Science

DETECTING
VIRUSES

CRISPR technologies move
toward diagnostics in the field

CONTENTS

376 & 409–418
Spatial entanglement

368
Rediscovering Colombia

CONTENTS

378 & 423

Weighing proteins

374 & 404

ON THE COVER

Portable diagnostic tools developed on the basis of CRISPR-Cas biology enable rapid, highly sensitive detection of pathogens in patient samples. A conceptual illustration of a Zika virus capsid drawn in the sand highlights the promise of deploying these tools for detection and tracking of disease in remote settings. See pages 381, 436, 439, and 444. *Illustration: Valerie Altounian/*Science*; Data: PDB ID 5IRE*

Published by AAAS

Published by AAAS

Evidence for opportunity

"Our nation is moving toward two societies, one black, one white—separate and unequal." So concluded a 1968 report by the Kerner Commission, established by U.S. President Lyndon Johnson to investigate the race riots of 1967. Not only did the report shine a spotlight on America's unfulfilled promises, it spurred action by politicians and policy-makers. Fifty years later, it is fair—and necessary—to ask if anything has changed. *Healing Our Divided Society*, the 2018 sequel to the Kerner report, argues sadly that gains of the 1970s and early 1980s are evaporating or reversing. But, noting the role of empirical evidence in bolstering past reforms, the new report suggests hopefully that "the quantity and sophistication of scientific information available today far exceeds what was available [in 1968]."

To be sure, there was important progress after the 1968 report: Education achievement gaps narrowed (mostly in the early grades), college participation and degree attainment rose for all groups, and average family wealth for black and Hispanic Americans increased. But the current picture is alarming: Income inequality has exploded; child poverty is unacceptably high, especially in racially concentrated neighborhoods; and black children face considerably lower chances of upward mobility than their white peers.

Thankfully, the editors of *Healing* are right: The U.S. now has more and better policy-relevant research and evidence that can help move the needle. For example, DeLuca, Clampet-Lundquist, and Edin have shown that at-risk adolescents with "identity projects" have increased hopes of staying on track and finishing school; Duncan and Murnane have shown that research-informed strategies and "sensible accountability" can level the playing field between poor and more advantaged youth; and Johnson and Jackson have shown that investments in programs such as Head Start, coupled with sustained K–12 funding, can break the cycle of poverty.

A key takeaway from such examples is that political will is necessary but insufficient without empirical evidence. The question is whether we can be confident in the supply of good research, in renewed political commitment, and in a revived appetite for evidence-informed policy at all levels of government.

Let's hope so. To be prepared, we must address worrisome trends. After decades of federal funding, the U.S. has a robust supply of doctoral-level scientists in education and related fields, but federal resources for their continuing work are meager and politically vulnerable; the odds of winning a first-time grant have fallen sharply. Private foundations mostly advance the public good, but few support general education research and some put advocacy ahead of evidence; hearing a foundation officer say, perhaps jokingly, "of course we support objective research...as long as it proves that charter schools work" may not inspire confidence.

To help address these converging trends, Congress should increase funding for behavioral and social sciences, and governments at all levels should consider new approaches to accessing evidence, perhaps modeled on the Congressional Budget Office, the (now defunct) Office of Technology Assessment, or successful state and local experiments, such as the Washington State Institute for Public Policy and research-practice partnerships in Long Beach (California), Chicago, Baltimore, New York, Houston, and elsewhere. Congress should reaffirm commitments to the National Academies of Sciences, Engineering, and Medicine and the National Academy of Education—reputable producers of nonpartisan and policy-relevant research—and should act on recommendations of the bipartisan Commission on Evidence-Based Policymaking. Private foundations should redouble their commitment to evidence over ideology and to investing in basic research as well as scrupulous evaluations of their investments.

If there is hope for restoring economic and educational opportunity, research is essential. The proven tradition of relying on science to make the world better cannot end on our watch.

–**Michael J. Feuer**

"...policy-relevant research... can help move the needle."

Michael J. Feuer is dean of the Graduate School of Education and Human Development, The George Washington University, Washington, DC, USA. mjfeuer@gwu.edu

IN BRIEF

Edited by **Jeffrey Brainard**

SPACE EXPLORATION

New NASA leader inherits long to-do list

A lunar space station could lead to a human mission to Mars.
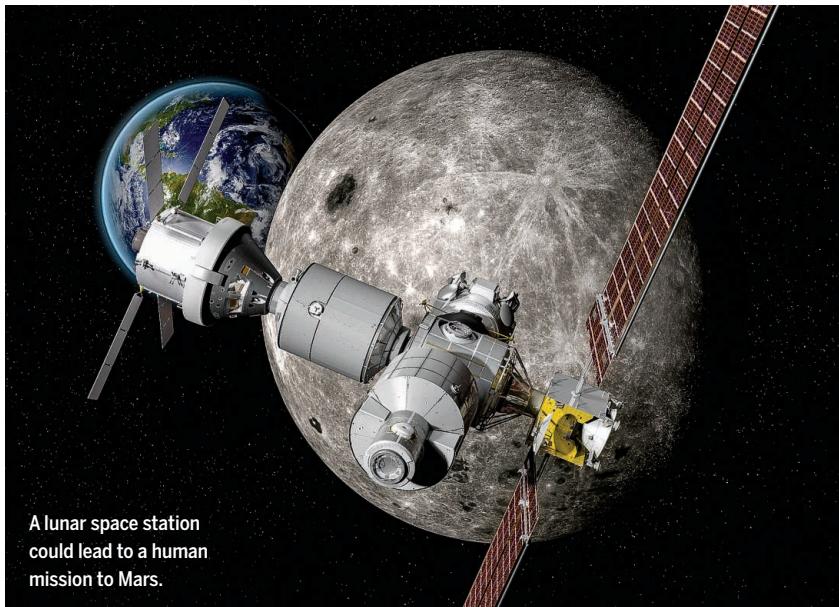
Last week, in a narrow, 50–49 vote, the U.S. Senate confirmed Representative Jim Bridenstine (R–OK), a former Navy pilot, to serve as NASA's 13th administrator. Bridenstine, the first politician to serve in the role, will have a full plate to start. He will manage the delayed launch of the $8 billion James Webb Space Telescope. He'll have to decide on a mission to retrieve a cache of rocks collected by the Mars 2020 rover. He's under pressure to resolve delays and finish NASA's heavy-lift rocket, the Space Launch System. And space experts hope he will flesh out the agency's focus on exploring the moon, including plans for a small orbiting space station called the Lunar Orbiting Platform-Gateway that's intended to support a human mission to Mars. Bridenstine faces skepticism from some members of Congress who question his qualifications and past statements skeptical of global warming. "There is simply no excuse for voting for someone so unqualified to run NASA," Senator Brian Schatz (D–HI) wrote following Bridenstine's confirmation. Winning over critics, and appointing a deputy with a technical background, should be his top priority, observers say.

Dengue vaccine on hold

PUBLIC HEALTH | The world's first dengue vaccine should be used only in people who have previously been infected with the mosquito-borne disease, the World Health Organization (WHO) said last week. Dengue is most dangerous when a person is infected a second (or subsequent) time, and giving the vaccine to people who have never had the disease can leave them vulnerable to a dangerously severe reaction if they are infected later. (The vaccine doesn't confer full protection from the virus.) The vaccine's maker, Sanofi Pasteur, issued a similar warning in November 2017, after new data revealed the risk to previously uninfected people. Because no rapid, reliable test for previous dengue infection is available, the new guidelines mean the vaccine can't be widely used. WHO said it hopes a test might be available in 2 to 3 years.

Cannabis-based drug wins nod

DRUG DEVELOPMENT | A U.S. Food and Drug Administration (FDA) advisory panel last week unanimously recommended that the agency approve the country's first marijuana-derived prescription medication. The drug, called Epidiolex, aims to treat two rare and severe forms of epilepsy in children known as Lennox-Gastaut and Dravet syndromes. GW Pharmaceuticals, based in Cambridge, U.K., makes the syrup, which contains purified cannabidiol (CBD), a pot extract. CBD has been shown to reduce monthly seizures by about 40% in patients who otherwise suffer as many as 100 daily. How CBD reduces seizures is unknown. Unlike tetrahydrocannabinol, another pot extract, CBD does not cause a high. FDA, which is expected to approve Epidiolex, has previously approved synthetic CBD to ease nausea in cancer patients. Medical marijuana is used to treat other ailments, but the dose from smoking or ingesting the plant is hard to consistently control.

Risks of air pollution grow

PUBLIC HEALTH | Air pollution from power plant smoke, dust, and other kinds of fine, inhalable particles is growing and is a leading contributor to deaths worldwide, a

The Turaco lineage, which includes this Knysna turaco from South Africa, has been difficult to place in the avian tree of life.

EVOLUTION

Birds will perch on an improved tree of life

Bird lovers can mark their calendars: The most comprehensive avian family tree is expected to be completed by 2022, thanks to The OpenWings Project launched this month at the American Ornithological Meeting in Tucson, Arizona. The $1.42 million effort will be the first to sequence selected DNA segments from all 10,560 bird species to establish how they are related and trace their evolutionary history. It should be finished long before a similar tree-building effort called the Bird 10,000 Genomes Project, which will sequence the whole genomes of all these species. The OpenWings Project will isolate about 5000 short pieces of DNA from each species, saving time and money by focusing on regions that are very highly conserved among all birds, says project co-leader Brant Faircloth from Louisiana State University in Baton Rouge. Harvard University evolutionary biologist Scott Edwards says OpenWings "will be a huge improvement over what we have now," including an avian tree published in *Science* in 2014 based on 40 whole genomes. But, "Ultimately, OpenWings will be a stepping stone to the grand tree that the whole genomes [will generate]."

new report says. In 2016, the most recent year for which data were available, 95% of the world's population lived in areas where airborne particulates exceeded levels recommended by the World Health Organization. From 2010 to 2016, the global average concentration grew by an estimated 18%, according to the report by the nonprofit Health Effects Institute in Boston and the Institute for Health Metrics and Evaluation at the University of Washington in Seattle. The largest increases came in developing countries such as Bangladesh, India, Nigeria, and Pakistan. Indoor burning of coal, wood, and dung for heating and cooking also contributed substantially to mortality, but these deaths fell 16% during those years as the practice declined. In deaths caused, these forms of indoor and outdoor air pollution rank ahead of other well-known health risk factors such as low physical activity, low birth weight, and poor sanitation, which are estimated to be major killers but did not crack this top 10 list.

EPA releases data access plan

SCIENCE POLICY | U.S. science groups are harshly criticizing a data access proposal unveiled this week by Environmental Protection Agency (EPA) head Scott Pruitt. It would, in general, require EPA to use only publicly available data and models to evaluate the health impacts of pollution when it develops any regulation that would cost more than $100 million annually. EPA said the move will ensure that "pivotal regulatory science" is "fully transparent." But critics say the vaguely worded plan would make it harder for regulators to use certain kinds of studies, including long-term air pollution and workplace exposure studies that aren't easy to reproduce or involve confidential health records.

Ethics plan for Africa genomics

GENOMICS | A consortium of genomics and public health researchers last week published the first ethical framework created by Africans to govern genomics research in Africa. The pact, which is non-binding, emphasizes recruiting African researchers to play a central role in these projects; many African institutions cannot afford the technology required for next-generation sequencing studies. The framework also calls on scientists who plan to use biological samples to obtain an ethical review from the country in which the samples will be collected and for the work to benefit the African people. The Human Heredity and Health in Africa Initiative, funded by the U.S. National Institutes of Health and the Wellcome Trust, produced the framework. Many researchers are interested in human genomes from Africa because of the continent's genetic diversity and high disease burden.

Salk suspends leading biologist

WORKPLACE | The Salk Institute for Biological Studies in San Diego, California, put prominent cancer scientist Inder Verma on administrative leave on 20 April and said it is investigating harassment allegations against him. In an email to Salk employees, Dan Lewis, chairman of the institute's board of trustees, wrote that the institute "recently" learned of allegations about Verma, who was hired by Salk in 1974. The institute began investigating him in February and on 12 March hired The Rose Group, an international employment law firm in San Diego, to conduct an outside probe, a Salk spokesperson wrote in an email. In a statement, Verma denied ever using his position at Salk "to take advantage of others. … I have never inappropriately touched, nor have I made any sexually charged comments, to anyone affiliated with the Salk Institute."

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

Top 10 contributors to deaths worldwide, 2016

High systolic blood pressure	10,455,860
Smoking	6,321,098
High blood glucose level	5,612,454
High body mass index	4,525,095
High total cholesterol	4,392,505
Inhaling fine particles (e.g. soot)	4,092,692
Alcohol use	2,814,642
Inhaling smoke from fuel burned indoors	2,576,361
Impaired kidney function	2,554,213
Diet low in whole grains	2,498,691

● Deaths linked to air pollution

Published by AAAS

ARCHAEOLOGY

Searching for a Stone Age Odysseus

Modern humans and even Neandertals may have plied the Mediterranean long ago

By **Andrew Lawler**, *in Washington, D.C.*

Odysseus, who voyaged across the wine-dark seas of the Mediterranean in Homer's epic, may have had some astonishingly ancient forerunners. A decade ago, when excavators claimed to have found stone tools on the Greek island of Crete dating back at least 130,000 years, other archaeologists were stunned—and skeptical. But since then, at that site and others, researchers have quietly built up a convincing case for Stone Age seafarers—and for the even more remarkable possibility that they were Neandertals, the extinct cousins of modern humans.

The finds strongly suggest that the urge to go to sea, and the cognitive and technological means to do so, predates modern humans, says Alan Simmons, an archaeologist at the University of Nevada in Las Vegas who gave an overview of recent finds at a meeting here last week of the Society for American Archaeology. "The orthodoxy until pretty recently was that you don't have seafarers until the early Bronze Age," adds archaeologist John Cherry of Brown University, an initial skeptic. "Now we are talking about seafaring Neandertals. It's a pretty stunning change."

