17*–*19*).

In the petrochemical industry, platinum (Pt) is frequently used in combination with acidic zeolites for the hydroconversion of linear alkanes to enhance the quality of liquid fuels (fig. S1), and its performance is influenced by the Pt nanoparticle (NP) size and distribution, metal-support interactions, and acidic properties (*20*–*23*). To describe the required Pt loading sufficient to maintain the bifunctional metal-acid balance in hydroconversion catalysts, Guisnet and coworkers proposed a widely applied parameter of the surface Pt sites to Brønsted acid sites ratio (n_{pt}/n_a) (*24*). To meet this criterion, the typical Pt loading for bifunctional catalysts is in the range of 0.3 to 3 wt % (*25*–*27*). For integrating two functional components, Weisz's intimacy criterion, which is often interpreted as "the closer the better," has been applied to spatial organization (*28*). Recently, we have shown at Pt loadings around 0.5 wt % that separation of Pt and acid sites at the nanoscale enhanced the isomer selectivity in hydroisomerization of linear alkanes, whereas placing Pt inside zeolite crystals with closest proximity promoted acid-catalyzed cracking or overcracking reactions because of the high concentrations of intracrystalline carbenium ions (*29*–*31*). With this concept of nanoscale intimacy, we found that lowering the Pt loading holds potential (*30*), which inspired us to study the lower limit of NMs for catalysis as a function of the Pt NP location.

We controlled the Pt location and loading on industrially relevant platinum-zeolite-alumina (Al_2O_3) composite catalysts to decrease the Pt loading for hydroisomerization while maintaining optimal performance. We used *n*-heptane as a model molecule relevant to gasoline upgrading. We used the one-dimensional (1D) zeolites HZSM-22 and HMOR to construct bifunctional catalysts (tables S1 and S2 and fig. S2). To provide mechanical stability and to avoid the high-pressure drop in industrial fixed-bed reactors, zeolite catalysts need to be shaped into composites—for example, as extrudates with the use of a binder such as clay (*32*). The location of Pt sites in these composite catalysts can be roughly classified into three types: (i) on the outer surface of zeolite crystals; (ii) inside the zeolite crystals; and (iii) on the binder (fig. S3). We controlled the Pt location on the composite catalysts through either ion exchange (IE) by using $Pt(NH_3)_4(NO_3)_2$ as Pt precursor or ion adsorption (IA) by using H_2PtCl_6 (fig. S4). The actual weight loading of Pt was determined by means of inductively coupled plasma (ICP) spectrometry (tables S3 and S4). Because all raw materials are based on commercial products and the IE and IA methods are commonly applied in industry, no major barriers are foreseen for the catalyst preparation on a large scale.

High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) imaging combined with ultramicrotomy can expose cross sections of zeolite crystals or shaped catalysts and enables the visualization of the spatial distribution of Al_2O_3 binder, zeolite crystals, and Pt sites (*29*). The zeolite-Al_2O_3 composite, by use of acetic acid as a peptizing agent in this work, exhibited a uniform distribution of zeolite crystals, which were coated and stabilized by the Al_2O_3 binder (Fig. 1A). By contrast, large agglomerates of zeolites and Al_2O_3 were observed over zeolite+Al_2O_3 composite without acetic acid (fig. S5). Because the kinetic diameter of the $Pt(NH_3)_4^{2+}$ (0.48 nm) is comparable with the HZSM-22 pore size (0.45 by 0.55 nm) and HMOR pore size (0.67 by 0.70 nm), the diffusion of $Pt(NH_3)_4^{2+}$ ions into the micropores during the IE process could be limited. The STEM images of Pt-HZSM-22 slices confirmed that after reduction, most Pt NPs were present on the external crystal surfaces, regardless of the Pt loading, and a small fraction was located inside the HZSM-22 crystals (Fig. 1B). The average sizes of Pt NPs for 0.2Pt-HZSM-22 (where 0.2 is Pt loading in weight percent) and 1.0Pt-HZSM-22 were 1.5 and 1.7 nm, respectively. For 0.2Pt-HMOR and 1.0Pt-HMOR, the average diameters of Pt NPs were 1.2 and 1.6 nm, respectively, and the Pt NPs were exclusively deposited inside HMOR crystals (Fig. 1C). The Pt NPs were larger than the micropores of the zeolite and were likely accommodated by the local destruction of micropores during NP growth, an effect confirmed in previous studies (*33*).

The zeolite powders containing Pt NPs were then mixed with boehmite and acetic acid to prepare bifunctional catalysts with a zeolite-to-Al_2O_3 weight ratio of 1/1. The TEM imaging revealed that for IE-prepared Pt-HZSM-22/Al_2O_3, most Pt NPs remained on the HZSM-22 crystal surfaces with only a few NPs inside the

[1]Materials Chemistry and Catalysis, Debye Institute for Nanomaterials Science, Utrecht University, 3584 CG Utrecht, Netherlands. [2]State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China. [3]Electron Microscopy Centre, Utrecht University, 3584 CG Utrecht, Netherlands. [4]Applied Sciences, bp Innovation and Engineering, BP plc, Saltend, Hull HU12 8DS, UK. [5]Applied Sciences, bp Innovation and Engineering, BP plc, Naperville, IL 60563, USA. [6]The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan.
*Corresponding author. Email: k.p.dejong@uu.nl
†These authors contributed equally to this work.

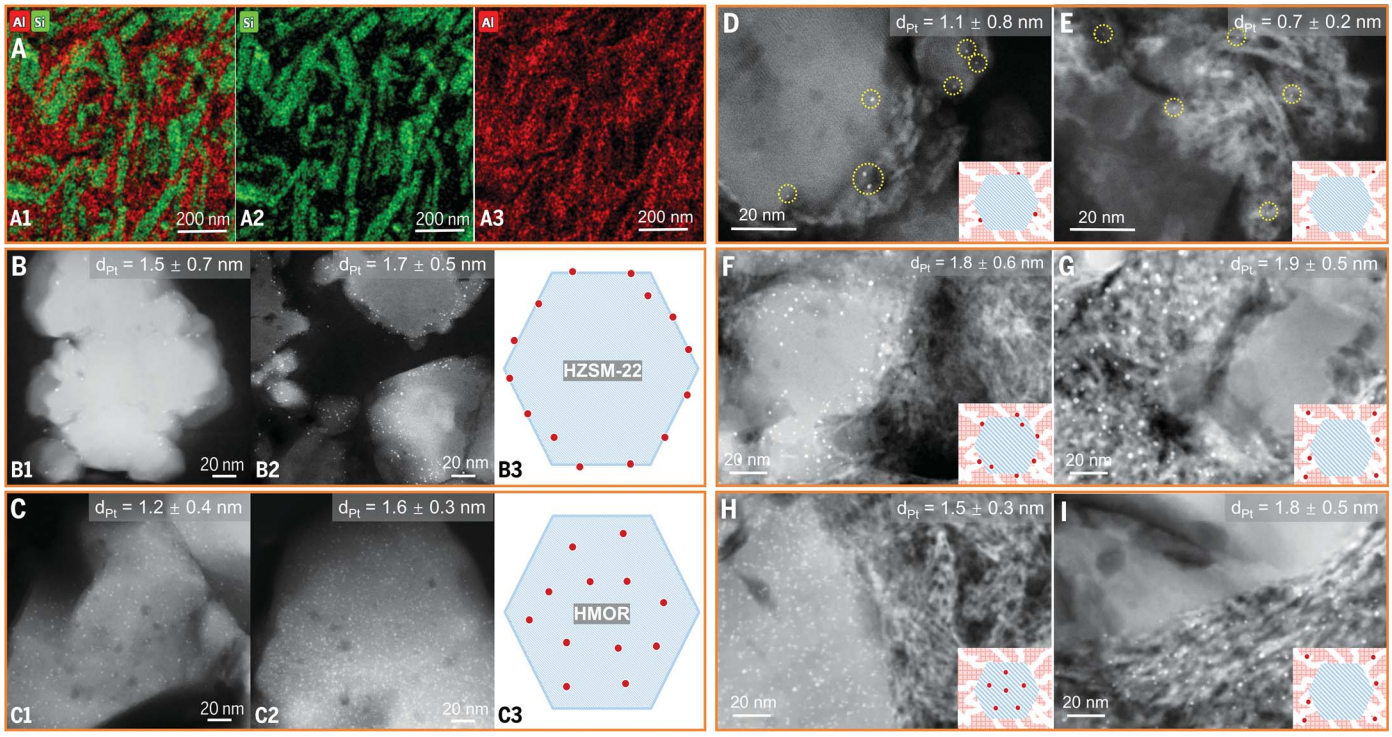


Fig. 1. HAADF-STEM and STEM–energy-dispersive x-ray (EDX) images of ultramicrotomed 70-nm-thick sections of catalysts. (**A**) HZSM-22/Al_2O_3 composite by using acetic acid as a peptizing agent. The images (A1) to (A3) show the location of silicon (green) and aluminum (red), which is indicative of the presence of the HZSM-22 and Al_2O_3 components. (**B**) (B1) 0.2Pt-HZSM-22, (B2) 1.0Pt-HZSM-22, and (B3) scheme for Pt distribution over HZSM-22. (**C**) (C1) 0.2Pt-HMOR, (C2) 1.0Pt-HMOR, and (C3) scheme for Pt distribution over HMOR. (**D** and **E**) 0.01Pt-HZSM-22/Al_2O_3 and 0.01Pt-Al_2O_3/HZSM-22, respectively. (**F** and **G**) 0.5Pt-HZSM-22/Al_2O_3 and 0.5Pt-Al_2O_3/HZSM-22, respectively. (**H** and **I**) 0.5Pt-HMOR/Al_2O_3 and 0.5Pt-Al_2O_3/HMOR, respectively. In the schematic representations, the red dots and blue hexagons indicate Pt NPs and zeolite crystals, respectively.

HZSM-22 crystals, and those NPs were near NPs present on HZSM-22 surfaces; no NPs were observed on the mesoporous Al_2O_3 binder (Fig. 1, D and F, and fig. S6A). Conversely, almost all of the Pt NPs were located on the Al_2O_3 binder for Pt-Al_2O_3/HZSM-22 prepared with the IA method (Fig. 1, E and G, and fig. S6B). Correspondingly, the Pt NPs were still exclusively present inside the HMOR crystals for IE-prepared Pt-HMOR/Al_2O_3 (Fig. 1H), and all of the Pt NPs were deposited on the Al_2O_3 binder for IA-prepared Pt-Al_2O_3/HMOR (Fig. 1I). With this methodology, we prepared a series of composite catalysts with different Pt loadings and locations (table S4). The Pt loading in all cases had a minor effect on the catalyst acidity (fig. S7).

The catalytic performance of HZSM-22–based catalysts in the hydroisomerization of *n*-C_7 was strongly affected by the Pt loading and distribution, especially at loading of 0.005 to 0.05 wt % (Fig. 2 and fig. S8). With an increase of Pt loading, the *n*-C_7 conversion compared at identical temperature and weight hourly space velocity (WHSV) increased first and then reached a plateau. Unexpectedly, with the Pt-on–HZSM-22 configuration, an ultralow Pt loading of 0.01 wt % was sufficient to maintain the *n*-C_7 conversion and selectivity of C_7 isomers (Fig. 2A). Such a low Pt loading is usually insufficient for catalyzing this reaction. However, 0.01Pt-Al_2O_3/HZSM-22 exhibited inferior activity and selectivity (Fig. 2B). The C_7 isomers consisted of monobranched and traces of dibranched C_7 isomers (mono-C_7 and di-C_7, respectively), and the main side products were propane, *n*-butane, and *iso*-butane from acid-catalyzed hydrocracking (*30*).

At similar conversion levels, 0.01Pt-HZSM-22/Al_2O_3 also provided higher selectivity of C_7 isomers than did 0.01Pt-Al_2O_3/HZSM-22 (fig. S9). Increasing the Pt loading >0.01 wt % decreased the selectivity of C_7 isomers, which was likely the result of more Pt NPs sited in the pore mouth region of HZSM-22, which would promote cracking reactions (*30*). For Pt-Al_2O_3/HZSM-22 with Pt-on-Al_2O_3, both the *n*-C_7 conversion and selectivity of C_7 isomers increased steadily and then reached a plateau as Pt loading increased to 0.05 wt % (Fig. 2B). With Pt-on-binder, therefore, a minimum loading of 0.05 wt % was required to sustain the *n*-heptane conversion and the selectivity of C_7 isomers, to ensure the equilibrium of the dehydrogenation-hydrogenation reactions.

Over HMOR-based catalysts, both the *n*-C_7 conversion and the selectivity of C_7 isomers increased with the Pt loading, and overall, the performance of catalysts with Pt inside HMOR crystals was more strongly influenced by the Pt loading (Fig. 2, C and D, and fig. S8). At the same Pt loading, the specific *n*-C_7 conversion over Pt-HMOR/Al_2O_3 was lower than the corresponding Pt-Al_2O_3/HMOR, probably because of pore blockage or poor accessibility of Pt NPs (*34*). With a 10-fold increase of Pt loading from 0.05 to 0.5%, the *n*-C_7 conversion increased from 58 to 71% with a selectivity of C_7 isomers ~80% over Pt-Al_2O_3/HMOR, whereas the *n*-C_7 conversion increased from 22 to 67% and the selectivity of C_7 isomers increased from 28 to 72% over Pt-HMOR/Al_2O_3. Regardless of the Pt loading and location over the composite catalysts, after more than 100 hours, we observed virtually no catalyst deactivation for both HZSM-22– and HMOR-based catalysts (figs. S10 and S11).

The dependence of the maximum yield of C_7 isomers, an important performance indicator in alkane hydroisomerization (*26*), on the platinum loading is summarized in Fig. 3. The acid density of HMOR is much higher than that of HZSM-22. Therefore, HMOR-based

catalysts required more Pt sites to obtain the same n_{Pt}/n_a ratio (fig. S12). A composite catalyst with Pt-on–HZSM-22 exhibited an optimum loading of only 0.01 wt % for maximizing the yield of C_7 isomers (≈64%) (Fig. 3A). For Pt-Al_2O_3/HZSM-22 with Pt-on-binder, the increase of Pt loading above 0.05 wt % maximized the yield of C_7 isomers toward 68%. Likewise, the maximum yield of C_7 isomers for Pt-Al_2O_3/HMOR went up with Pt loading, and relative to HZSM-22, a much higher loading of Pt (≈0.1 wt %) was needed to provide maximum yields of C_7 isomers (Fig. 3B). Disadvantageous for the yield of C_7 isomers was to confine the Pt inside HMOR crystals, which promoted the cracking reaction and prevented the conversion of n-C_7. The maximum yield of C_7 isomers over 0.05Pt-Al_2O_3/HMOR was near that over 0.5Pt-HMOR/Al_2O_3, suggesting that a rational placement of Pt could lead to a 10-fold reduction in its usage.

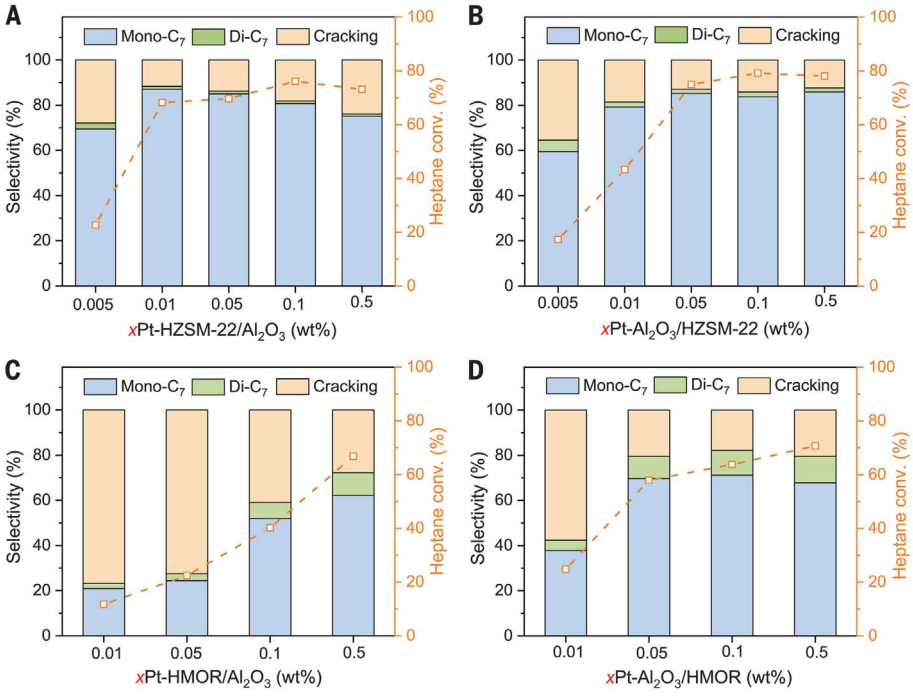


Fig. 2. Hydroconversion of *n*-heptane over bifunctional catalysts with different Pt loadings and locations. (**A**) Pt-HZSM-22/Al_2O_3 with Pt on the HZSM-22 crystals and (**B**) Pt-Al_2O_3/HZSM-22 with Pt on the Al_2O_3 binder. Data are partly from (*30*). (**C**) Pt-HMOR/Al_2O_3 with Pt inside the HMOR crystals and (**D**) Pt-Al_2O_3/HMOR with Pt on the Al_2O_3 binder. Data are partly from (*30*). Reaction conditions were pressure (P) = 10 bar, H_2/n-C_7 = 10/1, WHSV = 2.1 $g_{n\text{-}C7}$ g_{cat}^{-1} $hour^{-1}$, and temperature (T) = 320°C for HZSM-22–based catalysts and T = 270°C for HMOR-based catalysts.

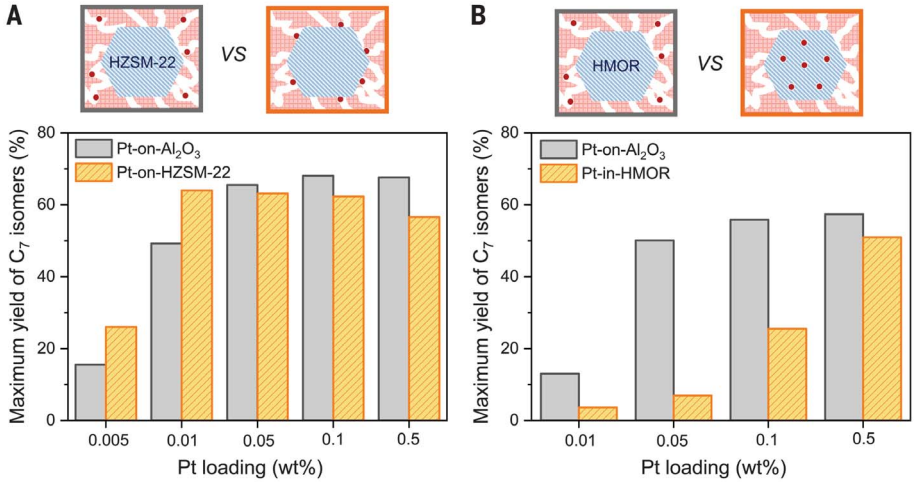


Fig. 3. The effect of Pt loading on the maximum yield of C_7 isomers. (**A**) Pt–HZSM-22–Al_2O_3 composite catalysts, with Pt NPs placed either on Al_2O_3 binder or on HZSM-22 crystals, and (**B**) Pt-HMOR-Al_2O_3, with Pt NPs placed either on Al_2O_3 binder or inside HMOR crystals. Reaction conditions were P = 10 bar, H_2/n-C_7 = 10/1, and WHSV = 2.1 $g_{n\text{-}C7}$ g_{cat}^{-1} $hour^{-1}$. The red dots and blue hexagons indicate Pt NPs and zeolite crystals, respectively.

In the low loading range (≤0.01 wt %), Pt placed on the Al_2O_3 binder was less effective than on HZSM-22 crystals (Figs. 2, A and B, and 3A). Highly dispersed Pt sites stabilized by the alumina (Pt–O–Al) could be difficult to reduce and not be sufficiently active for dehydrogenation of n-heptane. To investigate this, we intentionally prepared a SAC-Pt-Al_2O_3/HMOR catalyst with single-atom Pt dispersed on Al_2O_3, which was almost inactive in n-heptane hydroconversion (fig. S13). The x-ray absorption spectroscopy (XAS) shows the k^2-weighted (where k is photo-electron wave number), phase-corrected extended x-ray absorption fine structure (EXAFS) spectra of reduced Pt catalysts together with Pt standards (Fig. 4A). The scattering from the first shell of Pt–O led to the peak below 2.0 Å, whereas the scattering Pt–Pt in the Pt foil resulted in peaks at 2.4 and 2.9 Å (*7*). The scattering peak at 2.2 Å observed with a bond distance shorter than that of Pt–Pt may indicate the formation of Pt dimers (*35*).

Strong Pt–Pt scattering for 0.1Pt-HMOR and 0.1Pt-HZSM-22 with 0.1 wt % loadings could be observed (Fig. 4A), which is indicative of the formation of metallic Pt NPs or clusters. A Pt–O scattering peak was observed over 0.01Pt-HZSM-22/Al_2O_3 and 0.01Pt-Al_2O_3/HZSM-22, indicating that single-atom Pt (ions) was present within the two samples together with Pt NPs observed with electron microscopy (Fig. 1, D and E). More than half of Pt was metallic over 0.01Pt-HZSM-22/Al_2O_3, which exhibited higher reactivity in n-C_7 conversion than that of 0.01Pt-Al_2O_3/HZSM-22 that contained mainly oxidized Pt (fig. S14 and table S5). As the Pt loading on the Al_2O_3 binder increased from 0.01 to 0.05 wt %, the Pt–O scattering became negligible, and Pt–Pt scattering appeared, which suggested the formation of larger numbers of small Pt NPs or clusters in line with electron microscopy results (Fig. 1). The in situ x-ray photoelectron spectroscopy (Fig. 4B) further suggested reduced 0.1Pt-HZSM-22/Al_2O_3 and 0.1Pt-Al_2O_3/HZSM-22 to mainly contain metallic Pt (313.7 eV), whereas the atomically dispersed Pt on Al_2O_3 (SAC-0.2Pt-Al_2O_3) was in a partially oxidized state (314.6 eV) (*36*). The weaker Pt signals of 0.1Pt-HZSM-22/Al_2O_3 than those of 0.1Pt-Al_2O_3/HZSM-22 imply that a fraction of Pt NPs might be confined in the micropores of HZSM-22, causing cracking reactions at high (≥0.1 wt %) Pt loadings (Fig. 2A).

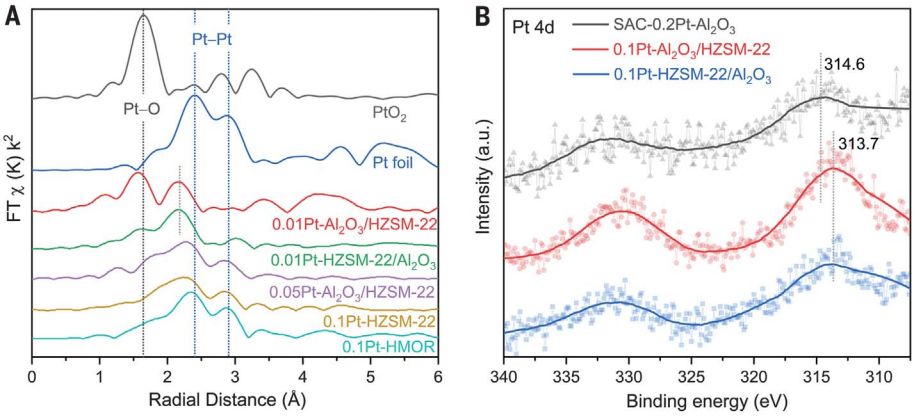


Fig. 4. Chemical state of Pt. (**A**) The k^2-weighted Pt L_3 edge EXAFS spectra in R space of bifunctional catalysts. Catalysts were reduced under 3.5 vol % H_2 flow at 350°C for 1 hour. PtO_2 and Pt foil were used as Pt standards. (**B**) In situ XPS spectra of reduced catalysts.

To increase the Pt utilization over bifunctional Pt-zeolite catalysts, we have varied the proximity between sites by controlling the spatial distribution of Pt NPs. Because of the low diffusivity of alkanes and alkenes, Pt sites should be placed in an accessible location in solid catalysts rather than in the micropores of a zeolite, such as on a mesoporous alumina binder or the outer surface of the zeolite crystals. Additionally, maintaining a separation between Pt and zeolite acid sites can preserve the metal and acid functions to the greatest extent through three main mechanisms: (i) limiting micropore blockage by metal clusters, (ii) enhancing accessibility of metal sites, and (iii) favoring pore mouth catalysis to reduce secondary cracking reactions (*29*). Furthermore, reducible Pt ensemble sites were confirmed to be more active as the dehydrogenation-hydrogenation function than Pt SAC, which was strongly bonded by the alumina binder, especially at low platinum loadings. Our findings generally agree with the classical criterion that a minimum amount of NMs is necessary to maintain the metal-acid balance and ensure a quasi-equilibration of the dehydrogenation-hydrogenation reactions, but this study confirms that the required loading for NMs is profoundly affected by how functional sites are arranged at the nanoscale.

REFERENCES AND NOTES

1. A. Wang, J. Li, T. Zhang, *Nat. Rev. Chem.* **2**, 65–81 (2018).
2. L. Chong *et al.*, *Science* **362**, 1276–1281 (2018).
3. R. T. Hannagan *et al.*, *Science* **372**, 1444–1447 (2021).
4. F. Zhang *et al.*, *Science* **370**, 437–441 (2020).
5. Z. Huang *et al.*, *Nat. Catal.* **3**, 170–178 (2020).
6. L. DeRita *et al.*, *Nat. Mater.* **18**, 746–751 (2019).
7. H. Jeong *et al.*, *Nat. Catal.* **3**, 368–375 (2020).
8. K. D. Rasmussen, H. Wenzel, C. Bangs, E. Petavratzi, G. Liu, *Environ. Sci. Technol.* **53**, 11541–11551 (2019).
9. A. E. Hughes, N. Haque, S. A. Northey, S. Giddey, *Resources* **10**, 93 (2021).
10. B. Qiao *et al.*, *Nat. Chem.* **3**, 634–641 (2011).
11. S. Mitchell, E. Vorobyeva, J. Pérez-Ramírez, *Angew. Chem. Int. Ed.* **57**, 15316–15329 (2018).
12. Z. Li *et al.*, *Chem. Rev.* **120**, 623–682 (2020).
13. H. Xiong, A. K. Datye, Y. Wang, *Adv. Mater.* **33**, e2004319 (2021).
14. J. D. Kistler *et al.*, *Angew. Chem. Int. Ed.* **53**, 8904–8907 (2014).
15. N. Kosinov, C. Liu, E. J. M. Hensen, E. A. Pidko, *Chem. Mater.* **30**, 3177–3198 (2018).
16. L. Liu *et al.*, *Nat. Mater.* **18**, 866–873 (2019).
17. S. M. Wu, X. Y. Yang, C. Janiak, *Angew. Chem. Int. Ed.* **58**, 12340–12354 (2019).
18. J. Zhang *et al.*, *Nat. Catal.* **1**, 540–546 (2018).
19. Z. Jin *et al.*, *Science* **367**, 193–197 (2020).
20. L. Liu, A. Corma, *Chem. Rev.* **118**, 4981–5079 (2018).
21. B. Zhang *et al.*, *J. Am. Chem. Soc.* **141**, 8185–8197 (2019).
22. T. W. van Deelen, C. Hernández Mejía, K. P. de Jong, *Nat. Catal.* **2**, 955–970 (2019).
23. L. I. van der Wal *et al.*, *ACS Catal.* **11**, 3842–3855 (2021).
24. G. E. Giannetto, G. R. Perot, M. R. Guisnet, *Ind. Eng. Chem. Prod. Res. Dev.* **25**, 481–490 (1986).
25. A. C. M. van den Broek, J. van Grondelle, R. A. van Santen, *J. Catal.* **167**, 417–424 (1997).
26. J. A. Martens *et al.*, *ChemSusChem* **6**, 421–425 (2013).
27. Z. Guo *et al.*, *Angew. Chem. Int. Ed.* **59**, 1548–1551 (2020).
28. P. B. Weisz, E. W. Swegler, *Science* **126**, 31–32 (1957).
29. J. Zečević, G. Vanbutsele, K. P. de Jong, J. A. Martens, *Nature* **528**, 245–248 (2015).
30. K. Cheng *et al.*, *Angew. Chem. Int. Ed.* **59**, 3592–3600 (2020).
31. J. Oenema *et al.*, *ACS Catal.* **10**, 14245–14257 (2020).
32. N. L. Michels, S. Mitchell, J. Pérez-Ramírez, *ACS Catal.* **4**, 2409–2417 (2014).
33. J. Zečević, A. M. J. van der Eerden, H. Friedrich, P. E. de Jongh, K. P. de Jong, *ACS Nano* **7**, 3698–3705 (2013).
34. M. Guisnet, F. Alvarez, G. Giannetto, G. Perot, *Catal. Today* **1**, 415–433 (1987).
35. U. Müller, K. Sattler, J. Xhie, N. Venkateswaran, G. Raina, *Z. Phys. D At. Mol. Clust.* **19**, 319–321 (1991).
36. T. Huizinga, H. F. J. van 't Blik, J. C. Vis, R. Prins, *Surf. Sci.* **135**, 580–596 (1983).

ACKNOWLEDGMENTS

N. Nikolopoulos, G. Totarella, and S. Zanoni are acknowledged for measuring and analyzing the N_2 physisorption data. **Funding:** This work has been supported by BP plc and Topsector Chemie consortia for Knowledge and Innovation (TKIs). K.P.d.J. acknowledges funding from the European Research Council (ERC), EU FP7 ERC Advanced Grant 338846. K.C. thanks the support from the National Key Research and Development Program of Ministry of Science and Technology 2020YFB0606401. H.Y. thanks the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers by JSPS. **Author contributions:** K.C. conducted the catalyst preparation; catalytic evaluation; sample characterizations by NH_3-TPD, SEM, and XRF; and drafted the manuscript. L.C.J.S. assessed the catalyst stability, performed additional characterizations, and helped revise the manuscript. L.I.v.d.W., J.D.M., N.L.V., and H.Y. performed STEM imaging. J.O. guided the catalyst preparation and maintained the reaction unit. G.S., T.R., P.E.d.J., and J.Z. contributed to the discussion of all the results. Z.X. and E.D. performed the XAS measurements and analyzed the data. Y.Z. performed XPS analysis. K.P.d.J. initiated and supervised the project and manuscript writing. All authors approved the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and images are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn8289
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S5
References (*37*–*40*)

Submitted 11 January 2022; resubmitted 24 March 2022
Accepted 18 May 2022
10.1126/science.abn8289

THERMOELECTRICS

High figure-of-merit and power generation in high-entropy GeTe-based thermoelectrics

Binbin Jiang[1,2]†, Wu Wang[1]†, Shixuan Liu[1], Yan Wang[1], Chaofan Wang[1], Yani Chen[1], Lin Xie[1], Mingyuan Huang[1], Jiaqing He[1,3]*

The high-entropy concept provides extended, optimized space of a composition, resulting in unusual transport phenomena and excellent thermoelectric performance. By tuning electron and phonon localization, we enhanced the figure-of-merit value to 2.7 at 750 kelvin in germanium telluride–based high-entropy materials and realized a high experimental conversion efficiency of 13.3% at a temperature difference of 506 kelvin with the fabricated segmented module. By increasing the entropy, the increased crystal symmetry delocalized the distribution of electrons in the distorted rhombohedral structure, resulting in band convergence and improved electrical properties. By contrast, the localized phonons from the entropy-induced disorder dampened the propagation of transverse phonons, which was the origin of the increased anharmonicity and largely depressed lattice thermal conductivity. We provide a paradigm for tuning electron and phonon localization by entropy manipulation, but we have also demonstrated a route for improving the performance of high-entropy thermoelectric materials.