Scholars long thought that the capability to construct and victual a watercraft and then navigate it to a distant coast arrived only with the advent of agriculture and animal domestication. The earliest known boat, found in the Netherlands, dates back only 10,000 years or so, and convincing evidence of sails only shows up in Egypt's Old Kingdom around 2500 B.C.E. Not until 2000 B.C.E. is there physical evidence that sailors crossed the open ocean, from India to Arabia.

But a growing inventory of stone tools and the occasional bone scattered across Eurasia tells a radically different story. (Wooden boats and paddles don't typically survive the ages.) Early members of the human family such as *Homo erectus* are now known to have crossed several kilometers of deep water more than a million years ago in Indonesia, to islands such as Flores and Sulawesi.

Ancient island-hopping

Recent finds in the Ionian and Aegean seas suggest that early modern humans and Neandertals may have voyaged to remote islands before 130,000 years ago.

Modern humans braved treacherous waters to reach Australia by 65,000 years ago. But in both cases, some archaeologists say early seafarers might have embarked by accident, perhaps swept out to sea by tsunamis.

In contrast, the recent evidence from the Mediterranean suggests purposeful navigation. Archaeologists had long noted ancient-looking stone tools on several Mediterranean islands including Crete, which has been an island for more than 5 million years, but they were dismissed as oddities.

Then in 2008 and 2009, Thomas Strasser of Providence College in Rhode Island co-led a Greek-U.S. team with archaeologist Curtis Runnels of Boston University and discovered hundreds of stone tools near the southern coastal village of Plakias. The picks, cleavers, scrapers, and bifaces were so plentiful that a one-off accidental stranding seems unlikely, Strasser says. The tools also offered a clue to the identity of the early seafarers: The artifacts resemble Acheulean tools developed more than a million years ago by *H. erectus* and used until about 130,000 years ago by Neandertals as well.

Strasser argued that the tools may represent a sea-borne migration of Neandertals from the Near East to Europe. The team used a variety of techniques to date the soil around the tools to at least 130,000 years old, but they could not pinpoint a more exact date. And the stratigraphy at the site is unclear, raising questions about whether the artifacts are as old as the soil they were embedded in.

At Stelida on the Greek island of Naxos, researchers have found stone tools perhaps made by Neandertals.

So other archaeologists were skeptical.

But the surprise discovery prompted researchers to scour the region for additional sites, an effort that is now bearing fruit. Possible Neandertal artifacts have turned up on a number of islands, including at Stelida on the island of Naxos. Naxos sits 250 kilometers north of Crete in the Aegean Sea; even during glacial times, when sea levels were lower, it was likely accessible only by watercraft. A Greek-Canadian team co-led by Tristan Carter of McMaster University in Hamilton, Canada, uncovered hundreds of tools embedded in the soil of a chert quarry. The hand axes and blades resemble the so-called Mousterian toolkit, which Neandertals and modern humans made from about 200,000 years ago until 50,000 years ago. These tools require a more sophisticated flaking method than Acheulean types do, including preparing a stone core before striking flakes off it.

Dating work on the artifacts is ongoing and Carter declined to comment pending publication. But Cherry says the Naxos evidence may be persuasive because it is well stratified, which means researchers should be able to date it more securely. "It is very convincing, because there are a lot more tools in situ," adds Strasser, who, like Cherry, was not involved in the dig. "It is a quarry site littered with Mousterian stone tools."

Other Paleolithic tools that appear to be Mousterian have been recovered on the western Ionian islands of Kefalonia and Zakynthos. The plethora of sites adds weight to the idea of purposeful settlement. "People are going back and forth to islands much earlier than we thought," Simmons says.

But determining which of today's islands were truly islands tens of thousands of years ago isn't easy, as it depends on local land movements as well as broader sea-level changes, says Nikos Efstratiou, an archaeologist at Aristotle University of Thessaloniki in Greece. On the Aegean island of Lemnos, his team found what he thinks is a Paleolithic hunting camp dating back more than 10,000 years. But he can't yet be sure when Lemnos was cut off from the mainland. Efstratiou adds that archaeologists need to better characterize the sorts of tools made on the mainland and the islands, so they can find links between the mainland and island peoples.

Other archaeologists are already reckoning with the possibility that humans and our cousins went to sea thousands of years earlier than had been thought. "We severely miscalculated," admits Runnels, who excavated at the Crete site. If his colleagues are right, he says, "the seas were more permeable than we thought." ∎

ASTRONOMY

Data trove helps pin down the shape of the Milky Way

Europe's Gaia satellite reveals motions for 1.3 billion stars

By **Daniel Clery**

"It's like waiting for Christmas," Vasily Belokurov of the University of Cambridge in the United Kingdom said last week. Days later, on 25 April, he and hundreds of other astronomers around the world got their hands on one of the biggest data dumps in the history of astronomy: the exact positions, motions, brightnesses, and colors of 1.3 billion stars in and around the Milky Way, gathered over the first 2 years of operation by the European Space Agency's (ESA's) €750 million Gaia satellite, which launched in 2013. The data—Gaia's second release—will fuel an explosion of studies about the structure of the Milky Way.

Our galaxy is hard to map because gas and dust veil large parts of it—and because we see it from the inside. The trove of stellar positions from Gaia will sharpen our picture of the galaxy's spiral structures, and the data on stellar motions will allow astronomers to wind the clock backward to see how the galaxy evolved over the past 13 billion years. Gaia's color and brightness information, meanwhile, will help astronomers classify stars by composition and pin down where different types are born.

Closer to home, the new data set contains information about 14,000 asteroids in our solar system. That's a small fraction of the roughly 750,000 known minor bodies, but Gaia provides orbit information 100 times more accurate than before and distinguishes asteroid colors, says University of Cambridge astronomer Gerry Gilmore, who heads the U.K. branch of Gaia's data processing consortium. That should help astronomers identify families of asteroids and trace how they relate to one another, shedding light on the solar system's past.

For cosmologists, the data set will improve distance measures to stars of known brightness such as Cepheid variables, crucial markers for measuring the expansion rate of the universe (*Science*, 10 March 2017, p. 1010). Exoplanet hunters hope that in later data releases Gaia will see thousands of stars shifting from side to side as Jupiter-size planets in distant orbits tug on them. "No one in the world knows what we'll find," says David Hogg of New York University in New York City.

The Gaia team released a first catalog in 2016 that held more than a billion stars, but it only provided motions for 2 million of them. It was a "sampler to get people used to handling Gaia-type data," Gilmore says. Even so, he says the data have led to an average of a paper per day. This time, Belokurov says he and his group have about 50 ideas to pursue, including an assessment of the distribution of mass across the Milky Way and the Large Magellanic Cloud (LMC), a nearby satellite galaxy. They may see the LMC gravitationally perturbing Milky Way objects, which would be a sign that the LMC is more massive than once thought. "There's going to be a complete science explosion," Belokurov says. "I'm planning on not sleeping for a week or two."

> ## "No one in the world knows what we'll find."
> **David Hogg**, New York University

Hogg has invited colleagues from around the world to New York City to analyze the data together. He plans to start by drawing up plots that were not previously possible. Graphing color versus brightness for white dwarf stars, for example, could illuminate how these stellar remnants change as they cool off and eventually become black stellar cinders.

The 450-strong Gaia consortium is already at work on an even bigger data release, planned for 2020. "There are very clear areas we can improve," says ESA's Timo Prusti, Gaia project scientist at the European Space Research and Technology Centre in Noordwijk, the Netherlands. Because the brightest stars saturate the detector, for example, the Gaia team plans to remeasure them in a special short-exposure observing mode.

Contrary to the usual practice in astronomy, Gaia scientists don't get a period of exclusive access to the data. They have to wait to study them after the release, just like everyone else. "It's brave and very admirable," Belokurov says. Gilmore says team members have been laying bets on how many papers will hit the preprint servers on Day One. "It's like going to a festival—the festival of Gaia," Belokurov says. ∎

ARCHAEOLOGY

Siberian sculpture is among the oldest monumental art

The 11,600-year-old Shigir Idol offers an enigmatic glimpse of hunter-gatherers' world view

By **Andrew Curry**

In 1894, gold prospectors digging up a peat bog near the Russian city of Yekaterinburg unearthed something bizarre: a carved wooden idol 5 meters long. Carefully smoothed into a plank, the piece was covered front and back with recognizable human faces and hands, along with zigzag lines and other mysterious details. It also had a recognizably human head, with its mouth open in an "o." For more than a century, the statue was displayed as a curiosity in a Yekaterinburg museum, assumed to be at most a few thousand years old.

This week, a paper published in the journal *Antiquity* argues that the statue was crafted from a single larchwood log 11,600 years ago, making it one of the world's oldest examples of monumental art. In age and appearance although not material, the authors write, the so-called Shigir Idol resembles the stone sculptures of Göbekli Tepe in Turkey, which are often cited as the first monumental ritual structures (*Science*, 18 January 2008, p. 278). Both monuments represent a leap beyond the naturalistic images of the ice age.

The idol also shows that large-scale, complex art emerged in more than one place—and that it was the handiwork of hunter-gatherers and not, as was once assumed, of later farming societies. "We have to conclude hunter-gatherers had complex ritual and expression of ideas. Ritual doesn't start with farming, but with hunter-gatherers," says Thomas Terberger, an archaeologist at the University of Göttingen in Germany and a co-author of the paper.

The first radiocarbon dating of the idol, in the 1990s, yielded a startlingly early date: 9800 years old. But many scholars rejected the result as implausibly old. They argued that hunter-gatherers couldn't have produced such a large sculpture, nor have had the complex symbolic imagination to decorate it.

New samples were taken in 2014. At a 2015 press conference in Yekaterinburg, team members announced (before the results were peer reviewed), that these samples revealed even older dates, moving the age of the sculpture back 1500 years, to a time when the world was still transitioning out of the last ice age.

The new dates come from samples taken from the core of the log, uncontaminated by earlier efforts to conserve the wood. "The further you go inside, the older [the date] becomes—it's very indicative some sort of preservative or glue was used" after discovery, says Olaf Jöris, an archaeologist at the Monrepos Archaeological Research Centre and Museum for Human Behavioural Evolution in Neuwied, Germany,
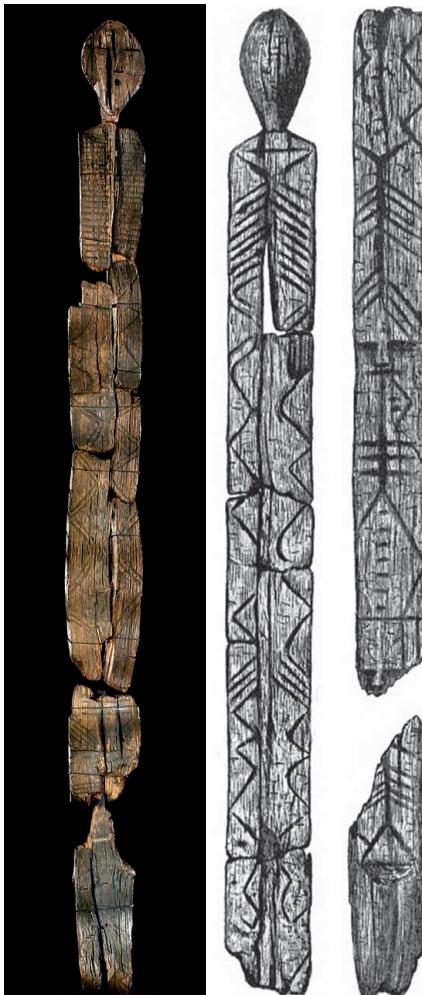
Carved markings cover both front and back of the Shigir Idol, which was originally 5 meters tall.

who wasn't involved with the study. An antler carving discovered near the original find spot in the 19th century yielded similar dates, adding credibility to the result.

The date places the statue at a time when forests were spreading across a warmer, postglacial Eurasia. As the landscape changed, art did, too, perhaps as a way to help people come to grips with the unfamiliar forest environments they were navigating, says Peter Vang Petersen, an archaeologist at The National Museum of Denmark in Copenhagen who was not involved with the study. "Figurative art in the Paleolithic and naturalistic animals painted in caves and carved in rock all stop at the end of the ice age. From then on, you have very stylized patterns that are hard to interpret," Petersen says. "They're still hunters, but they had another view of the world."

At a conference in Yekaterinburg last year, experts debated the meaning of the Shigir symbols, comparing them to other art from the period and more recent ethnographic examples. The most similar finds from that time are those at Göbekli, more than 2500 kilometers away, where hunter-gatherers gathered for rituals and carved similar stylized animals on stone pillars more than 5 meters high.

Terberger sees a more recent parallel: the totem poles of the Pacific Northwest, meant to honor gods or venerate ancestors. Co-author and archaeologist Mikhail Zhilin of the Russian Academy of Sciences in Moscow says the idol might depict local forest spirits or demons. Petersen suggests that the zigzag carvings could be a kind of "Keep out!" sign intended to mark a dangerous or taboo space.

The society that carved the idol is starting to come out of the shadows. Equipped with pumps and special equipment, Zhilin has returned to Shigir and another bog site about 50 kilometers away to excavate finds buried several meters deep in the waterlogged soil. He and his team have found hundreds of small bone points and daggers from the same time period, along with elk antlers carved with animal faces.

They've also found ample evidence of prehistoric carpentry: stone adzes, other woodworking tools, and even part of a pine log smoothed with an adze. "They knew how to work wood perfectly," Zhilin says. The idol is a reminder that stone wasn't the only material people in the past used to make art and monuments—just the one most likely to survive, possibly skewing our understanding of prehistory. "Wood normally doesn't last," Terberger says. "I expect there were many more of these and they're not preserved." ■

Andrew Curry is a journalist in Berlin.