Energy waste during the conversion and operation processes is a serious problem globally and has attracted increasing attention (*1*–*4*). Thermoelectric (TE) technology provides the prospect for harvesting waste heat from scattered heat sources, which suffer from low conversion efficiency (*5*–*8*). Conversion efficiency is related primarily to the dimensionless figure-of-merit [$zT = S^2\sigma T/(\kappa_e + \kappa_L)$] of thermoelectric materials, where S, σ, T, κ_e, and κ_L denote the Seebeck coefficient, electrical conductivity, absolute temperature, carrier, and lattice thermal conductivities, respectively (*9*–*12*). To pursue high TE performance, the electrical and thermal transport properties should be simultaneously decoupled and optimized. Generally, successful strategies for improving TE performance require high electrical transport properties from the ordered band structure and low lattice thermal conductivity from the disordered atomic arrangement, following the phonon-glass electron-crystal concept (*13*–*15*). The recently proposed high-entropy strategy provides an extended composition range, in which a stabilized crystalline structure and disordered atomic arrangement can be simultaneously realized (*4*, *16*–*19*). Possible Anderson localization of the electrons in the disordered system, however, may deteriorate the electrical transport properties (*6*). Furthermore, the increased crystal symmetry of low-symmetry materials due to the entropy-driven structural stabilization also may strengthen phonon propagation. Therefore, the challenge is to avoid possible electron localization while achieving the localized distribution of phonons in high-entropy materials.

GeTe belongs to the IV-VI chalcogenides and stabilizes as a rhombohedral structure at room temperature by distorting high-temperature cubic symmetry (*20*–*22*). IV-VI chalcogenides with high TE performance typically exhibit octahedral-like coordination, described as a cationic atom surrounded by six anionic atoms. In these materials, however, each atom contains only three valence electrons on average because of the small *sp*-hybridization resulting from the large energy difference between the *s*-orbitals and the *p*-orbitals (*23*, *24*). The sharing of three valence electrons to form six bonds results in delocalized electrons, unlike localized electron distribution in covalent bonding, which has been called metavalent or resonant bonding (*25*, *26*). We demonstrated that delocalized electrons from increased crystal symmetry and localized phonons from entropy-induced disorder could coexist in high-entropy GeTe-based materials, resulting in a zT value of 2.7 at 750 K (Fig. 1A) (*11*, *20*–*22*, *27*–*33*). The atomic electric field and charge density of the GeTe-based materials, which were determined by differential phase contrast (DPC) imaging in scanning transmission electron microscopy (STEM), clearly showed the increased delocalization of electrons from decreasing rhombohedral distortion, which was responsible for improved band convergence and power factor. Moreover, the localized phonons from entropy-induced disorder largely damped the propagation of transverse phonons and increased the anharmonicity, resulting in ultralow lattice thermal conductivity. The high performance of the GeTe-based materials improved the conversion efficiencies of the fabricated single and segmented TE modules to 10.5 and 13.3% at a temperature difference of 506 K (Fig. 1B), respectively. These experimental values were among the highest in the entire TE community (*4*, *34*–*45*), indicating an important step toward TE applications.

Solubility in alloys can be largely increased by increasing the elemental species based on entropy-driven structural stabilization, which has been summarized as high-entropy materials (*4*, *16*). Because of the well-known disadvantage of high lattice thermal conductivity in GeTe, this high-entropy phenomenon can be used to lower its lattice thermal conductivity by increasing the phonon scattering. We used Ag, Sb, and Pb as the alloying elements because of their high limited solubility. We found obvious impurity peaks in the x-ray powder diffraction (XRD) patterns for $Ge_{0.88}Pb_{0.12}Te$, $Ge_{0.89}Ag_{0.11}Te$ and $Ge_{0.87}Sb_{0.13}Te$ samples when the number of components was three, and these peaks were indexed to PbTe, Ag_2Te, and Sb_2Te_3, respectively (Fig. 2A). When the number of components increased to five, the second phases were eliminated in all of the samples, indicating the stabilization phenomenon of the phase structure dominated by the increasing entropy. The single phase was stable even when we further introduced Cd, Mn, and Sn, which formed a more complicated composition with six components. These results proved that the high-entropy phenomenon could be used to obtain single-phase compositions with stable structures, offering the possibility of increasing the solubility limit.

The ternary chalcogenide Ge-Sb-Te is a typical phase-change material with an amorphous structure (*23*–*25*), which typically exhibits substantially deteriorated electrical transport properties because of the damaged periodicity of its atomic arrangement. To further differentiate the crystalline high-entropy composition from the amorphous form and to confirm the sites of the alloyed elements, we conducted energy-dispersive spectroscopy (EDS) analysis at an atomic scale using STEM. We performed EDS mapping of $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ and other samples (Fig. 2, B to G, and fig. S1). The atomic arrangements of all five elements showed good periodicity, indicating the intuitive appearance of the crystalline high-entropy composition. We found that the Ge, Ag, Sb, and Pb atoms were located in the same site, whereas the Te atom was located solely in another site, demonstrating that possible antisite defects were absent even if Sb and Te had a similar nature (*46*). The homogeneous distribution of all five elements further confirmed the single phase (fig. S2), illustrating that high entropy should provide the driving force for stabilizing different elements, with content that exceeded the solubility limit of one structure.

[1]Shenzhen Key Laboratory of Thermoelectric Materials, Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China. [2]School of Materials and Energy, University of Electronic Science and Technology of China, Chengdu 610054, China. [3]Guangdong-Hong Kong-Macao Joint Laboratory for Photonic-Thermal-Electrical Energy Materials and Devices, Southern University of Science and Technology, Shenzhen 518055, China.
*Corresponding author. Email: hejq@sustech.edu.cn
†These authors contributed equally to this work.

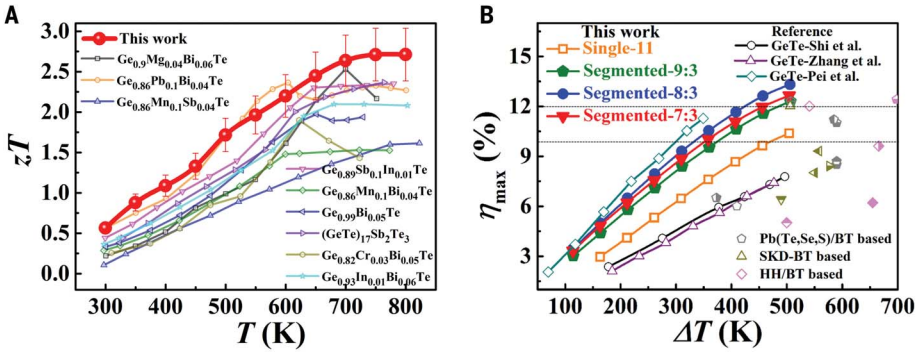


Fig. 1. High physical properties of the GeTe-based TE materials and modules. (**A**) Temperature dependence of *zT* values for the high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ sample in this work, for which the reported results in the literature (*11*, *20*–*22*, *27*–*33*) are also included for comparison. (**B**) Maximum conversion efficiency (η_{max}) as a function of ΔT for the fabricated GeTe-based TE modules in this work. The reported results of the GeTe-based, Pb(Te,Se,S)/BT–based, SKD/BT–based, and HH/BT–based materials from the literature (*4*, *34*–*45*) are also included for comparison. BT, SKD, and HH represent Bi_2Te_3, Skutterudite and half-Heusler materials, respectively.

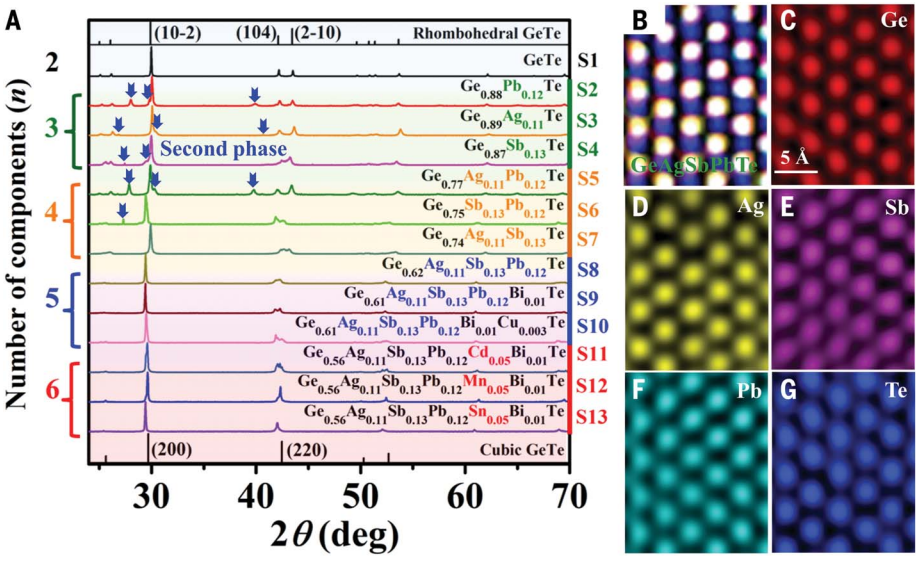


Fig. 2. Formation of the high-entropy GeTe-based materials. (**A**) XRD patterns of GeTe-based materials with different numbers of components. Blue arrows indicate the impurity peaks of the second phases. The nominal compositions of the GeTe-based materials labeled with S1 to S13 are shown by their chemical formulas. (**B** to **G**) Overall EDS mapping image and individual Ge (C), Ag (D), Sb (E), Pb (F), and Te (G) EDS mapping images of the high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ sample.

In our previous work, we proved that a high-entropy composition should have ultralow lattice thermal conductivity due to the increased lattice strains while maintaining the electrical transport path by stabilizing the crystal structure (*4*). This was also the original motive for introducing different elements into GeTe to form the high-entropy materials. We measured the TE properties of a series of samples that were formed by increasing the elemental species, the nominal compositions of which are listed in detail in Fig. 2A (labeled as S1 to S13), and determined the electrical conductivity and Seebeck coefficient (fig. S4). The electrical conductivity decreased with increasing elemental species, whereas the Seebeck coefficient followed the opposite trend, the obvious result of the decreased carrier concentration. The calculated power factor ($PF = \sigma S^2$) at a low temperature range for the high-entropy samples was much larger than that for the low-entropy samples with multiple phases and pure GeTe (Fig. 3A). This was unusual and could not be completely explained by the eliminated electron scattering around the phase boundary. As expected, the lattice thermal conductivity κ_L was largely depressed to an ultralow value (~0.3 $Wm^{-1}K^{-1}$) throughout the temperature range, indicating a temperature-independent trend (Fig. 3B). This phenomenon demonstrated that point defect scattering and the Umklapp scattering for the phonons were very strong in this material system (*47*). Because of the largely improved *PF* and depressed κ_L values at a low temperature range, the *zT* values improved to 0.6 at room temperature and to 2.7 at 750 K for the $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ sample (Fig. 3C). We calculated the average *zT* in the 300 to 750 K range, which showed a highly comparable value >1.7 for the $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ sample because of the largely improved TE performance at the low temperature range (Fig. 3D). The high-performance high-entropy composition also shows a good reproductivity and thermal stability (figs. S6 and S7).

Because of the largely improved *zT* values in the high-entropy GeTe-based materials, we fabricated a TE module to realize a high experimental conversion efficiency. We fabricated four TE modules, which were labeled as *11*, *7:3*, *8:3*, and *9:3*, respectively (Fig. 3F). The module labeled as *11* was a single module with a leg length of 11 mm, which consisted of p-type $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ and n-type $Pb_{0.997}In_{0.003}Te_{0.996}I_{0.004}$. The module labeled as *7:3* was a segmented module with a high-temperature leg length of 7 mm and a low-temperature leg length of 3 mm. The high-temperature legs consisted of p-type $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ and n-type $Pb_{0.997}In_{0.003}Te_{0.996}I_{0.004}$, and the low-temperature legs consisted of commercial p-type $Bi_{0.5}Sb_{1.5}Te_3$ and n-type $Bi_2Te_{2.7}Se_{0.3}$. The construction of the modules labeled *8:3* and *9:3* was understood to be analogous to the ones labeled *7:3*. We also measured the output properties by the commercial Power Generation Efficiency Characteristics Evaluation System (PEM-2, Advance Riko, Japan) (Fig. 3, E and F, and figs. S8 to S12).

We found that the open-circuit voltage was largely improved by using the Bi_2Te_3-based materials as low-temperature legs (Fig. 3E), which exhibited a positive effect on the segmented module. Because of the improved open-circuit voltage, which was related to the Seebeck coefficient, the output power (*P*) calculated by $P = IU$ was also improved (Fig. 3E). The output power reached a maximum value of 3.5 W for the *7:3* module with a hot-side temperature of 800 K. Subsequently, the maximum value of the output power decreased with increasing leg length because of the increased internal resistance of the module. The increased internal resistance of the module was illustrated by the increased slope of the current-dependent voltage curve (Fig. 3E). Although the increased leg length exhibited decreased output power, the heat flow (*Q*) also decreased. Then the conversion efficiency (η) would have an optimized value with increasing

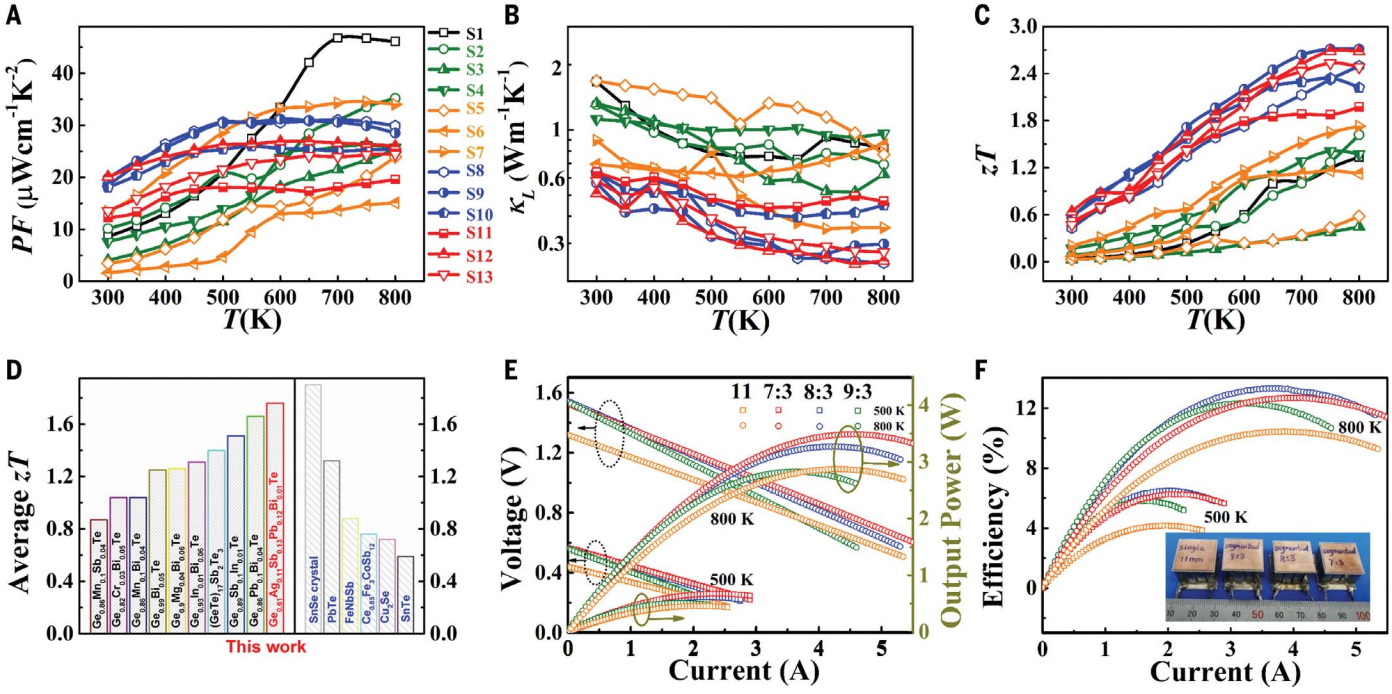


Fig. 3. TE performance of the GeTe-based materials and modules. (**A** to **C**) Temperature dependences of the power factor *PF* (A), lattice thermal conductivity κ_L (B), and *zT* (C) values of all samples in this work. (**D**) Comparison of the average *zT* value in this work with the reported results for GeTe, SnSe, PbTe, FeNbSb, $Ce_{0.85}Fe_3CoSb_{12}$, Cu_2Se, and SnTe in the literature (*5*, *11*, *13*, *17*, *20*–*22*, *29*–*35*, *41*, *61*, *62*). (**E** and **F**) Current dependences of the output voltage and output power (E) and conversion efficiencies (F) for the fabricated *11*, *7:3*, *8:3*, and *9:3* modules.

leg length due to the relation of $\eta = P/(P + Q)$. The conversion efficiency reached a maximum value of 13.3% for the segmented *8:3* module at a hot-side temperature of 800 K, which was much higher than 10.5% for the single *11* module (Fig. 3F). Compared with the high conversion efficiency in the literature (tables S2 and S3), our GeTe-based module realized a higher value than the others (PbTe, GeTe, Skutterudite, and half-Heusler). The origin of this high conversion efficiency should be from the optimized *zT* over the whole temperature range.

Strong alloying scattering and Anderson localization will always occur for electrons in disordered materials, typically resulting in the deteriorated electrical transport properties (*6*). In our experiment, we introduced Ag, Sb, and Pb atoms into the Ge sites to maintain the ordered Te sublattice. Because the contribution of Te-5p orbitals to the valence maximum (Σ, η, *Z*, and *L* bands) is much higher than that of Ge-4p orbitals (fig. S13), Anderson localization from disordered Ge sites would have less effect on the electrical transport properties of p-type GeTe. The weakened scattering around the phase boundary and Anderson localization of the electrons were the possible reasons for the well-maintained *PF* in the high-entropy materials; however, the improved *PF* with increasing entropy (Fig. 3A) was surprising. The electrical transport properties were closely related to the localization of electrons, and were visualized by charge density distribution. The atomic structures of the GeTe and $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ samples along the [110] zone axis were revealed by high-angle annular dark-field (HAADF) imaging in STEM (Fig. 4, A and B). With the *Z*-contrast nature of HAADF imaging (where *Z* indicates the atomic number), the brighter spots in the HAADF images of both samples indicated Te sites and thelower-intensity spots indicated Ge sites. The Ge sites incorporated with alloying elements in high-entropy sample exhibited higher intensity compared with the same pristine sites. We also found that the Ge atoms deviated from the geometric center of the rhombus formed by Te atoms in the GeTe sample, resulting in a polarized electric field and the origin of ferroelectricity (*48*). The deviation of Ge atoms from the geometric center is defined as a variable value of δ (Fig. 4C). The δ value decreased with increasing elemental species (Fig. 4F), as calculated from atomic-resolution HAADF images of the four GeTe-based samples (fig. S14). This was consistent with the XRD results (Fig. 2A) where the (2-10) and (104) peaks in the rhombohedral structure gradually converged into one (220) peak in the cubic structure. The increased crystal symmetry from the rhombohedral to the cubic structure by the formation of high-entropy materials was the origin of the weakened ferroelectricity (*48*).

The tuned distortions quantified by the variable δ in the GeTe-based materials potentially caused redistribution and changes to electron localization. We further applied DPC-STEM imaging to obtain the atomic electric field and charge density of the GeTe-based materials (*49*). The electron probe deflected by the electric field of the specimen caused a shift in the center of mass of the diffraction pattern, which was determined by the segmented detector in DPC-STEM (fig. S15) and used to calculate the projected electric field with Ehrenfest's theorem (*50*, *51*). We observed coupling between the electric field of the Ge and Te sites in the pristine sample (Fig. 4D), although the coupling effect of the electric field in these two sites was largely reduced in the high-entropy sample (Fig. 4E). The projected electric field images of the GeTe-based samples were further converted to projected charge density maps according to Gauss's law, as shown in Fig. 4, G and H. In the octahedron formed by one Ge atom and six Te atoms, we observed three shorter bonds and three longer bonds (Fig. 4J). The electrons were highly localized and formed shorter bonds because of symmetry breaking in the rhombohedral structure with a distorted lattice. The fair features in the gaps of the coupled Ge and Te atoms (shorter bonds) in the charge density map of the GeTe sample (Fig. 4G) provided an indicator for the existence of localized electrons. By forming high-entropy materials with more element species, the charge density in the gaps between the Ge sites and

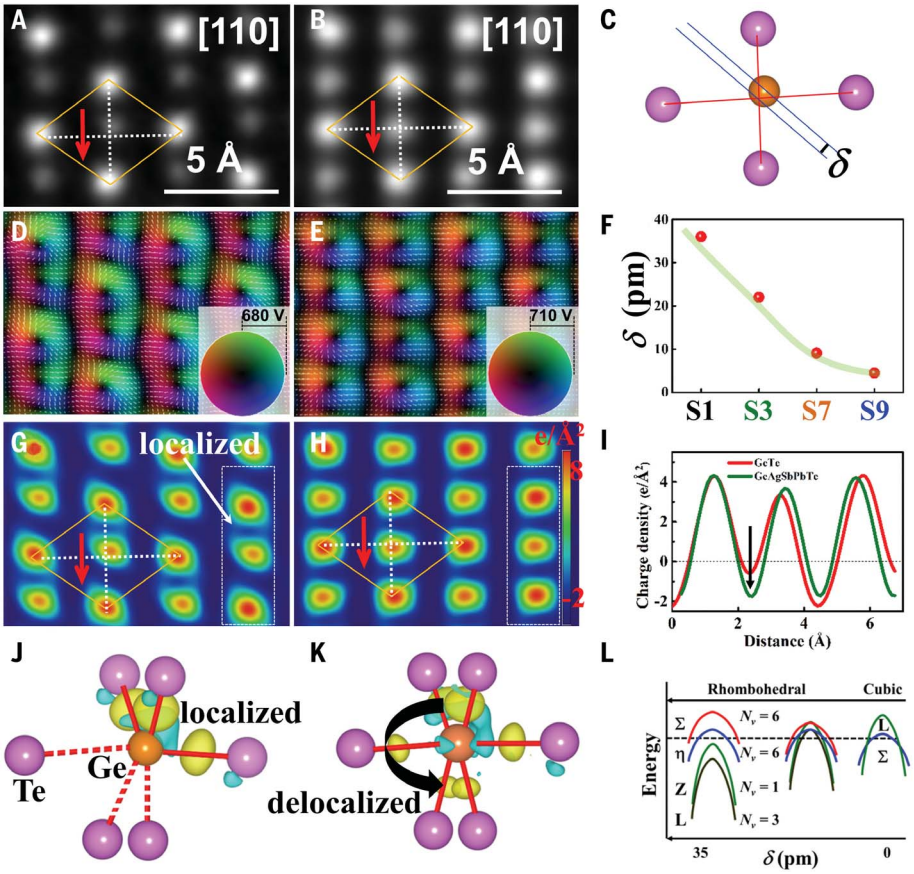


Fig. 4. Delocalizing the distribution of electrons. (**A** and **B**) STEM-HAADF images of the GeTe (A) and high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ (B) samples. The orange rhombus indicates the Te atoms and the red arrow represents the deviation direction of the Ge atom. (**C**) Sketch map of the deviation value (δ) from the geometry center. (**D** and **E**) Electrical field distributions of the the GeTe (D) and high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ (E) samples. (**F**) Changed deviation values (δ) with increasing entropy. (**G** and **H**) Charge distribution of the GeTe (G) and high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ (H) samples. (**I**) Distance dependence of the charge density of the white-boxed region in (G) and (H). (**J** and **K**) Schematic diagrams of the localized (J) and delocalized (K) distribution of electrons calculated by density functional theory calculations, The red solid line indicate short bonds and the red dotted lines indicate long bonds. (**L**) Schematic diagram of band convergence by tuning the deviation value δ.

surrounding Te sites was quite similar (Fig. 4H), demonstrating a delocalized distribution of electrons. With the incorporation of alloying elements (Fig. 4K), localized electrons could transfer from the shorter bonds to the longer bonds, resulting in delocalized electron distribution and similar bond features. This phenomenon was also observed by comparing the charge densities between the Ge and Te sites (Fig. 4I). We also calculated the electron localization function to show the distribution of electrons (fig. S16), which demonstrated the decreased value (delocalization) by forming high-entropy composition. Because of the localized electrons around the shorter bonds, there was a large difference in electron distributions between the shorter and longer bonds, resulting in a split valence maximum. The well-distributed six bonds with similar features enable the split valence maximums to converge, which should be the origin of the high *PF* over a wide temperature range. The delocalized electrons decrease the difference of p-orbital bonding in different orientations, resulting in a small band difference between the light and heavy bands (*23*, *25*). We calculated the band structures of low-entropy GeTe and high-entropy $Ge_5AgSbPbTe_8$ (fig. S17), which illustrated the increased equivalent degenerated valleys of the band structure (N_v) by delocalizing electrons. The increased N_v will increase the effective mass of carriers by the term of $m^* = (N_v)^{2/3} m_b^*$ and contribute to a high Seebeck coefficient (*20*). The change in the band structure was also expressed by our schematic diagram (Fig. 4L) and explained as slight symmetry breaking by Pei *et al.* (*20*).

Overall, the increased crystal symmetry will increase lattice thermal conductivity because of the improved periodicity in undistorted atomic arrangements (*52*). For example, the lattice thermal conductivity of rhombohedral GeTe decreased with increasing temperature due to strengthened Umklapp scattering, whereas cubic GeTe exhibited an increased value after the phase transition (Fig. 3B, black line) (*20*, *47*). The average long-range crystal symmetry was increased with increasing entropy, as shown by the decreased deviation value of δ (Fig. 4F). The disordered distribution of different elemental species in one atomic site caused fluctuations in the atomic environment and the breaking translational symmetry at an atomic scale. This indicated that many different octahedrons were composed of the six Te atoms, which were centered on the Ge, Ag, Sb, or Pb atoms. Because of fluctuations in the atomic environment, the lattice also showed changed parameters in different regions, resulting in large lattice strains. We used geometric phase analysis based on the high-resolution STEM-HAADF images (fig. S18, A and B) to measure the lattice strains of the high-entropy GeTe-based materials (Fig. 5, A and B, and fig. S18C). The spatial distribution of lattice strains in the high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ sample showed typical gathered regions on the nanoscale, whereas the low-entropy GeTe sample showed small fluctuations from normal background noise. We counted the statistical distribution of lattice strains to quantify the effect of high entropy (main plots in Fig. 5, A and B, and fig. S18C), which indicated a broadened statistical distribution and larger root mean square σ in the high-entropy sample compared with the low-entropy sample.

The large lattice strains demonstrated that the symmetry of the crystal structure broke down, which changed the distribution and propagation of the phonons. We used the Raman spectra to measure the changes in optical phonons (Fig. 5C), which were the direct expression of the changed atomic environment and force constant (*53*). The formation of high-entropy materials was expected to break the center of inversion symmetry in GeTe, resulting in the development of a net dipole moment and the broadening of Raman peaks or new active Raman modes (*54*). The two intrinsic Raman peaks largely broadened and a new low-frequency mode at 60 cm^{-1} appeared when the composition changed from GeTe to high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ (Fig. 5C) because of the loss of center of inversion symmetry with different features of the Ge, Ag, Sb, and Pb atoms (*55*). The vibrating atoms experienced distribution of chemical environments, leading to fluctuations in the vibration frequencies. These Raman results suggested that the phonon mode split with the breakdown of the center of inversion symmetry in the octahedron. Thus, the propagation of

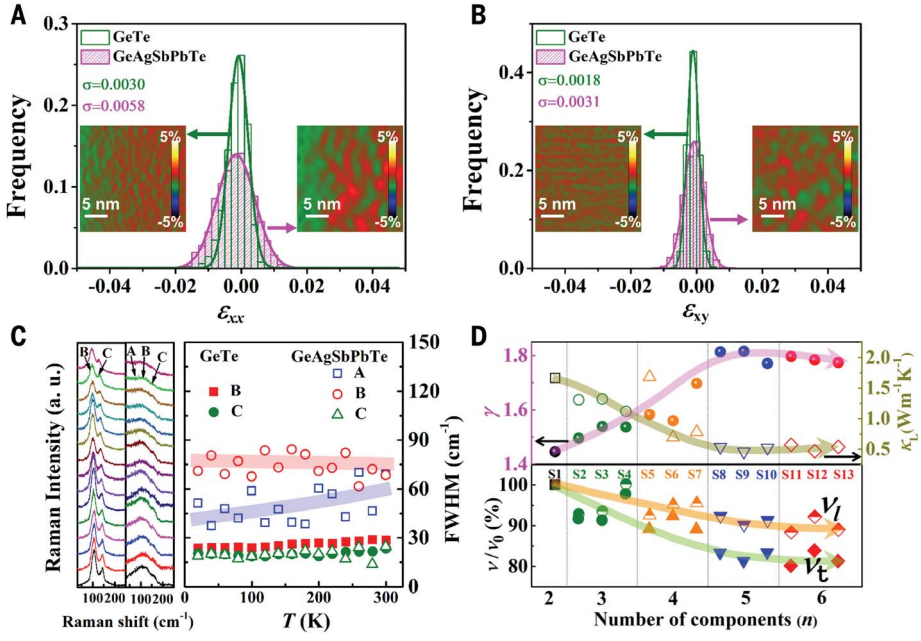


Fig. 5. Localized distribution of phonons. (**A** and **B**) Distribution of normal (A) and shear strains (B) in the GeTe and high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ samples. Left images and green lines are from GeTe samples, and right images and pink lines are from the $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ sample. The value of σ is the root mean square. (**C**) Raman intensity and full width at half maximum (FWHM) of the GeTe (left) and high-entropy $Ge_{0.61}Ag_{0.11}Sb_{0.13}Pb_{0.12}Bi_{0.01}Te$ (middle) samples. (**D**) Measured sound velocities of the transverse (v_t) and longitudinal (v_l) waves and the calculated Gruneisen parameter (γ) and lattice thermal conductivity (κ_L) values at room temperature, where v_0 is the measured sound velocity of the GeTe sample.

heat between the split phonon modes was scattered because of the different vibrational frequencies. This phenomenon of broadening and excessive phonon modes could be explained by the (quasi)localized phonons, which arose from phonon damping related to structural disorder (*56*, *57*). Unlike selective Anderson electron localization, the phonon localization was nonselective in the disordered system. The holes were transported mainly in the Te sublattice in the p-type GeTe, whereas the phonons were definitely transported in both the Ge and Te sublattices.

Phonon damping typically would strengthen scattering for phonons and decrease the mean free path. In a high-entropy system with large disorder, phonon damping should be sufficiently strong, resulting in a small mean free path comparable to the wavelength and approaching the Loffe-Regel limit (*58*). Phonons with frequencies above the Loffe-Regel limit should be substantially damped and no longer propagate (*56*). A disordered structure should dampen transverse phonons, but not longitudinal phonons, because the transverse rather than the longitudinal phonon mode would be more sensitive to disorder (*59*). As a result, disordered systems typically show softer transverse phonon modes compared with longitudinal systems. In this work, we measured the sound velocities of all samples (Fig. 5D). The transverse sound velocity v_t decreased more quickly than the longitudinal sound velocity v_l with an increasing degree of disorder (entropy), illustrating more softened transverse phonons compared with the longitudinal system. Because the Gruneisen parameter (γ) was closely proportional to the ratio of v_l/v_t, the decrease in v_t from the softened transverse phonons would increase γ (*60*). Thus, the anharmonicity demonstrating the interactions among the different branches of phonons would be largely strengthened (Fig. 5D), resulting in depressed and temperature-independent lattice thermal conductivity (*47*).

In conclusion, we promoted the TE performance of GeTe-based materials and modules by forming a high-entropy composition. The distribution of electrons was highly delocalized because of the increased crystal symmetry driven by increasing entropy in the disordered system, which maintained good electrical properties. The localized phonons from the disorder resulted in a strong damping effect on the propagation of transverse phonons, which was responsible for the strengthened anharmonicity and depressed lattice thermal conductivity. The demonstrated strategy of delocalized electrons and localized phonons in the high-entropy materials provided the possibility of simultaneously optimizing the electrical and thermal transport properties.

REFERENCES AND NOTES

1. J. Mao *et al.*, *Science* **365**, 495–498 (2019).
2. G. Tan, L. D. Zhao, M. G. Kanatzidis, *Chem. Rev.* **116**, 12123–12149 (2016).
3. J. He, T. M. Tritt, *Science* **357**, eaak9997 (2017).
4. B. Jiang *et al.*, *Science* **371**, 830–834 (2021).
5. B. Qin *et al.*, *Science* **373**, 556–561 (2021).
6. S. Roychowdhury *et al.*, *Science* **371**, 722–727 (2021).
7. G. J. Snyder *et al.*, *Adv. Mater.* **32**, e2001537 (2020).
8. K. Xia *et al.*, *Energy Environ. Sci.* **12**, 1568–1574 (2019).
9. H. Bai *et al.*, *Nat. Commun.* **12**, 7207 (2021).
10. T. R. Wei *et al.*, *Science* **369**, 542–545 (2020).
11. J. Dong *et al.*, *Energy Environ. Sci.* **12**, 1396–1403 (2019).
12. Z. Dong *et al.*, *Nat. Commun.* **13**, 35 (2022).
13. H. Liu *et al.*, *Nat. Mater.* **11**, 422–425 (2012).
14. X. Zhou *et al.*, *Mater. Today* **21**, 974–988 (2018).
15. X. L. Shi, J. Zou, Z. G. Chen, *Chem. Rev.* **120**, 7399–7515 (2020).
16. N. Dragoe, D. Bérardan, *Science* **366**, 573–574 (2019).
17. L. Hu *et al.*, *Adv. Energy Mater.* **8**, 1802116 (2018).
18. R. Liu *et al.*, *Adv. Mater.* **29**, 1702712 (2017).
19. B. Jiang *et al.*, *Mater. Today Phys.* **5**, 20–28 (2018).
20. J. Li *et al.*, *Joule* **2**, 976–987 (2018).
21. M. Hong *et al.*, *Adv. Mater.* **30**, 1705942 (2018).
22. Z. Zheng *et al.*, *J. Am. Chem. Soc.* **140**, 2673–2686 (2018).
23. M. Cagnoni, D. Führen, M. Wuttig, *Adv. Mater.* **30**, e1801787 (2018).
24. M. K. Brod, M. Y. Toriyama, G. J. Snyder, *Chem. Mater.* **32**, 9771–9779 (2020).
25. M. Wuttig, V. L. Deringer, X. Gonze, C. Bichara, J. Y. Raty, *Adv. Mater.* **30**, e1803777 (2018).
26. S. Lee *et al.*, *Nat. Commun.* **5**, 3525 (2014).
27. Z. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5332–5337 (2018).
28. S. Perumal *et al.*, *Joule* **3**, 2565–2580 (2019).
29. D. Wu *et al.*, *J. Am. Chem. Soc.* **136**, 11412–11419 (2014).
30. M. Hong *et al.*, *Joule* **4**, 2030–2043 (2020).
31. J. Shuai, Y. Sun, X. Tan, T. Mori, *Small* **16**, e1906921 (2020).
32. X. Xu, L. Xie, Q. Lou, D. Wu, J. He, *Adv. Sci. (Weinh.)* **5**, 1801514 (2018).
33. T. Xing *et al.*, *Adv. Mater.* **33**, e2008773 (2021).
34. T. Xing *et al.*, *Energy Environ. Sci.* **14**, 995–1003 (2021).
35. Z. Bu *et al.*, *Energy Environ. Sci.* **14**, 6506–6513 (2021).
36. G. Bai *et al.*, *Adv. Energy Mater.* **11**, 2102012 (2021).
37. P. Jood, M. Ohta, A. Yamamoto, M. G. Kanatzidis, *Joule* **2**, 1339–1355 (2018).
38. X. Hu *et al.*, *Energy Environ. Sci.* **9**, 517–529 (2016).
39. B. Jiang *et al.*, *Energy Environ. Sci.* **13**, 579–591 (2020).
40. Q. Zhang *et al.*, *Energy Environ. Sci.* **10**, 956–963 (2017).
41. C. Fu *et al.*, *Nat. Commun.* **6**, 8144 (2015).
42. P. Qiu *et al.*, *Joule* **3**, 1538–1548 (2019).
43. Y. Xing *et al.*, *Joule* **4**, 2475–2483 (2020).
44. Y. Xing *et al.*, *Energy Environ. Sci.* **12**, 3390–3399 (2019).
45. B. Jiang *et al.*, *Nat. Commun.* **12**, 3234 (2021).
46. T. Zhu, L. Hu, X. Zhao, J. He, *Adv. Sci. (Weinh.)* **3**, 1600004 (2016).
47. D. T. Morelli, V. Jovovic, J. P. Heremans, *Phys. Rev. Lett.* **101**, 035901 (2008).
48. D. Di Sante, P. Barone, R. Bertacco, S. Picozzi, *Adv. Mater.* **25**, 509–513 (2013).
49. W. Gao *et al.*, *Nature* **575**, 480–484 (2019).
50. N. Shibata *et al.*, *Nat. Phys.* **8**, 611–615 (2012).
51. K. Muller *et al.*, *Nat. Commun.* **5**, 1–8 (2014).
52. G. J. Snyder, E. S. Toberer, *Nat. Mater.* **7**, 105–114 (2008).
53. A. Pawbake *et al.*, *Phys. Rev. Lett.* **122**, 145701 (2019).
54. S. Acharya, J. Pandey, A. Soni, *Appl. Phys. Lett.* **109**, 133904 (2016).
55. K. S. Andrikopoulos, S. N. Yannopoulos, A. V. Kolobov, P. Fons, J. Tominaga, *J. Phys. Chem. Solids* **68**, 1074–1078 (2007).
56. H. Shintani, H. Tanaka, *Nat. Mater.* **7**, 870–877 (2008).
57. S. Ren *et al.*, *Nat. Commun.* **12**, 5755 (2021).
58. J. E. Graebner, B. Golding, L. C. Allen, *Phys. Rev. B* **34**, 5696–5701 (1986).
59. X. Y. Li *et al.*, *Phys. Rev. Lett.* **124**, 225902 (2020).
60. T. T. Jia, G. Chen, Y. S. Zhang, *Phys. Rev. B* **95**, 155206 (2017).
61. Y. Wu *et al.*, *Joule* **3**, 1276–1288 (2019).
62. P. Zong *et al.*, *Energy Environ. Sci.* **10**, 183–191 (2017).

ACKNOWLEDGMENTS

Funding: We thank the SUSTech Core Research Facilities for use of the STEM instrument. This work was supported by the Guangdong Province Program (grant no. 00201517), the Guangdong-Hong Kong-Macao Joint Laboratory (grant no. 2019B121205001), the National Natural Science Foundation of

China (grant nos. 52002167, 11874194, and 11934007), the Science and Technology Innovation Committee Foundation of Shenzhen (grant nos. KQTD2016022619565991 and ZDSYS20141118160434515), and special funds (grant no. G02206302). **Author contributions:** B.J. and J.H. designed this work. B.J. and S.L. synthesized the samples, fabricated the modules, and performed the transport property measurements. W.W. performed the STEM observations and analysis. Y.W. and L.X. performed the calculations. C.W., Y.C., and M.H. performed the Raman measurements. B.J., W.W., and J.H. wrote the manuscript, which was edited by all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq5815
Materials and Methods
Figs. S1 to S19
Tables S1 to S3
References (*63*–*87*)

Submitted 18 April 2022; accepted 1 June 2022
10.1126/science.abq5815

CHEMISTRY

Dilution-induced gel-sol-gel-sol transitions by competitive supramolecular pathways in water

Lu Su[1,2]†, Jesús Mosquera[1,3]†, Mathijs F. J. Mabesoone[1], Sandra M. C. Schoenmakers[1], Cyprien Muller[1], Marle E. J. Vleugels[1], Shikha Dhiman[1], Stefan Wijker[1], Anja R. A. Palmans[1], E. W. Meijer[1,4]*

Fascinating properties are displayed by synthetic multicomponent supramolecular systems that comprise a manifold of competitive interactions, thereby mimicking natural processes. We present the integration of two reentrant phase transitions based on an unexpected dilution-induced assembly process using supramolecular polymers and surfactants. The co-assembly of the water-soluble benzene-1,3,5-tricarboxamide (BTA-EG_4) and a surfactant at a specific ratio yielded small-sized aggregates. These interactions were modeled using the competition between self-sorting and co-assembly of both components. The small-sized aggregates were transformed into supramolecular polymer networks by a twofold dilution in water without changing their ratio. Kinetic experiments show the in situ growth of micrometer-long fibers in the dilution process. We were able to create systems that undergo fully reversible hydrogel-solution-hydrogel-solution transitions upon dilution by introducing another orthogonal interaction.

The design of interactive and adaptive soft materials requires a complex composition of chemical components for which nature is often a source of inspiration (*1*). Multicomponent systems formed by multiple associative and dissociative interactions are key signatures of biological matter, but the competitive interplay of all these interactions makes the understanding of certain natural processes difficult (*2*, *3*). Synthetic systems, because of their simplicity, can unravel some of these challenges while yielding new functionalities (*4*–*7*). Although this field is still in its infancy, several examples have highlighted its potential for applications in catalysis, electronics, and medicine (*8*–*11*), with the development of the lipid nanoparticle–encapsulated mRNA vaccine as an obvious successful highlight in this respect (*12*).

Liquid-liquid phase separations are often accompanied by the occurrence of reentrant phase transitions (RPTs), which cause fascinating phenomena in these multicomponent systems. Here, changes in temperature and/or addition of one of the components give rise to a recurrence of the original phase (*11*, *13*, *14*). This phenomenon was recently found to play a critical role in the formation of intracellular membraneless organelles, which are mainly systems based on liquid-liquid phase separation (*15*–*17*). A broader perspective on phase transitions in molecular systems with many components under thermodynamic equilibrium has been presented by Jacobs and Frenkel (*18*). Their statement—that only minor adjustments to the strengths of intermolecular interactions are required to regulate the formation of different domains with specific compositions—is relevant for our current study.

Two frequently used units to construct aqueous synthetic supramolecular systems are surfactants and supramolecular monomers (*6*, *19*–*24*). Monomer **BTA-EG_4** undergoes supramolecular polymerization in water to afford one-dimensional aggregates at micromolar concentration driven by hydrogen bonding and hydrophobic effects, and is made water-soluble by hydrophilic tetraethylene glycol groups (Fig. 1A and fig. S6) (*24*). Dynamic hydrogels are formed at concentrations above 8 mM (figs. S7 and S8) (*25*). The cationic surfactant octyltrimethylammonium bromide (OTAB) forms spherical micelles with a hydrodynamic diameter ~2 nm above its critical micelle concentration (CMC = ~297 mM at 20°C) (*26*). An unexpected finding sparked our attention when studying orthogonality by combining these two units. We observed that the addition of OTAB provoked the full disruption of the **BTA-EG_4** hydrogel. Surprisingly, the resulting solution was transformed into a transparent hydrogel upon dilution, followed by reentering the solution phase with further dilution, giving rise to a supramolecular sol-gel-sol RPT in water. The first two states were visualized by cryogenic transmission electron microscopy (cryo-TEM) (Fig. 1B) and small-angle x-ray scattering (SAXS) (fig. S9). This process was fully reversible by increasing the concentration and led us to investigate the underlying mechanism of this hitherto undescribed dilution-induced supramolecular polymerization in water. It is combined with the development of two integrated RPTs (gel-sol-gel-sol) upon dilution with the supramolecular units shown in Fig. 1.

To elucidate the pathways that direct the sol-gel transition upon dilution, we investigated the interaction between **BTA-EG_4** and a library of surfactants at micromolar concentrations (figs. S10 to S16). We illustrate our detailed explorations using a combination of spectroscopic, cryo-TEM, light scattering, isothermal titration calorimetry, and theoretical techniques (Fig. 2 and figs. S1 to S5, S17 to S21) by taking the cationic surfactant cetrimonium bromide (CTAB, CMC ≈ 1 mM) as an example. Nuclear magnetic resonance (NMR) and cryo-TEM measurements of **BTA-EG_4** (250 μM) were performed in the presence of 0, 1, 2, and 3 eq. of CTAB (Fig. 2, A and B). As a result of the formation of micrometer-sized polymers in D_2O, no resonances from **BTA-EG_4** were observed in the NMR, whereas the sample containing 250 μM CTAB showed sharp and well-split peaks. CTAB peaks disappeared when mixed in equimolar ratio with **BTA-EG_4** in NMR, and cryo-TEM confirmed the presence of micrometer-long aggregates (Fig. 2B). Together, these results indicate that all the surfactants are integrated in the polymers. However, when more equivalents of CTAB were applied, we observed a gradual increase in intensity and resolution of the peaks associated with **BTA-EG_4**. At 2 eq. of CTAB, the ethylene glycol peak of **BTA-EG_4** appeared; ^{1}H diffusion-ordered spectroscopy (DOSY) NMR (fig. S17) revealed that all the surfactants were still embedded, but in much smaller assemblies. These smaller aggregates were confirmed by cryo-TEM images showing fibers less than 500 nm in length. At 3 eq. of CTAB, both signals corresponding to the short **BTA-EG_4**/CTAB fibers and **BTA-EG_4**/CTAB micelles were observed in DOSY NMR; cryo-TEM showed small spherical aggregates and occasionally short fibers. Ultraviolet-visible (UV-vis) spectroscopy confirmed fiber disruption and the formation of small BTA aggregates (Fig. 2C). This transition was dependent on the overall concentration as well as the ratio of both components (Fig. 2D), with more surfactants needed at lower BTA concentrations. The saturation of the polymer with surfactants destabilized the structure and destroyed the fibers at high concentrations of surfactant, most probably as a result of electrostatic and/or geometric constraints in the structures formed (Fig. 2E).

Studies with other surfactants revealed three relevant conclusions: (i) The affinity between the surfactant and **BTA-EG_4** is determined by the length of the aliphatic chain of the surfactant, (ii) **BTA-EG_4** fibers are unstable below the CMC of the surfactants, and (iii) the transition is dependent on both the overall concentration and the ratio of the components.

[1]Institute for Complex Molecular Systems, Laboratory of Macromolecular and Organic Chemistry, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands. [2]Leiden Academic Centre for Drug Research, Leiden University, 2333 CC Leiden, Netherlands. [3]Centro de Investigacións Científicas Avanzadas, Universidade da Coruña (CICA), A Coruña, Spain. [4]School of Chemistry and RNA Institute, University of New South Wales, Sydney, NSW 2052, Australia.
*Corresponding author. Email: e.w.meijer@tue.nl
†These authors contributed equally to this work.

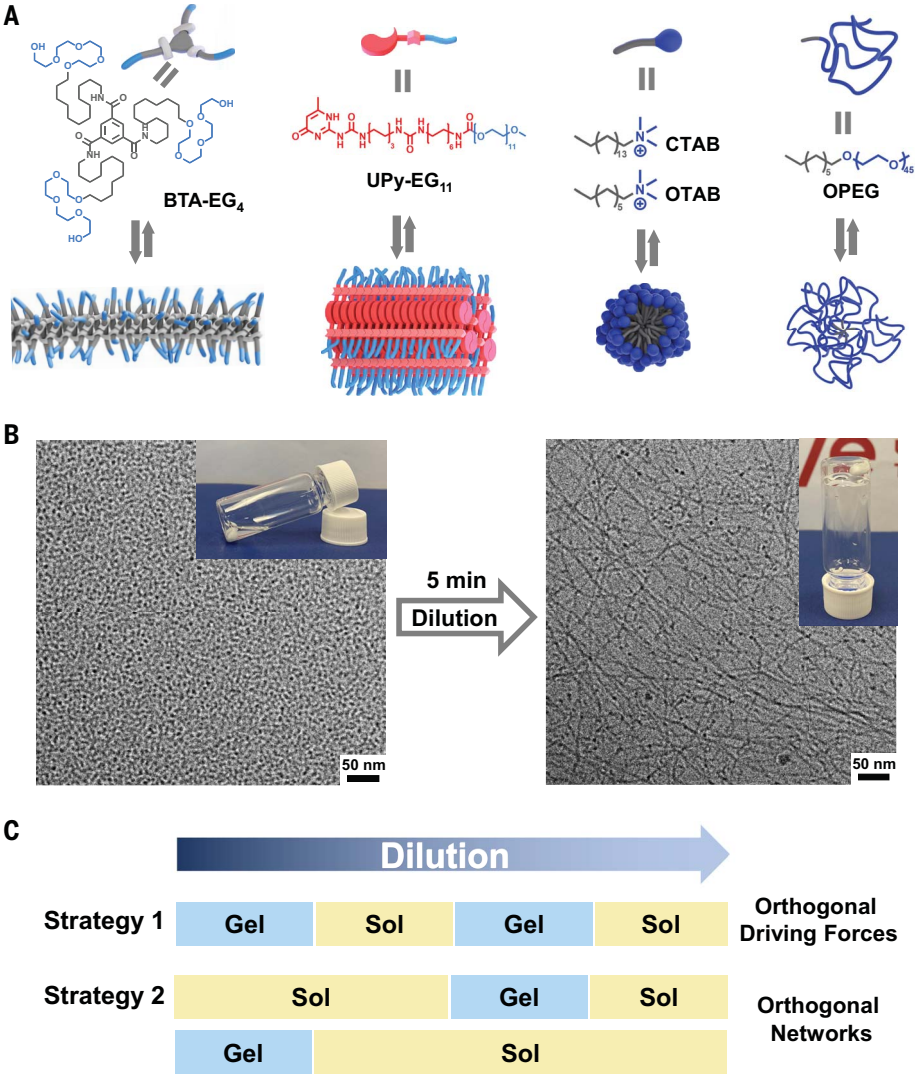


Fig. 1. Dilution-induced gelation in water. (**A**) Supramolecular components and their individual assemblies in water. (**B**) Cryo-TEM images of dilution-induced gelation in a **BTA-EG$_4$**/OTAB system. Left: A solution of **BTA-EG$_4$** (4 wt %, 31 mM) and OTAB (6.5 eq., 202 mM) shows small aggregates with a diameter of 5 to 8 nm, suggesting the presence of micellar **BTA-EG$_4$**/OTAB complexes. Right: A transparent hydrogel of **BTA-EG$_4$** (2 wt %) and OTAB (101 mM), generated upon twofold dilution with MQ-H_2O, shows an entangled **BTA-EG$_4$** fiber network. Scale bar, 50 nm. (**C**) Two strategies for dilution-induced gel-sol-gel-sol transitions by competitive supramolecular pathways in water.

From these results, we hypothesize that the sol-gel transition upon dilution is driven by the concentration-dependent complexation of surfactants to **BTA-EG$_4$** polymers. To corroborate this hypothesis, we constructed a thermodynamic mass-balance model that describes competition among supramolecular polymerization, micellization, complexation of CTAB to the polymers, and polymer collapse (Fig. 2, E and F, figs. S1 to S5, and supplementary text). We subsequently studied the dilution-induced effect of a solution of **BTA-EG$_4$** with 3 eq. of CTAB, in which only small aggregates exist. Upon dilution, peaks corresponding to both **BTA-EG$_4$** and CTAB gradually broadened and weakened in NMR (Fig. 2G) while the characteristic peaks (211 and 227 nm) in UV-vis spectra related to fiber structure appeared (Fig. 2H), as corroborated by cryo-TEM (fig. S21). We hypothesize that CTAB is released when the overall concentration decreases, thereby enabling the elongation of the BTA fiber to eventually achieve micrometer-long structures. The dilution-induced supramolecular polymerization became faster upon adding more water (fig. S20).

With an understanding of the mechanisms underlying the dilution-induced supramolecular polymerization in water, an integration of two RPTs was developed by introducing an additional orthogonal interaction through two different strategies (Fig. 1C). The first strategy uses the intrinsic property of a polymeric surfactant to lead to further aggregation at higher concentration, whereas the second strategy takes advantage of an orthogonal supramolecular double network with a different gelation concentration window.

The polymeric surfactant, octyl polyethylene glycol (OPEG, 2 kDa; Fig. 1A and figs. S22 and S23) was synthesized with a C8-spacer. The interaction of OPEG with **BTA-EG$_4$** at micromolar concentration was evaluated by UV-vis, NMR, and cryo-TEM, showing a behavior similar to that observed for CTAB (figs. S24 and S25). Interestingly, an optimized **BTA-EG$_4$**:OPEG molar ratio of 2:3 ([**BTA-EG$_4$**] = 80 mM) in Milli-Q (MQ)–H_2O gave rise to an opaque **Gel 1** with a mesh size of ~10 to 20 nm as indicated by cryo-TEM (Fig. 3A) and SAXS (fig. S36). The densely packed network was composed of wormlike fibers of fused **BTA-EG$_4$**/OPEG micelles. Rheological experiments showed a storage modulus, G', of up to 2 kPa with a complex viscosity, η^*, above 2 kPa·s, as well as a fast and near-complete self-healing behavior (Fig. 3, C to E). When diluting this gel with water to [**BTA-EG$_4$**] = 35 mM, a solution was obtained after equilibrium, owing to the dissociation of the fused micelles as depicted by cryo-TEM and SAXS. In line with the morphology transition, η^* dropped to ~1 Pa·s. When this solution was further diluted to [**BTA-EG$_4$**] = 27 mM, which is below the CMC of OPEG (~42 mM; fig. S26), a transparent and dynamic **Gel 2** was formed with G' of ~5 Pa and η^* of ~10 Pa·s. This gelation was triggered by the dissociation/release of OPEG from the **BTA-EG$_4$**/OPEG micellar complex, which enabled the elongation of the BTA fibers and led to the formation of the typical, entangled **BTA-EG$_4$** fiber network. With further dilution to [**BTA-EG$_4$**] = 0.8 mM and still the same **BTA-EG$_4$**:OPEG molar ratio of 2:3, the hydrogel changed into a viscous solution showing η^* of ~0.4 Pa·s and micrometer-long individual fibers. A detailed characterization of this gel-sol-gel-sol system (cryo-TEM, SAXS, and rheology) is presented in figs. S27 to S36.

The good biocompatibility of **BTA-EG$_4$** (*27*) provides further opportunity for application of this multicomponent system in more complex media. Phosphate-buffered saline and DMEM/F12 basal medium were used for the whole process, showing RPT behavior identical to that observed in MQ-H_2O (fig. S37). Additionally, when physiological temperature was applied, G' of **Gel 1** became reduced by a factor of 8 (G' ~ 250 Pa), probably owing to a fastened dissociation of **BTA-EG$_4$**/OPEG micelles (fig. S38).

The second strategy relies on an orthogonal supramolecular network formed by water-soluble supramolecular polymers based on the ureidopyrimidinone unit (**UPy-EG$_{11}$**, Fig. 1A), which is less dynamic than **BTA-EG$_4$** polymers (*28*).

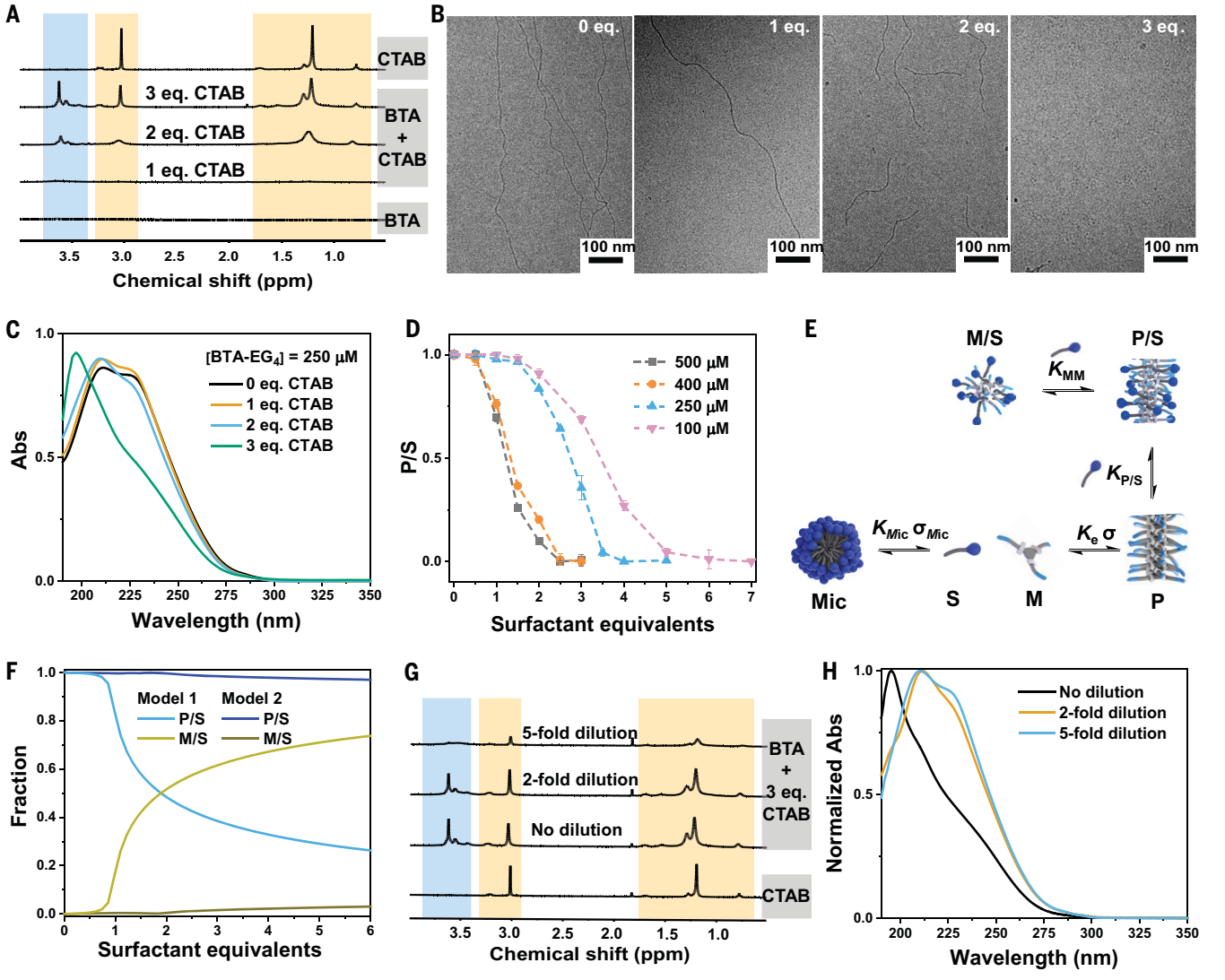


Fig. 2. Interaction and dilution-induced assembly of BTA-EG$_4$ and CTAB in water. (**A** to **C**) ^{1}H NMR spectra (400 MHz, D_2O; peaks corresponding to BTA and CTAB are in blue and yellow zones, respectively) (A), cryo-TEM images (scale bar, 100 nm) (B), and UV-vis spectra (C) of **BTA-EG$_4$** (250 μM) in the presence of 0, 1, 2, and 3 eq. of CTAB, respectively. (**D**) Experimental fraction curves of **BTA-EG$_4$**/CTAB polymer obtained by normalizing the UV absorbance at 227 nm. The concentration of **BTA-EG$_4$** was held constant (500, 400, 250, or 100 μM, respectively) while CTAB content was varied. (**E**) Thermodynamic model for the cooperative, competitive pathways, with M, P, S, and Mic representing **BTA-EG$_4$** monomer, **BTA-EG$_4$** polymer, free surfactant, and surfactant micelle, respectively, and K_i and σ_i the equilibrium constant and cooperativity parameter of process *i*. (**F**) Simulated fraction curves of P/S and M/S with 1 eq. (model 1) and 2 eq. (model 2) of [CTAB]/[**BTA-EG$_4$**] accommodations, in which the concentration of **BTA-EG$_4$** was held constant (250 μM) while the surfactant content was varied. (**G** and **H**) ^{1}H NMR spectra (400 MHz, D_2O; peaks corresponding to BTA and CTAB are in blue and yellow zones, respectively) (G) and UV-vis spectra (H) of [**BTA-EG$_4$**]/[CTAB] (1/3) upon dilution, with [**BTA-EG$_4$**]$_{initial}$ = 250 μM. CTAB absorption in all UV-vis spectra was removed for clarity.

In water, **UPy-EG$_{11}$** gives rise to micrometer-long bundled fibers with a width of ~5 to 14 nm. **UPy-EG$_{11}$**/CTAB interactions were confirmed by ^{1}H NMR (fig. S39). However, **UPy-EG$_{11}$** and **UPy-EG$_{11}$**/CTAB mixtures (up to 5 eq. of CTAB) showed almost identical UV spectra and cryo-TEM images, indicating that CTAB binds with **UPy-EG$_{11}$** bundles yet does not markedly disrupt the supramolecular polymer. When mixed, **BTA-EG$_4$** and **UPy-EG$_{11}$** form self-sorted homopolymers. Hence, **BTA-EG$_4$**, **UPy-EG$_{11}$**, and a surfactant make an ideal combination to construct an interpenetrating double network consisting of different gelation windows. As depicted in Fig. 3B, F to H, an opaque hydrogel **Gel 1′** was fabricated with a **BTA-EG$_4$:UPy-EG$_{11}$**:OTAB molar ratio of 3.2:1:21.2 at [**BTA-EG$_4$**] of 93 mM in MQ-H_2O. **Gel 1′** is mainly stabilized by the **UPy-EG$_{11}$** network, whereas **BTA-EG$_4$** exists as **BTA-EG$_4$**/OTAB micelles. Upon dilution to [**BTA-EG$_4$**] = 46.7 mM, the **UPy-EG$_{11}$** network was not dense enough to construct a three-dimensional network that tightly held all the water. As the length of the **BTA-EG$_4$** fibers was not sufficient to form a network because of interaction with the surfactant, a transparent solution with a low η* of ~1 Pa·s was accomplished. Further dilution yielded the sol-gel-sol transition as in the first strategy. All results of the detailed characterization are presented in figs. S40 to S43.

After exploring the different states at specific concentrations, we tested methods for tracking the formation of these polymers upon dilution. We used total internal reflection fluorescence microscopy (TIRFM) to trace the in situ formation of the supramolecular polymeric fibers after the addition of droplets of concentrated solution to an aqueous film (Fig. 4).