CANCER

Is genome-guided cancer treatment hyped?

New tailored therapies can work wonders, but the pace of development is slow

By **Jocelyn Kaiser**

For people with advanced cancer who are running out of options, many cancer centers now offer this hope: Have your tumor's genome sequenced, and doctors will match you with a drug that targets its weak spot. But this booming area of cancer treatment has critics, who say its promise has been oversold. Last week, two prominent voices in the field faced off in a sometimes-tense debate on what's often called precision oncology at the annual meeting of the American Association for Cancer Research (AACR) in Chicago, Illinois. Their dispute threw a splash of cold water on a meeting packed with sessions on genome-based cancer treatments.

On one side was David Hyman, a 30-something oncologist at Memorial Sloan Kettering Cancer Center in New York City and a leader of clinical studies testing gene-targeted cancer drugs. Genomics won't help most cancer patients at this point, he acknowledged, but many do clearly benefit: "I think this is certainly not hype," Hyman said.

Countering Hyman was another young oncologist, Vinay Prasad of Oregon Health & Science University in Portland, a prolific author of provocative journal articles and outspoken Twitter commentator with more than 17,000 followers. Prasad dispenses sharp criticism about various issues in medicine, from drug prices to conflicts of interest. In the case of precision oncology, he argues that far fewer patients will benefit than proponents suggest. "When you look at all of the data, it's a sobering picture," he said.

Targeted drugs for cancer date back to Herceptin, approved for breast cancer in 1998, and the 2001 approval of the leukemia drug Gleevec, both of which have saved many lives. So far, the U.S. Food and Drug Administration (FDA) has approved 31 targeted therapies for various cancers. Like Gleevec, most work by blocking mutant cancer-driving proteins, although one type harnesses the immune system to fight tumors.

Hyman stressed that growing numbers of patients qualify for the treatments. At Sloan Kettering, which has tested more than 25,000 patients' tumors, 15% can be matched with one of the FDA-approved drugs, and 10% with a drug in clinical trials, he said. Another 10% to 15% or so have a DNA change that matches a potential drug tested in animals. Other cancer centers are reporting similar match rates.

Trials known as "basket" studies are likely to expand the patient pool by showing that these drugs can work against multiple tumor types that carry the same key DNA changes. In February, for example, Hyman and co-authors reported in *The New England Journal of Medicine* that a drug aimed at genes known as tropomyosin receptor kinase (TRK) fusions works for many tumor types.

A lucky few

The portion of U.S. advanced cancer patients who can be matched with a Food and Drug Administration–approved drug based on their tumor's genome is growing slowly, and only some will see their cancer shrink.

- ● Eligible for genome-guided drug
- ● Expected to benefit from drug
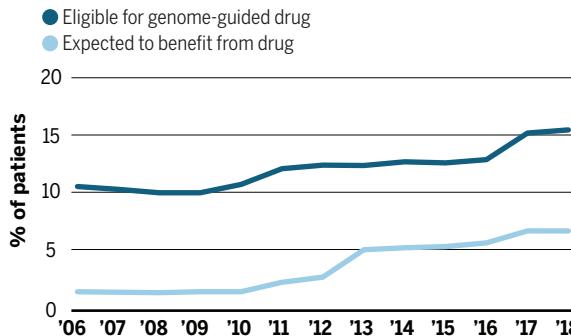
FDA endorsed the first "tissue-agnostic" cancer treatment last year—an immunotherapy drug that is now approved for any advanced solid tumor with flaws in genes that repair DNA. More are expected soon. And although TRK fusions and the DNA repair mutations occur in only a small fraction of patients with a particular cancer type, when tallied across cancers, such drugs can help many patients. As the list of targeted drugs grows, it makes sense to test tumors with genome-wide assays, Hyman said. "If we don't test people broadly, we will miss patients who have alterations for which there is now an approved therapy."

But Prasad sees a glass half-empty when he combines data on how many patients respond to genome-based drugs. He agrees with Hyman that the pool of potential patients is growing. In his latest analysis, published online last week in *JAMA Oncology*, he and his co-authors found that in 2018, 15.4% of 610,000 U.S. patients with metastatic cancer were eligible for an FDA-approved, genome-guided drug. But he also found that because the drugs shrink tumors in only some eligible patients, just 6.6% likely benefited. And many patients relapse after a couple of years on the drugs.

Even the new drugs being tested in basket trials have just a 22% response rate overall, which is no better than chemotherapy, Prasad argued in an unpublished analysis he presented. At the current slow, steady rate of new FDA drug approvals, it would take more than 200 years for all patients to be helped, he said.

He's also troubled by a recent federal decision to allow Medicare to cover a genome sequencing test that analyzes tumors for 324 cancer genes. Prasad thinks these test results could lead some oncologists to prescribe a "bad drug" that matches a mutation but hasn't been proven to work for that cancer. He argues that the testmaker should have first conducted a clinical trial comparing how long patients live after receiving a drug suggested by a genome test to patients whose treatment is based only on conventional pathological analysis of their tumor. "You could answer the question in just 1 year," he said during his AACR talk.

A testy discussion followed. Moderator José Baselga of Sloan Kettering and Hyman criticized Prasad's analysis of the basket trials, saying these exploratory studies are not meant to give firm data on response rates. Baselga also suggested oncologists would be sued if they didn't order genomic testing for lung cancer patients, who often can benefit from a targeted drug. Hyman added that it's "paternalistic" to say physicians can't handle tumor genome data.

The two young cancer researchers found a few points of agreement, however. Prasad praised the basket trials that Hyman helps lead for generating rigorous evidence supporting some targeted drugs: "We actually think similarly," he said. And Hyman conceded that "there's a lot of work [still] to be done." ∎

An automatic weather balloon launcher in Fairbanks, Alaska. The machines need reloading every 12 days.

METEOROLOGY

Robotic weather balloon launchers spread in Alaska

U.S. weather service union fears that automation threatens jobs and sacrifices forecasters' local knowledge

By **Julia Rosen**

Last Thursday, just before 3 p.m., things began stirring inside the truck-size box that sat among melting piles of snow at the airport in Fairbanks, Alaska. Inside, software ran checks on instruments to measure atmospheric temperature, humidity, and pressure; a tray slid into place; and a nozzle began filling a large balloon with gas. Finally, the roof of the box yawned open and a weather balloon took off into the sunny afternoon, instruments dangling. The entire launch was triggered with the touch of a button, 5 kilometers away at an office of the National Weather Service (NWS).

The flight was smooth, just one of hundreds of twice-daily balloon launches around the world that radio back crucial data for weather forecasts. But most of those balloons are launched by people; the robotic launchers, which are rolling out across Alaska, are proving to be controversial. NWS says the autolaunchers will save money and free up staff to work on more pressing matters. But representatives of the employee union question their reliability, and say they will hasten the end of Alaska's remote weather offices, where forecasting duties and hours have already been slashed. "The autolauncher is just another nail in their coffin," says Kimberly Vaughan, a union steward in Juneau.

Autolaunchers have operated around the world for decades, but NWS has just begun to use them. Its demonstration project in Alaska began last October, in Kodiak. Fairbanks is the second site to get one, and the agency will install them at 11 other locations, ending with Nome in 2020, says Susan Buchanan, an NWS spokesperson in Silver Spring, Maryland.

Margaret Mooney, a former NWS meteorologist who is now the director for education and public outreach at the University of Wisconsin's Cooperative Institute for Meteorological Satellite Studies in Madison, says automation makes sense. "I remember walking outside at 3 a.m. when the temperature was 20 below zero, thinking there must be a better way to do weather observations, and sure enough, there is," she says. Moreover, she adds, measurements from Alaska are indispensable for national forecasts because they are "upstream for most of the United States."

Once deployed across the state, the $1.2 million machines, built by Finnish company Vaisala, will save about 8 hours of forecaster time a day—and about $1 million a year at NWS, Buchanan says. That's because the agency tries to staff each remote site with three people, but job vacancies mean overworked employees are shuffled around the vast state to keep up. "We have a difficult time recruiting people to go to these locations," Buchanan says. Recently, some stations have skipped scheduled launches.

Now, Buchanan says, NWS will need just one person at each remote site, to serve as a community liaison and to reload the autolauncher every 12 days. The other staff will relocate to bigger offices, like the ones in Anchorage or Fairbanks, where they can retrain for missions such as forecasting sea ice conditions and volcanic ash hazards, she says. The agency also plans to scale back office space and housing at the remote sites.

Vaughan argues, however, that the plan perpetuates a loss of local knowledge and jobs, adding that the vacancy problem is self-inflicted. She says few people want to work at the remote stations because the remaining positions are temporary and low-paying. She also worries that the machines could malfunction and, without human backups, miss launches.

Dan Sobien, president of the Washington, D.C.–based NWS Employees Organization, fears that if autolaunchers spread to other states, they could help pave the way for further consolidation. He says slow hiring has left roughly 560 positions unfilled nationwide, and he suspects the technology may factor into potential plans to reduce hours and duties at some of NWS's 122 forecasting offices. "In order to do that, they can't have people in there launching balloons," Sobien says.

Buchanan says NWS has funding for only eight autolaunchers outside of Alaska, and that talk of proliferation is premature. But President Donald Trump's administration has called for cutting 355 additional positions from NWS in its 2019 budget proposal for the National Oceanic and Atmospheric Administration (NOAA), which oversees NWS, and it wants to improve the agency's flexibility and efficiency. "Technology and better business practices at the weather service will allow us to absorb the reductions in people in this budget," acting NOAA Administrator Tim Gallaudet said in Washington, D.C., at an 11 April congressional budget hearing. ∎

Julia Rosen is a journalist in Portland, Oregon.

DEVELOPMENTAL BIOLOGY

Chronicling embryos, cell by cell, gene by gene

Single-cell sequencing tracks cells' gene activity during zebrafish and frog embryogenesis

By **Elizabeth Pennisi**

One of biology's great mysteries is how a single fertilized egg gives rise to the multitude of cell types, tissues, and organs that fit together to make a body. Now, a combination of single-cell sequencing technologies and computational tools is providing the most detailed picture yet of this process. In three papers online in *Science* this week, researchers report taking multiple snapshots of gene activity in most of the cells in developing zebrafish or frog embryos. They then assembled those data, taken at intervals of just minutes to hours, into coherent, cell-by-cell histories of how those embryos take shape.

"My first reaction was, 'Wow!'" says developmental biologist Robert Zinzen of the Berlin Institute for Medical Systems Biology. Just last week, two other papers online in *Science* traced cell-by-cell gene activity in planaria, simple flatworms, as they regenerated after being cut into pieces. In vertebrates, "the complexity is much higher," Zinzen notes.

Yet the researchers managed to track the emerging identities of thousands of cells and their progeny. "I think the future of development will be to routinely single-cell sequence embryos," says Detlev Arendt, an evolutionary developmental biologist at the European Molecular Biology Laboratory in Heidelberg, Germany.

All these studies started by gently dissolving embryos of different stages in special solutions, then shaking or stirring them to free individual cells. For each cell, the researchers then determined the sequences of all the strands of messenger RNA (mRNA), which reflect the genes being transcribed.

At Harvard University, teams led by Allon Klein, Marc Kirschner, and Sean Megason focused on zebrafish and frogs, two vertebrates that developmental biologists have studied for decades. In their fish study, Klein and Megason analyzed some 92,000 zebrafish cells, compiling mRNA data from seven different embryo stages. Their group started with 4-hour embryos and ended 24 hours after fertilization—the point at which the basic organs have begun

to appear. Each cell's pattern of gene activity reveals the developmental direction it is headed, and, ultimately, its final identity.

To trace how the cells and their progeny changed over time, the researchers equipped some of the single-cell fish embryos with genetic tracers: many tiny pieces of unique DNA, injected into the embryos' cytoplasm. As the cells repeatedly divided in the growing embryo, these barcodes found their way into the nucleus and were incorporated in the chromosomes. By the end of the experiment, each lineage of cells ended up with a distinctive combination of

In a close-up of a zebrafish embryo, green and red fluorescence marks cells specializing into different tissues. (Blue highlights cellular DNA.)

barcodes. By merging this information with the gene activity profiles, the research team could trace cell fate through time to see how a fertilized egg gave rise to a variety of specialized cells, such as heart, nerve, and skin.

In a separate study, a team led by Harvard developmental biologist Alexander Schier created its own computational method to trace cells in maturing zebrafish. After the group sampled cells every 45 minutes over 9 hours of early embryo growth and sequenced the cells' mRNA, software reconstructed the biography of each cell by taking the gene activity of fully differentiated cells and analyzing which cells in the next-oldest embryo had the most similar gene activity profile. The system worked backward through each embryo stage, all the way to the base of the tree—the starting, undifferentiated cell.

"That was computationally very intense," says Schier, noting that the reconstruction showed the initial one-celled embryo gave rise to 25 main cell types.

The analyses led to some surprises. Developmental biologists had thought that once a cell starts down a path to becoming, say, a muscle cell, it would not stray. But some zebrafish cells switched midstream to a different type, as indicated by changes in their gene activity, Schier and his colleagues report. "The picture is a lot more complex" than we thought, Megason says.

In the frog *Xenopus tropicalis*, Kirschner and Klein did single-cell RNA sequencing at 10 embryonic stages between 5 and 22 hours after fertilization. Their team ultimately read the mRNA of 137,000 cells. The gene activity data showed that even when the frog embryo appears to be an undifferentiated blob, its cells have begun to take on their eventual identities, say as a tail bud.

When Klein, Kirschner, and Megason compared the results for the frog and the zebrafish, they found surprising differences. For example, the developmental routes of certain cell types varied by species. And although the activity of key transcription factor genes was similar in common cell types, the activity of other genes in some cell types differed more than the researchers expected between the two species.