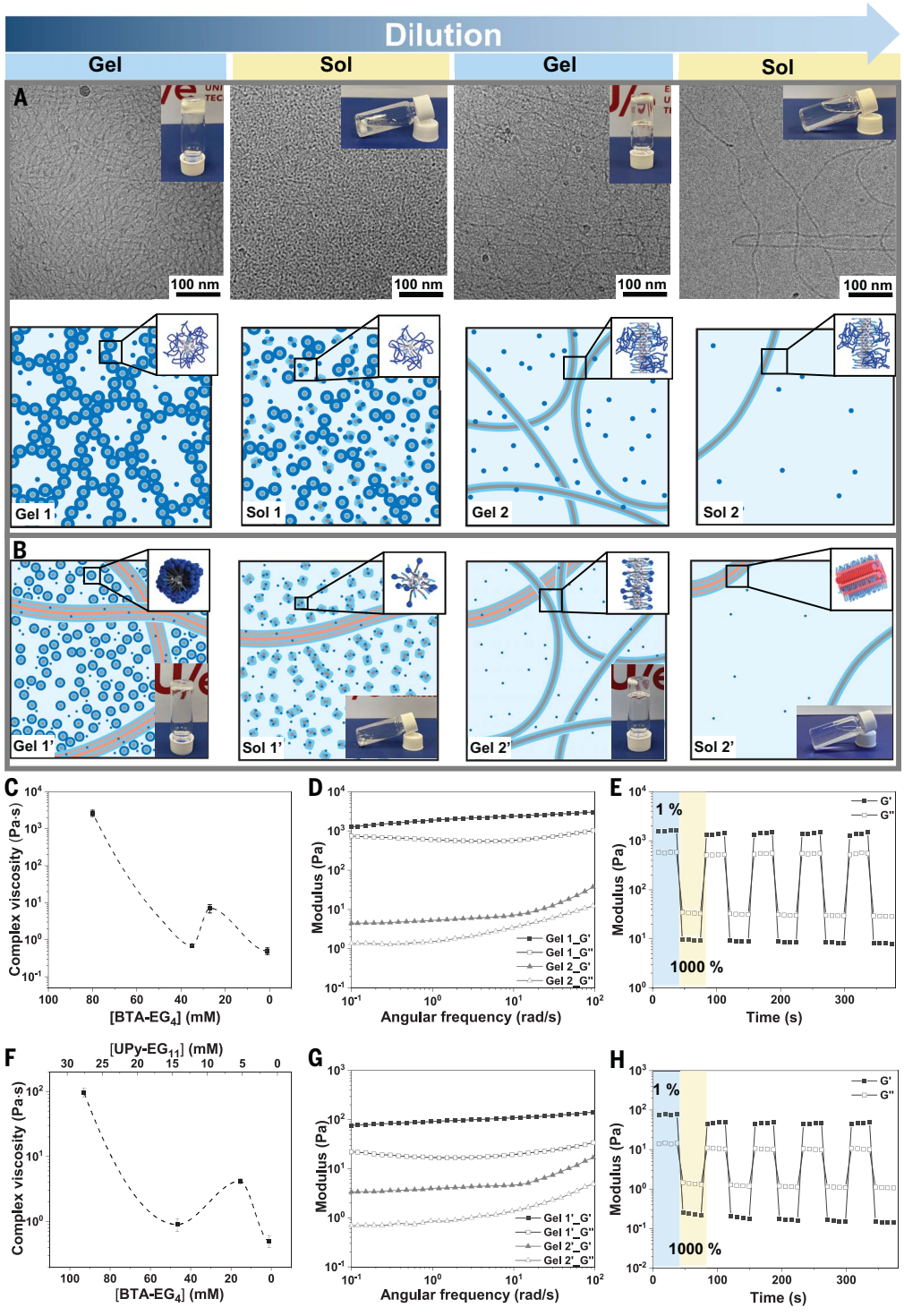


Fig. 3. Dilution-induced gel-sol-gel-sol transitions in water by using a two-component (BTA-EG_4/OPEG) or three-component (BTA-EG_4/UPy-EG_{11}/OTAB) system, respectively. (**A**) Cryo-TEM images and schematic illustrations of gel-sol-gel-sol transitions of a **BTA-EG_4**/OPEG system with their macroscopic images and molecular-level illustrations as insets above and below. (**B**) Schematic illustrations of gel-sol-gel-sol transitions of a **BTA-EG_4/UPy-EG_{11}/**OTAB system, displayed as in (A). (**C** and **F**) Complex viscosity points (ω = 1 rad/s, 1% strain, 20°C) corresponding to the gel-sol-gel-sol transitions in the **BTA-EG_4**/OPEG system (C) and the **BTA-EG_4/UPy-EG_{11}/**OTAB system (F), with the dashed line fitted through the cubic spline curve. (**D** and **G**) Frequency-dependent oscillatory rheology (1% strain, 20°C) of the **BTA-EG_4**/OPEG system (D), showing that **Gel 1** ([**BTA-EG_4**] = 80 mM) has a stronger network than **Gel 2** ([**BTA-EG_4**] = 27 mM), and the **BTA-EG_4/UPy-EG_{11}/**OTAB system (G), showing that **Gel 1′** ([**BTA-EG_4**] = 93 mM) has a stronger network than **Gel 2′** ([**BTA-EG_4**] = 15.5 mM). (**E** and **H**) Step-strain measurements with applied oscillatory strain alternated between 1% and 1000% for 30-s periods (ω = 1 rad/s, 20°C) in **Gel 1** of the **BTA-EG_4**/OPEG system (E) and **Gel 1′** of the **BTA-EG_4/UPy-EG_{11}/**OTAB system (H).

(See figs. S44 to S46 for details of the experimental setup.) When an aqueous droplet of [**BTA**]:[CTAB] (1:3, [**BTA**] = 250 μM), containing 5% **BTA-Cy3** for visualization, was pipetted into the aqueous film and thus 100-fold diluted in situ, supramolecular fibers started to grow at several positions out of droplets (Fig. 4, A to C). The relatively slow process in thin films enabled us to follow the growth over time. Some of the fibers became many tens of micrometers long (Fig. 4A). In some cases, the fibers were connected to other droplets (Fig. 4C). In other cases, the fibers both grew and shrank, and the pulling of droplets by the contracting fibers could be observed (Fig. 4D and movies S1 to S4). We ascribe the growth and the shrinkage to concentration gradients from the droplet and against the droplet, respectively—a phenomenon similar to the dynamic interconnected networks formed by Marangoni flows in the mesoscale positioning of amphiphiles when extruded from the droplets (*29*). These results emphasize the critical role of concentrations in multicomponent systems and are useful to understand several phenomena in biology, where RPTs are critically important to understand cellular condensates.

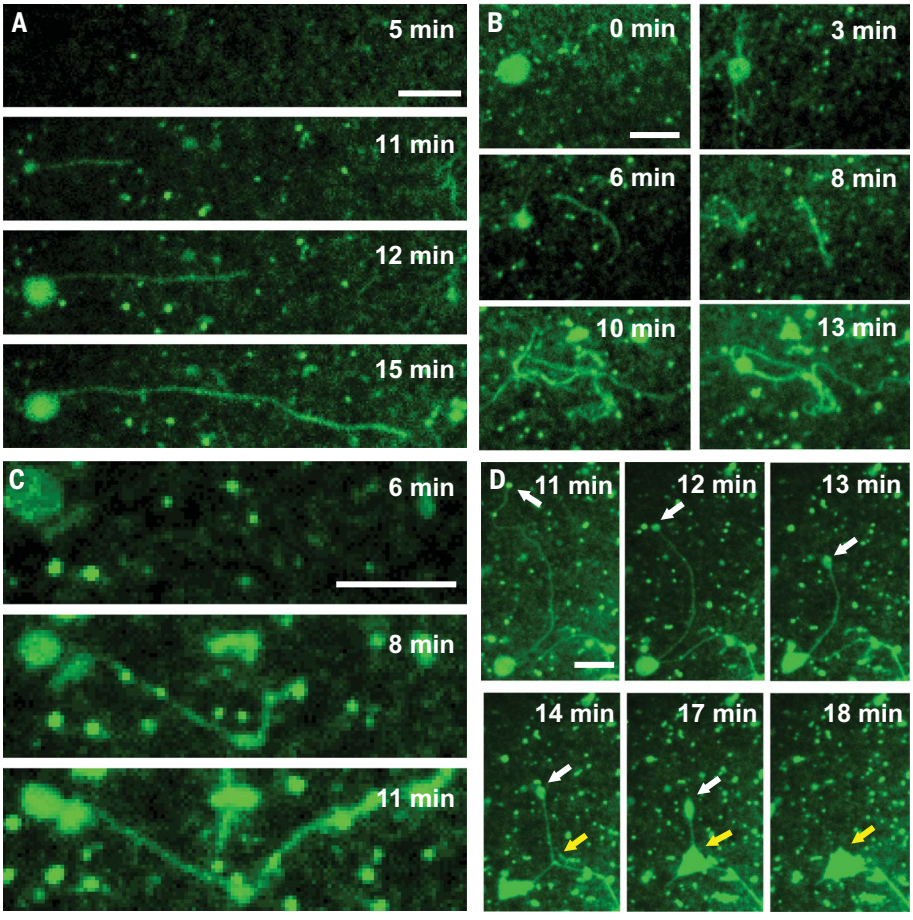


Fig. 4. Kinetic in situ formation of fibers upon dilution. (**A** to **C**) TIRFM images from movies S1 and S2 upon addition of **BTA-EG$_4$**/CTAB droplets to a film of PBS over time. The droplets in time adhere to the surface and fibers are formed after several minutes, using 1% **BTA-biotin**. The green fluorescence is the result of 5% **BTA-Cy3**; the higher the intensity, the higher the concentration of **BTA**. The increase of the droplet size results from an increased adhesion of the droplet to the surface. The increase in length of the fibers is the result of dilution-induced self-assembly/polymerization by diffusion of aggregates from the densely packed droplet to the continuous water film. (**D**) Snapshots from movies S3 and S4 of the shrinkage (white arrow) and connection (yellow arrow) of the fibers due to concentration gradients. Scale bar, 5 μm.

REFERENCES AND NOTES

1. R. Merindol, A. Walther, *Chem. Soc. Rev.* **46**, 5588–5619 (2017).
2. K. Simons, E. Ikonen, *Nature* **387**, 569–572 (1997).
3. W. Baumeister, J. Walz, F. Zühl, E. Seemüller, *Cell* **92**, 367–380 (1998).
4. G. Vantomme, E. W. Meijer, *Science* **363**, 1396–1397 (2019).
5. J.-F. Lutz, J.-M. Lehn, E. W. Meijer, K. Matyjaszewski, *Nat. Rev. Mater.* **1**, 16024 (2016).
6. T. Aida, E. W. Meijer, S. I. Stupp, *Science* **335**, 813–817 (2012).
7. E. Mattia, S. Otto, *Nat. Nanotechnol.* **10**, 111–119 (2015).
8. Q. Wan, W. P. To, X. Chang, C. M. Che, *Chem* **6**, 945–967 (2020).
9. S. Li *et al*., *J. Am. Chem. Soc.* **140**, 10794–10802 (2018).
10. D. A. Wilson, R. J. M. Nolte, J. C. M. van Hest, *Nat. Chem.* **4**, 268–274 (2012).
11. F. Helmich *et al*., *Angew. Chem. Int. Ed.* **49**, 3939–3942 (2010).
12. N. Pardi, M. J. Hogan, F. W. Porter, D. Weissman, *Nat. Rev. Drug Discov.* **17**, 261–279 (2018).
13. E. A. Clark, J. E. G. Lipson, *Polymer* **53**, 536–545 (2012).
14. T. M. Hermans *et al*., *Nat. Nanotechnol.* **4**, 721–726 (2009).
15. P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, A. A. Deniz, *Angew. Chem. Int. Ed.* **56**, 11354–11359 (2017).
16. J. A. Riback *et al*., *Nature* **581**, 209–214 (2020).
17. Y. Shin, C. P. Brangwynne, *Science* **357**, eaaf4382 (2017).
18. W. M. Jacobs, D. Frenkel, *Biophys. J.* **112**, 683–691 (2017).
19. R. J. Brea *et al*., *J. Am. Chem. Soc.* **140**, 17356–17360 (2018).
20. M. P. Hendricks, K. Sato, L. C. Palmer, S. I. Stupp, *Acc. Chem. Res.* **50**, 2440–2448 (2017).
21. J. Kang *et al*., *Science* **347**, 646–651 (2015).
22. G. González-Rubio *et al*., *Science* **368**, 1472–1477 (2020).
23. A. Heeres *et al*., *J. Am. Chem. Soc.* **125**, 14252–14253 (2003).
24. C. M. A. Leenders *et al*., *Chem. Commun.* **49**, 1963–1965 (2013).
25. E. Vereroudakis *et al*., *ACS Cent. Sci.* **6**, 1401–1411 (2020).
26. R. Zieliński, *J. Colloid Interface Sci.* **235**, 201–209 (2001).
27. S. Varela-Aramburu *et al*., *Biomacromolecules* **21**, 4105–4115 (2020).
28. S. I. S. Hendrikse *et al*., *Chem. Commun.* **53**, 2279–2282 (2017).
29. A. van der Weijden, M. Winkens, S. M. C. Schoenmakers, W. T. S. Huck, P. A. Korevaar, *Nat. Commun.* **11**, 4800 (2020).

ACKNOWLEDGMENTS

We thank D. Frenkel, T. de Greef, R. Cardinaels, and P. Korevaar for discussions. **Funding:** Supported by NWO (TOP-PUNT grant 10018944); the Dutch Ministry of Education, Culture and Science (Gravitation program 024.001.035); and the European Commission (SYNMAT-788618-1 and H2020-MSCA-IF-2017-794016). C.M. gratefully thanks the French National Research Agency (ANR-17-EURE-0016). **Author contributions:** J.M. and C.M. found the effect serendipitously and performed the first series of experiments. J.M., L.S., and E.W.M. designed the study, conceived the overall set of experiments, and wrote the manuscript. L.S., M.E.J.V., S.D., and S.M.C.S. contributed to all analyses and the synthesis of the components. S.W. and L.S. did the SAXS and S.M.C.S. performed all the cryo-TEM experiments. S.D. and L.S. performed all the optical microscopy experiments and M.F.J.M. did the modeling. All authors critically revised the manuscript, and they all approve the manuscript. **Competing interests:** The authors declare 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.abn3438
Materials and Methods
Supplementary Text
Figs. S1 to S46
References (*30*–*39*)
Movies S1 to S4
MAPLE Files: Derivation_model_1eq.mw, Derivation_model_2eq.mw
MATLAB Scripts and Function

Submitted 19 November 2021; accepted 11 May 2022
10.1126/science.abn3438

PALEONTOLOGY

Ultrastructure reveals ancestral vertebrate pharyngeal skeleton in yunnanozoans

Qingyi Tian[1,2], Fangchen Zhao[2]*, Han Zeng[2], Maoyan Zhu[2,3], Baoyu Jiang[1]*

Pharyngeal arches are a key innovation that likely contributed to the evolution of the jaws and braincase of vertebrates. It has long been hypothesized that the pharyngeal (branchial) arch evolved from an unjointed cartilaginous rod in vertebrate ancestors such as that in the nonvertebrate chordate amphioxus, but whether such ancestral anatomy existed remains unknown. The pharyngeal skeleton of controversial Cambrian animals called yunnanozoans may contain the oldest fossil evidence constraining the early evolution of the arches, yet its correlation with that of vertebrates is still disputed. By examining additional specimens in previously unexplored techniques (for example, x-ray microtomography, scanning and transmission electron microscopy, and energy dispersive spectrometry element mapping), we found evidence that yunnanozoan branchial arches consist of cellular cartilage with an extracellular matrix dominated by microfibrils, a feature hitherto considered specific to vertebrates. Our phylogenetic analysis provides further support that yunnanozoans are stem vertebrates.

Debate over the origin and early evolution of vertebrates revolves around the pharynx (*1*–*4*), but there is little high-resolution information available in the critical part of the vertebrate tree—its stem (*5*). Yunnanozoans from the early Cambrian Chengjiang fauna (~518 million years old) of China (*6*) are the oldest relatives of crown-group vertebrates, although their phylogenetic position remains sharply debated—they have been variously classified as stem vertebrates, cephalochordates, hemichordates, or even stem deuterostomes on the basis of disputed anatomical interpretations (*7*–*13*). Three species have been erected in the clade, *Yunnanozoon lividum* Hou, Ramsköld, and Bergström, 1991 (*7*), *Haikouella lanceolata* Chen, Huang, and Li, 1999 (*10*), and *H. jianshanensis* Shu *et al*., 2003 (*11*), but the latter two species were subsequently revised as junior synonyms of *Y. lividum* (*13*) (Fig. 1, A and B). This revised classification is adopted in this paper. Seven pairs of bilaterally symmetrical branchial arches have long been recognized in the pharynx of yunnanozoans (Fig. 1C). The first pair of mouth-surrounding arches were once interpreted as blood vessels in the upper and lower lips (*10*, *12*) or oral skirts (*11*), but others have argued for their close similarity to the remaining arches (*13*). Each branchial arch consists of a bamboo-like bar formed by evenly spaced, transverse septa and two rows of lanceolate gill filaments that attach to both sides of each segment (fig. S1B) (*10*). The stacked discoid structures in the bamboo-like bar were compared to those in the cellular cartilages of embryonic vertebrates (*8*, *10*), but this claim lacks further anatomical support (*12*). The arches attach to slightly bent horizontal rods at both their dorsal and ventral ends (*8*, *11*). The dorsal and ventral horizontal rods were previously interpreted as blood vessels, margins of the endostyle trough, or simply folds (*8*, *11*, *12*, *14*). These debates call for further study based on new specimens. Here, we reexamine the pharyngeal skeleton of yunnanozoans on the basis of 127 newly collected specimens and provide high-resolution anatomical and ultrastructural correlation with that of vertebrates.

Newly collected specimens confirm that the branchial arches comprise bamboo-like bars and lanceolate gill filaments (Fig. 1, D to F, and fig. S1, A and B). Dorsoventrally preserved specimens show that the gill filaments are external to the bars (Fig. 1H and fig. S5, B, D, and G). All the bars consist of ~25 straight, first-order septa that form a series of stacked discoid structures measuring ~170 μm wide and 140 μm thick (Fig. 1, D and E, and fig. S1B). Each discoid structure is subdivided into two to four cellular chambers by irregular second-order septa (N = 1 to 3) between the two neighboring first-order septa, which are straight and regularly organized (Fig. 1, D and E). X-ray computed microtomography (micro-CT) of two different animals reveals that the cellular chambers are 23 to 84 μm in diameter (N = 43) (Fig. 2 and fig. S2).

The arches are the most commonly preserved structures in yunnanozoans (*11*), with abundant carbonaceous residues (Fig. 3I and fig. S3, N, O, and Q). Under scanning electron microscopy (SEM), these residues comprise microfibrils that are thin, slightly wavy, largely parallel, and densely packed (Fig. 3, D and F, and fig. S3). Locally, the microfibrils are connected by short, parallel cross-linkages forming a lattice (Fig. 3F and fig. S3I). The microfibrils are extremely fine, best measured with transmission electron microscopy to be 12 ± 2 nm in diameter (N = 16) (Fig. 3, G and H). Carbonaceous particles locally aggregate forming a "beads-on-a-string" appearance with 50- to 60-nm intervals (Fig. 3H).

We sought and found the first branchial arch, lateral to a stout, dark circumoral structure (description is given in the supplementary text section of the supplementary materials) at the rostral end of the body. The first branchial arch is often fully or partly obscured by the underlapping circumoral structure, but close inspection reveals that the first arch, identical to the remaining arches, has bamboo-like segments delineated by first-order septa (fig. S1, F, L, and M), attached gill filaments (fig. S1, D, F, and H), and microfibrils in its matrix (fig. S3, C and D).

The seven branchial arches are aligned subparallel to each other and slightly inclined toward the rostral end (Fig. 1, C and F). The arches are connected by dorsoventrally curved, parenthesis-shaped horizontal rods at both ends (Fig. 1, F to H, and figs. S4 and S5; see also figure 16 in (*13*); measurements are given in supplementary text). These dorsal and ventral horizontal rods run subparallel to each other through most of the length of the pharynx then gradually converge between the seventh pair of arches and the gonads (fig. S5H). Locally, the rods show poorly preserved, putative bamboo-like structures (figs. S4D and S5, F and G). No gill filaments occur on the rods, but their matrix does contain microfibrils, as in the branchial arches (fig. S3S). The branchial bars articulate with the horizontal rods at joints, evidenced by the common presence of disarticulation between the two structures (figs. S1I and S5F). The horizontal rods are not to be confused with nearby dark lines called ventral and dorsal thin horizontal lines ("vtl" in fig. S5, B, C, and G; and "dtl" in fig. S1, K and N). These thinner lines might be longitudinal blood vessels, because they occur in the regions of the aortae of vertebrates (*10*–*12*).

Our findings can be interpreted in a comparative biological context. In vertebrates, chondrocytes are stacked parallel to the arch axis during chondrogenesis, forming the classical stacked discoid structures of pharyngeal arches in extant vertebrates (*15*, *16*). The arrangement, size, shape, and number of cellular chambers in the branchial bars of yunnanozoans match the chondrocytes in the pharyngeal arches of embryonic and primitive vertebrates and in the oral cirri of cephalochordates (Fig. 4D).

The microfibrils in the branchial arches of yunnanozoans have characteristics of fibrillin microfibrils, including their diameter of ~12 nm (*17*, *18*), wavy appearance, parallel bundling, and cross-linkages (*19*). They differ notably

[1]State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering and Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing 210023, China. [2]State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology and Center for Excellence in Life and Palaeoenvironment, Chinese Academy of Sciences, Nanjing 210008, China. [3]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.
*Corresponding author. Email: fczhao@nigpas.ac.cn (F.Z.); byjiang@nju.edu.cn (B.J.)

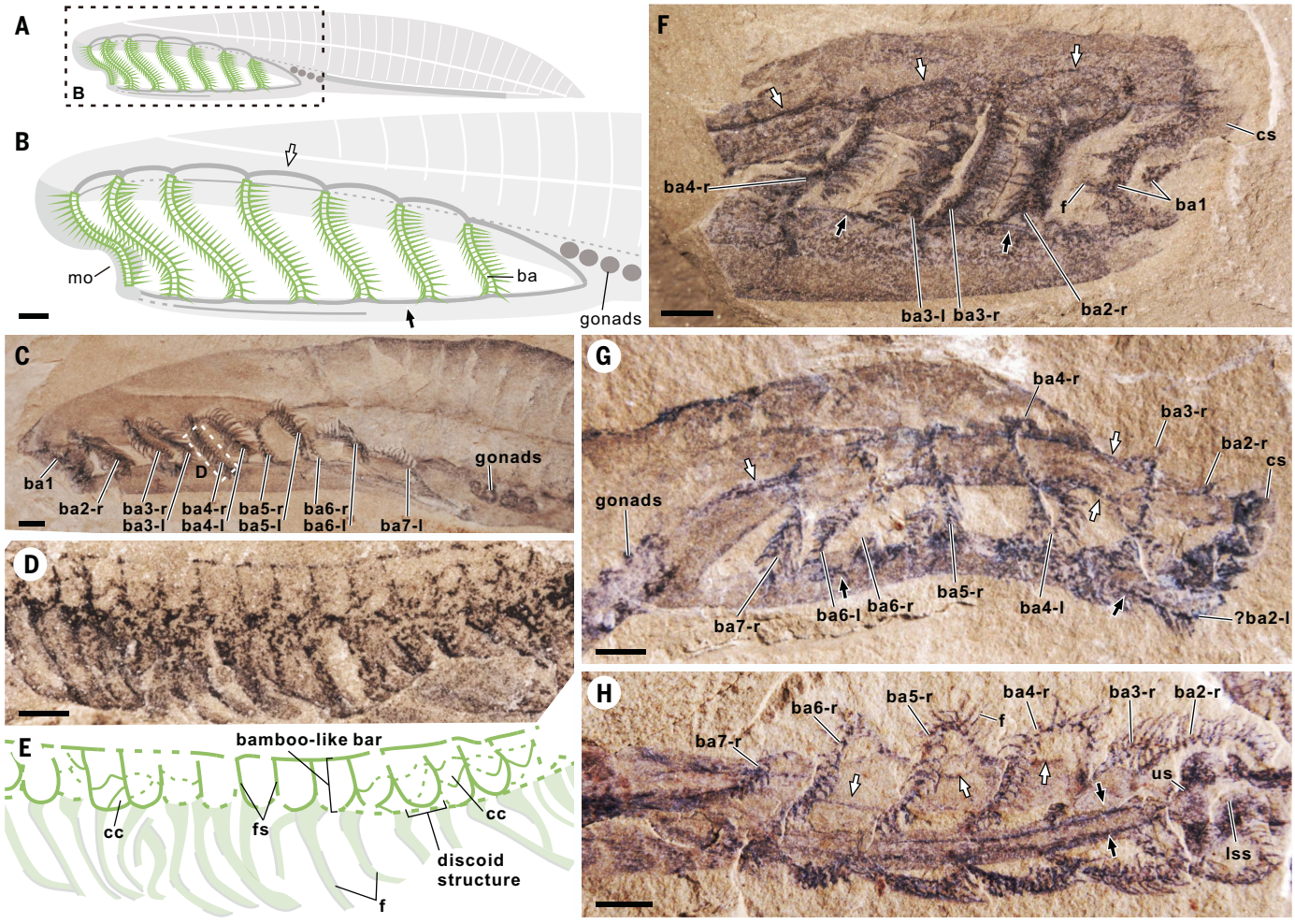


Fig. 1. Pharyngeal structures of yunnanozoans. (**A** and **B**) Line drawings of a yunnanozoan (A) and its pharynx (B). (**C** to **E**) An overview of the branchial arches in specimen NIGP 176256a (C). An enlargement (D) and line drawing (E) of the boxed area in (C) show detailed anatomy of each arch. (**F** to **H**) Lateral [(F) and (G)] and ventral views (H) show the dorsoventrally curved, parenthesis-shaped, horizontal rods in both the dorsal (white arrows) and ventral (black arrows) sides in specimens NIGP 176267a (F), EC00086a (G), and EC00308a (H). Abbreviations: ba1–7, branchial arches 1 to 7 (with "-r" and "-l" meaning right and left, respectively); cc, cellular chamber; cs, C-shaped structure; f, filament; fs, first-order septa; lss, lobe-shaped structure; mo, mouth opening; us, U-shaped structure. Scale bars: 1 mm in (B), (C), and (F) to (H); and 200 μm in (D).

from collagen fibrils, which have thicker diameters ranging from 20 to 500 nm and repeating banding patterns (every ~67 nm) instead of beads (*20*). Unlike collagen fibrils that are the major component of cartilages in jawed vertebrates (gnathostomes) (*20*, *21*), fibrillin microfibrils are the dominant extracellular matrix in some parts of cyclostome skeletons. The prominent examples occur in the first two pharyngeal arches of larval lampreys (*17*) and in the posterior lingual cartilage of a hagfish (*18*). Similar matrix microfibrils also characterize the pharyngeal skeletons of extant cephalochordates (*21*). The common presence of microfibrils further confirms that yunnanozoan branchial arches are cartilaginous, and this cartilage shares both anatomical and ultrastructural characteristics with that of primitive vertebrates and cephalochordates. The attached gill filaments with matrix microfibrils (fig. S3I) are possibly cartilaginous gill ray–like structures, as previously interpreted (*12*).

Turning next to the horizontal rods that connect the branchial arches in yunnanozoans: The presence of the microfibrils and putative septa again indicate their cartilaginous nature. As skeletal elements, the horizontal rods perhaps compare to the subchordal and hypobranchial bars in the branchial basket of lampreys (Fig. 4, B and C) (*15*). Comparable horizontal elements also existed in the stem hagfish *Myxinikela* (*22*) and the fossil jawless fish *Euphanerops* (*23*).

These new findings reveal that the branchial skeleton of yunnanozoans shares three characteristics broadly present in crown-group vertebrates: dominance of microfibrils in the matrix of a cartilage with chondrocytes, branchial arches with stacked discoid structures, and presence of the horizontal bars at both ends of the branchial arches. We ran phylogenetic analyses on a newly combined set of characters, both without and with the character modifications demanded by the findings of this study (fig. S6). These analyses support that yunnanozoans are the earliest branching stem vertebrates (Fig. 4A and fig. S6B).

The branchial arches of yunnanozoans also shed light on the early evolution of the vertebrate pharyngeal arches. Because the first arch in yunnanozoans is lateral to the circumoral structure and identical to the remaining arches, it completes a serial pattern of similar arches that is also seen in another stem vertebrate, *Metaspriggina* (Fig. 4) (*5*). Yunnanozoans are consistent with the conventional hypothesis that every pharyngeal

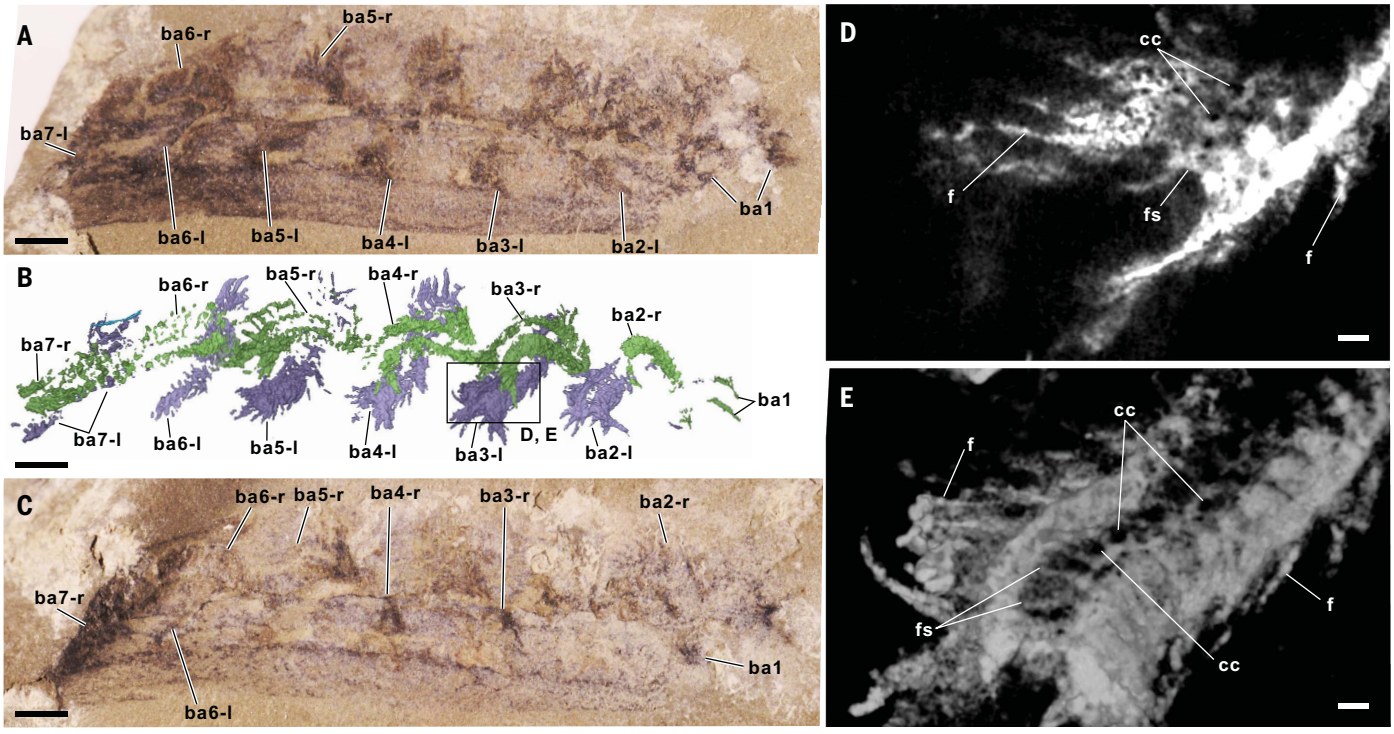


Fig. 2. X-ray computed microtomography of the branchial arches. (**A** to **C**) Photomicrographs of part (A) and mirrored counterpart (C) of specimen NIGP 176268 and its three-dimensional micro-CT reconstruction (B). (**D** and **E**) A selected slice and rendered image of the third left branchial arch show the cellular chambers. Abbreviations as in Fig. 1. Scale bars: 1 mm in (A) to (C), and 100 μm in (D) and (E).