Both zebrafish teams also tracked gene activity in fish that had a mutation expected to seriously disturb development. The two groups' different mutations completely eliminated specific cell types—presumably those directly affected by the disrupted gene—but most other cells differentiated almost normally. This "is just the tip of the iceberg," in terms of analyzing the developmental effects of mutations, Arendt says.

Such studies could also provide a recipe book for stem cell scientists and tissue engineers who want to make new cell types, just as embryos do. The new results, says developmental biologist David Kimelman of the University of Washington in Seattle, "are a real tour-de-force effort and a major accomplishment in being able to understand one of the fundamental questions in developmental biology." ∎

IMAGE: JEFFREY FARRELL

FEATURES

PEACE DIVIDEND

Colombian scientists race to study once-forbidden territory before it is lost to development—or new conflict

By **Lizzie Wade**, *in Triganá, Colombia;*
Photography by **Juan Cristobal Cobo**

The rare travelers intrepid enough to find their way here to northern Colombia come for quiet beaches and coral reefs. Not Camilo Montes. After enduring a choppy motorboat ride across the Gulf of Urabá to this ramshackle village, he and three other geologists head into the jungle, seeking clues to a profound geological event that transformed the Americas.

The team hikes for 3 hours before reaching a spot where a creek spills gently over a mass of smooth, dark rock. They call out GPS coordinates and use rock hammers to chip off flecks, which they examine with magnifying glasses. There's no doubt about it: The rock and its pistachio-colored inclusions were formed as magma welled up and created the mountains of the Panama arc, which forms Panama and northern Colombia. "It's like a photo of when the magma was crystallizing," Montes says, bagging several kilograms of samples to take back to his lab. It's exactly what he came here hoping to find.

Montes, a geologist at the University of the North (Uninorte) in Barranquilla, Colombia, has studied rocks like these his entire career. He investigates the formation of Panama's volcanoes and their collision with South America, which linked the Americas and allowed ecosystems separated for millions

On his first trip to Colombia's Gulf of Urabá, once occupied by guerrillas, geologist Camilo Montes hunts for clues to when the Americas collided.
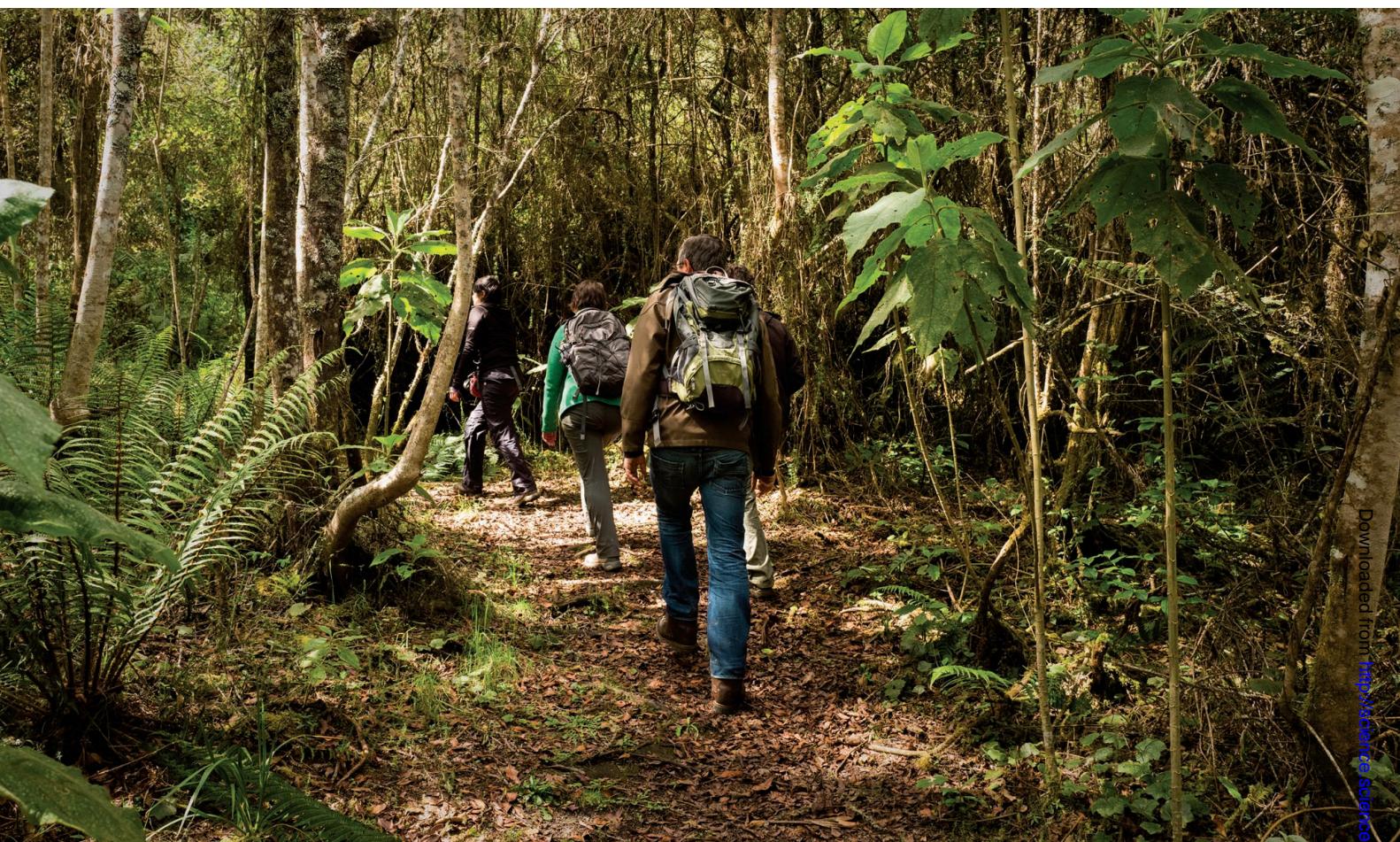
Biologist Juan Manuel Posada and his team are monitoring the Encenillo forest, secondary growth that is thriving on the site of a former limestone mine.

of years to mingle. For decades, scientists have thought that the land bridge between the continents formed about 3.5 million years ago. Citing fossil evidence and the ages of volcanic rocks, Montes and others argue that the continents joined before 10 million years ago (*Science*, 19 July 2013, p. 230). If confirmed, the older date would change both how scientists analyze fossils from the two continents and how they calibrate the molecular clocks used to estimate when species diverged. "If the 3.5-million-year date is wrong, it could upend everything," says Susana Caballero, a biologist at the University of the Andes (Uniandes) in Bogotá. Dating the crystalline inclusions, Montes says, may influence that scientific debate.

BEFORE VENTURING HERE, Montes had limited his fieldwork in this geological collision zone to Panama, where telltale rocks are also uncovered. For decades, the Urabá region, just across the border in Montes's home country, was occupied by the Revolutionary Armed Forces of Colombia (FARC), a leftist

guerrilla group that had waged war on the Colombian state since 1964. "Those who went in never came out," Montes says. That all changed with a stroke of a pen on 26 September 2016, when Colombian President Juan Manuel Santos signed a peace deal with the FARC. Guerrilla fighters shuttered their jungle camps and handed over their weapons.

Now, Montes and other Colombian scientists are rushing in, exploring the geology of

Geologist Felipe Lamus Ochoa studies volcanic rock near Triganá, Colombia.

their country, its wealth of species, and how its ecosystems are coping with stresses such as deforestation and climate change. Those forays are risky: Vast areas haven't yet been cleared of land mines (see sidebar, p. 371), and drug traffickers, paramilitary groups, and non-FARC armed insurgents plague the countryside. But the researchers are seduced by the prospect of prying scientific secrets from huge swaths of land that are no longer off-limits. Urabá's geology, Montes says, "is all a blank slate."

MUCH OF COLOMBIA'S BIODIVERSITY is also unexplored. The nation is thought to contain nearly 10% of Earth's known species. It has the most bird species of any country and trails only Brazil, nearly eight times larger, in amphibians. The wellspring of that diversity is a unique geography: In Colombia, the Andes mountain system fractures into three ranges, crucibles of speciation that have spawned hundreds of distinct ecosystems. "Colombia has an enormous responsibility to conserve these species," says Andrés

Link, a biologist at Uniandes who studies monkey behavior. But to protect its species, researchers need to inventory them.

Until now, surveys have been difficult, and prospects for long-term studies were even bleaker. As a student in 2002, Link studied monkey populations from a research station in Macarena national park, 200 kilometers south of Bogotá. At the time, Macarena was a FARC stronghold, and hostilities were intensifying. Link was at the station that January when locals, following FARC orders, urged the scientists to leave. Biologist Adriana Sánchez Andrade, another Uniandes student then, also found that critical areas of Colombia were off-limits for her work on plant ecology. She opted to do fieldwork in Ecuador. "It was unbelievable, not to be able to go out and explore my own country," she says.

Now that researchers are more free to explore, they are finding that the long conflict had an ecological upside: suppressing development in guerrilla-held territory. FARC fighters deforested some areas to plant coca, but their camps were well hidden and had a light environmental footprint. The violence also deterred farmers and ranchers from encroaching on the wilderness.

In the months since peace took hold, the Colombian government has sponsored collecting trips to those ecological terrae novae. The Alexander von Humboldt Biological Resources Research Institute, an environmental nonprofit in Bogotá, has run four such trips since August 2016 and is planning a fifth. "We've found new species on almost all the expeditions," says Javier Barriga Bernal, a Humboldt biologist and expedition coordinator. The team has collected thousands of samples and is putting the data online. "This is the most optimistic moment I've ever experienced in Colombia," says Mailyn Gonzalez Herrera, who spearheads Humboldt's genetic research. "It's a privilege and a dream for us as scientists to be able to discover our own megadiverse country."

But the once-occupied ecosystems won't remain pristine for much longer, Link says. "These areas haven't just opened up for scientists," he says. "They've opened up for the whole machinery of development." Deforestation is picking up nationwide, including in Colombia's Amazon region, as farmers clear forest for pastures and crops. In 2016, the last year for which statistics are available, deforestation in Colombia jumped 44%. "The country is experiencing more environmental deterioration than ever before," Link says.

He is now studying spider monkeys in Magdalena Medio, an area in central Colombia with rampant deforestation over the past several years. Link has seen the monkeys' habitat shrink and fragment, making the fruit they depend on hard to find. He has

Cleaning up the killing fields

BOGOTÁ—Fernando Cristancho watches nervously as a remote-controlled, Jeep-like contraption with six wheels inches across a denuded field laid with land mines. The nuclear physicist and his colleagues at the National University of Colombia here built the vehicle themselves, and this foray at a military test facility near Bogotá is its debut outside the lab. Extending from its frame is a metal arm that sweeps in a semicircle as the machine advances. The arm carries a neutron source and detectors designed to record the number and speed of the neutrons ricocheting off hidden objects such as mines—revealing them from a safe distance, Cristancho hopes.

The mines at the test field are disarmed. But across Colombia, land mines killed or injured more than 11,500 people during a decades-long war between the government and the Revolutionary Armed Forces of Colombia (FARC), and a 2016 peace deal has not eliminated the threat. Mines continue to kill or maim dozens each year, and in rural areas, they keep villagers who fled during the conflict from returning home. To tackle the scourge, Colombia has committed to ridding itself of land mines by 2021.

"There's still a huge amount to work to do," says Nestor Peña, an electrical engineer at the University of the Andes (Uniandes) here. So far, only 11% of an estimated 52 square kilometers littered with mines has been cleared. The going is slow, he says:

Low-tech land mines riddle rural Colombia, and most can't be found with metal detectors.

"Techniques that are well-established for conventional wars haven't worked in Colombia." Worldwide most land mines are military-grade and can be spotted with metal detectors. In Colombia, the FARC and other armed groups crafted low-tech improvised explosive devices (IEDs) from PVC pipes, plastic soda bottles, and other hard-to-detect containers packed with explosives and attached to a detonator. "These IEDs are designed to be hard to find," says Chris Ince, the Colombia program manager for the HALO Trust, a humanitarian demining agency here. Colombia's rugged terrain, wet climate, and lush forests also make demining a stiff challenge.

Protocols to find and disarm those sorts of homemade mines require painstakingly clearing vegetation and then removing, millimeter by millimeter, the top 15 centimeters of soil to expose mines. That process is invasive—and possibly harmful—to Colombia's rich biodiversity (see main story, p. 368). One fragile ecosystem is the páramo, an otherworldly high-altitude meadow that includes Chingaza national park, near the capital. The park's eastern half, riddled with mines, is closed to visitors. Demining there could damage the thick, porous soil that collects and filters the freshwater that Bogotá relies on.

So Cristancho and others are seeking less arduous and destructive ways to do the job. Electrical engineers at the National University are testing a stationary detector featuring ground-penetrating radar, which could find mines from a distance through thick vegetation. At Uniandes, Peña and his colleagues try to help mine-sniffing dogs work more efficiently. Upon catching a whiff of an explosive, the dog sits—and for now, its handler must be close enough to see it, making for slow progress when land mine density is low, as it usually is in Colombia. The Uniandes engineers are testing devices that can inform handlers what their dogs are doing up to 300 meters away, allowing the dogs to roam.

"The combination of various techniques is what's going to increase efficiency and decrease the time [demining] takes," says Diego Torres, a nuclear physicist working with Cristancho at the National University. But Ince points out that Colombia's 2021 target doesn't leave much time to vet and deploy techniques.

At the proving ground, Cristancho's neutron-based minesweeper detects every one of the neutered mines. "The thing works," he says, with relief. But he worries about how the vehicle will perform when up against the different moisture levels across Colombia's diverse soils. No new demining technology has ever been tested in such challenging conditions, Cristancho says. "There are no research protocols for this." *—Lizzie Wade*

A landscape opens up

A 2016 peace treaty signed by Colombia and the Revolutionary Armed Forces of Colombia (FARC) allows scientists to study some of the most biodiverse and geologically significant territory in the world.

Gulf of Urabá

La Miel (Panama)
Sapzurro
Triganá
Barranquilla

Magdalena Medio and Bajo Cauca

Chingaza national park

PANAMA

Encenillo Biological Reserve

Bogotá

COLOMBIA

Pacific Ocean

Macarena national park

● FARC presence in 2013
● FARC presence in 2002

0   250
Km

Geologists take notes on strata exposed in Colombia, near the Gulf of Urabá. For the first time, they can directly compare the geology to rocks in Panama.

also spotted several albino spider monkeys—a sign, he says, of inbreeding, which becomes more common as populations shrink.

The environmental destruction is in vivid relief on the geologists' journey in Urabá. On their trek to the smooth, dark rock, the landscape becomes hilly. In one entrancing swath of forest, towering ceiba trees loom and hooting howler monkeys leap from branch to branch in the understory. But over the crest of the next hill, cornstalks wither along the trail, and the valley below is carpeted with grass where the land has been cleared for cattle ranching.

"HERE'S THE ENTRANCE," says Juan Manuel Posada, pointing to an apparently impenetrable thicket. The biologist tucks his tall frame between branches and vines and slips into the forest of the Encenillo Biological Reserve, where he and others are studying the ability of denuded landscapes to regenerate. Until 1992, this area, about a 90-minute drive from Bogotá, was a limestone mine. Depressions—former pits, now filled in—are visible below the overgrowth. The FARC captured La Calera, a town along the road to Encenillo, in 1994, severing access to the reserve.

Now, taking advantage of what he calls "the post-postconflict" that Bogotá has enjoyed since the FARC was driven away in 2003, Posada is chronicling how secondary forests are rebounding in this and 45 other plots around the capital. His team from Del Rosario University in Bogotá is tracking pioneer species, how long reforestation takes, and whether secondary forests in the Andes sequester as much carbon as primary forests. At Encenillo, dashes of red paint mark trees in a 20-square-meter plot that Posada and his colleagues have monitored since 2013. They measure size, growth rate, root mass, rate of photosynthesis, carbon storage, and other indicators of ecosystem health.

A mature forest at this latitude and elevation—about 2800 meters—would be so thick with tall trees that little light would reach the forest floor. Here, patches of blue sky are visible through the canopy. But the forest is on the mend, and wildlife is coming back. Spectacled bears, an emblematic but vulnerable and rarely seen Andean species, have even been spotted. "There's a lot of biodiversity even after centuries of human intervention," Posada says. His research, he says, supports safeguarding areas even after

they are no longer pristine.

Still, Encenillo and other ecosystems face a reckoning as the climate changes. Posada is tracking that trend in another landscape once occupied by guerrillas: the rainy, high-altitude páramo of Chingaza national park, a treeless ecosystem that supplies 80% of Bogotá's water.

A few years ago, Sánchez Andrade, now Posada's colleague at Del Rosario University, spent 18 months collecting data on temperature, precipitation, cloud cover, and how each variable affected the growth rate and photosynthesis of plants native to the páramo. She and Posada hope to anticipate how the ecosystem may respond to climate change. Its plants should be resilient to rising temperatures, Sánchez Andrade says, because they can withstand the páramo's daily 20°C swings. But if climate change were to curtail the páramo's nearly year-round rainy season, the consequences could be devastating for the ecosystem—and for the millions of Bogotá residents who rely on the aquifer. "Everything is so uncertain

The moist, high-altitude terrain of Chingaza national park, once a guerrilla stronghold, supplies 80% of Bogotá's water.

right now," Sánchez Andrade says. But at least, she says, she can be here to study it.

"BIENVENIDOS A PANAMA." The hand-painted wooden sign greets Montes and his team of geologists at the border between Sapzurro, Colombia, and La Miel, Panama, a short boat ride from Triganá. To Montes's chagrin, the geology here turns out to be less compelling. It's dominated by sedimentary rocks that built up gradually, rather than volcanic rocks that provide a precise time marker. But that's not the only reason the area might not be ideal for his research.

The sleepy, beach-town feel of Sapzurro and La Miel belies a humanitarian crisis engulfing the border. Elsewhere, Colombia is struggling to absorb more than half a million refugees from Venezuela's deepening calamity. Here, migrants come from as far away as Asia and Africa, hoping to cross into Central America on their way to the United States or Canada. The turmoil attracts human traffickers to Urabá.

The region also has long been a popular route for drug smugglers. It connects the remote mountains of Antioquia, a state where farmers grow coca for processing into cocaine, to the Caribbean Sea. "The stars look beautiful here," a guide tells the geologists. "But some of them aren't stars—they're drones," which drug cartels use to monitor the gulf. The FARC, like other rebel groups in Colombia, funded its activities through the cocaine trade. Paramilitary groups, cartels, and smaller guerrilla outfits have filled the vacuum. "Now it's pretty anarchic," says Pablo Stevenson, an ecologist at Uniandes who studies primates. In the past couple of years, coca production in Colombia has soared, according to the United Nations Office on Drugs and Crime. That trend could usher in more violence and more land mines—and restrict access again as cartels and paramilitary groups expand production and occupy more territory.

Warning signs are flashing elsewhere. Farther inland, Stevenson wants to reopen Macarena station. He has made three trips to the area since 2015 and has observed healthy populations of monkeys there (*Science*, 21 October 2016, p. 271). But negotiations to turn the station's lights back on haven't been smooth. Despite the support of some villag-

ers hoping for jobs, armed groups in the area haven't agreed to allow the researchers back in. "The optimism I had at the beginning of the year is diminishing," Stevenson says. Barriga Bernal, who coordinates the Humboldt expeditions, also has faced resistance. "Distrust is a generalized feeling in places where the conflict lasted nearly 60 years … and violence is still a daily occurrence," he says. "In these areas, the postconflict era is still far away."

Many scientists also see a looming threat in Colombia's presidential elections next month. A conservative, Iván Duque Márquez, leads in the polls, vowing to modify the peace accord to jail former FARC fighters. Some fear that Duque, if elected, "could blow up the peace process," says Uninorte's Felipe Lamus Ochoa, a geologist working with Montes. In that event, further field-work in Urabá would be unlikely. For now, Montes has seized on what may be a fleeting opportunity; earlier this week, he was back in Urabá with dozens of students. "The access has improved," he says. "But who knows how long it will last?" ∎

INSIGHTS

PERSPECTIVES

REGENERATION

Regenerating tissues

Studies of regeneration inform us about how to position organs

By **Elly M. Tanaka**

Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Vienna, Austria. Email: elly.tanaka@imp.ac.at

Within the first days to weeks of embryonic development, communication between different cells places the brain and eyes within an emerging head-to-tail, top-to-bottom, and left-to-right body axis (*1*). Remarkably, two other biological situations can serve up heads on a platter: Some types of flatworm (planaria) can regenerate the head after body transection, and miniature brains can form from aggregates of mouse or human stem cells (organoids) (*2–5*). Axial self-organization also operates during regeneration in vertebrates such as the salamander (one species of which is the axolotl). Although they cannot regenerate a head, they can faithfully regenerate whole appendages such as an arm or leg.

These examples highlight the remarkable self-generating ability of the body axis communication system and the competence of cells to organize themselves according to (re)generating cues. Substantial progress on understanding the basic principles and the molecular underpinnings of regeneration in the flatworm and salamander, including the recent sequencing of their genomes, has provided important insights and tools for understanding these processes (*6–8*), which could bring us a step closer to engineering human organs. On page 404 of this issue, Atabay *et al*. (*9*) and Hill and Petersen (*10*) have tricked the planaria into aberrant eye placement, revealing a fascinating underlying organ-placement system.

When an adult flatworm is cut in half, the headless stump regenerates the instructions for making the head and for placing the brain and eyes correctly and in a way that connects with the remaining structures—an incredible feat. The cuts can be oblique to the main body axis or a small piece 1/300th of the body can be excised, and an appropriately patterned, intact animal will emerge—with body parts and position scaled down to the size of the remaining body fragments. This means that when a flatworm is decapitated, the head part readjusts the size and position of the head while regenerating the tail—further highlighting the amazing properties of axial self-organization.

Regeneration relies on the synchronous coordination of cells to form spatial signaling centers that produce morphogens—extracellular signaling molecules that instruct the placement of surrounding cells to form different cell types based on concentration and time-dependent

The axolotl salamander, shown here, is a model of limb regeneration.

signaling (*11*). In the case of organoids, the formation of organizing centers is currently spontaneous and difficult to predict or control, whereas in the regenerating flatworm or salamander limb, body amputation initiates repatterning in a process that uses landmarks existing in the remaining stump as a guide (*6*, *12*, *13*). These regenerating animals therefore represent valuable models for understanding the emergent properties of morphogen systems and how they may be controlled during spontaneous patterning in organoids.

The morphogens that instruct stem cells to make a head versus a tail in flatworms are similar to those in other animals, including humans, which reinforces the usefulness of regenerative worms to understand principles of tissue self-organization. Along the head-to-tail axis, Wnt morphogens suppress head formation, whereas bone morphogenetic protein (BMP)–type morphogens affect top-to-bottom (dorsal-ventral) cell differentiation. Adult flatworms harbor a large number of stem cells, called neoblasts, throughout their body that are constantly turning over tissues. To maintain homeostasis, neoblasts continuously receive instructions from morphogens that are expressed by a scaffold of muscle cells running head-to-tail and laterally that encase the stem cells. High levels of Wnts are produced by tail muscles, whereas low levels are found in head muscles. Reflecting this tissue turnover, when cells are prevented from responding to Wnts through inactivation of β-catenin, the worms slowly, over weeks, transform into multiheaded animals (*2*).

What happens when a flatworm is cut in half? The removal of tissue simultaneously causes some neoblasts to accumulate at the cut site, some of which differentiate into tip cells, and the remaining muscle cells start resetting the gradient of Wnt expression scaled to fit the total size of the new piece. The planaria starts a process of both differentiating the new tip tissue and replacing the old tissue to form a small-sized planaria with appropriately proportioned organs. The preexisting information in the adult planaria that ensures gradient directionality so that heads will grow from tail fragments and vice versa is currently unknown.

To understand how neoblasts find their way to a site of eye organogenesis, the main approach of Atabay *et al.* was to remove the posterior portion of the body, leaving the head fragment. After letting the body morphogen gradient readjust, they surgically ablated one of two eyes. In planaria, eye progenitors start the process of differentiation posterior to the location of eye formation; a stream of cells migrate to the final eye field that Atabay *et al.* call the tar-get zone (TZ). The regenerated eye formed closer to the tip of the head compared with the remaining eye, suggesting that newly formed eyes read the body gradient information and identify the TZ. If the tip of the head fragment was removed so as to further alter the properties of the body axis, the de novo eye shifted backward with respect to the uninjured eye. Interestingly, the undamaged eye stayed in place, indicating that the preexisting eye dominantly attracts new eye progenitors, preventing de novo eye formation on that side. Indeed, when the authors
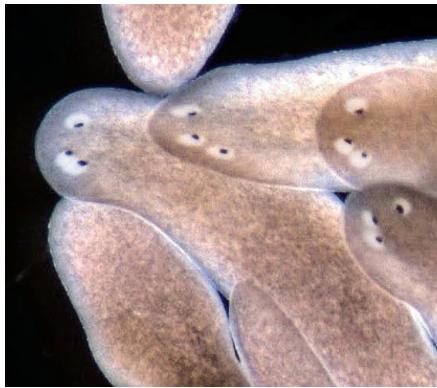
Three eyes can be generated by altering body axis in regenerating types of planaria.

left a remnant of eye tissue during ablation, the eye formed at the "old" location rather than nucleating a new eye.

In a complementary approach, Hill and Petersen prevented the synthesis of the Wnt signaling inhibitor *notum* in unamputated animals, which shifts Wnt signaling forward along the body so that an extra set of eyes emerged anterior to the normal set after several weeks. Interestingly, both sets of eyes were smaller than expected. The new, anterior eyes were connected to the brain, demonstrating that the nervous system had shifted its organization according to the new Wnt gradient imposed by *notum* knockdown.

In both papers, if the "old" eye was ablated, a new eye was not regenerated in its place because it was presumably outside of the new TZ. This led Atabay *et al.* to propose a new concept: the targetable zone (TAZ), broader than the TZ, in which an already formed eye can be maintained but not formed de novo. Indeed, transplantation of eyes to different locations in the body showed that locations far from the TZ, such as the tail, could not stably sustain an eye, whereas locations closer to the TZ could. The TAZ is likely defined as a region where eye progenitors roam toward the TZ. These cells can be recruited to an existing eye that is slightly outside the TZ.

In the future, a deeper understanding of how cells nucleate eyes de novo and whether these properties are the same or different to the attractiveness of preformed eyes, including molecular identification of the attraction cues, will be important. Also fascinating will be to understand how a preexisting eye shifts position slowly over time in a regenerating head fragment. Interestingly, Hill and Petersen also used knockdown of *wntP-2*, which resulted in extra pharynx formation, showing that such rules for organ placement may be general or often used.