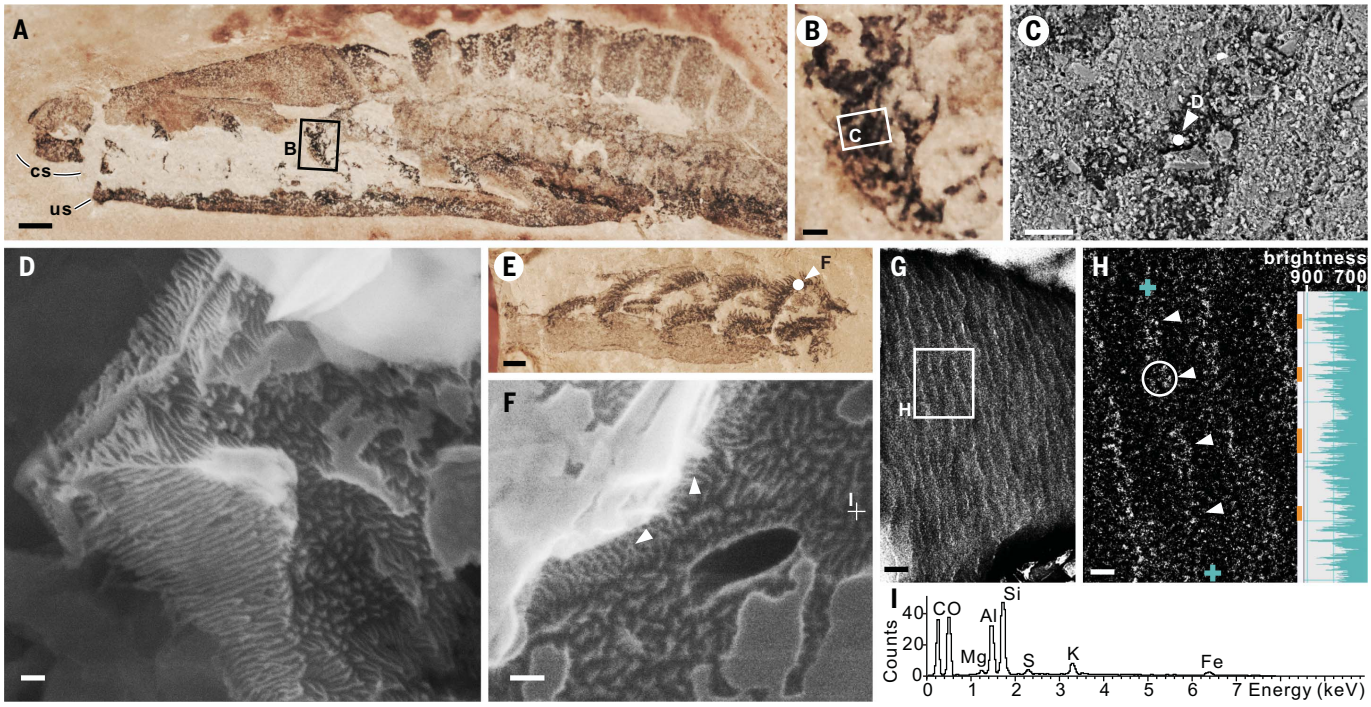


Fig. 3. Ultrastructure of the branchial arches of yunnanozoans. (**A**) An overview of the branchial arches of specimen NIGP 176258a. Abbreviations as in Fig. 1. (**B**) Enlarged view of the fifth arch [the boxed area in (A)]. (**C**) Backscattered scanning electron microscopy image of a straight septum [the boxed area in (B)]. (**D**) Secondary electron image shows that the carbon residues consist of bundles of microfibrils. (**E**) An overview of the branchial arches of specimen NIGP 176255. (**F**) Secondary electron image of the white point area in (E) shows the cross-linkages between the microfibrils (white arrowheads). (**G** and **H**) Transmission electron images show the microfibrils and the possible beads at 50-nm intervals (white arrowheads, one bead is circled) in specimen NIGP 176263. The inset in (H) is the brightness profile between the two cyan crosses, with positions of the possible beads marked by orange bars. (**I**) Energy dispersive spectrometry spectrum based on SEM at the white cross in (F). Scale bars: 1 mm in (A) and (E); 200 μm in (B); 50 μm in (C); 100 nm in (D), (F), and (G); and 20 nm in (H).

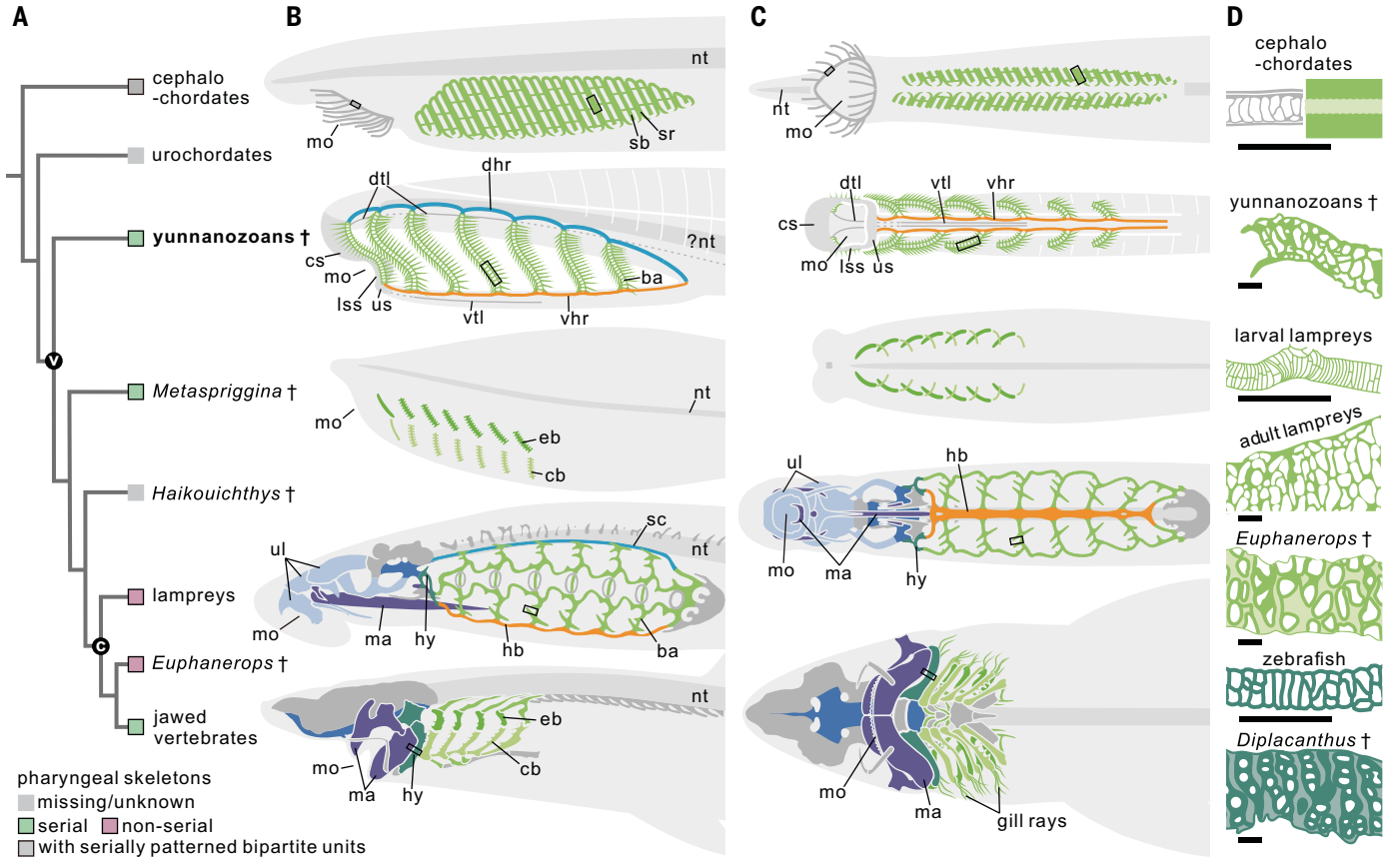


Fig. 4. Phylogenetic relationships and morphological comparisons of pharyngeal skeletons among representative clades of chordates. (**A**) Simplified consensus tree from Bayesian analyses (from fig. S6B), with morphological state of pharyngeal skeletons shown at the tips. This state is indicated as "serial" (green box), meaning the first arch resembles the other arches to form a series, or as "nonserial" (purple box), meaning the first arch has extensive skeleton and distinct from other arches. Cephalochordates are colored in gray, indicating that they have two alternate units of serially patterned arches. Encircled "V" indicates total-group vertebrates; encircled "C" means crown vertebrates. Fossil taxa are indicated with a single-dagger symbol. (**B** and **C**) Left and ventral views of the pharyngeal skeletons. (**D**) Chondrocyte arrangement in the boxed areas of the pharyngeal skeletons in (B) and (C). The left half in cephalochordates is from oral cirri, whereas the right half is from pharyngeal skeletons. Abbreviations: ba, branchial arches; cb, ceratobranchials; dhr/vhr, dorsal/ventral horizontal rod in yunnanozoans; dtl/vtl, dorsal/ventral thin horizontal line in yunnanozoans; eb, epibranchials; hb, hypobranchial bars in lamprey; hy, hyoid skeleton; ma, mandibular skeleton; nt, notochord; sc, subchordal bar in lamprey; sr/sb, skeleton in primary/secondary bar in cephalochordates; ul, upper lip cartilages in lamprey. For other abbreviations, see Fig. 1. Scale bars: 50 μm in (D). Images in (B) and (C) are not scaled. References and additional explanation of the drawings are available in supplementary text.

arch of the ancestral vertebrates was a typical branchial arch (*24*).

The yunnanozoan basket, with its successive, unjointed branchial arches that are connected by dorsal and ventral horizontal rods, is cyclostome-like (Fig. 4, B and C). Indeed, many researchers claim that the ancestral vertebrates had such an external basket (*3*) and not the deeper, jointed, and less interconnected arches seen in jawed fishes. This scenario, however, is challenged by the isolated and bipartite branchial bars in the higher stem vertebrates from the Cambrian, *Metaspriggina* and its relatives, the branchial skeletons of which are more gnathostome-like (Fig. 4) (*5*). This either implies that the cyclostome basket was secondarily derived and thus is merely convergent on the original, yunnanozoan basket—which is unlikely, given the real similarities we found here—or else it implies that *Metaspriggina* acquired a pre-gnathostome condition, having already lost the ancestral basket and gained the more separated arches.

REFERENCES AND NOTES

1. J. Mallatt, *J. Zool.* **204**, 169–183 (1984).
2. J. Mallatt, *Zoolog. Sci.* **25**, 990–998 (2008).
3. T. Miyashita, *Biol. Rev. Camb. Philos. Soc.* **91**, 611–657 (2016).
4. C. Hirschberger, V. A. Sleight, K. E. Criswell, S. J. Clark, J. A. Gillis, *Mol. Biol. Evol.* **38**, 4187–4204 (2021).
5. S. C. Morris, J.-B. Caron, *Nature* **512**, 419–422 (2014).
6. C. Yang, X.-H. Li, M. Zhu, D. J. Condon, J. Chen, *J. Geol. Soc. London* **175**, 659–666 (2018).
7. X. Hou, L. Ramskold, J. A. N. Bergstrom, *Zool. Scr.* **20**, 395–411 (1991).
8. J.-Y. Chen, J. Dzik, G. Edgecombe, L. Ramsköld, G. Q. Zhou, *Nature* **377**, 720–722 (1995).
9. D. Shu, X. Zhang, L. Chen, *Nature* **380**, 428–430 (1996).
10. J.-Y. Chen, D.-Y. Huang, C.-W. Li, *Nature* **402**, 518–522 (1999).
11. D. Shu *et al.*, *Science* **299**, 1380–1384 (2003).
12. J. Mallatt, J. Y. Chen, *J. Morphol.* **258**, 1–31 (2003).
13. P.-Y. Cong, X.-G. Hou, R. J. Aldridge, M. A. Purnell, Y. Z. Li, *Palaeontology* **58**, 45–70 (2015).
14. X. Hou, P. Cong, Y. Li, *Acta Palaeontol. Sin.* **48**, 402–413 (2009).
15. W. M. Martin, L. A. Bumm, D. W. McCauley, *Dev. Dyn.* **238**, 3126–3138 (2009).
16. C. B. Kimmel *et al.*, *Dev. Biol.* **203**, 245–263 (1998).
17. G. M. Wright, J. H. Youson, *Am. J. Anat.* **165**, 39–51 (1982).
18. G. M. Wright, F. W. Keeley, J. H. Youson, D. L. Babineau, *Am. J. Anat.* **169**, 407–424 (1984).
19. E. C. Davis, R. A. Roth, J. E. Heuser, R. P. Mecham, *J. Struct. Biol.* **139**, 65–75 (2002).
20. D. E. Birk, P. Brückner, in *The Extracellular Matrix: an Overview*, R. P Mecham, Ed. (Biology of Extracellular Matrix Series, Springer, Berlin Heidelberg, 2011), pp. 77–115.
21. G. M. Wright, F. W. Keeley, P. Robson, *Cell Tissue Res.* **304**, 165–174 (2001).
22. T. Miyashita, *Can. J. Zool.* **98**, 850–865 (2020).
23. P. Janvier, M. Arsenault, *Geodiversitas* **29**, 143–216 (2007).
24. G. De Beer, *The Development of the Vertebrate Skull* (Clarendon Press, 1937).

ACKNOWLEDGMENTS

We acknowledge M. Benton for comments and Z. Yang and Y. Wang for discussion. We also acknowledge the following individuals for their laboratory assistance: J. Tang, C. Wang, Y. Fang, and S. Wu at the Nanjing Institute of Geology and

Palaeontology, Chinese Academy of Sciences (NIGPAS); J. Chen, H. Liu, P. Jiang, and Z. Xia at the State Key Laboratory for Mineral Deposits Research, Nanjing University; G. Wang at Shanghai Jiao Tong University; and B. Wang, X. Lei, X. Chen, L. Hu, Z. Yin, R. Wu, T. Zhao, W. Zhang, and X. Zhao. We are grateful to two anonymous reviewers for their constructive comments and suggestions. **Funding:** Funding was provided by the Strategic Priority Research Program (B) of the Chinese Academy of Sciences [XDB26000000 (B.J., F.Z., and M.Z.)], the National Science Foundation of China [42288201 (B.J.); 41921002 and 42072006 (F.Z. and M.Z.)], and the Fundamental Research Funds for the Central Universities [0206-14380137 (B.J.)]. **Author contributions:** Conceptualization: F.Z. and B.J. Fossil collection: F.Z. and M.Z. Investigation: Q.T., F.Z., H.Z., and B.J. Funding acquisition: F.Z., M.Z., and B.J. Writing – original draft: Q.T., F.Z., and B.J. Writing – review & editing: Q.T., F.Z., H.Z., M.Z., and B.J. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The specimens studied here are all stored in NIGPAS. The measurements and datasets for the phylogenetic analysis are available in the supplementary materials. Computed tomography data are deposited in MorphoSource at https://www.morphosource.org/projects/000378422?locale=en.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm2708
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S4
References (*25*–*64*)
MDAR Reproducibility Checklist
Data S1 to S3

Submitted 5 September 2021; accepted 17 May 2022
10.1126/science.abm2708

STRUCTURAL BIOLOGY

Conformational selection guides β-arrestin recruitment at a biased G protein–coupled receptor

Andrew B. Kleist[1,2]†‡, Shawn Jenjak[1]†, Andrija Sente[3], Lauren J. Laskowski[4], Martyna Szpakowska[5], Maggie M. Calkins[4], Emilie I. Anderson[4], Lisa M. McNally[4], Raimond Heukers[6]§, Vladimir Bobkov[6], Francis C. Peterson[1], Monica A. Thomas[1,2]¶, Andy Chevigné[5], Martine J. Smit[6], John D. McCorvy[4], M. Madan Babu[7,8], Brian F. Volkman[1]*

G protein–coupled receptors (GPCRs) recruit β-arrestins to coordinate diverse cellular processes, but the structural dynamics driving this process are poorly understood. Atypical chemokine receptors (ACKRs) are intrinsically biased GPCRs that engage β-arrestins but not G proteins, making them a model system for investigating the structural basis of β-arrestin recruitment. Here, we performed nuclear magnetic resonance (NMR) experiments on $^{13}CH_3$-ε–methionine–labeled ACKR3, revealing that β-arrestin recruitment is associated with conformational exchange at key regions of the extracellular ligand-binding pocket and intracellular β-arrestin–coupling region. NMR studies of ACKR3 mutants defective in β-arrestin recruitment identified an allosteric hub in the receptor core that coordinates transitions among heterogeneously populated and selected conformational states. Our data suggest that conformational selection guides β-arrestin recruitment by tuning receptor dynamics at intracellular and extracellular regions.

Once thought of as G protein–coupled receptor (GPCR) off-switches, β-arrestins are now known to coordinate diverse, G protein–independent signaling responses (*1*). In a phenomenon called biased signaling,some GPCR ligands (biased ligands) preferentially recruit β-arrestins or G proteins (*2*). Because they select one pathway, and therefore a specific functional outcome, biased ligands have demonstrated advantages over conventional ligands in preclinical (*3*–*5*) and clinical (*6*, *7*) testing. More than 30% of US Food & Drug Administration–approved drugs target GPCRs (*8*), so understanding the mechanistic basis by which GPCRs recruit β-arrestins has important implications for drug development.

Recent GPCR–β-arrestin structures (*9*–*11*) and biophysical studies of β-arrestin–biased ligands (*12*–*14*) provide mechanistic insights into β-arrestin recruitment. For instance, structural comparisons of GPCRs bound to antagonists and β-arrestin–biased ligands suggest that ligand-specific conformational changes govern β-arrestin recruitment (*12*, *13*). Despite this progress, how ligand-binding pocket changes are transmitted to the β-arrestin interface remains poorly understood. Other findings have complicated our understanding of β-arrestin recruitment by GPCRs. The identification of multiple, distinct β-arrestin–recruiting GPCR conformations suggests there may be multiple conformational solutions to β-arrestin recruitment (*15*), but raises the question of how such different conformations could elicit similar outcomes. Conversely, nearly identical conformations of β-arrestin– and G protein–bound GPCRs (*9*–*11*) suggest that conformational changes alone might not account for β-arrestin recruitment. Finally, some studies of β-arrestin recruitment use ligands that provoke residual signaling at G proteins (*16*), making it challenging to isolate the specific molecular changes leading to β-arrestin recruitment.

Given these complexities, studies of β-arrestin recruitment could benefit from functionally decoupled receptors that exclusively recruit β-arrestin. Atypical chemokine receptors (ACKRs) represent one such naturally occurring system. ACKR3 is an intrinsically β-arrestin–biased GPCR that recruits β-arrestin but does not activate G protein (Fig. 1A) (*17*). Here, we performed nuclear magnetic resonance (NMR) studies of ACKR3, showing that β-arrestin recruitment at an intrinsically biased GPCR is guided by tuning its conformational equilibrium.

RESULTS

In agreement with prior studies, ACKR3 recruits β-arrestin but does not activate G protein signaling (Fig. 1, A and B; fig. S1A; and table S1). To investigate the mechanisms underlying β-arrestin recruitment at ACKR3, we first measured β-arrestin-2 recruitment in response to a panel of ACKR3 agonists, including the endogenous chemokine CXCL12 [median effective concentration (EC_{50}) = 0.75 nM] (*18*), the small molecule CCX777 (EC_{50} = 0.95 nM; maximum effect (E_{max}, % of CXCL12) = 75 ± 2%] (*19*), and a recently described peptide LIH383 (EC_{50} = 4.8 nM; E_{max} = 83 ± 1%) (Fig. 1C) (*20*). We also identified a potent, extracellular-targeting, ACKR3-competitive antagonist nanobody called VUN701 [median inhibitory concentration (IC_{50}) = 1.47 μM] using nanobody phage display (Fig. 1C; fig. S1, B to F; and materials and methods). These four ligands display a range of activities for β-arrestin-2 recruitment, allowing us to sample inactive (i.e., VUN701-bound) and active (i.e., CCX777-, LIH383-, and CXCL12-bound) β-arrestin–recruiting states of ACKR3 (Fig. 1D).

NMR spectroscopy allows simultaneous characterization of receptor conformation at multiple sites (*21*). We previously purified $^{13}CH_3$-ε-Met–labeled, ACKR3-bound CCX777 in lauryl maltose neopentyl glycol/cholesteryl hemisuccinate (MNG/CHS) micelles for NMR studies (*22*), and we used this method for other ACKR3-ligand complexes here (Fig. 1E, figs. S2 and S3, and materials and methods). As with other GPCRs, copurification with ligands was necessary to produce sufficient quantities of ACKR3 for NMR studies (*23*). Structural homology modeling of ACKR3 (fig. S4 and materials and methods) demonstrates eight native Met residues (excluding the N-terminal Met) distributed in ACKR3's tertiary structure for NMR labeling. NMR labels serve as sensitive probes reporting on regional changes in GPCR conformation and dynamics (*24*, *25*). In this setting, $Met212^{5x39}$ and $Met138^{3x46}$ [where the superscripts indicate GPCRdb nomenclature (*26*)] are well positioned to report on conformational changes associated with β-arrestin recruitment at ACKR3 because they are located in key regions of the ligand-binding pocket (*27*, *28*) and intracellular effector binding interface (*29*), respectively.

Ligand interactions with TM5 play key roles in GPCR activation (*27*, *30*), but how conformational changes in TM5 lead to β-arrestin recruitment is unclear. NMR spectra for the four ligand-bound states show chemical shift perturbations (CSPs) for the $Met212^{5x39}$ peak, which remains colinear along the ^1H axis but variable along the ^{13}C axis (Fig. 2, A and B). This indicates that ACKR3 exists along a two-state equilibrium, with the peaks at either extreme defining the conformational end points (*31*). In the CCX777-bound state, downfield peak positions [~18.2 to 19 parts per million (p.p.m.)] indicate that $^{13}CH_3$ is in a *trans* rotamer (end point 1), whereas in the CXCL12- and LIH383-bound states, upfield peak positions

[1]Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [2]Medical Scientist Training Program, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [3]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [4]Department of Cell Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, WI 53226, USA. [5]Immuno-Pharmacology and Interactomics, Department of Infection and Immunity, Luxembourg Institute of Health (LIH), L-4354 Esch-sur-Alzette, Luxembourg. [6]Amsterdam Institute for Molecular and Life Sciences, Division of Medicinal Chemistry, Faculty of Sciences, Vrije Universiteit, 1081 HZ Amsterdam, Netherlands. [7]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. [8]Center for Data Driven Discovery, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
*Corresponding author. Email: bvolkman@mcw.edu
†These authors contributed equally to this work. ‡Present address: Harriet Lane Pediatric Residency Program, Johns Hopkins Children's Center, Baltimore, MD 21287, USA. §Present address: QVQ Holding BV, 3584 CL Utrecht, Netherlands. ¶Present address: Department of Anesthesiology and Critical Care Medicine, Johns Hopkins Medicine, Baltimore, MD 21287, USA.

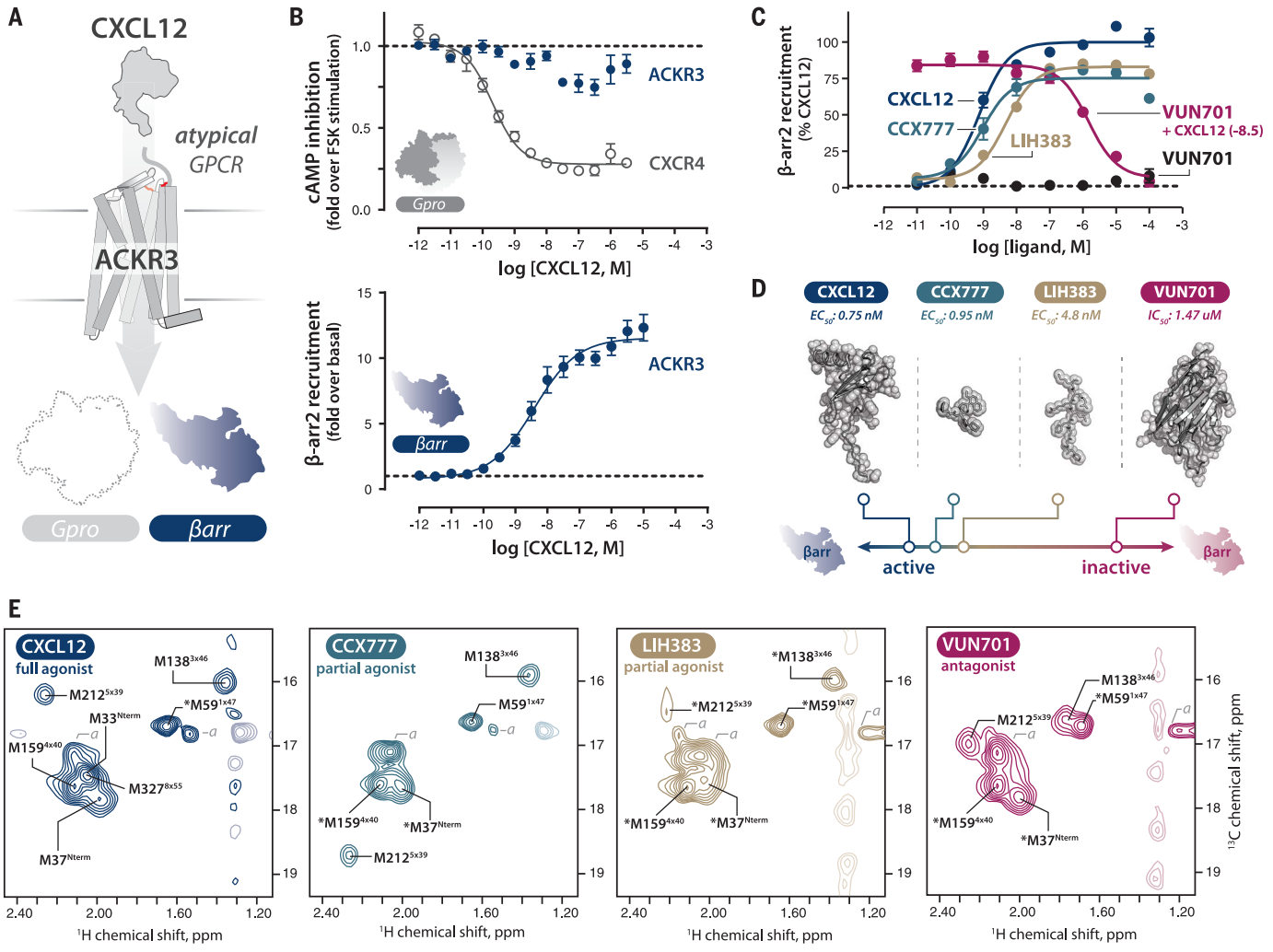


Fig. 1. β-arrestin–biased signaling at ACKR3 and structural characterization by NMR. (**A**) CXCL12 activates β-arrestin but not G protein at ACKR3. (**B**) CXCL12-mediated cAMP inhibition of CXCR4 (positive control) and ACKR3 as shown by Glosensor assay (EC_{50} = 0.21 nM, CXCR4) (top). CXCL12-mediated β-arrestin-2 recruitment to ACKR3 as shown by Tango assay (EC_{50} = 3.9 nM, ACKR3) (bottom). N = 3 in triplicate in both assays. Error bars indicate SEM. (**C**) β-arrestin recruitment to ACKR3 as shown by Nano-BiT assay. VUN701 dose response alone could not be fit (black circles). Purple circles reflect VUN701 dose response with CXCL12 at 3.2 nM . See text and table S1 for EC_{50} and E_{max}. All conditions, N = 3 in duplicate. Error bars indicate SEM. (**D**) Summary of ligand potency at β-arrestin recruitment. Structures are not to scale. CXCL12 PDB: 2KEC; nanobody from PDB 6KNM used to represent VUN701, and LIH383 and CCX777 were modeled in PyMol. LIH383 sequence is FGGFMRRK (*20*). The chemical structure of CCX777 is shown in (*19*). (**E**) ^{1}H-^{13}C heteronuclear single quantum coherence NMR spectra of WT-ACKR3 with various ligands at 310 K. Assigned Met residues are labeled. Asterisk denotes inferred assignments from other ligand-bound states (see the materials and methods). Negative contour peaks are shown in semitransparency and dashed lines. Peaks marked "*a*" encompass natural abundance peaks from buffer and detergent components (see also figs. S2G and S3). All spectra are shown at the same contour except LIH383, which was lowered to represent $M212^{5x39}$.

(~16 to 16.5 p.p.m.) indicate that $^{13}CH_3$ is in a *gauche* rotamer (end point 2) (Fig. 2B and fig. S5, A and B) (*22*, *32*–*34*).

By contrast, in VUN701-bound ACKR3, the $Met212^{5x39}$ peak appears between the two end points at the random coil $^{13}CH_3$ position, indicating fast conformational exchange between *gauche* and *trans* rotamers in the inactive, β-arrestin–nonrecruiting state (Fig. 2B). Indeed, CSPs of $Met212^{5x39}$ from $^{13}CH_3$ random coil vary by ligand type (fig. S5A), suggesting that β-arrestin recruitment is associated with decreasing exchange among $Met212^{5x39}$ rotamers irrespective of which χ_3 rotameric state is selected. Antagonists such as VUN701 may fail to constrain the region sampled by $Met212^{5x39}$, preserving conformational heterogeneity (seen by increased exchange relative to the active state) and precluding TM5-binding pocket stabilization required for ACKR3 activation. Binding data show that Ala mutagenesis of $Met212^{5x39}$ similarly affects CXCL12 and VUN701, suggesting that NMR changes between inactive and active states likely represent functionally important regional conformational changes, as opposed to differences in direct ligand contacts at the probe residue (fig. S5, C to E, and table S1). Indeed, $Met212Ala^{5x39}$ has equivalent efficacy for β-arrestin recruitment as wild type (WT) (fig. S5F), indicating that $Met212^{5x39}$ reports on but does not itself mediate functional changes in the TM5 region.

Does this occur in other GPCRs? We identified all ligand-contacting residues among structures of (i) GPCRs bound to β-arrestin–biased ligands (or full-length β-arrestin; jointly called β-arresting–recruiting states) and (ii) inactive-state structures of the same GPCRs (fig. S6, A and B). Using contact network analysis (*29*, *35*), we found that ligands in β-arrestin–recruiting states make similar numbers of contacts with residues in TM5 but significantly

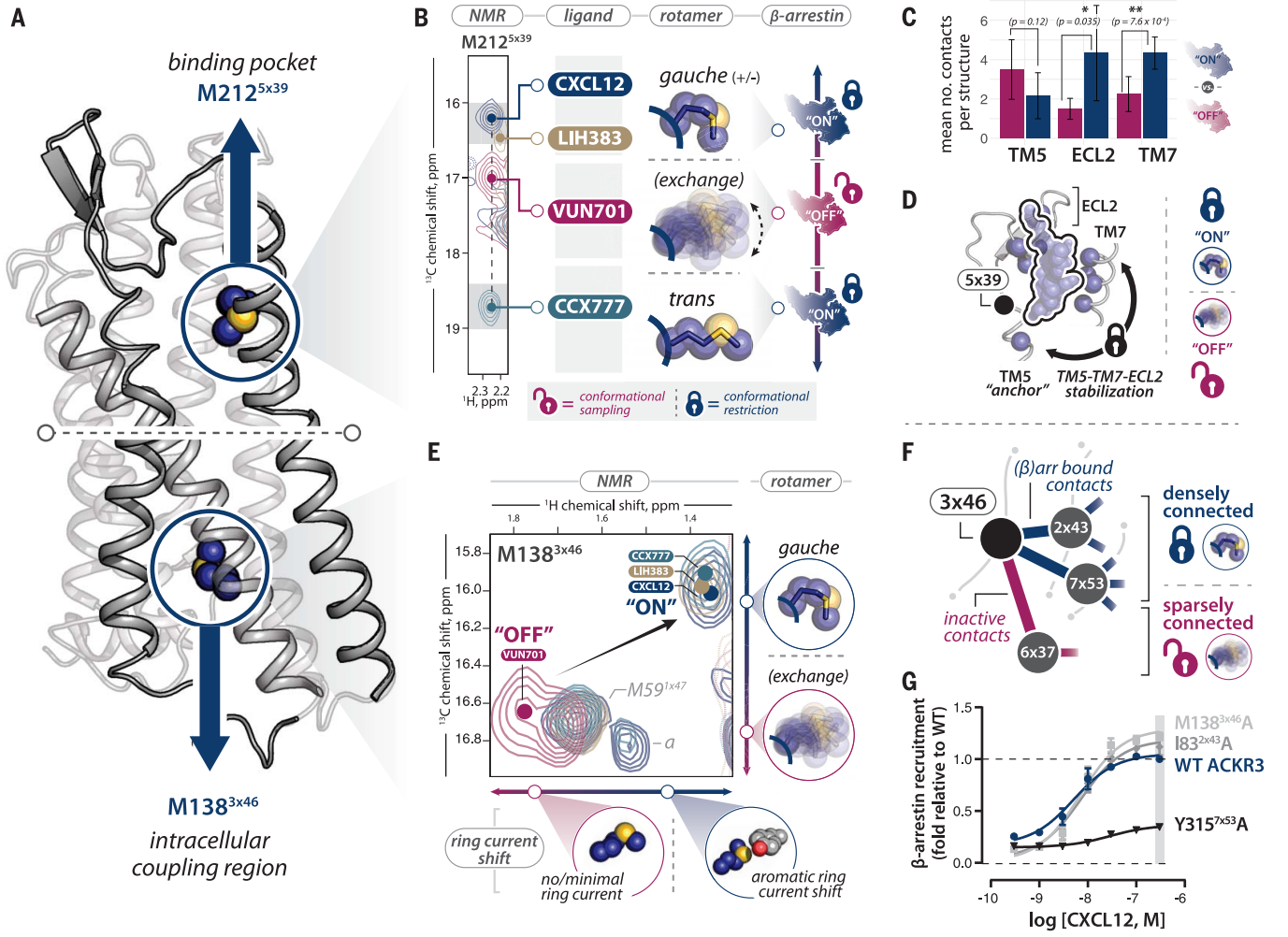


Fig. 2. Conformational changes in the ligand-binding pocket and intracellular region characterize the β-arrestin–recruiting state. (**A**) ACKR3 model depicting M212^{5x39} and M138^{3x46} probes. (**B**) Overlay of M212^{5x39} peaks from ligand-bound ACKR3 complexes at 310 K. The 16 to 16.5 and 18.5 to 19 p.p.m. peaks (^{13}C) correspond to *gauche* and *trans* rotameric states, respectively. Ligand-specific β-arrestin activity is depicted at right. Open and closed locks depict conformational sampling and restriction, respectively (bottom). (**C**) Ligand-residue interactions were compared between antagonist and β-arrestin–biased ligand-bound GPCRs (see the materials and methods). Shown is a comparison of the mean number of ligand contacts with TM5, ECL2, and TM7 residues. *P < 0.05; **P < 0.005, unpaired t test. (**D**) AT$_1$R contacts with the ligand TRV023 (β-arrestin–biased; black outline) in TM5, ECL2, and TM7 are shown as blue spheres (PDB 6OS1). The position of NMR probe 5x39 in AT$_1$R is shown. Stabilization of TM5-ECL2-TM7 by biased ligands is depicted as a lock (right). (**E**) Met138^{3x46} peaks in all four ligand-bound states. Upfield peak positions (^{1}H: ~1.3 p.p.m.) of M138^{3x46} among agonist-bound states supports ring-current shifts due to aromatic side chain interactions. Peaks are marked "*a*" as in Fig. 1. (**F**) 3x46 contact residues at 2x43 and 7x53 exclusively in active-state complexes and 6x37 exclusively in inactive-state complexes (see the materials and methods). (**G**) β-arrestin recruitment of WT ACKR3 versus Ala mutants of 3x46-contacting mutants by NanoBiT (N = 3). See also table S1.

more contacts with residues in ECL2 and TM7 than do antagonists (Fig. 2C). Thus, TM5 may anchor agonist interactions with ECL2 and TM7 (*27*, *36*), creating a tripartite lock involving the 5x39 region (Fig. 2D and fig. S6C), whereas antagonists fail to stabilize all three lock points (ECL2, TM5, and TM7) simultaneously.

Large-scale conformational changes underlying GPCR activation are driven by local rearrangements of residues at key positions in the GPCR structure (so-called microswitch residues (*37*). We next examined Met138^{3x46}, a microswitch residue in the β-arrestin–coupling region (Fig. 2A) (*29*). In the three agonist-bound states (CXCL12-, CCX777-, and LIH383-bound), the Met138^{3x46} NMR peak has nearly identical chemical shifts (Fig. 2E), indicating that all three agonists elicit a shared conformation of Met138^{3x46}. The upfield position (~1.3 p.p.m.) of Met138^{3x46} in the ^{1}H dimension likely reflects ring current shifts caused by proximity to an aromatic side chain (*33*, *38*, *39*). These data indicate that the agonist-specific conformations of Met212^{5x39} (i.e., *trans* versus *gauche*) in the TM5-binding pocket are funneled into a common conformation at the β-arrestin–coupling region sampled by M138^{3x46}.

By contrast, the inactive, VUN701-bound ACKR3 state is characterized by CSPs of the Met138^{3x46} peak in both ^{1}H and ^{13}C values (Fig. 2E), indicating diminished ring-current shift (^{1}H dimension), depletion of the *gauche* rotamer (^{13}C dimension), and more *gauche*/*trans* exchange (^{13}C dimension). As in the binding pocket, VUN701 causes the smallest CSP from random coil $^{13}CH_3$-ε-Met versus the three agonists, suggesting that agonists regulate β-arrestin recruitment by decreasing

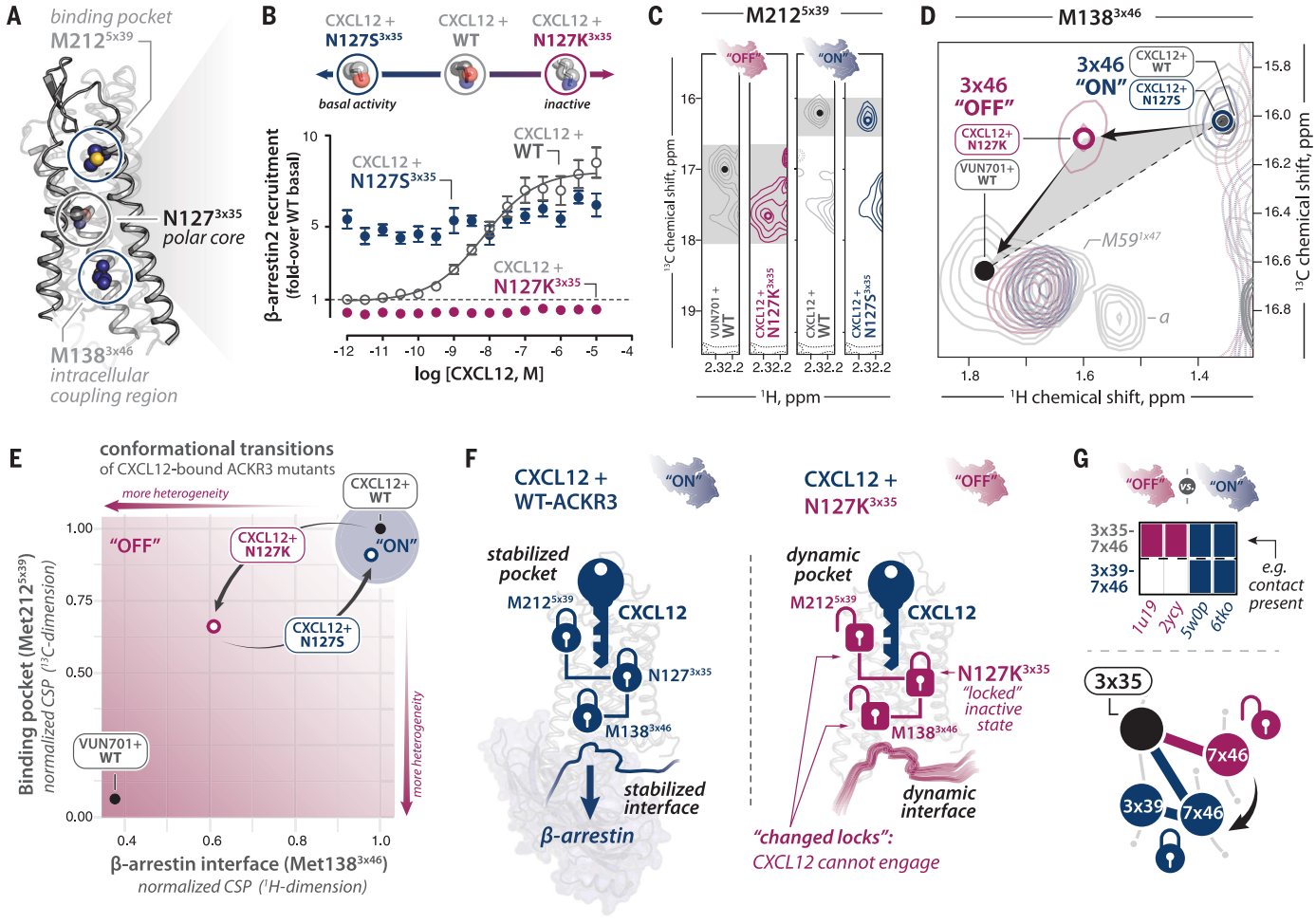


Fig. 3. Mutational inactivation of β-arrestin recruitment through the conserved polar network. (**A**) Position of Asn127^{3x35} in the ACKR3 model relative to Met212^{5x39} and Met138^{3x46}. (**B**) β-arrestin-2 recruitment with CXCL12 for the ACKR3 mutants Asn127Lys^{3x35} and Asn127Ser^{3x35} as shown by the Tango assay. All conditions, *N* = 3 in triplicate. Error bars indicate SEM. See also table S1. (**C**) Close-up view of ^{1}H-^{13}C-HSQC in the Met212^{5x39} region (right), highlighting VUN701-WT-ACKR3 (first panel), CXCL12-ACKR3 Asn127Lys^{3x35} (second panel), CXCL12-WT-ACKR3 (third panel), and CXCL12-ACKR3 Asn127Ser^{3x35} (fourth panel). (**D**) Overlay of ^{13}C-HSQC in the Met138^{3x46} region for WT-ACKR3-VUN701, ACKR3-Asn127Lys^{3x35}-CXCL12, ACKR3-Asn127Ser^{3x35}-CXCL12, and WT-ACKR3-CXCL12 complexes. A shaded triangle suggests peak noncolinearity. Peaks are marked "*a*" as in Fig. 1. See fig. S8A and legend for the Met138^{3x46} assignment method. (**E**) Normalized CSPs for Met212^{5x39} in the ^{13}C dimension (*y* axis) and Met138^{3x46} in the ^{1}H dimension (*x* axis) from random coil for Asn127^{3x35} mutants and CXCL12- and VUN701-bound WT-ACKR3. Arrows depict transitions among CXCL12-bound WT-ACKR3 and mutant ACKR3. (**F**) Depiction of differences between WT-ACKR3 and Asn127Lys^{3x35} in CXCL12-bound states. Despite being bound to CXCL12 (blue key), Asn127Lys^{3x35} locks ACKR3 in the inactive state, abrogating CXCL12's effects on the Met212^{5x39} and Met138^{3x46} probes. (**G**) Comparison of residue-residue interactions AT_1R–β-arrestin–biased ligand and AT_1R-antagonist bound states (top) reveals the formation of a 3x39–7x46 interaction in the β-arrestin state (bottom).

conformational heterogeneity in this region of the β-arrestin–coupling region relative to the inactive state (fig. S7A).

What is the role of 3x46, a known microswitch involved in G protein activation (*29*), in β-arrestin recruitment among other GPCRs? We calculated all intramolecular, residue-residue interactions among structures of (i) β-arrestin-bound and (ii) inactive-state structures of GPCRs with resolved side chains (β_1AR and rhodopsin) (fig. S7, B and C), finding that 3x46 contacts 2x43 and 7x53 in β-arrestin-bound structures and 6x37 in inactive structures (Fig. 2F and fig. S7, D to F). Of note 7x53 is a highly conserved Tyr among class A GPCRs (*37*), and 7x53 and 6x37 have been shown to stabilize GPCRs in G protein signaling (7x53) and inactive states (6x37), respectively (*29*).

To investigate whether 3x46 and active-state contacts from GPCR-β–arrestin complexes (2x43 and 7x53) might play a role in β-arrestin recruitment to ACKR3, we tested Met138Ala^{3x46}, Ile84Ala^{2x43}, and Tyr315Ala^{7x53} mutants for β-arrestin recruitment (Fig. 2G and table S1). Although Met138Ala^{3x46} and Ile84Ala^{2x43} minimally affected ACKR3 function, Tyr315Ala^{7x53} almost completely abolished β-arrestin recruitment, demonstrating the essential role for this residue on ACKR3 activation. The close proximity of 3x46 to Tyr^{7x53} (Fig. 2F and fig. S7, D to F) and the pronounced ring current shift (^{1}H) of Met138^{3x46} (Fig. 2E) in active but not inactive states of ACKR3 suggest that Met138^{3x46} reports on dynamic alterations in Tyr315^{7x53} associated with β-arrestin recruitment (*33*, *38*, *39*). Active-state interactions between 3x46 and 7x53 may thus lock this region in place, decreasing local conformational heterogeneity and increasing the likelihood of β-arrestin engagement. Indeed, receptors contact the β-arrestin finger loop, a key structural region anchoring GPCR-β–arrestin interactions, directly at or in the vicinity of 3x46 and 7x53 in β_1AR- and rhodopsin-arrestin complexes (*11*, *40*). By contrast, a more heterogeneous

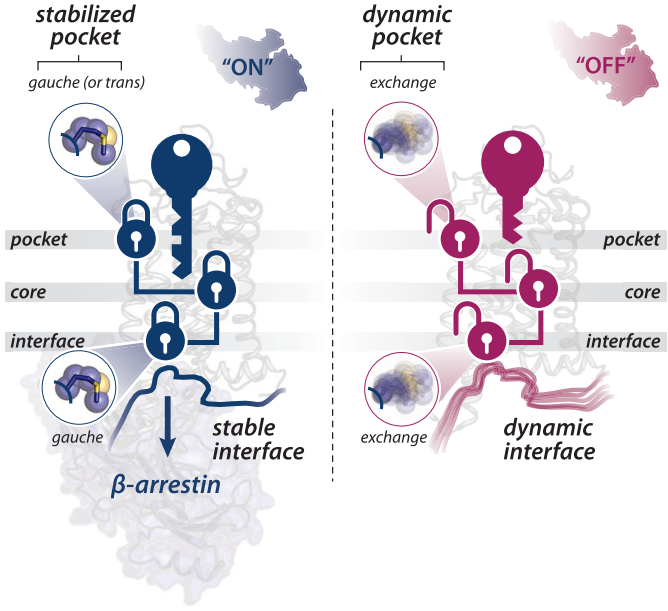


Fig. 4. Dynamic control of β-arrestin recruitment to ACKR3. Allosteric regulation of GPCR β-arrestin activation by coordinated transitions in conformational heterogeneity. Ligands constrain (agonists) or promote (antagonists) conformational heterogeneity by altering intermolecular (ligand-residue) and intramolecular (residue-residue) interactions throughout the GPCR structure.

inactive state (characterized by rotameric exchange and a diminished ring current shift at Met138^{3x46}) may reflect absent interactions with 7x53 and a lower propensity for β-arrestin engagement (Fig. 2F).

To what extent are dynamic changes in extracellular (Met212^{5x39}) and intracellular (Met138^{3x46}) probes allosterically linked through the receptor core during ACKR3 activation? Conserved polar core residues, including the residue Asn^{3x35} (*12*, *41*, *42*), coordinate a sodium ion (inactive state) and a water network (active state) to regulate GPCR activation and bias (Fig. 3A) (*42*). We investigated the effects of known inactivating and constitutively activating mutations of conserved polar core residue Asn127^{3x35} on ACKR3 activation. Consistent with prior results, the Asn127Lys^{3x35} mutation showed complete inactivity in β-arrestin recruitment (Fig. 3B), whereas Asn127Ser^{3x35} acts as a constitutively active mutant (Fig. 3B and table S1) (*43*).

How do Asn127^{3x35} mutants affect the conformation and stability of the binding pocket and intracellular probes as observed by NMR chemical shifts? In the context of CXCL12, the Asn127Lys^{3x35} mutation shifts the Met212^{5x39} peak downfield relative to its position in the WT-ACKR3 complex (Fig. 3C and fig. S8A), resembling the VUN701-bound state characterized by *trans*/*gauche* rotamer exchange at Met212^{5x39}. This reveals that the receptor core mutation exerts long-range, allosteric effects that enhance conformational exchange in the extracellular ligand-binding pocket. The Asn127Lys^{3x35} mutation shifts the Met138^{3x46} peak downfield in both the ^{1}H and ^{13}C values to an intermediate position between those in CXCL12- and VUN701-bound states (Fig. 3D). This peak is not colinear with CXCL12 and VUN701, suggesting the occupancy of a third, distinct conformation in this ACKR3 mutant. Although Met138^{3x46} in this mutant occupies a *gauche* rotamer (^{13}C: ~16.1 p.p.m.), its ^{1}H value is closer to that of VUN701-bound ACKR3, suggesting fewer local aromatic interactions resulting in a diminished ring current shift. We speculate that this downfield ^{1}H shift reflects increased conformational exchange between both *gauche* rotamers (i.e., + and −) in the mutant state, which would weaken interactions with an aromatic side chain compared with a single *gauche* rotamer (i.e., + or −) in the active state (fig. S8B). Regardless, pronounced downfield ^{1}H shifts and peak broadening of Met138^{3x46} in the CXCL12-bound, inactivating Asn127Lys^{3x35} mutant reveal that, as in the ligand-binding pocket, the conserved core mutation abrogates β-arrestin-2 recruitment by increasing conformational heterogeneity at a key position in the β-arrestin–coupling region (Fig. 3E). Thus, a single mutation in the receptor core, Asn127Lys^{3x35}, destabilizes ACKR3 at extracellular and intracellular sites, overriding the effects of CXCL12 to decrease conformational exchange present in WT ACKR3 (Fig. 3F).

In the CXCL12-bound, constitutively active Asn127Ser^{3x35} mutant receptor, Met212^{5x39} and Met138^{3x46} peaks overlay those in the CXCL12-WT-ACKR3 state (Fig. 3, C to E, and fig. S8A) despite this state's unresponsiveness to CXCL12, indicating that the activating mutation promotes conformational homogeneity (no rotameric exchange) extracellularly and intracellularly, even as it decouples allosteric transmembrane communication. By contrast, VUN701 decreases the elevated basal β-arrestin recruitment of ACKR3 Asn127Ser^{3x35}, suggesting that it acts as an inverse agonist at this mutant (fig. S8C and table S1). As shown by NMR, the VUN701-bound Asn127Ser^{3x35} mutant reverts to the inactive (i.e., *trans*/*gauche* interconverting) state at Met212^{5x39}, and the Met138^{3x46} peak is diminished, suggesting transition to a more dynamic state at both probes (fig. S8, D to G).

Among analyzed GPCR structures (i.e., those bound to β-arrestin or β-arrestin–biased ligands), AT$_1$R is most closely related to ACKR3 and is the only one that shares Asn^{3x35}. Calculation of intramolecular, residue-residue interactions of AT$_1$R bound to antagonists and β-arrestin–biased agonists revealed that whereas 3x35 interacts with 7x46, a key residue for β-arrestin recruitment (*12*) in both states, 7x46 interacts with 3x39, which is also involved in β-arrestin recruitment (*12*), only in the active state (Fig. 3G). This rearrangement in turn shifts the register of TM3 and TM7 relative to one another in the region above 3x46 and 7x53 (Fig. 3G), suggesting a mechanism by which 3x35 mutations in ACKR3 might be transmitted to the β-arrestin–coupling region as observed at the intracellular probe Met138^{3x46}. In effect, Asn^{3x35} mutations in ACKR3 may function by promoting (active) or disrupting (inactive) 3x39–7x46 interactions, resulting in decreased (active) or increased (inactive) conformational exchange at the intracellular probe.

DISCUSSION

Some studies have identified distinct conformational changes associated with β-arrestin recruitment (*12*, *13*), but others have shown multiple β-arrestin competent conformations (*14*). How can β-arrestin be both sensitive to and tolerant of GPCR conformational changes?

The results presented here help to reconcile this apparent contradiction by revealing that whereas β-arrestin is tolerant of diverse GPCR conformations in some parts of the receptor, it may have more stringent conformational requirements in other parts. At TM5 of the ligand-binding pocket, we found that multiple conformational solutions (*15*) are compatible with β-arrestin recruitment. These multiple-binding-pocket conformations are funneled into a single active conformation as monitored at the intracellular probe Met138^{3x46}. NMR and structural evidence for active-state interactions between 3x46 and 7x53 suggest that the stringent intracellular conformational requirements comprise the intracellular regions of TM3 and TM7. Indeed, although studies of AT$_1$R bound to multiple β-arrestin agonists have revealed diverse intracellular conformations, most variation was in TM5 and TM6, with only modest variation in TM7 (*14*).

Conformational control, although important, may not fully account for β-arrestin recruitment. The observation that intracellular and

extracellular probes both transition between conformational heterogeneity (inactive) and homogeneity (active) suggests that β-arrestin recruitment is governed in part by tuning the conformational spectrum sampled by the receptor. The regulation of β-arrestin recruitment in this manner could help to explain the apparent agnosticism of β-arrestin for specific GPCR conformations (*14*): Alteration of receptor stability in key regions may tune a GPCR's ability to couple β-arrestin irrespective of conformation (e.g., at $Met212^{5x39}$). Likewise, globally different GPCR conformations may be similarly tuned to couple β-arrestins by decreasing conformational heterogeneity of the same key epitopes of the β-arrestin interface.

Ligand-specific modulation of the conformational spectrum observed at NMR probes supports a role for conformational selection in β-arrestin recruitment. In this model, a preexisting spectrum of conformations is narrowed upon perturbation of the system (*44*), in this case by addition of various ligands. In the ligand-binding pocket, the inactive conformation of $M212^{5x39}$ is a composite of multiple active conformations, suggesting that β-arrestin recruitment is governed not by switching between distinct on and off conformations, but rather by selecting one of many active conformations (active) versus facilitating rapid exchange between them (inactive). The same is only partially true intracellularly: $Met138^{3x46}$ exists as a composite of *trans*/*gauche* rotamers in the inactive state that is narrowed to *gauche*-only rotamers in the active state. Nevertheless, $Met138^{3x46}$ only experiences ring current shift effects in the active state. Although this might suggest that the active-state conformation of $Met138^{3x46}$ is not accessible to ACKR3 in the inactive state, it could also reflect sensitivity limitations of ^{13}C compared with other NMR nuclei (*45*).

How might different ligands enhance or suppress changes in conformational plasticity to regulate β-arrestin? β-arrestin agonists might be described as stabilizing a particular set of conformations that facilitate β-arrestin binding (Fig. 4). Indeed, contact network analysis of inactive- and active-state structures of GPCRs reveals enhanced intermolecular (ligand-residue) and intramolecular (residue-residue) interactions in β-arrestin active states, suggesting that β-arrestin agonists might stabilize specific conformations by organizing denser contact networks. Antagonists fail to fasten these locks, leaving them open (or conformationally heterogeneous), which might disfavor β-arrestin recruitment (Fig. 4).

Whereas NMR data and structural analysis point to the association of β-arrestin recruitment with increased stability, other studies show that G protein activation is associated with conformational heterogeneity at the intracellular region (*46*). There are several possible reasons for this difference. Principally, compared with extensive GPCR–G protein interface (mediated primarily through the α5-helix of G protein), GPCR-β-arrestin interactions are mediated by a smaller, less structured finger loop. Other reasons include dependence of G protein (but not β-arrestin) engagement on outward, destabilizing motions of TM6 (*14*, *25*), and requirement of G protein (but not β-arrestin) coupling to be linked to GTP hydrolysis.

We propose roles for receptor conformations and conformational equilibria in regulating β-arrestin recruitment, but other aspects of β-arrestin coupling must also be considered. For instance, the observed conformational changes might lead to differences in GPCR kinase recruitment between antagonists and agonists, resulting in distinct phosphorylation patterns that alter GPCR functional properties (*47*). Indeed, β-arrestin interaction with phosphorylated GPCRs in the absence of core engagement has been shown to be sufficient for GPCR internalization and β-arrestin–mediated signaling (but not G protein desensitization) (*48*). Nevertheless, other studies have correlated the extent of allosteric coupling between agonists and β-arrestin with a particular ligand's efficacy in β-arrestin recruitment (*49*), suggesting at least some role for GPCR conformational changes dictating β-arrestin engagement.

In summary, our results show that coordinated, allosterically linked changes in receptor dynamics regulate β-arrestin recruitment to the intrinsically biased receptor ACKR3. Our results provide a framework with which to understand the molecular changes required for β-arrestin recruitment and provide insights that may facilitate the design of biased therapeutics.

REFERENCES AND NOTES

1. Y. K. Peterson, L. M. Luttrell, *Pharmacol. Rev.* **69**, 256–297 (2017).
2. D. Wootten, A. Christopoulos, M. Marti-Solano, M. M. Babu, P. M. Sexton, *Nat. Rev. Mol. Cell Biol.* **19**, 638–653 (2018).
3. A. Manglik *et al.*, *Nature* **537**, 185–190 (2016).
4. C. L. Schmid *et al.*, *Cell* **171**, 1165–1175.e13 (2017).
5. L. M. Slosky *et al.*, *Cell* **181**, 1364–1379.e14 (2020).
6. N. K. Singla *et al.*, *Pain Pract.* **19**, 715–731 (2019).
7. N. Singla *et al.*, *J. Pain Res.* **10**, 2413–2424 (2017).
8. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, D. E. Gloriam, *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
9. D. P. Staus *et al.*, *Nature* **579**, 297–302 (2020).
10. W. Huang *et al.*, *Nature* **579**, 303–308 (2020).
11. Y. Lee *et al.*, *Nature* **583**, 862–866 (2020).
12. L. M. Wingler *et al.*, *Science* **367**, 888–892 (2020).
13. C. M. Suomivuori *et al.*, *Science* **367**, 881–887 (2020).
14. L. M. Wingler *et al.*, *Cell* **176**, 468–478.e11 (2019).
15. L. M. Wingler, R. J. Lefkowitz, *Trends Cell Biol.* **30**, 736–747 (2020).
16. C. Wang *et al.*, *Science* **340**, 610–614 (2013).
17. S. Rajagopal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 628–632 (2010).
18. J. M. Burns *et al.*, *J. Exp. Med.* **203**, 2201–2213 (2006).
19. M. Gustavsson *et al.*, *Nat. Commun.* **8**, 14135 (2017).
20. M. Meyrath *et al.*, *Nat. Commun.* **11**, 3033 (2020).
21. I. R. Kleckner, M. P. Foster, *Biochim. Biophys. Acta* **1814**, 942–968 (2011).
22. A. B. Kleist *et al.*, *Methods Cell Biol.* **149**, 259–288 (2019).
23. X. Zhang, R. C. Stevens, F. Xu, *Trends Biochem. Sci.* **40**, 79–87 (2015).
24. R. Sounier *et al.*, *Nature* **524**, 375–378 (2015).
25. J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wüthrich, *Science* **335**, 1106–1110 (2012).
26. V. Isberg *et al.*, *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
27. J. D. McCorvy *et al.*, *Nat. Struct. Mol. Biol.* **25**, 787–796 (2018).
28. T. Warne, C. G. Tate, *Biochem. Soc. Trans.* **41**, 159–165 (2013).
29. A. J. Venkatakrishnan *et al.*, *Nature* **536**, 484–487 (2016).
30. S. G. Rasmussen *et al.*, *Nature* **469**, 175–180 (2011).
31. B. F. Volkman, D. Lipson, D. E. Wemmer, D. Kern, *Science* **291**, 2429–2433 (2001).
32. A. S. Solt *et al.*, *Nat. Commun.* **8**, 1795 (2017).
33. G. L. Butterfoss *et al.*, *J. Biomol. NMR* **48**, 31–47 (2010).
34. Y. Kofuku *et al.*, *Nat. Commun.* **3**, 1045 (2012).
35. M. Kayikci *et al.*, *Nat. Struct. Mol. Biol.* **25**, 185–194 (2018).
36. M. Szpakowska *et al.*, *Br. J. Pharmacol.* **175**, 1419–1438 (2018).
37. R. Nygaard, T. M. Frimurer, B. Holst, M. M. Rosenkilde, T. W. Schwartz, *Trends Pharmacol. Sci.* **30**, 249–259 (2009).
38. R. E. London, B. D. Wingad, G. A. Mueller, *J. Am. Chem. Soc.* **130**, 11097–11105 (2008).
39. S. J. Perkins, K. Wüthrich, *Biochim. Biophys. Acta* **576**, 409–423 (1979).
40. Y. Kang *et al.*, *Nature* **523**, 561–567 (2015).
41. G. Fenalti *et al.*, *Nature* **506**, 191–196 (2014).
42. V. Katritch *et al.*, *Trends Biochem. Sci.* **39**, 233–244 (2014).
43. N. Montpas *et al.*, *J. Biol. Chem.* **293**, 893–905 (2018).
44. T. R. Weikl, F. Paul, *Protein Sci.* **23**, 1508–1518 (2014).
45. L. Ye, N. Van Eps, M. Zimmer, O. P. Ernst, R. S. Prosser, *Nature* **533**, 265–268 (2016).
46. A. Manglik *et al.*, *Cell* **161**, 1101–1111 (2015).
47. A. Sente *et al.*, *Nat. Struct. Mol. Biol.* **25**, 538–545 (2018).
48. T. J. Cahill3rd *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2562–2567 (2017).
49. R. T. Strachan *et al.*, *J. Biol. Chem.* **289**, 14211–14224 (2014).

ACKNOWLEDGMENTS

We thank A. Dishman for comments on the manuscript and J. Campbell and colleagues at Chemocentryx for supplying CCX777. **Funding:** This work was supported by the National Institutes of Health (grant F30CA196040 to A.B.K.; grant R01AI058072 to B.F.V.; grant F30HL134253 to M.A.T.; grant R35GM133421 to J.D.M.; and grant T32 GM080202 to the Medical Scientist Training Program at Medical College of Wisconsin to A.B.K. and M.A.T.); the State of Wisconsin Tax Check-Off Program for Cancer Research and the Medical College of Wisconsin Cancer Center (B.F.V.); the Luxembourg National Research Fund (Pathfinder "LIH383," INTER/FWO "Nanokine" grant 15/10358798 to A.C.); INTER/FNRS grants 20/15084569, PoC "Megakine" 19/14209621, AFR-3004509, and PRIDE 11012546 "NextImmune" to A.C.; F.R.S.-FNRS-Télévie grants 7.4593.19, 7.4529.19, and 7.8504.20 to A.C.; European Union's Horizon 2020 MSCA Program (grant 641833 ONCORNET and 860229 ONCORNET2.0 to M.J.S.); American Lebanese Syrian Associated Charities (ALSAC grant to M.M.B.); and the UK Medical Research Council (MRC grant MC_U105185859 to M.M.B. and A.S.). **Author contributions:** A.B.K. and B.F.V. conceived of the project. B.F.V. supervised the project. A.B.K. and S.J. made constructs, purified protein, and collected and analyzed NMR data. A.B.K., S.J., and B.F.V. analyzed data, interpreted results, and wrote the manuscript. A.B.K. and S.J. produced the figures. A.B.K. and A.S. performed contact network analysis. A.B.K., A.S., and M.M.B. interpreted contact network analysis data. M.J.S., R.H., and V.B. identified VUN701 and performed nanobody screening and characterization. F.C.P. cloned VUN701, optimized pulse sequences, and oversaw NMR data collection and analysis. M.A.T. made and validated ACKR3 homology model. A.C. identified and provided LIH383 and provided input on binding and functional assays. A.C. and M.S. performed and analyzed fluorescence binding assays and some β-arrestin and binding assays. L.J.L., M.M.C., E.I.A., and L.M.M. performed and G protein and some β-arrestin assays under the supervision of J.D.M.. All authors provided comments on the manuscript. **Competing interests:** B.F.V. and F.C.P. have an ownership interest in Protein Foundry, LLC. R.H. is CSO of QVQ Holding BV. A patent application has been filed on "Novel Selective ACKR3 Modulators and Uses Thereof" (applicant: Luxembourg Institute of Health; inventors: A.C. and M.S.; PCT application no.:

PCT/EP2020/061981). **Data and materials availability:** CCX777 was supplied through a materials transfer agreement (MTA) with Chemocentryx. LIH383 was supplied to B.F.V. through a MTA with A.C. VUN701 was supplied to B.F.V. through a MTA with M.J.S. All data are available in the main text or the supplementary materials. NMR peak assignments are deposited in the Biological Magnetic Resonance Bank (BMRB) under entries 51451 to 51454.