It is unlikely that we will regenerate heads or eyes soon, but such principles will be relevant for engineering tissues in vitro or contemplating regeneration of other body parts, as seen in salamanders. After limb amputation anywhere along the length, cells from the stump regenerate the missing part. Unlike planaria, the old limb is not rescaled through stem cell–mediated tissue turnover, but cells accumulate at the amputation plane to form a bud that grows to reform the correct skeletal elements. It is likely that scaling of gradients occurs across this bud. Another similarity is the requirement for cells and information from multiple axial locations to come together to signal the location where the tip of the limb should grow (*6*). Last, skeleton formation involves the self-aggregation of chondrocytes in different size clusters depending on position to make upper and lower arm and hand bones (*14*). This needs to happen in coordination with the bone at the amputation plane, but can also happen de novo in the absence of a bone template. Developmental studies have proposed models of morphogen action for bone condensation, and such signaling events would have to be tuned by position-dependent amputation signals and according to animal size in order to yield appropriate patterning (*15*). Understanding how injury events, position, and size are integrated to control morphogen expression will be greatly facilitated by the ease of genome manipulation in these animals (*7*, *8*). ∎

REFERENCES

1. C. P. Petersen, P. W. Reddien, *Cell* **139**, 1056 (2009).
2. S. Owlarn, K. Bartscherer, *Regeneration (Oxf.)* **3**, 139 (2016).
3. S. A. Elliott, A. Sánchez Alvarado, *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 301 (2013).
4. M. Eiraku *et al.*, *Cell Stem Cell* **3**, 519 (2008).
5. M. Eiraku *et al.*, *Nature* **472**, 51 (2011).
6. E. M. Tanaka, *Cell* **165**, 1598 (2016).
7. S. Nowoshilow *et al.*, *Nature* **554**, 50 (2018).
8. M. A. Grohme *et al.*, *Nature* **554**, 56 (2018).
9. K. D. Atabay *et al.*, *Science* **360**, 404 (2018).
10. E. M. Hill, C. P. Petersen, *eLife* **7**, e02238 (2018).
11. N. Takata *et al.*, *Nat. Commun.* **8**, 1339 (2017).
12. E. G. Butler, *J. Morphol.* **96**, 265 (1955).
13. S. V. Bryant, L. E. Iten, *Dev. Biol.* **50**, 212 (1976).
14. R. P. Gould *et al.*, *Exp. Cell Res.* **72**, 325 (1972).
15. J. Raspopovic *et al.*, *Science* **345**, 566 (2014).

QUANTUM ENTANGLEMENT

Split, but still attached

Entanglement survives between spatially separated atoms of an atomic cloud after it expands

By **Daniel Cavalcanti**

Recent years have witnessed the beginning of the second quantum revolution, in which an impressive degree of control over quantum systems has led to several applications in quantum communication, computation, and sensing, along with new host materials reaching commercial success. A key driver behind many of these applications is entanglement, a form of correlation that can develop between quantum systems that is stronger than any type of correlation that can exist between the macroscopic systems we deal with in our everyday life. The creation, manipulation, storage, and detection of entanglement have posed some of the biggest challenges to quantum physicists. On pages 409, 413, and 416 of this issue, Fadel *et al.* (*1*), Kunkel *et al.* (*2*), and Lange *et al.* (*3*), respectively, describe three independent experiments in which entanglement is observed in a system composed of thousands of ultracold atoms. More importantly, the entanglement is observed between atoms occupying different spatial regions, which paves the way to new applications of these systems.

Entanglement is a very fragile property. Generating it requires highly precise operations and very low levels of noise. There are generally two approaches for entangling particles. The first requires the ability to control each particle and to entangle them one by one by generating suitable interactions. Using this strategy, physicists have managed, for example, to produce entangled states of up to 10 photons (*4*) and 20 ions (*5*). The second approach involves confining the particles and applying controlled global operations in order to make them interact collectively and evolve into an entangled state. This idea has been used to entangle thousands of atoms in Bose-Einstein condensates (BECs), a state of matter at such extremely low temperatures that all atoms behave collectively (*6*). Although this method has been used to entangle a huge number of particles, the lack of control over individual particles means

that these systems are not applicable to the majority of quantum-information tasks (see the figure, top).

The three studies in this issue used the second approach and demonstrated an important step toward having more control over the generated entanglement (see the figure, bottom). They generated ultracold atomic clouds and split them into different spatial regions, where they could then perform measurements. By making observations of local properties of each separated region, they demonstrate the existence of entanglement between the atoms belonging to these separated cloud regions. In order to prove the existence of entanglement in their systems, each group of authors developed their own strategy, based on two key concepts of quantum mechanics—namely, the uncertainty principle and quantum steering.

The uncertainty principle states that there are pairs of quantum observables O_1 and O_2 that cannot be determined simultaneously with arbitrary precision, no matter how sophisticated the implemented experimental setup is. The standard Heisenberg uncertainty relation states that $\Delta O_1 \times \Delta O_2 \geq k$, where ΔO quantifies the uncertainty one has in the measurement of O and k is a constant that depends on the observables in question. Examples of such pairs are the position and momentum of a particle, or different spin directions of an electron.

For systems composed of many parties (here, the different atomic cloud regions), there are relations between the uncertainties of observables applied to each party that are satisfied by all nonentangled states (*7*, *8*). Lange *et al.* showed that measurements of spin observables applied on the spatially separated atomic clouds violate one of these relations, proving that the two clouds are entangled.

Fadel *et al.* and Kunkel *et al.* also make use of the uncertainty principle but combine it with yet another interesting property of quantum mechanics, that of quantum steering. This concept was first discussed by Schrödinger in 1935 (*9*) and more recently has emerged as a new form of quantum nonlocality with interesting properties (*10*, *11*). Schrödinger noticed that if two observers—say, Alice and Bob—share a system of two spin-½ particles in a maximally entangled state $|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle$, and Alice projects her system into the state $a|\uparrow\rangle + b|\downarrow\rangle$, Bob's system immediately collapses to $a*|\uparrow\rangle + b*|\downarrow\rangle$, where * stands for complex conjugation. This result implies that by adjusting the coefficients a and b, Alice can prepare different states on Bob's system; that is, she can steer it.

In order to understand how steering and the uncertainty principle can be used to detect entanglement, consider this particular example. Suppose that in every run of an experiment, Alice and Bob share a maximally entangled spin system $|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle$. In each round, they will both either measure their spins along the x or z directions. According to the uncertainty principle, after averaging over all runs, Bob will observe that the uncertainties of his two measurements satisfy $\Delta x \times \Delta z \geq 1$. Notice, however, that because of steering, when Alice measures z and observes that the spin direction is $|\uparrow\rangle$, she knows for sure that the result of the z

Ultracold atom entanglement

It is difficult to address and control individual atoms in Bose-Einstein condensates. Fadel *et al.*, Kunkel *et al.*, and Lange *et al.* show that entanglement is preserved between atoms when these atomic clouds are split up.

The whole cloud
Previous experiments on entanglement were global measurements on the entire atomic cloud that could not address individual atoms.

Atom

The cloud split
Local measurements on atomic clouds that were split demonstrate the existence of entanglement and quantum steering between separated atoms.

Institut de Ciències Fotòniques (ICFO), The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain. Email: daniel.cavalcanti@icfo.eu

measurement on Bob's system is also |↑⟩ (the same certainty is achieved if she observes |↓⟩ instead). Equivalently, if Alice measures x and observes spin direction |→⟩ = |↑⟩ + |↓⟩, she knows that Bob's x measurement will also provide |→⟩ (similarly for |←⟩).

Thus, from Alice's perspective, there is no uncertainty in Bob's measurements because she knows exactly what Bob's outcome will be, given that he applies the same measurement as hers. This example illustrates how steering allows for an apparent violation of the uncertainty principle that can only happen if Alice and Bob share an entangled state. This basic idea was used by Fadel et al. and Kunkel et al., although in a more complex fashion. They demonstrated how spin measurements on the atoms residing on one side of the cloud led to a reduction in the uncertainty of measurements applied to the other side.

The articles by Fadel et al., Kunkel et al., and Lange et al. address an important requirement for several applications in

> "...they demonstrate the existence of entanglement between the atoms belonging to these separated cloud regions."

quantum information and sensing: namely, addressability. The three papers demonstrate the flexibility of BECs in the generation and detection of entanglement by dealing with different system sizes (from a few hundred to a few thousand atoms), using different entanglement criteria to detect entanglement, and showing entanglement between different regions of the atomic cloud. Besides the foundational aspects of these results, the techniques developed could lead to future applications, such as the estimation of local properties of BECs or quantum steering–based tasks (11). ∎

REFERENCES

1. M. Fadel, T. Zibold, B. Décamps, P. Treutlein, *Science* **360**, 409 (2018).
2. P. Kunkel *et al.*, *Science* **360**, 413 (2018).
3. K. Lange *et al.*, *Science* **360**, 416 (2018).
4. X.-L. Wang *et al.*, *Phys. Rev. Lett.* **117**, 210502 (2016).
5. N. Friis *et al.*, arXiv:1711.11092 [quant-ph] (2017).
6. L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, P. Treutlein, arXiv:1609.01609v2 [quant-ph] (2017).
7. L.-M. Duan, G. Giedke, J. I. Cirac, P. Zoller, *Phys. Rev. Lett.* **84**, 2722 (2000).
8. R. Simon, *Phys. Rev. Lett.* **84**, 2726 (2000).
9. E. Schrödinger, *Math. Proc. Camb. Phil. Soc.* **31**, 555 (1935).
10. H. M. Wiseman, S. J. Jones, A. C. Doherty, *Phys. Rev. Lett.* **98**, 140402 (2007).
11. D. Cavalcanti, P. Skrzypczyk, *Rep. Prog. Phys.* **80**, 024001 (2017).

10.1126/science.aat4590

IMMUNITY

Disrupting metabolism to treat autoimmunity

Dimethyl fumarate, a treatment for multiple sclerosis, inhibits Warburg metabolism

By **Mai Matsushita**[1] *and* **Edward J. Pearce**[1,2]

Autoimmune and inflammatory diseases are diverse conditions caused by inappropriate and prolonged activation of immune cells with associated ongoing production of inflammatory mediators that cause tissue damage. In 2013, dimethyl fumarate (DMF), a methyl ester of fumaric acid used to treat psoriasis (an autoimmune skin condition), was approved for the treatment of multiple sclerosis (MS), a demyelinating autoimmune disease (1). Although this drug is now first-line treatment for relapsing remitting MS, its mechanism of action is elusive (1, 2). On page 449 of this issue, Kornberg et al. (3) provide evidence that the beneficial effects of DMF are related to its ability to inhibit glyceraldehyde-3-phosphate dehydrogenase (GAPDH)—a central enzyme in glucose metabolism (glycolysis)—and, in so doing, inhibit the development and function of inflammatory immune cells, highlighting the promise of targeting metabolism to modulate immune responses.

[1]Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. [2]Faculty of Biology, University of Freiburg, Freiburg, Germany. Email: pearceed@ie-freiburg.mpg.de

When activated, immune cells undergo profound alterations in metabolism that are integral to their changing bioenergetic and biosynthetic needs. Increases in glycolysis are widely acknowledged to be a hallmark of immune cell activation (4). Targeting this pathway with small-molecule inhibitors, such as 2-deoxyglucose, has indicated that enhanced glycolysis is required for cellular functions that occur upon activation (4, 5). Glycolysis serves numerous functions in cellular biology. Pyruvate, produced from glucose by glycolysis, can feed the tricarboxylic acid (TCA) cycle, and therefore mitochondrial oxidative phosphorylation (OXPHOS) (see the figure), which together couple adenosine triphosphate (ATP) production with redox balance. However, glucose is uniquely capable of additionally supporting Warburg metabolism (or aerobic glycolysis), in which pyruvate is converted to lactate through a process that is coupled to ATP production in the cytoplasm. Warburg metabolism is a mark of immune cells that have the potential to cause inflammation (4, 5). Various glycolysis intermediates upstream of pyruvate provide the initiating molecules for ancillary pathways that assume greater importance as cells move from quiescence into growth, proliferative, migratory, and/or

Inhibition of GAPDH blocks inflammation

Resting immune cells depend on mitochondrial respiration to produce ATP. In autoimmunity, Warburg metabolism is increased, supporting the inappropriate production of inflammatory mediators that fight infection and cancer. DMF inhibits GAPDH and thus Warburg metabolism, increasing OXPHOS and inhibiting inflammatory cytokine production, allowing the emergence of cells with anti-inflammatory properties.
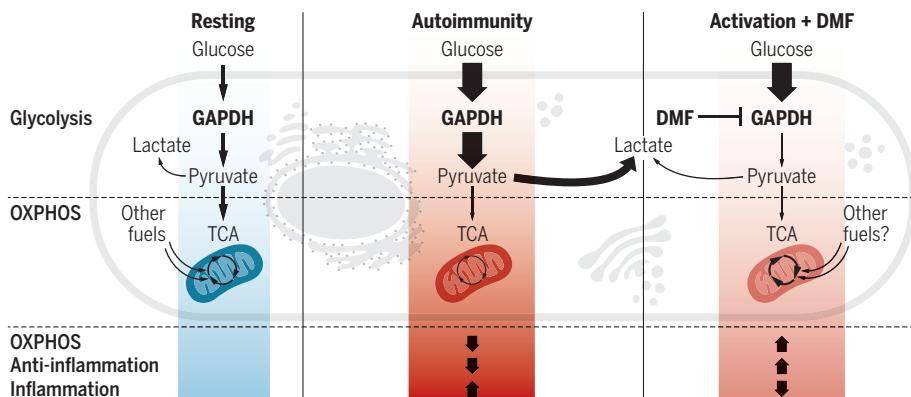
secretory states, which are typical of immune cell activation (6).