SUPPLEMENTARY MATERIALS

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

Submitted 16 May 2021; accepted 7 June 2022
10.1126/science.abj4922

ACTUATORS

A processable, high-performance dielectric elastomer and multilayering process

Ye Shi[1,2]†, Erin Askounis[1]††, Roshan Plamthottam[1]†, Tom Libby[3], Zihang Peng[1], Kareem Youssef[1], Junhong Pu[1], Ron Pelrine[3]§, Qibing Pei[1]*

Dielectric elastomers (DEs) can act as deformable capacitors that generate mechanical work in response to an electric field. DEs are often based on commercial acrylic and silicone elastomers. Acrylics require prestretching to achieve high actuation strains and lack processing flexibility. Silicones allow for processability and rapid response but produce much lower strains. In this work, a processable, high-performance dielectric elastomer (PHDE) with a bimodal network structure is synthesized, and its electromechanical properties are tailored by adjusting cross-linkers and hydrogen bonding within the elastomer network. The PHDE exhibits a maximum areal strain of 190% and maintains strains higher than 110% at 2 hertz without prestretching. A dry stacking process with high efficiency, scalability, and yield enables multilayer actuators that maintain the high actuation performance of single-layer films.

Dielectric elastomers (DEs) earned their artificial muscle moniker because of their large electrically induced actuation strain, high energy density, fast response speed, and mechanical compliancy, all of which reproduce or in some aspects exceed the multifunctional performance of natural muscles (*1*, *2*). When a DE film is sandwiched between compliant electrodes, it acts as a deformable capacitor and is known as a dielectric elastomer actuator (DEA) (*3*). Under a voltage, the electric field across the DE generates a strong electrostatic interaction between the electrodes, known as a Maxwell stress, which compresses the film in the thickness direction and expands it in area. On the basis of the working mechanisms, a high-performance DE should have sufficiently high elastic strains, a large dielectric constant, high dielectric strength, and an actuation stability without premature failure.

It has been challenging to tailor the electromechanical properties of an elastomer to meet all these requirements, and the selection of DE materials is therefore limited (*4*). Commercial 3M very high bond (VHB) acrylate tapes and silicone elastomer resins are the most widely used DE materials. VHBs, however, suffer from high viscoelastic losses, whereas silicones tend to exhibit low maximum strains and dielectric strength (*5*). Additionally, many soft conventional elastomers exhibit a long stress-strain plateau (*5*, *6*), and their performance is limited by electromechanical instability (EMI) (*3*, *7*). Under constant voltage, the electric field increases as the DE film thickness decreases, resulting in progressively greater strain that proceeds until failure or induces unstable snap-through. EMI on some DEs can be suppressed by applying a constant strain before actuation, known as prestretching (*8*, *9*). This, however, requires a rigid frame to maintain the applied strain.

The contradiction between low driving voltages and high energy output also limits DEA applications. Multilayer DEAs are able to scale up force and energy outputs without increasing voltages (*2*, *10*, *11*), but currently used wet stacking methods—where an uncured DE film is deposited on a cured DE—show low efficiency, poor scalability, low yield, and suboptimal actuation performance as a result of layer nonuniformity. In addition, wet techniques are not applicable for precured films, such as VHBs (*12*).

There have been important advances in the development of DE materials, including polyacrylates with optimized cross-linking networks (*13*), bottle-brush polymers (*6*), interpenetrating networked elastomers (*14*), and high–dielectric constant elastomers (*15*, *16*). A bimodal-networked DE material was reported to suppress EMI without prestretching (*8*). However, it suffered from low maximum strains (<90% areal strain) and high viscoelastic loss, which limited its energy and power output.

We built a bimodal-networked elastomer using two cross-linkers with different chain lengths and tailored its electromechanical properties for high actuation performance. DE films were fabricated through solution processing and cured under ultraviolet (UV) light. The long chain segment in the bimodal network ensures large elongation, and the second relatively short chain segment raises the modulus at modest strains to resist the rapidly increased Maxwell stress during actuation and suppress EMI (Fig. 1A) (*8*, *17*). Figure 1B shows the molecular structures of reactants. Detailed formulations and nomenclature of samples can be found in table S1. Butyl acrylate (BA) and isobornyl acrylate (IBOA) were selected as comonomers to lower the modulus and improve the toughness of copolymers, respectively (*18*). They were also important to reduce the viscosity of prepolymer solutions (*19*). CN9021, a urethane diacrylate (UDA) with a high molecular weight, was selected as the flexible long-chain cross-linker, and propoxylated neopentyl glycol diacrylate (PNPDA) was used as the short-chain cross-linker (fig. S1). 2,2-dimethoxy-2-phenylacetophenone (DMPA) and benzophenone (BP) were used as cophotoinitiators to ensure complete curing through the bulk and surface of the film.

We developed systematic synthetic strategies to tune the bimodal system's stress-strain responses and viscoelasticity. First, short-chain cross-linkers with softer and more-extended chains were used to replace previously explored short and stiff molecules, such as hexanediol diacrylate (HDDA), thus achieving more tunable stretchability and tensile strength (Fig. 1C and figs. S2 and S3) (*8*). The cross-link density/gel fraction (fig. S4) as well as the stress-strain responses were further controlled by changing the concentrations of the short-chain cross-linker, PNPDA (Fig. 1C) (*8*). Without PNPDA, the DE shows a similar long stress-strain plateau to that of non-prestretched VHB (fig. S5) (*5*, *6*). With the bimodal network structure, our DE stiffens after a critical stretch ratio in its non-Gaussian region (*20*), which can suppress EMI. This critical stretch ratio shifts to a smaller value as the cross-link density increases. Second, the concentration of hydrogen bonds was optimized in the DE network to modify the viscoelasticity while maintaining its stress-strain relationship. A small amount of acrylic acid (AA) comonomer (2.5 parts of weight) was added, which provides side groups to form hydrogen bonds with themselves, as well as with the –NH– groups on the CN9021 and PNPDA cross-linkers (fig. S6) (*21*). The mechanical loss factor of DE, a measurement of the viscoelasticity, was then decreased from ~0.22 to ~0.11 at room temperature and low frequencies (fig. S7) without changing the strain stiffening behavior (table S2) and high elasticity (fig. S8). After AA is added, hydrogen bonds partially replace covalent cross-links in the network. Hydrogen bonds act as weak, physical cross-links and can dynamically dissociate, leading to a lower glass transition temperature (T_g) (fig. S9) and higher chain mobility in the network (*22*). As more AA is added, hydrogen bonds become highly concentrated, and the density of covalent bonds decreases (fig. S7), which results in inhibited actuation performance (fig. S10). At frequencies above 20 Hz, the storage modulus and loss factor of DE also increase rapidly (fig. S11).

The static actuation performance of DEs was tested in a diaphragm configuration (Fig. 1D)

[1]Department of Materials Science and Engineering, University of California, Los Angeles, Los Angeles, CA, USA. [2]ZJU-UIUC Institute, Zhejiang University, Hangzhou, China. [3]SRI International, Menlo Park, CA, USA.
*Corresponding author. Email: qpei@seas.ucla.edu
†These authors contributed equally to this work.
‡Present address: Exponent, Inc., Menlo Park, CA, USA.
§Present address: Pelrine Innovations, Longmont, CO, USA.

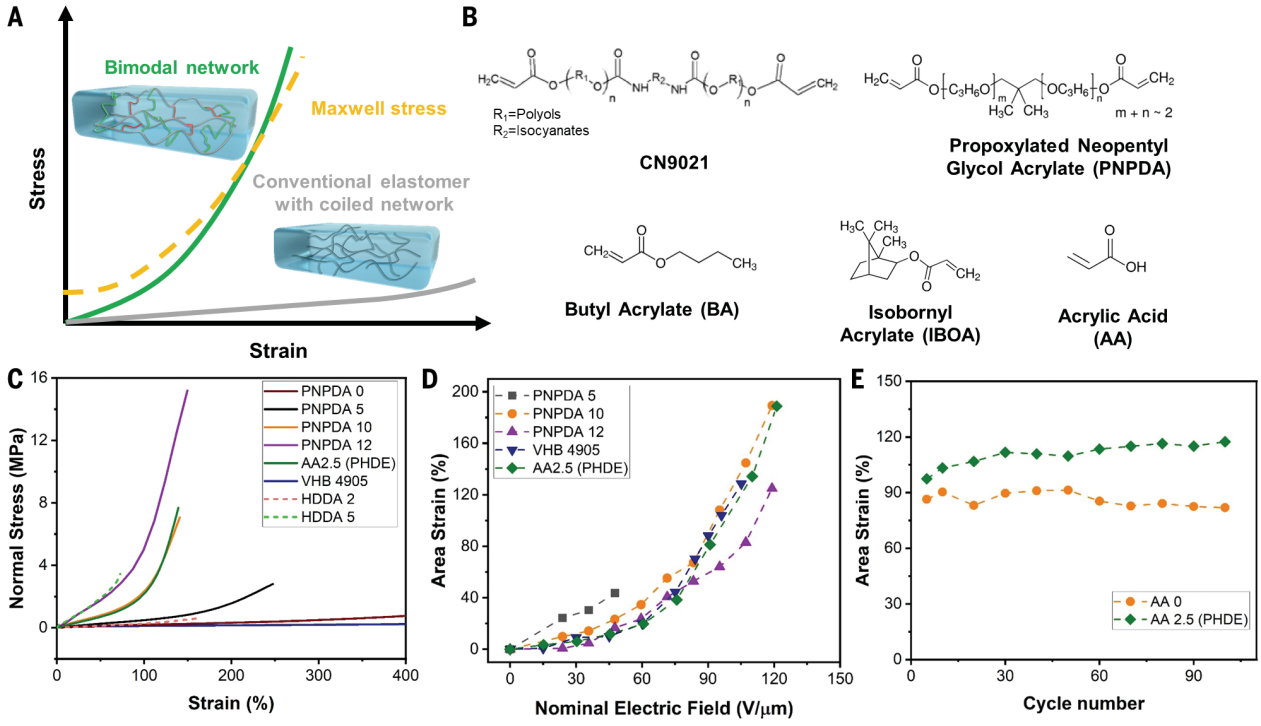


Fig. 1. Synthesizing bimodal-networked DEs with soft cross-linkers and hydrogen bonds for high-performance actuation. (**A**) Schematic illustration of the design rules for DE with bimodal network structure. (**B**) Molecular structures of acrylates and diacrylates used for the synthesis of bimodal-networked DE. (**C**) Stress-strain relationships of PHDE, bimodal DE materials with different concentrations (weight %) of PNPDA or HDDA, and VHB 4905 without prestretching. The complete curves of PNPDA 0 and VHB 4905 are shown in fig. S5. (**D**) Static actuation of PHDE, bimodal DE materials with different concentrations of PNPDA, and VHB 4905 with 300% biaxial prestretching, measured on a diaphragm. Carbon grease was used as the compliant electrodes. (**E**) Cyclic actuation of PHDE and bimodal DE without AA under a 2-Hz square wave voltage of 4.5 kV for 100 cycles.

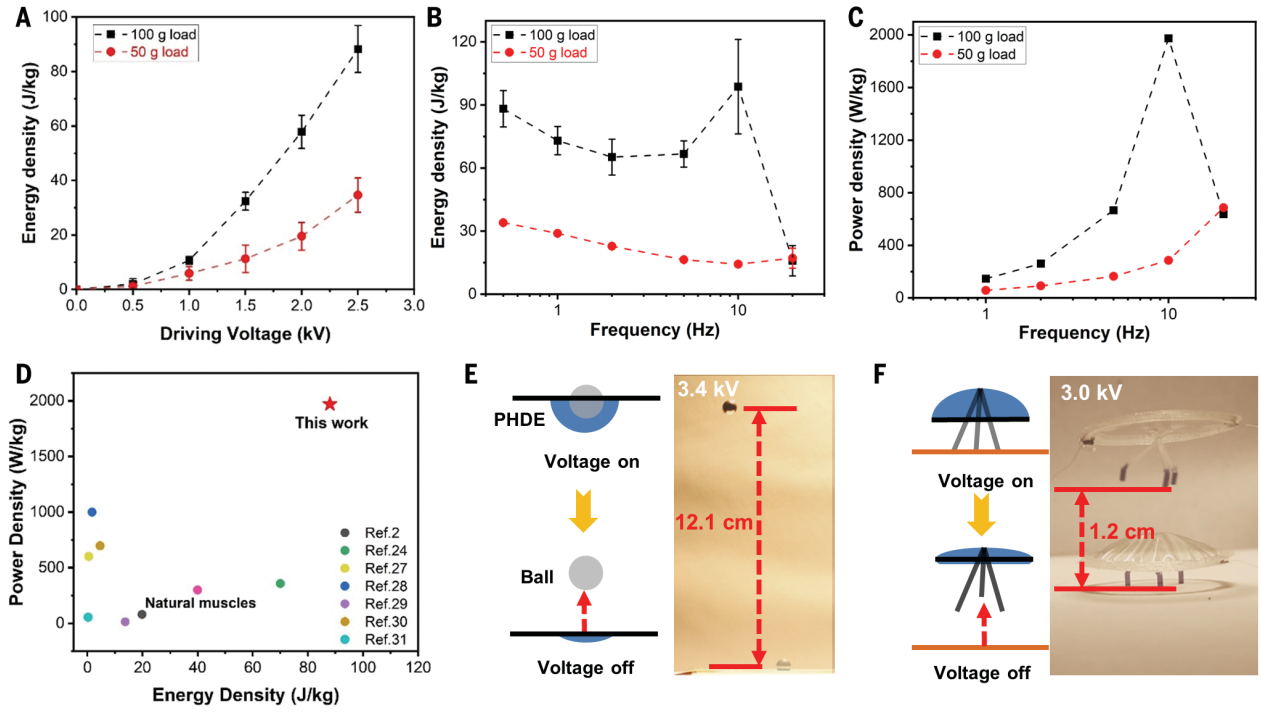


Fig. 2. Energy density and power density properties of PHDE. (**A**) Energy density of single-layer PHDE films measured at specified loads. (**B**) Energy density at different frequencies. The driving voltage is 2.5 kV. (**C**) Average power density of single-layer PHDE tested in pure-shear mode with different loads. (**D**) Performance comparison of PHDE with natural muscles and other DEAs reported in the literature. (**E**) Ball toss by the PHDE actuator. (**F**) A jumping robot fabricated with a PHDE actuator.

(*23*). A moderate ratio of short-chain cross-linker to long-chain cross-linker was found to be ideal for high actuation strains. The DE with 10 wt % of PNPDA exhibited a maximum strain of 189% with EMI fully suppressed, as indicated by the modeling study (fig. S12). After the small amount of AA was added, the DE maintained this high and stable actuation and improved its response speed owing to reduced viscoelasticity. Under a 2-Hz square wave voltage, the DE with AA exhibited an average areal actuation of 110%, which is superior to those of the DE without AA and of other previously reported DEs (Fig. 1E, movie S1, and table S3). Thus, the DE with 10 wt % of PNPDA and an additional 2.5 parts of AA is finalized as the processable, high-performance dielectric elastomer (PHDE) composition for film preparation and device fabrication. PHDE showed high dielectric strength with an apparent electrical breakdown field of 330 V/μm, low leakage currents (fig. S13), a high dielectric constant of 5.35, and a small dielectric dissipation factor of 0.014 at 1 kHz (fig. S14).

The force output, energy density, and power density of the single-layer PHDEs were tested in a pure-shear mode (fig. S15) (*23*, *24*). The film measured 6 cm by 1 cm by 40 μm and ~0.030 g of active material. Under preloaded isometric conditions, PHDE was able to output a stable force of ~0.95 N (~0.40 MPa stress) at 2.5 kV with a 1.2 N preload without prestretching (fig. S16A). The force maintained ~0.67 N at 2 Hz and ~0.4 N at 20 Hz (fig. S16B). Under isotonic conditions, PHDE films performed a specific work of ~35 J/kg with a 50-g load and ~88 J/kg with a 100-g load during contraction at 2.5 kV (Fig. 2A and movie S2). This maximum energy density is more than twice that of natural muscles (0.4 to 40 J/kg). PHDE films maintained energy densities of ~67 J/kg at 5 Hz and ~16 J/kg at 20 Hz with a 100-g load (Fig. 2B and movie S2), and the corresponding power densities were calculated to be ~670 W/kg and ~640 W/kg, respectively (Fig. 2C) (*24*). The peak energy and power density at 10 Hz was caused by resonant responses of planar actuator systems (*24*, *25*). These power densities are about one order of magnitude as high as those of natural muscles (*26*) and can be achieved in the 10^1-Hz range or lower. At frequencies greater than or equal to 50 Hz, the films were not able to effectively perform work. Notably, the blocked force and energy density can be tuned by adjusting the concentration of short-chain cross-linkers for different applications (figs. S17 and S18). PHDE exhibited high stability when performing work (fig. S19). The maximum energy and power output densities of PHDE are superior to those of natural muscles and DEAs developed in recent years (Fig. 2D) (*2*, *24*, *27*–*31*), which enable lightweight, energy-dense actuation systems. A PHDE film was demonstrated to toss an aluminum ball, which is 20 times as heavy as itself, to 12.1 cm high (Fig. 2E and movie S3), reaching an energy transfer efficiency of ~10% (fig. S20). A miniature jumping robot based on a single-layer PHDE actuator reached a height of 1.2 cm, which was ~67% of the theoretical height of this jumping system (Fig. 2F, fig. S21, and movie S4).

To scale up the energy and power outputs of DEAs at low voltages, we developed a dry stacking method to fabricate multilayer DEAs. This method was demonstrated with PHDE but is applicable for other DEs, including preformed DE films. As illustrated in Fig. 3A, to fabricate an N-layer PHDE stack, N-1 films are cast on glass, and one film is prepared on polyethylene terephthalate (PET), with uniform thickness. Two substrates were selected to provide different mechanical properties and adhesions, thus avoiding delamination during stacking. Carbon nanotube (CNT) electrodes and a thin layer of uncured acrylic polymer precursor acting as a binding layer are sprayed on the glass/PHDE films. The PET/PHDE is aligned onto a glass/PHDE/CNT/precursor and laminated in a vacuum laminator. After UV curing the binding layer, the glass is removed and the two-layer PHDE stack, which is now on PET, is aligned onto another glass/PHDE/CNT/precursor to add the third PHDE layer. These processes are repeated until the stack is fabricated. Figure 3B and fig. S22 show the pictures of a 10-layer PHDE stack with an array of 20 actuators. Its cross section was examined by scanning electron microscopy (SEM), and the image in Fig. 3C shows uniform thickness of PHDE layers and solid sealing between them. The interfacial electrode with interpenetrating structure allowed strong bonding between layers, which was proved by a swelling test (fig. S23), and the stress-strain curve and dynamic moduli of a 10-layer PHDE stack are very close to those of single-layer films (fig. S24). The dielectric constant of the PHDE

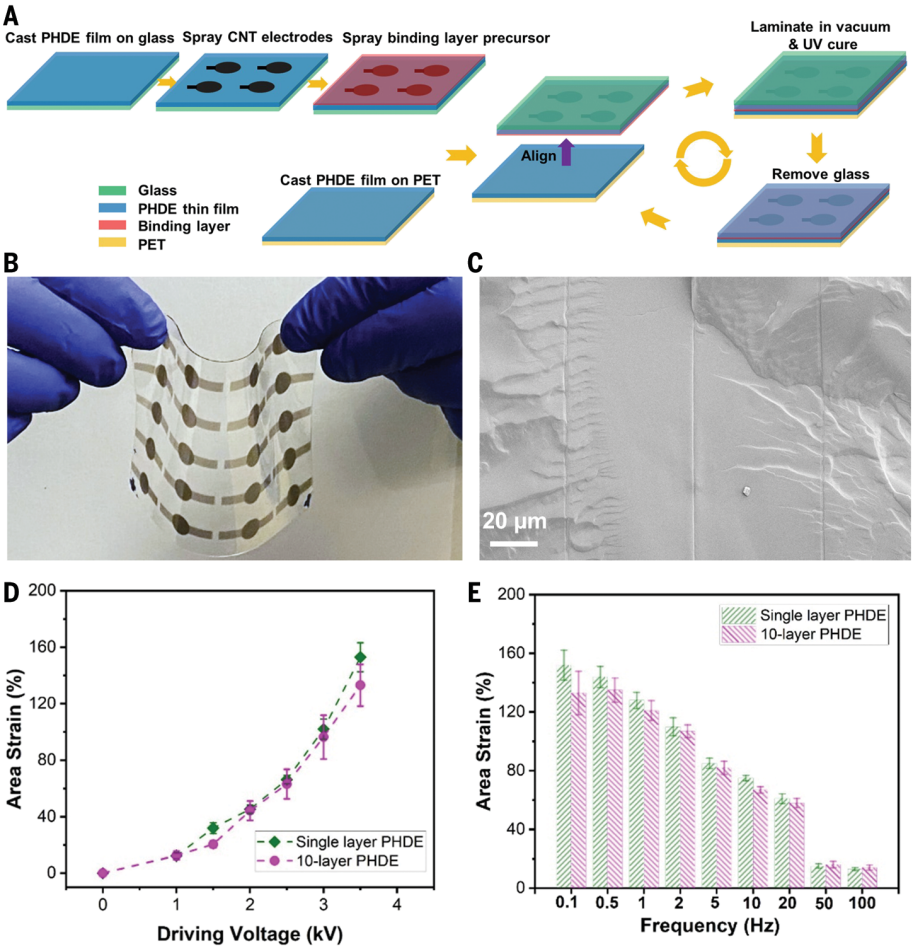


Fig. 3. Multilayer PHDE actuators enabled by dry stacking method. (**A**) Diagram of dry stacking for multilayer PHDE actuator fabrication. (**B**) Photo of a 10.16 cm by 12.7 cm (4 inch by 5 inch) 10-layer PHDE stack with 20 actuators. (**C**) SEM images of cross-sectional view of a PHDE stack. (**D**) Static actuation of single-layer and 10-layer PHDE films, tested on a diaphragm. CNT was used as compliant electrodes. (**E**) Frequency response of single-layer and 10-layer PHDE actuators under 3.5 kV (square wave).

stack is measured to be 5.37, and its dissipation factor increases to 0.020 at 1 kHz, possibly because of the dielectric loss at interfaces.

Compared with wet stacking methods (*2*, *10*, *11*), our dry stacking method shows many advantageous features. It is compatible with established, large-scale materials processing techniques, and the film dimensions and thickness can be easily tuned (fig. S25). Additionally, this method enables a high yield by screening individual films before stacking (fig. S26). Because film preparation, electrode deposition, and lamination can be conducted in parallel, the process is very efficient. The efficiency can be further improved by modularly dry stacking PHDE stacks on each other, as shown in fig. S27. The actuation strain of 10-layer PHDE actuators is comparable to that of single-layer actuators at the same low driving voltages (Fig. 3D). The strain is slightly lower at higher voltages, possibly because of greater stiffness from the inactive border surrounding the 10-layer actuators (movie S5) and decreased conductivity of CNT electrodes at highly stretched states. Moreover, the multilayer actuator maintained high response speed and reached strains of ~110% at 2 Hz and ~60% at 20 Hz (Fig. 3E). Additionally, our dry stacking method can linearly scale up the force, energy, and power outputs with the number of layers and experiences less than a 20% reduction (fig. S28).

The multilayer PHDE stacks enable new actuator designs. We fabricated spider actuator, which were identified as a potential DEA design but never realized, in part because they are not compatible with high prestretching. The basic concept uses radially symmetric inclined legs to couple axial force to radial loading of a circle of DE film, giving the linkage nonlinear mechanical advantage that can be tuned to maximize performance. We built one instantiation of this linkage, tuned for high axial force, and tested it with four-layer PHDE stacks with an active area of ~4.8 cm^2. As shown in Fig. 4A and movie S6, the spider actuator successfully lifted a 200-g load with ~3.0-mm linear displacement at 2.5 kV, reaching a specific work of ~74 J/kg.

We also designed multifunctional roll actuators. As shown in Fig. 4B, a two-layer PHDE strip was rolled into a free-standing tube with a total of ~18 active layers. The roll was first characterized as a linear actuator (fig. S29) and lifted a 100-g load by 1.4 mm at 0.5 Hz, which translated to an output energy density of 15.6 J/kg (Fig. 4C and movie S7). The displacement at 5 Hz was 1.1 mm (fig. S29C and movie S7), with a power density calculated to be 122 W/kg. These values are lower than those determined in the pure-shear mode and the spider actuators, possibly because the inactive layers restrict its linear strain, and the circumferential strain is not coupled.

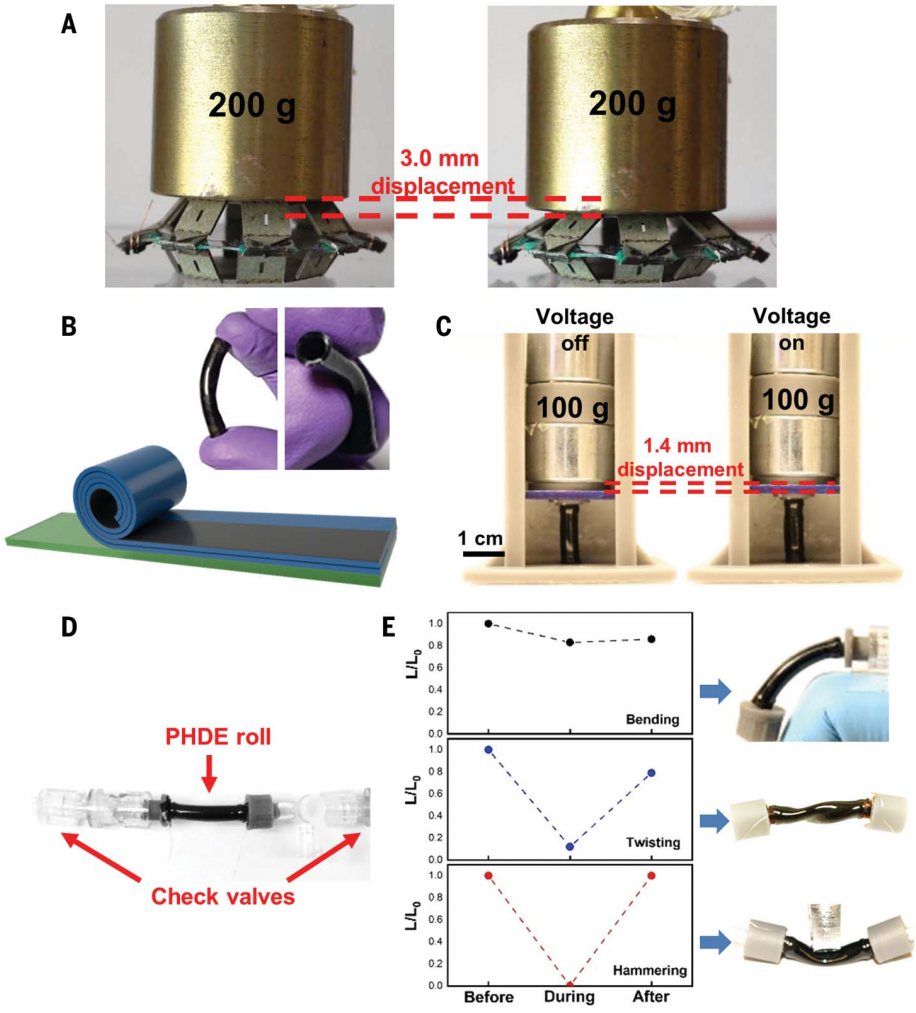


Fig. 4. Actuators enabled by multilayer PHDE. (**A**) A PHDE spider actuator lifts a 200-g load at 2.5 kV. (**B**) Schematic illustration of fabrication of PHDE roll actuators. The inset shows the photos of a PHDE roll. (**C**) A PHDE roll actuator lifts a 100-g load at 0.5 Hz. (**D**) Photograph of a PHDE roll pump. (**E**) Normalized flow rates (L/L_0) during various mechanical manipulations of the pump. The pump is driven at 2.5 kV at 10 Hz while bent at 90°, twisted axially, and hammered. The initial flow rates and the recovery flow rates are also shown.

However, the roll is made entirely of polymers (except for a small amount of conductive materials). It is highly robust and remains functional after they are bent, twisted, and hammered (fig. S30 and movie S8). We further integrated two check valves to build a tubular pump from a PHDE roll (Fig. 4D). The flow rate (L) could be tuned by controlling the driving voltage and frequency. The peak flow rate reached ~20.4 mL/min at 2.5 kV and 10 Hz, which translated to a specific flow rate of ~100,000 mL/kg per minute (fig. S31 and movie S9). This value is about two orders of magnitude higher than that of a miniature pump (TCS Micropumps, MGD 1000S) (*32*). Owing to its robust nature, the PHDE roll pump showed nearly full performance recovery after it was bent, twisted, or hammered during operation (Fig. 4E). Its tunable flow rate, compactness, light weight, low noise, and high mechanical compliancy make it a promising device for flexible medical implants and soft robotics.

REFERENCES AND NOTES

1. R. Pelrine, R. Kornbluh, Q. Pei, J. Joseph, *Science* **287**, 836–839 (2000).
2. M. Duduta, E. Hajiesmaili, H. Zhao, R. J. Wood, D. R. Clarke, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2476–2481 (2019).
3. Z. Suo, *Acta Mech. Solida Sin.* **23**, 549–578 (2010).
4. L. J. Romasanta, M. A. Lopez-Manchado, R. Verdejo, *Prog. Polym. Sci.* **51**, 188–211 (2015).
5. Y. Qiu, E. Zhang, R. Plamthottam, Q. Pei, *Acc. Chem. Res.* **52**, 316–325 (2019).
6. M. Vatankhah-Varnoosfaderani *et al.*, *Adv. Mater.* **29**, 1604209 (2017).
7. X. Zhao, Z. Suo, *Phys. Rev. Lett.* **104**, 178302 (2010).
8. X. Niu *et al.*, *J. Polym. Sci. B Polym. Phys.* **51**, 197–206 (2013).
9. L. Jiang, A. Betts, D. Kennedy, S. Jerrams, *J. Phys. D Appl. Phys.* **49**, 265401 (2016).

10. M. Duduta, R. J. Wood, D. R. Clarke, *Adv. Mater.* **28**, 8058–8063 (2016).
11. S. Rosset, H. R. Shea, *Appl. Phys. Rev.* **3**, 031105 (2016).
12. G. Kovacs, L. Düring, S. Michel, G. Terrasi, *Sens. Actuators A Phys.* **155**, 299–307 (2009).
13. L.-J. Yin *et al*., *Nat. Commun.* **12**, 4517 (2021).
14. S. M. Ha, W. Yuan, Q. Pei, R. Pelrine, S. Stanford, *Adv. Mater.* **18**, 887–891 (2006).
15. C. Ellingford, C. Bowen, T. McNally, C. Wan, *Macromol. Rapid Commun.* **39**, e1800340 (2018).
16. P. Caspari, S. J. Dünki, F. A. Nüesch, D. M. Opris, *J. Mater. Chem. C* **6**, 2043–2053 (2018).
17. C. Tugui *et al*., *J. Mater. Chem. C* **3**, 8963–8969 (2015).
18. M. Hayashi, S. Matsushima, A. Noro, Y. Matsushita, *Macromolecules* **48**, 421–431 (2015).
19. P. K. T. Oldring, Ed., *Chemistry & Technology of UV & EB Formulations for Coatings, Inks & Paints*, vol. 2 (SITA Technology, 1991).
20. M. Rubinstein, S. Panyukov, *Macromolecules* **35**, 6670–6686 (2002).
21. H. Guo, Y. Han, W. Zhao, J. Yang, L. Zhang, *Nat. Commun.* **11**, 2037 (2020).
22. C. L. Lewis, K. Stewart, M. Anthamatten, *Macromolecules* **47**, 729–740 (2014).
23. F. Carpi *et al*., *Smart Mater. Struct.* **24**, 105025 (2015).
24. E. Acome *et al*., *Science* **359**, 61–65 (2018).
25. C. Cao *et al*., *Soft Sci.* **1**, 1 (2021).
26. S. I. Rich, R. J. Wood, C. Majidi, *Nat. Electron.* **1**, 102–112 (2018).
27. Y. Chen *et al*., *Nature* **575**, 324–329 (2019).
28. X. Ji *et al*., *Sci. Robot.* **4**, eaaz6451 (2019).
29. H. S. Jung *et al*., *Int. J. Control. Autom. Syst.* **15**, 25–35 (2017).
30. S. Xu, Y. Chen, N.-P. Hyun, K. P. Becker, R. J. Wood, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103198118 (2021).
31. H. Zhao *et al*., *Adv. Funct. Mater.* **28**, 1804328 (2018).
32. V. Cacucciolo *et al*., *Nature* **572**, 516–519 (2019).