GAPDH occupies a central position in the glycolysis pathway. Intriguingly, GAPDH can be inhibited through the succination of active site cysteines by the TCA cycle intermediate fumarate (7). Although the role of this process under physiological conditions is unclear, Kornberg et al. reasoned that the therapeutic benefits of DMF may result from its ability to mediate similar effects on GAPDH. Previously, DMF was shown to succinate key cysteines in tubulin (8). Moreover, succination of KEAP1 (kelch-like ECH-associated protein 1) effectively allows release and therefore activation of NRF2 (nuclear factor erythroid 2–related factor 2) from the inactive KEAP1-NRF2 complex (7). NRF2 is a transcriptional regulator of genes encoding antioxidant proteins, and it has been argued that this is a primary mechanism of DMF action to prevent inflammation-associated damage (9). However, the beneficial effects of DMF still occur in NRF2-deficient mice with experimental autoimmune encephalomyelitis (EAE; a model of MS) (10). It has become clear that DMF can succinate >2400 cysteine residues in numerous proteins in T cells, which may be integral to the ability of the drug to modulate immune cell activation (2). Therefore, Kornberg et al. examined the effects of DMF on GAPDH in activated immune cells.

They found that DMF causes succination of GAPDH active-site cysteines in immune cells from the blood of DMF-treated MS patients and in splenic and intestinal tissues from DMF-treated mice. GAPDH active-site cysteine succination irreversibly inhibited enzymatic activity. The effects of DMF are equivalent to those of heptelidic acid, a selective GAPDH inhibitor that also targets the active-site cysteine by direct modification (11, 12). DMF reduced the production of lactate in activated macrophages. Consistent with a critical role for Warburg metabolism in macrophage activation, DMF treatment significantly inhibited the production of inflammatory mediators such as interleukin-1β (IL-1β) by activated cells. Importantly, DMF had no effect on lactate production by resting macrophages nor on cellular pathways upstream of the switch to Warburg metabolism in macrophage activation. The authors also found that DMF promotes OXPHOS independently of glycolysis inhibition, possibly by directly fueling the TCA cycle.

Activation-associated increases in glycolysis are integral to effector functions of T helper 17 (T_H17) cells and T_H1 cells, which are proinflammatory, but not forkhead box protein 3 (FOXP3)$^+$ regulatory T (T_{reg}) cells, which inhibit inflammation (4). Kornberg et al. show that DMF reduces the development and viability of T_H1 and T_H17 cells and in-

hibits their ability to produce inflammatory cytokines when activated. Notably, DMF accentuates T_{reg} cell development and switches the commitment of T_H17 cells to the FOXP3$^+$ T_{reg} cell lineage. These findings parallel clinical observations that DMF causes a reduction in T_H17 and T_H1 cell populations in MS patients. A bias toward T_{reg} cell responses and anti-inflammatory macrophages may explain the efficacy of DMF in autoimmune diseases (1). The success of DMF in treating MS therefore results from its ability to inhibit GAPDH in proinflammatory immune cells with Warburg metabolism, although it remains possible that succination of cysteines in other proteins contributes to the effects of DMF on immune cells (2).

Recent advances in anti-inflammatory therapeutic approaches, such as the development of antibodies blocking inflammatory cytokines and inhibitors of Janus kinases, are revolutionizing the treatment of autoimmune conditions (13). Nevertheless, subgroups of patients remain refractory to these therapies. The results presented by Kornberg et al. strongly support aerobic glycolysis as a therapeutic target for regulating inflammation. The authors provide proof of principle that targeting GAPDH active-site cysteines can ameliorate inflammatory disease by using heptelidic acid to successfully diminish EAE severity in mice, a result that echoes earlier reports that DMF can itself diminish EAE disease symptoms (10). Perhaps surprisingly, despite the importance of glycolysis and of GAPDH in this pathway in all cell types, DMF does not have broad deleterious effects in vivo. This parallels findings that inhibition of GAPDH by vitamin C or heptelidic acid affects cancer cells (with Warburg metabolism), but not normal cells (14, 15), and probably reflects that GAPDH is rate limiting only during Warburg metabolism (11). Therefore, the exciting possibility exists that targeting key enzymes within metabolic pathways enhanced in activated inflammatory immune cells will provide new therapeutic options for autoimmune diseases. ∎

REFERENCES
1. E. A. Mills et al., Front. Neurol. 9, 5 (2018).
2. M. M. Blewett et al., Sci. Signal. 9, rs10 (2016).
3. M. D. Kornberg et al., Science 360, 449 (2018).
4. M. D. Buck et al., Cell 169, 570 (2017).
5. L. A. O'Neill et al., J. Exp. Med. 213, 15 (2016).
6. D. J. Puleston et al., Cell Metab. 26, 131 (2017).
7. M. Blatnik et al., Ann. N. Y. Acad. Sci. 1126, 272 (2008).
8. G. G. Piroli et al., Biochem. J. 462, 231 (2014).
9. R. A. Linker et al., Brain 134, 678 (2011).
10. U. Schulze-Topphoff et al., Proc. Natl. Acad. Sci. U.S.A. 113, 4777 (2016).
11. M. V. Liberti et al., Cell Metab. 26, 648 (2017).
12. K. Sakai et al., Biochim. Biophys. Acta 1077, 192 (1991).
13. K. F. Baker et al., Ann. Rheum. Dis. 77, 175 (2018).
14. J. Yun et al., Science 350, 1391 (2015).
15. A. A. Shestov et al., eLife 3, e03342 (2014).

10.1126/science.aat4984

BIOPHYSICS

Weighing one protein with light

A light-scattering method allows rapid measurement of the mass of individual proteins

By **Steven F. Lee** *and* **David Klenerman**

Mass spectrometry is a crucial tool for the identification of proteins and other biomolecules in biological samples. In recent years, methods for measuring the mass of individual molecules have been reported, but these methods cannot follow the assembly of protein complexes in real time and are limited to relatively large protein complexes. On page 423 of this issue, Young et al. (1) use light scattering to detect and

"...Young et al.'s light-scattering approach allows video-rate imaging in a parallel fashion of smaller unlabeled biological molecules in biologically relevant environments."

measure the mass of individual unlabeled biomolecules that they had adsorbed from solution to a glass surface. The method allows for the direct detection of proteins as they assemble into functional complexes or misfolded aggregates.

One of the most widely applied methods to detect and characterize individual biological molecules is single-molecule fluorescence imaging (2). This method requires labeling—that is, the nonperturbative attachment of a suitable, small (diameter ~1 nm) fluorophore to the molecule of interest. Typically, irreversible photobleaching limits the observation time in fluorescent imaging unless the fluorophore repeatedly

Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. Email: dk10012@cam.ac.uk

binds and unbinds from the target molecule, as is the case in DNA-PAINT experiments (*3*). Other methods use attachment of more photostable but physically larger quantum dots (5 to 10 nm) or gold nanoparticles (>20 nm), which are potentially more perturbative, to allow imaging for extended times.

Only a few methods can detect and image unlabeled individual molecules. Atomic force microscopy is one such method, in which a proximal probe is scanned over the sample. This method is generally applied to larger biological assemblies, and recent advances in instrument design have allowed close to video-rate imaging of, for example, myosin V (520 kDa) walking on an actin filament (*4*). Individual molecules have also been weighted directly in a vacuum with a nano-electromechanical resonator. This method is, however, currently limited to larger molecules such as immunoglobulin M antibodies (950 kDa) (*5*). Raman scattering enhanced by the presence of a gold nanoparticle or surface can also be used to detect individual molecules directly and has been combined with nonlinear laser spectroscopy so as to further increase the sensitivity (*6*). In addition, scientists have used techniques such as photo-thermal contrast imaging in order to image single nonfluorescent azo dye molecules via the refractive effect of the heat that they release, but this is limited to solvents with poor heat conduction, such as glycerol (*7*).

In contrast to these previous methods, Young *et al.*'s light-scattering approach allows video-rate imaging in a parallel fashion of smaller unlabeled biological molecules in biologically relevant environments. Building on the authors' own previous work (*8*) and that of others (*9*), the approach is based on detecting the weak scattering of light from a single molecule. They show that individual biomolecules between 50 and 800 kDa can be accurately weighed and that assembly of a protein complex, an actin filament, can be followed in real time. They also follow the aggregation of the protein α-synuclein, which is associated with Parkinson's disease, on a lipid membrane.

The fact that light scattering can accurately measure the mass of unlabeled assemblies of proteins will be especially useful to study the formation of larger assemblies, a key biological process. This is because most current methods use labeling. Even if the labeling efficiency is relatively high at 90%, this would mean that only 35% of protein complexes made from 10 proteins will be fully labeled. This is not just an issue for fluorescently tagged molecules; even genetically encoded fluorescent proteins do not fold with 100% efficiency into a detectable fluorescently active state (*10*). However, fluorescence-based methods still have important

Measuring the mass of unlabeled biomolecules

Using light scattering, Young *et al.* can detect small changes in the mass of individual proteins without the need for fluorescent tags. The method allows video-rate detection and is applicable to both individual proteins and protein complexes.
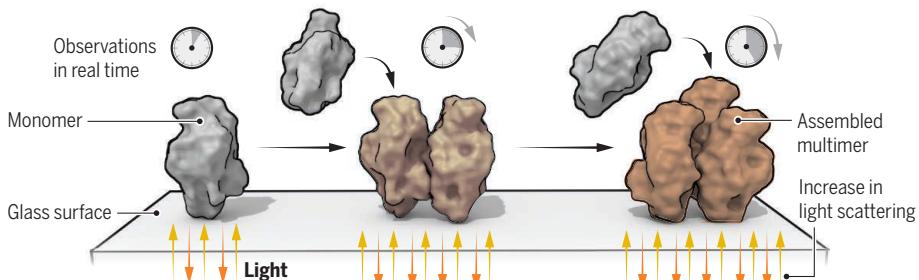
Complex assembly
The method can be used to detect mass changes associated with the real-time assembly of protein complexes.
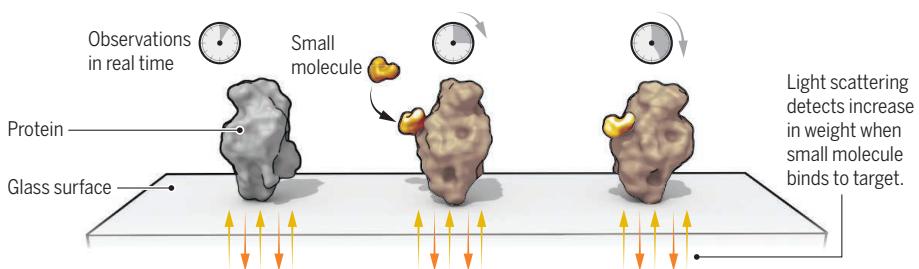
Drug screening
Another possible application is in drug screening, allowing the detection of mass changes caused by binding of small molecules.

advantages when studying interactions in complex environments or cells, where labels can be used to selectively identify the molecule of interest. Furthermore, Förster resonance energy transfer–based methods, in which the separation between a fluorescent donor and acceptor molecules can be measured in the distance range of 1 to 10 nm, provide structural information on biological complexes and conformational changes to be followed in complex cellular backgrounds.

Single-molecule detection by means of light scattering opens up a range of possible experiments. For basic research, it allows the direct assembly of protein complexes to be followed at the single-complex level, providing detailed information about the kinetics and thermodynamics of assembly (see the figure). This will be particularly important in protein aggregation, in which it has proved difficult to directly follow the assembly of the small soluble protein aggregates (oligomers) that are thought to play a key role in the initiation and spread of neurodegenerative diseases. Methods that can detect individual aggregates are essential because these aggregates are present at low abundance and are highly heterogeneous and dynamic, preventing the use of more standard ensemble-based biochemical methods.

There are also many potential applications in drug screening (see the figure). As Young *et al.* show, the method allows the direct detection of binding of a biotinylated peptide to streptavidin via the change in mass. Direct detection of binding to target molecules will allow the label-free screening of drug molecules in a simple and universal fashion in small volumes.

Another possible application is in microfluidic devices, which allow molecules of different sizes to be rapidly separated by diffusion. Currently, fluorescence is used for detection (*11*). If light scattering detection was used instead, this would allow for the rapid preparation and characterization of biological complexes with much higher sensitivity and require much smaller volumes than those of current technologies. ∎

REFERENCES

1. G. Young *et al.*, *Science* **360**, 423 (2018).
2. S. Shashkova, M. C. Leake, *Biosci. Rep.* **37**, BSR20170031 (2017).
3. A. Auer, M. T. Strauss, T. Schlichthaerle, R. Jungmann, *Nano Lett.* **17**, 6428 (2017).
4. T. Ando, *Biophys. Rev.* **9**, 421 (2017).
5. M. S. Hanay *et al.*, *Nat. Nanotechnol.* **7**, 602 (2012).
6. S. Yampolsky *et al.*, *Nat. Photon.* **8**, 650 (2014).
7. A. Gaiduk, M. Yorulmaz, P. V. Ruijgrok, M. Orrit, *Science* **330**, 353 (2010).
8. J. Ortega Arroyo *et al.*, *Nano Lett.* **14**, 2065 (2014).
9. M. Piliarik, V. Sandoghdar, *Nat. Commun.* **5**, 4495 (2014).
10. G. Vámosi *et al.*, *Sci. Rep.* **6**, 33022 (2016).
11. E. V. Yates *et al.*, *Nat. Chem.* **7**, 802 (2015).