ACKNOWLEDGMENTS

We acknowledge R. Kornbluh of SRI International for valuable suggestions in designing the ball-tossing and jumping-robot experiments. **Funding:** This work was supported by the Defense Advanced Research Projects Agency (DARPA), award no. HR001119C0043. **Author contributions:** Y.S., E.A., R.Pl., and Q.P. conceived the idea. E.A., Y.S., and R.Pl. performed materials synthesis, characterizations, and thin film preparation. Y.S. and R.Pl. performed device fabrication and characterizations, including multilayer stacks, planar jumping robots, and roll actuators. T.L. fabricated and tested spider actuators. K.Y., Z.P., and J.P. helped with film fabrication and material characterizations. Q.P. directed the project. Y.S., E.A., R.Pl., R.Pe., T.L., and Q.P. designed the actuators, analyzed the data, discussed the results, and wrote the manuscript. **Competing interests:** Q.P., E.A., and Y.S. are applying for patents related to the described work. The other authors declare that they have no competing interests. **Data and materials availability:** The data that support the findings of this study are available in the main text and the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn0099
Materials and Methods
Supplementary Text
Figs. S1 to S31
Tables S1 to S3
References (*33*–*44*)
Movies S1 to S9

Submitted 27 October 2021; resubmitted 27 March 2022
Accepted 1 June 2022
10.1126/science.abn0099

By **Hannah Gurholt**

Revising my mental health checklist

"Struggling Ph.D. student checklist, December 2021: Anxiety. Caffeine addiction. Depression. Emotional support cat that then causes stress. Imposter syndrome. Stress twitches. Hair loss." When I recently came across these words, I couldn't believe I had written them just 6 months earlier. Since then, I have tried three different antidepressants and acquired an unruly number of pill cutters. I have experienced migraines, drowsiness, weight loss, weight gain, and fatigue. At times I have still felt numb, lonely, and frustrated. But I have also felt hopeful and reinvigorated, confident in the knowledge that I am on the path of healing.

My troubles started early in graduate school. In addition to the typical stressors of starting a Ph.D., my advisers decided to move to a new university. They wanted me to go with them, whereas many of my classmates and other professors wanted me to stay. I felt pulled in two directions, and as a devout people-pleaser, my ever-present anxiety began to rear its ugly head.

I started seeing a therapist to help unpack the loaded question of what I wanted to do with my life. I shared my hesitancy about medication, having lost a beloved family member to suicide while they were taking antidepressants. My therapist and I decided we would explore every other option. We practiced coping mechanisms for when I felt particularly anxious. We discussed how to respond to the never-ending questions from colleagues about whether I was moving. I left each session with a positive feeling in my stomach.

But despite our efforts, anxiety and depression continued to creep up on me. With classes being held online due to the pandemic, I spent my days at a desk in the corner of my bedroom, searching the familiar walls for any distraction from reading papers. I spent nights lying awake with my thoughts. When I did sleep, I woke to frightening anxiety. I battled nausea, appetite loss, fatigue, panic attacks, and lack of motivation. I lost my eyebrow hair. My period became irregular to the point of spotting every day, with anemia close behind. My friends and family were concerned about my rapid weight loss. When a panic attack forced me to miss an important meeting with my advisers, I scheduled an emergency session with my therapist. She could tell everything she needed to know from my eyes, glassy and sunken with defeat. It was time to consider medication.

I went to a doctor, and with her support, I decided to take the plunge and start on antidepressants. After the appointment, I felt hopeful—an emotion I hadn't experienced in months. I called my parents and assured them I would be all right.

"Antidepressants aren't magic pills ... but they helped me when nothing else could."

My optimism was temporarily dampened when, shortly after starting the medicines, I was hit with intense migraines. But soon I began to see real benefit. I was sleeping again. My nausea dissipated and my appetite returned. My eyebrows grew back. My smile felt more genuine and my laughter less forced. I no longer felt like an imposter in my own body. I picked up my old hobbies of running, reading, drawing, and doing puzzles, and developed new ones, including cleaning and being an obnoxious cat mom. I returned to therapy with newfound light in my eyes. I began to work from local coffee shops, which helped keep me motivated, focused, mentally charged, and caffeinated. I visited my advisers' new university and finally decided to move with them, lifting a massive weight off my shoulders.

That isn't to say everything was suddenly perfect. The first medications made me feel emotionally numb, so my doctor and I tried a different one, which hit me with extreme drowsiness and weight gain. But my doctor and I will keep working to find the medications that work best for me, and I will continue therapy as well. Antidepressants aren't magic pills, and the side effects are no joke. But they helped me when nothing else could.

These days, I have a new checklist. Healing Ph.D. student checklist, July 2022: Drowsiness. Caffeine addiction. Appetite changes. Weight fluctuations. Newfound motivation. Transparency. Hope. ■

Hannah Gurholt is a Ph.D. student at Washington University in St. Louis.

ILLUSTRATION: ROBERT NEUBECKER

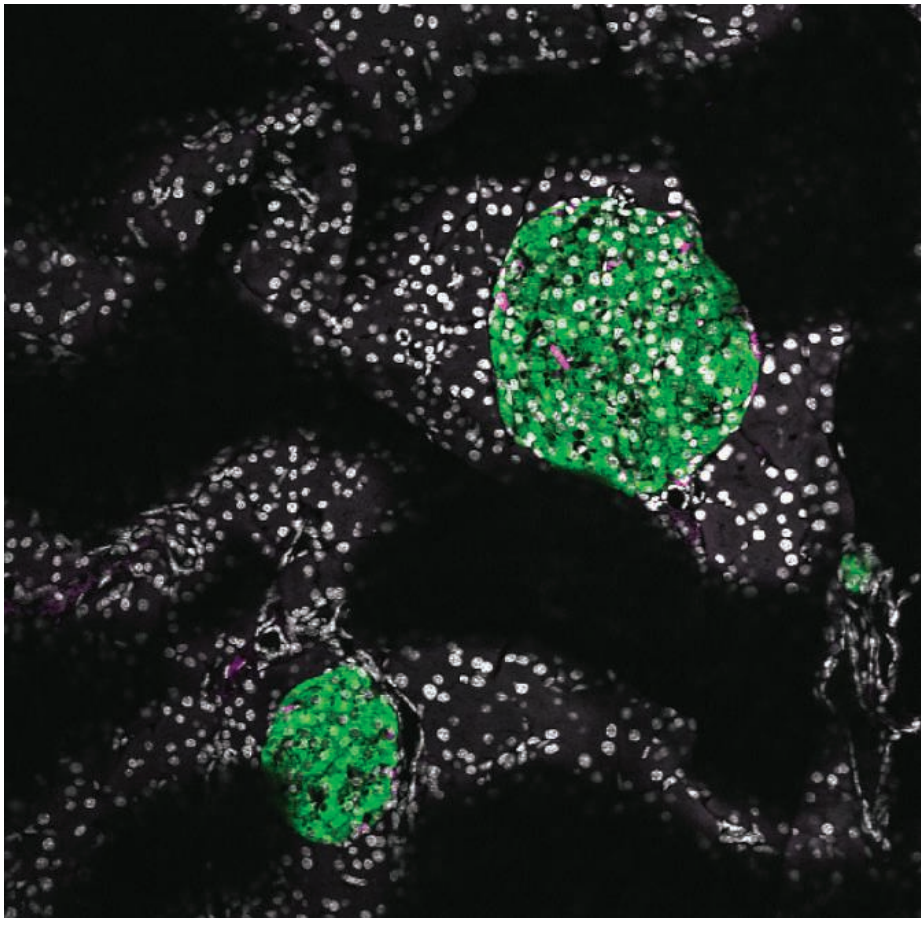
PRIZE ESSAY

GRAND PRIZE WINNER

Jennifer Hampton Hill

Jennifer Hampton Hill received an undergraduate degree from California State Polytechnic University Humboldt and her PhD in biology from the University of Oregon in Eugene in Dr. Karen Guillemin's laboratory. She is currently finishing a postdoctoral fellowship at the University of Utah in Salt Lake City in the laboratories of Drs. June Round and Charles Murtaugh. Her research focuses on host-microbe interactions that shape pancreatic development and health. **www.science.org/doi/10.1126/science.abq6051**

NOSTER | Science MICROBIOME PRIZE



Micrograph of mouse pancreatic islets composed of insulin-producing β cells and other hormone-secreting cells

PHOTOS: (LEFT TO RIGHT) COURTESY OF RACHEL MERRILL; JENNIFER HAMPTON HILL

MICROBIOME

From bugs to β cells

Do ancient microbial irritants offer early life protection against diabetes?

By **Jennifer Hampton Hill**

One-hundred years ago, the discovery of the hormone insulin by Frederick Banting and Charles Best transformed the prognosis of type 1 diabetes (T1D). Practically overnight, a diagnosis of T1D went from meaning certain suffering and death to being a manageable, albeit difficult, condition that could result in a long and full life. Over the century since that Nobel Prize–winning discovery, there has been monumental progress in the understanding of the basic etiology of diabetes, which is now known to be an autoimmune disease characterized by the destruction of pancreatic insulin-producing β cells (see the figure). However, a cure still evades discovery, and its realization will require a dual approach that tempers autoimmunity while simultaneously restoring endogenous insulin production. Focusing on the latter, our work revealed that the resident microbiome is a promising source of previously undiscovered modulators of host β cells.

Animals evolved within a microbial world, and it is plausible that early multicellular life used microbial cues to gather information about the environment, e.g., local nutrient availability, to set their metabolisms (*1*, *2*). Many animals, including fruit flies and mice, have reduced metabolic rates when grown in the absence of microbes—an artificial situation that could be interpreted by these organisms

Departments of Pathology and Human Genetics, University of Utah, Salt Lake City, UT, USA.
Email: jennifer.hill@path.utah.edu

as a dearth of nutrients (*3*). We investigated whether animals use microbial cues to tune the number of cells producing the conserved metabolic regulator, insulin. To optimize energy needed for development of these cells (*4*), animals need to match insulin production capacity to environmental nutrient availability. Thus, we hypothesized that hosts incorporate information from their resident microbiomes to accomplish this optimization.

We leveraged a vertebrate zebrafish model to study development of pancreatic β cells with or without microbiota, by comparing sterile or germ-free (GF) larvae and their conventionally reared (CV), microbe-carrying counterparts. We found that CV larvae had significantly more β cells than GF larvae, despite both groups having equivalent nutrient supplies from their yolk sacks (*5*). Taking a reductionist approach, we systematically added back individual zebrafish gut bacteria and their secreted products until we identified a single protein that was sufficient to restore GF β cell mass (*5*). The protein we identified was previously unknown, and we named it β cell expansion factor A (BefA).

Searches for BefA-related sequences uncovered homologs in a handful of bacteria of the *Enterobacteriaceae* family commonly found in healthy mammalian microbiomes after birth, when β cells normally undergo substantial expansion (*5*, *6*). To test whether BefA might elicit similar responses in mammalian species, we analyzed β cell development in gnotobiotic mouse models. Not only did we find that newborn GF mice had the same paucity of β cells as GF larval fish, but purified BefA was sufficient to increase their developing β cell mass (*7*).

We noted that purified BefA elicited its effect in GF zebrafish and mice whether it was delivered orally or intraperitoneally, suggesting that it might have traveled to the pancreas (*7*). We found that BefA could induce proliferation directly in primary islet cultures from both hosts (*7*). Although small microbial metabolites have frequently been shown to disseminate systemically, this is the first observation of the far-reaching effects of a much larger protein secreted by gut bacteria (*8*).

To better understand how BefA affects β cells, we resolved the protein's crystal structure. This structure revealed a previously unknown protein fold for a putative lipid-binding domain that was sufficient for BefA's activity (*7*). We next determined that BefA was capable of binding to membrane lipids and perturbing lipid bilayers—a trait common to antimicrobial proteins (AMPs), like the Reg proteins, secreted in the gut to prevent incursion by microbes (*9*). Notably, these proteins were originally discovered in pancreatic islets for their capacity to regenerate β cells, but this capacity has not been connected with their antimicrobial activity. We explored this by creating amino acid substitutions at several predicted lipid-binding residues in BefA. One of these substitutions decreased BefA's ability to both disrupt membranes and stimulate β cell turnover (*7*). Conversely, Reg3 was sufficient to induce a potent increase in β cell turnover from mouse islets, illustrating both the necessity and sufficiency of membrane perturbation in β cell expansion (*7*).

The hygiene hypothesis posits that certain skyrocketing diseases, such as T1D, are the result of changing societal practices that have reduced microbial exposures and decreased microbiome diversity (*10*). As a result, humans may have lost microbes that were relied on in the past to perform important functions, such as to stimulate developmental programs (*11*). In the case of T1D, decreased gut microbiome diversity is characteristic of and can be predictive of disease onset, suggesting that at-risk children may be missing specific microbes with disease-mitigating functions (*12*, *13*). The BefA-containing *Enterobacteriaceae* bacteria are abundant colonizers of the healthy infant gut during peak neonatal β cell expansion (*6*, *14*). Failure to develop a sufficient β cell mass in early childhood can lead to metabolic disorders, which suggests that missing out on microbial cues, such as BefA, may predispose individuals to diabetes. If so, fortification of these microbial activities in children who carry risk alleles for T1D could be a strategy to prevent or delay disease.

Bacteria did not evolve BefA to prevent T1D in their hosts. Long before animals developed Reg proteins, bacteria were making AMPs to compete with other microbes. Animal cells have always had to sense and defend themselves against bacterial membrane disrupters. But information could be gleaned amidst the onslaught. Because bacterial growth and competition is most pronounced in nutrient-rich environments, the intensity of membrane attack could be used to gauge metabolic needs. Perhaps insulin-producing cells in early animals were particularly sensitive in response to AMPs because damage to these precious cells would jeopardize survival. Possibly, animals evolved their own Reg-like proteins, with the dual functions of protecting against microbial invaders and stimulating nutrient utilization. Understanding current declining microbiome diversity as a loss of ancient developmental cues allows us to imagine future strategies to correct the environmental information conveyed during human pancreas development to ward off T1D.

FINALIST

Apollo Stacy

Apollo Stacy received his undergraduate degree from Washington University in St. Louis and his PhD from The University of Texas at Austin. Currently a postdoctoral fellow at the National Institutes of Health, Apollo will be starting his laboratory at the Cleveland Clinic Lerner Research Institute in 2022. His research investigates how host-derived metabolites shape the ecological balance between commensals and pathogens and thus host susceptibility to inflammatory disease. **www.science.org/doi/10.1126/science.abq6054**

FINALIST

Irina Leonardi

Irina Leonardi received undergraduate degrees from ETH Zürich and a PhD from the University of Zürich. She performed her postdoctoral work in the laboratory of Dr. Iliyen Iliev at Weill Cornell Medicine. Her research investigated the cellular mechanisms of fungal recognition in the gut and the local and systemic consequences of intestinal fungal colonization. She is currently the Scientific Communications Lead at Immunai in New York. **www.science.org/doi/10.1126/science.abq6056**

PHOTOS: (TOP TO BOTTOM) APOLLO STACY; COURTESY OF IRINA LEONARDI

REFERENCES AND NOTES

1. T. J. Wiles, K. Guillemin, *Curr. Opin. Microbiol.* **54**, 87 (2020).
2. T. C. G. Bosch, M. McFall-Ngai, *Curr. Top. Dev. Biol.* **141**, 399 (2021).
3. G. Storelli *et al.*, *Cell Metab.* **14**, 403 (2011).
4. L. C. Murtaugh, *Development* **134**, 427 (2007).
5. J. H. Hill, E. A. Franzosa, C. Huttenhower, K. Guillemin, *eLife* **5**, e20145 (2016).
6. B. E. Gregg *et al.*, *J. Clin. Endocrinol. Metab.* **97**, 3197 (2012).
7. J. H. Hill *et al.*, bioRxiv 10.1101/2022.03.24.485696 (2022).
8. J. H. Hill, J. L. Round, *Cell* **184**, 2796 (2021).
9. S. Mukherjee, L. V. Hooper, *Immunity* **42**, 28 (2015).
10. D. P. Strachan, *BMJ* **299**, 1259 (1989).
11. M. J. Blaser, S. Falkow, *Nat. Rev. Microbiol.* **7**, 887 (2009).
12. A. D. Kostic *et al.*, *Cell Host Microbe* **17**, 260 (2015).
13. P. Zheng, Z. Li, Z. Zhou, *Diabetes Metab. Res. Rev.* **34**, e3043 (2018).
14. C. J. Stewart *et al.*, *Nature* **562**, 583 (2018).

10.1126/science.abq6051

PRIZE ESSAY

FINALIST

Apollo Stacy

Apollo Stacy received his undergraduate degree from Washington University in St. Louis and his PhD from The University of Texas at Austin. Currently a postdoctoral fellow at the National Institutes of Health, Apollo will be starting his laboratory at the Cleveland Clinic Lerner Research Institute in 2022. His research investigates how host-derived metabolites shape the ecological balance between commensals and pathogens and thus host susceptibility to inflammatory disease. **www.science.org/doi/10.1126/science.abq6054**

NOSTER | Science MICROBIOME PRIZE

MICROBIOME

Remembrance of infections past

Host-derived metabolite trains the microbiota to develop colonization resistance

By **Apollo Stacy**

The COVID-19 pandemic reminds us that viruses and other infectious pathogens pose a constant threat to life and health. Fortunately, vaccines can bolster our defenses by eliciting immunological memory. Once established, memory cells can respond more rapidly and robustly to a secondary encounter with a pathogen. Although memory was previously thought to be a trait restricted to adaptive immunity, innate immunity can also undergo "training" to achieve an enhanced memory-like state (*1*). Likewise, other cellular systems across the body are increasingly recognized for their capacity to develop memory (*2*, *3*).

Host defense against pathogens relies on synergy between the immune system and microbiota, the diverse microbial communities that colonize mucosal surfaces. In addition to stimulating immunity against pathogens, these communities can directly compete with and exclude pathogens. This microbiota-dependent "colonization resistance" is vital to human health (*4*). Further, antibiotics-induced damage to the microbiota can allow for the expansion of antibiotic-resistant pathogens. Collectively, these pathogens cause a staggering 700,000 deaths worldwide per year (*5*). Thus, a clear need exists for therapies to sustain or enhance colonization resistance.

Like the immune system, the microbiota is highly dynamic, maturing in composition and function as we age, in part through beneficial interactions between the microbiota, diet, and immune system. Undoubtedly, disturbances to the microbiota also affect its maturation (*6*). For example, pathogens can remodel infection sites to access nonfermentable substrates, through high-energy respiration (*7*), and thereby outcompete the microbiota (*8*). But despite the frequency and potentially long-term effects of infections on the microbiota, colonization resistance is most often examined in laboratory mice reared under artificially pathogen-free environments, which raises the question: How do pathogens affect the microbiota's resistance to subsequent infections?

Laboratory of Host Immunity and Microbiome, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. Email: apollo.stacy@nih.gov

One hypothesis is that host organisms and their endogenous microbiota (metaorganisms) have coevolved to harness pathogen-induced disturbances to their advantage. I sought to address this hypothesis as a postdoctoral fellow in the laboratory of Y. Belkaid at the National Institutes of Health. We discovered that like the immune system, the microbiota's resistance to pathogens can be trained (*9*). Mechanistically, this "metaorganism memory" results from the host nourishing the microbiota with the sulfur-containing amino acid taurine. In turn, the microbiota converts taurine to hydrogen sulfide, a noxious gas that inhibits pathogen respiration. We also discovered that taurine can be coopted into a microbiota-directed therapy. Thus, our exploration into how infection alters colonization resistance unveiled a strategy to "immunize" the microbiota (*9*).

To assess how prior infection affects colonization resistance, we used two complementary mouse models: (i) mice harboring the microbiota of wild mice (*10*), because they encounter more infections than their laboratory counterparts, and (ii) mice transiently infected with a food-borne pathogen. Aligning with our hypothesis, both models displayed enhanced microbiota-dependent resistance to gut infection with the antibiotic-resistant pathogen *Klebsiella pneumoniae*. Through sequence-based profiling of the gut microbiota's composition and functional potential, we then identified in both models an expansion of microbial taxa that degrade the metabolite taurine. As predicted from the expansion of these taxa, metabolite profiling confirmed that taurine is indeed enriched in the intestinal tract of mice after infection.

In the gut, taurine is often found conjugated to bile acids—a class of liver-synthesized metabolites that aid digestion (*11*). Supporting their role as the source of elevated taurine after infection, further metabolite profiling revealed an enrichment of taurine-conjugated bile acids in transiently infected mice. Together, these results support a model in which, in response to infection, the host supplies taurine by means of bile acids to the microbiota, driving an expansion of taurine-degrading taxa that heighten resistance to subsequent infection.

Our model next led us to consider that taurine alone, without prior infection, would

PHOTO: APOLLO STACY

be sufficient to enhance colonization resistance. Agreeing with our prediction, supplementing mice with taurine enhanced their microbiota-dependent resistance against *K. pneumoniae* and *Citrobacter rodentium*, a mouse model for pathogenic *E. coli*.

To gain insight into how taurine enhances resistance, we functionally profiled the gut microbiota of taurine-supplemented mice, revealing that taurine markedly induces dissimilatory sulfite reductase (DSR), an enzyme that performs a key step in the degradation of taurine (*12*). A major byproduct of DSR activity is hydrogen sulfide, and we found that microbiota from taurine-supplemented mice released heightened levels of sulfide. These results suggested a role for taurine-derived sulfide in enhancing colonization resistance.

To further elucidate how taurine enhances resistance, we performed a high-throughput transposon mutant screen (*7*) on *K. pneumoniae* in transiently infected mice and found that these mice block aerobic respiration by *K. pneumoniae*. Because aerobic respiration is a primary target of sulfide (*13*), we then reasoned that taurine-derived sulfide, rather than taurine itself, directly mediates colonization resistance.

Supporting this model, *K. pneumoniae* in taurine-supplemented mice could no longer outcompete an isogenic mutant incapable of performing aerobic respiration. Because boosting sulfide with taurine could enhance resistance, we assessed whether depleting sulfide would impair resistance. We treated mice with bismuth, a known sulfide sequestrant (*14*), and as anticipated, these mice became highly susceptible to *K. pneumoniae*, confirming the pathogen-restricting role of sulfide in our model.

Despite the promise of microbiome research for the treatment of human diseases, microbiota-directed therapies are still lacking. In my recent work, a key revelation is that we can mine how the host regulates the microbiota as a natural source of microbiota-directed therapies. By probing one such mechanism, we discovered that taurine, a host-derived metabolite that trains the microbiota, can be co-opted into a potential therapy to combat infections.

Because taurine leverages a protective function (sulfide production) that is endogenous to the microbiota, similar metabolite-based approaches applied to humans could overcome the obstacles that hinder probiotic therapies, for which success relies on introducing completely foreign, exogenous taxa to the microbiota (*15*). An open question is whether training of the microbiota, through use of taurine or other metabolites, occurs at body sites outside the gut. Future studies are on the horizon.

REFERENCES AND NOTES

1. S. Bekkering, J. Domínguez-Andrés, L. A. B. Joosten, N. P. Riksen, M. G. Netea, *Annu. Rev. Immunol.* **39**, 667–693 (2021).
2. S. B. Larsen *et al.*, *Cell Stem Cell* **28**, 1758–1774.e8 (2021).
3. C. A. Thaiss *et al.*, *Nature* **540**, 544 (2016).
4. E. G. Pamer, *Science* **352**, 535 (2016).
5. United Nations (UN), "No Time to Wait: Securing the future from drug-resistant infections" (UN, 2019).
6. J. L. Sonnenburg, E. D. Sonnenburg, *Science* **366**, eaaw9255 (2019).
7. A. Stacy, D. Fleming, R. J. Lamont, K. P. Rumbaugh, M. Whiteley, *mBio* **7**, e00782-16 (2016).
8. B. M. Miller, A. J. Bäumler, *Annu. Rev. Immunol.* **39**, 1 (2021).
9. A. Stacy *et al.*, *Cell* **184**, 615.e17 (2021).
10. S. P. Rosshart *et al.*, *Cell* **171**, 1015.e13 (2017).
11. A. Wahlström, S. I. Sayin, H.-U. Marschall, F. Bäckhed, *Cell Metab.* **24**, 41 (2016).
12. H. Laue, M. Friedrich, J. Ruff, A. M. Cook, *J. Bacteriol.* **183**, 1727 (2001).
13. E. Forte *et al.*, *Sci. Rep.* **6**, 23788 (2016).
14. F. L. Suarez, J. K. Furne, J. Springfield, M. D. Levitt, *Gastroenterology* **114**, 923 (1998).
15. J. Suez, N. Zmora, E. Segal, E. Elinav, *Nat. Med.* **25**, 716 (2019).

10.1126/science.abq6054

PRIZE ESSAY

FINALIST

Irina Leonardi

Irina Leonardi received undergraduate degrees from ETH Zürich and a PhD from the University of Zürich. She performed her postdoctoral work in the laboratory of Dr. Iliyen Iliev at Weill Cornell Medicine. Her research investigated the cellular mechanisms of fungal recognition in the gut and the local and systemic consequences of intestinal fungal colonization. She is currently the Scientific Communications Lead at Immunai in New York.

www.science.org/doi/10.1126/science.abq6056

NOSTER | Science MICROBIOME PRIZE

MICROBIOME

Beyond the gut

Mycobiota modulate immunity and behavior

By **Irina Leonardi**

The microbiota of the gut has been implicated in the development of a variety of pathologies. Though microbiota research predominantly focuses on bacteria, the mammalian gut is home to a diverse community of fungi, viruses, and archaea (*1*). Recent research on the interaction between the host and intestinal fungi—collectively known as the mycobiota—has led to a flurry of exciting discoveries. Next-generation sequencing of the mycobiota has revealed the existence of pronounced changes in the composition of gut fungal communities in a variety of conditions. These changes occur not only in diseases directly affecting the intestine but also in extra-intestinal diseases such as type 1 diabetes and autism spectrum disorders (*1*).

We and others have further investigated the associations between mycobiota composition and therapeutic responsiveness (*2*–*4*). Fecal microbiota transplantation (FMT) has long been used to treat patients with intestinal *Clostridioides difficile* infections (CDIs) (*5*). In CDIs, where the intestinal microbiota is dominated by *C. difficile*, high levels of the gut commensal fungus *Candida* are associated with a lack of responsiveness to FMT (*3*). Is *Candida* consistently associated with a negative FMT outcome in multiple diseases? In a multicenter collaboration, we characterized the mycobiota of individuals who received FMT for ulcerative colitis (UC) (*6*). Our analysis of the patients' mycobiota revealed that those with high levels of *Candida* were more likely to benefit from FMT treatment (*4*). Further, we noted that amelioration of disease was associated with a decrease in *Candida* following FMT. These findings suggest that a careful analysis of individuals' mycobiota composition might help identify subsets of patients more likely to benefit from FMT (*4*) and that the effect of the mycobiota on its human host is highly context dependent.

Multiple factors—including the composition of the underlying microbiota, therapeutic regimen, diet, and genetic background—likely influence the effect of the gut microbiota (*7*). A reductionist approach can help us identify factors that influence the mycobiota's effect and facilitate the development of targeted therapeutic approaches. We developed mouse models to aid in our discovery of the mechanisms of fungal recognition in the gut, which revealed that the mycobiota can influence the immune system both locally and systemically (*8*, *9*).

Our reliance on the sequencing of fecal material to characterize the mycobiota has hindered our ability to evaluate the impact of gut fungal colonization on human health, as this approach poorly reflects the complexity of the bacterial communities along the gastrointestinal tract (*10*, *11*). Thus, we decided to characterize the spatial biogeography of the gastrointestinal mycobiota by next-generation sequencing of highly variable regions of the fungal ribosomal DNA (*12*). We identified a strong distinction between the community of fungi free-floating in the intestinal lumen and those able to closely associate with the intestinal mucosa. This finding made us wonder whether lumen- and mucosa-associated fungi would have distinct effects on their host.

To address this question, we colonized wild-type mice with two model consortia using fungi representative of each location and characterized the immune responses induced by intestinal colonization at the local and systemic level. T helper cells producing type 17 cytokines (T_H17) are crucial players in antifungal immune responses during both systemic infection and mucosal colonization (*13*). Our analysis showed that only the mucosal consortium induced a strong local and systemic T_H17 response, consistent with previous work indicating that the adhesion of bacteria to the intestinal epithelium is necessary for the activation of T_H17 responses (*11*). Further analysis revealed subtle locus-specific differences in the induced responses: The local response was characterized by the production of the cytokines interleukin-17A (IL-17A), IL-17F, and IL-22, whereas the systemic response was dominated by IL-17A. We next investigated the functional relevance of these distinct immune responses at the local and systemic level. We showed that intestinal colonization with mucosa-associated fungi, but not lumen-associated fungi, protected mice against intestinal injury. The cytokines IL-22 and IL-17 are known to play a key role in the maintenance and repair of the intestinal epithelium (*14*). Using mice

Immunai, New York, NY, USA.
Email: irina.leonardi@immunai.com

PHOTO: COURTESY OF IRINA LEONARDI

lacking either IL-22 or IL-17A, we demonstrated that the production of IL-22 by T_H17 cells was necessary for the protective effect exerted by mucosal fungi, whereas IL-17A was dispensable.

Having tested some of the local effects of fungi in the intestinal mucosa, we next turned our attention to their systemic consequences. Inspired by research demonstrating the existence of a gut-brain axis that enables the microbiota to modulate its host's behavior (*7*, *15*), we wondered whether fungi might exert similar effects. We were particularly interested to explore the association between gut mycobiota dysbiosis and autism spectrum disorder. We subjected wild-type mice to a behavioral phenotyping panel and found that gut colonization with mucosa-associated fungi promoted social behavior in mice. We hypothesized that IL-17A, a cytokine known to possess neuromodulatory properties (*15*), might play a mechanistic role. We used a transgenic mouse model to specifically disrupt IL-17 receptor signaling in neurons and demonstrated that IL-17 sensing by neurons mediates the effect of mucosa-associated fungi on social behavior (*12*). These results suggest that mucosa-associated fungi can affect their host along the gut-brain axis. Collectively our findings reveal the dual effect that mucosal-associated fungi exert on the integrity of the intestinal barrier and the neuroimmune modulation of their host's behavior.

Although mycobiota dysbiosis is observed in a disparate collection of diseases (*1*), causal evidence remains scarce. The identification of distinct local and systemic responses induced by fungi colonizing the intestinal mucosa is an important step toward understanding the consequences of mycobiota dysbiosis. Our efforts are now focused on understanding the mechanisms that drive site-specific responses to gut fungi with the hope of revealing their role in the pathogenesis of these diseases.

REFERENCES AND NOTES

1. I. D. Iliev, I. Leonardi, *Nat. Rev. Immunol.* **17**, 635 (2017).
2. S. L. Shiao *et al.*, *Cancer Cell* **39**, 1202 (2021).
3. T. Zuo *et al.*, *Nat. Commun.* **9**, 3663 (2018).
4. I. Leonardi *et al.*, *Cell Host Microbe* **27**, 823 (2020).
5. M. N. Quraishi *et al.*, *Aliment. Pharmacol. Ther.* **46**, 479 (2017).
6. S. Paramsothy *et al.*, *Lancet* **389**, 1218 (2017).
7. C. X. Yap *et al.*, *Cell* **184**, 5916 (2021).
8. I. Leonardi *et al.*, *Science* **359**, 232 (2018).
9. X. Li *et al.*, *Cell Host Microbe* **24**, 847 (2018).
10. K. Atarashi *et al.*, *Science* **358**, 359 (2017).
11. C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg, *Cell Host Microbe* **21**, 433 (2017).
12. I. Leonardi *et al.*, *Cell* **185**, 831 (2022).
13. A. Scheffold, P. Bacher, S. LeibundGut-Landmann, *Curr. Opin. Microbiol.* **58**, 116 (2020).
14. M. F. Neurath, *Nat. Immunol.* **20**, 970 (2019).
15. M. D. Reed *et al.*, *Nature* **577**, 249 (2020).

10.1126/science.abq6056