10.1126/science.aat5851

CHEMISTRY

Plastics recycling with a difference

A novel plastic with useful properties can easily be recycled again and again

By **Haritz Sardon**[1] and **Andrew P. Dove**[2]

Since the synthesis of the first synthetic polymer in 1907, the low cost, durability, safeness, and processability of polymers have led to ever-expanding uses throughout the global economy. Polymers, commonly called plastics, have become so widely used that global production is expected to exceed 500 million metric tons by 2050. This rising production, combined with rapid disposal and poor mechanisms for recycling, has led to the prediction that, by 2050, there will

in the environment for centuries. More than 40 years after the launch of the recycling symbol, only 5% of plastics that are manufactured are recycled, mainly mechanically into lower-value secondary products that are not recycled again and that ultimately find their way to landfills or pollute the environment (1). With these materials being lost from the system, there is a constant need for the generation of new plastics, mostly from petrochemical sources, thus further depleting natural resources. Although there has been a substantial effort to develop biodegradable plastics, with polylactide arguably the most

Chemical recycling methods have been classified into two groups. In the first, plastic waste is converted directly into products with high added value (7). For example, Hedrick and co-workers have pioneered the depolymerization of aromatic polycarbonates and polyesters directly into bisphenol-type monomers that can be used for the preparation of high-value poly(aryl ether sulfones) (8, 9). In the second, plastic waste is depolymerized back to the starting material and then repolymerized to yield virgin-like plastics (10, 11). For example, Hong and Chen have shown that poly(γ-butyrolactone) can be quantitatively depolymerized back into the initial γ-butyrolactone by simply heating the bulk material (12). However, plastics that can be so easily depolymerized lack suitable mechanical and thermal properties to be widely useful. This is the central paradox of the plastic problem: Although we desire our plastics to be readily recyclable or not to persist in the environment, they must also be sufficiently robust to function in their desired application.

In the search for an intrinsically recyclable plastic with robust mechanical properties, Zhu et al. have designed a variant of the five-membered γ-butyrolactone that bears a fused cyclohexyl ring with defined stereochemistry (see the figure). The steric hindrance and stereoregularity that result after polymerization yield

Repeatedly recyclable polymers

Zhu et al. report production of a plastic that can be recycled repeatedly through chemical methods without loss of function. Blending of the two enantiopure polymers yields a plastic that can withstand higher temperatures, expanding its usefulness further.

be more plastic in the sea than fish (1). On page 398 of this issue, Zhu et al. (2) report an important step toward addressing this problem with the synthesis of a plastic with mechanical properties comparable to those of commercially available plastics, but with an intrinsically infinite recyclability.

The production of synthetic plastics is far from being sustainable (3, 4). Most plastics are produced for single-use applications, and their intended use life is typically less than 1 year. Yet the materials commonly persist

successful example, the mechanical and thermal properties of these materials still need to be improved to be substitutes for a wider range of existing materials properties (3).

In the past decade, an alternative sustainable strategy has been proposed in which the plastic never becomes waste. Instead, once used, it is collected and chemically recycled into raw materials for the production of new virgin plastics with the same properties as the original but without the need for further new monomer feedstocks (5). This strategy not only helps to address the environmental issues related to the continual growth of disposed plastics over the world but may also reduce the demand for finite raw materials by providing a circular materials economy (6).

a semicrystalline polymer that has better thermal and mechanical properties than the unsubstituted derivative. Furthermore, retention of the five-membered γ-lactone core ensures that the thermal recyclability is preserved. The monomer incorporates the cyclohexyl ring fused to the α- and β-positions but leaves the γ-position unsubstituted. The results suggest that maintaining the γ-position unsubstituted is essential to obtain a monomer with sufficient ring strain to be readily polymerized.

Moreover, by judicious choice of catalyst, the authors were able to retain high stereoregularity and prepare highly isotactic polymers in which all stereocenters are aligned along the same side of the polymer chain.

[1]POLYMAT, University of the Basque Country UPV/EHU, Paseo Manuel de Lardizabal 3, 20018, Donostia–San Sebastian, Spain. [2]School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. Email: a.dove@bham.ac.uk

These polymers are semicrystalline and possess mechanical properties that compare well with those of polylactide.

Furthermore, the authors found that by blending two enantiomerically pure polymers of opposite stereochemistry in a 1:1 stoichiometric ratio, superior materials could be obtained. Because of a phenomenon known as stereocomplexation (13), the materials that result from the simple blending process have melting temperatures that are more than 75°C higher than those of either constituent polymer, potentially enabling use in high-temperature applications.

Plastics will continue to be critical for addressing the continuing demands of our society. New polymeric materials will, for example, be needed for energy generation and storage, to address healthcare needs, for food conservation, and for providing clean water. The circular materials economy will require implementation of an appropriate infrastructure to underpin collection and sorting of plastic that has reached the end of its first life before it reaches the environment (4). Beyond the design of new materials, this will require collaboration across scientific and nonscientific disciplines as well as political and public will to ensure success.

Studies such as that of Zhu et al., in which disposed plastics can be infinitely recycled without deleterious effects on their properties, can lead to a world in which plastics at the end of their life are not considered as waste but as raw materials to generate high-value products and virgin plastics. This will both incentivize recycling and encourage sustainability by reducing the requirement for new monomer feedstocks. Current chemical recycling processes are expensive and energetically unfavorable, and further advances in monomer and polymer development and catalyst design are required to facilitate the implementation of economically viable sustainable polymers (14). ∎

REFERENCES

1. Ellen MacArthur Foundation, "The new plastics economy: Rethinking the future of plastics and catalysing action," 2017; www.ellenmacarthurfoundation.org/publications/the-new-plastics-economy-rethinking-the-future-of-plastics-catalysing-action.
2. J.-B. Zhu et al., Science 360, 398 (2018).
3. A.-C. Albertsson, M. Hakkarainen, Science 358, 872 (2017).
4. E. MacArthur, Science 358, 843 (2017).
5. M. Hong, E. Y.-X. Chen, Green Chem. 19, 3692 (2017).
6. D. K. Schneiderman, M. A. Hillmyer, Macromolecules 50, 3733 (2017).
7. J. M. García et al., Science 344, 732 (2014).
8. G. O. Jones et al., Proc. Natl. Acad. Sci. U.S.A. 28, 7722 (2016).
9. F. Gardea et al., Macromol. Chem. Phys. 215, 2260 (2014).
10. J. P. Brutman, G. X. De Hoe, D. K. Schneiderman, T. N. Le, M. A. Hillmyer, Ind. Eng. Chem. Res. 55, 11097 (2016).
11. M. Hong, E. Y.-X. Chen, Nat. Chem. 8, 42 (2016).
12. M. Hong, E. Y.-X. Chen, Angew. Chem. Int. Ed. 55, 4188 (2016).
13. M. J. Stanford, A. P. Dove, Chem. Soc. Rev. 39, 486 (2010).
14. X. Zhang, M. Fevre, G. O. Jones, R. M. Waymouth, Chem. Rev. 118, 839 (2018).

10.1126/science.aat4997

VIROLOGY

Next-generation diagnostics with CRISPR

CRISPR-Cas biology promises rapid, accurate, and portable diagnostic tools

By **Daniel S. Chertow**

Rapid and accurate identification of infectious diseases is essential to optimize clinical care and guide infection control and public health interventions to limit disease spread both in highly specialized medical centers and remote health care settings. The ideal diagnostic test would be inexpensive, accurate, and provide a result rapidly, allowing for point-of-care use on multiple specimen types without need for technical expertise, ancillary equipment, or power. Such a test for highly pathogenic viruses that emerge in remote settings but might spread globally (for example, Ebola virus and Middle East respiratory syndrome coronavirus) would aid in early case detection and isolation, limiting disease spread and facilitating timely care (1). The sentinel discovery that prokaryotes (bacteria and archaea) have heritable adaptive immunity mediated through CRISPR and CRISPR-associated (Cas) proteins has led to transformative advances in molecular biology, most notably in gene editing (2). On pages 436, 444, and 439 of this issue, Chen et al. (3), Myhrvold et al. (4), and Gootenberg et al. (5), respectively, highlight how evolving insights into CRISPR-Cas biology are also revolutionizing the field of molecular diagnostics for infectious diseases, through detection of Zika virus (ZIKV), Dengue virus (DENV), and human papillomavirus (HPV) in human samples, and noninfectious diseases, such as detection of gene mutations in circulating cell-free DNA from lung cancer patients.

Prokaryotes store genetic elements from infectious agents (phages, plasmids, or transposons) in genomic loci called CRISPR arrays as memories for adaptive immunity. Cas proteins facilitate adaptive immunity through the processes of adaptation, CRISPR RNA (crRNA) generation, and interference (6). During adaptation, foreign genetic material is processed and selected for integration into the CRISPR array, providing a recall element during recurrent infection. Pre-crRNA is transcribed as a long precursor and processed into mature form as crRNA to guide Cas proteins to cleave complementary sequences of foreign elements (interference) to degrade and eliminate those elements. By uncovering the structural and functional components of these diverse systems, new tools, including those applicable to molecular diagnostics, are emerging (see the figure).

Chen et al. report the discovery that when CRISPR-Cas12a proteins cleave double-stranded DNA (dsDNA) in a sequence-specific manner, they induce robust nonspecific single-stranded DNA (ssDNA) trans-cleavage. The authors apply this observation to develop a rapid and accurate test to detect carcinoma-associated HPV types 16 and 18 from clinical specimens. HPV dsDNA is extracted from anal swabs and amplified through isothermal preamplification by recombinase polymerase amplification (RPA) (7), a method that is rapid and does not require specialized

> *"Future work will expand upon the range of diagnostic applications for infectious and noninfectious diseases..."*

equipment. A Cas12a-crRNA complex binds to and cleaves target HPV dsDNA, which activates trans-cleavage of ssDNA. A fluorescent reporter coupled to ssDNA generates a fluorescent signal upon cleavage. This new approach, called DNA endonuclease-targeted CRISPR trans reporter (DETECTR), offers a promising platform for rapid and accurate detection of cervical cancer–associated HPV subtypes that, consistent with World Health Organization recommendations, might augment screening programs worldwide (8).

Myhrvold et al. introduce a new approach to release and protect from degradation viral nucleic acids from clinical specimens, bypassing the need for nucleic acid extraction in molecular diagnostics. This method,

Critical Care Medicine Department, NIH Clinical Center, and the Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. Email: chertowd@cc.nih.gov

called HUDSON (heating unextracted diagnostic samples to obliterate nucleases), is a process of heat and chemical reduction that inactivates the high amounts of ribonucleases (RNases) found in body fluids and then lyses viral particles by disrupting the viral envelope, thereby releasing nucleic acids into solution. The authors combine HUDSON with SHERLOCK (specific high-sensitivity enzymatic reporter unlocking), a Cas13-based nucleic acid detection platform described previously (9). Similar to DETECTR, SHERLOCK combines RPA and an RNA-guided Cas13 that induces collateral cleavage of nucleic acids. In SHERLOCK, RNA coupled to a fluorescent reporter is cleaved, producing a fluorescent signal that is amplified though enzymatic activity, which enhances test sensitivity. By combining HUDSON and SHERLOCK, the authors develop a sensitive and specific diagnostic platform to detect the flaviviruses DENV and ZIKV directly from body fluids (urine, saliva, serum, plasma, and whole blood), with limited sample preparation or equipment, that provides a result within 1 to 2 hours. DENV and ZIKV cocirculate in many areas of Central and South America and have similar clinical presentation (10). ZIKV infection during pregnancy predisposes to severe congenital anomalies and is sexually transmissible, emphasizing the need for accurate diagnosis among pregnant women and their sexual partners (11). HUDSON combined with SHERLOCK reliably differentiates between DENV, ZIKV, and another flavivirus, yellow fever virus (YFV), which cocirculate and cause severe parallel epidemics in Brazil (10). The platform can also reliably distinguish between four closely related DENV serotypes and can detect single-nucleotide polymorphisms (SNPs) among ZIKV isolates. This could be applied to detect SNPs that confer antimicrobial resistance [for example, among *Mycobacterium tuberculosis* (TB) isolates] or enhance pathogen virulence or transmissibility [for example, among highly pathogenic avian influenza A (H5N1) viruses], allowing for better tracking of emerging pathogens. Finally, the authors show that the fluorescent readout of SHERLOCK can be replaced with a visual readout on a paper test-strip, suitable for point-of-care field application.

Gootenberg *et al.* introduce SHERLOCK version 2 (v2). This improved assay allows for detection of three ssRNA targets and one dsDNA target in a single reaction. The authors biochemically characterize 17 CRISPR-Cas13a and -Cas13b enzymes, se-

Application of CRISPR diagnostics

Next-generation diagnostics applying CRISPR-Cas biology will facilitate early disease detection and intervention.

Clinical encounter
Resource-poor setting

Sample collection
Urine, blood, stool

Infection control
Optimal clinical care

DIAGNOSTIC TEST
The ideal test is inexpensive, accurate, provides a result rapidly, and can be used on multiple specimen types without technical expertise, ancillary equipment, or power.

1 Prepare sample, release and protect nucleic acids
Method: HUDSON

2 Amplify DNA and RNA
Method: RPA

3 Accurately detect target and amplify signal
Method: SHERLOCK, SHERLOCKv2, and DETECTR

Target
Cas-crRNA detection of target
Cleavage
Reporter

lecting three with distinct cleavage preferences, that when combined with a Cas12a enzyme and RPA accurately detect ZIKV ssRNA, synthetic ssRNA, DENV ssRNA, and synthetic dsDNA by visual readout in less than 90 minutes. A potential application of multitarget RNA and DNA detection by SHERLOCKv2 could be a rapid and accurate diagnostic test for pneumonia pathogens. DNA and RNA viruses alone or in combination with bacterial infection cause pneumonia, a leading killer of children worldwide (12). An accurate and affordable point-of-care diagnostic for pneumonia would allow for early and targeted use of antibiotics in remote settings.

Another feature of SHERLOCKv2 is quantitative and sensitive target detection. This could be applied to a portable and accurate test to monitor viral load among HIV patients receiving antiviral therapy in resource-limited settings, substantially improving global HIV care (13). SHERLOCKv2 was also used to detect mutations in cell-free DNA from the blood of non–small cell lung cancer patients by fluorescence- and lateral flow–based readouts, further expanding the potential applications to liquid biopsy. Finally, in proof-of-concept in vitro experiments, the authors successfully apply SHERLOCKv2 both as a gene-editing therapeutic that corrects a gene mutation that predisposes to colon cancer, and as a diagnostic to concur-

rently determine the proportion of genes successfully edited.

These emerging diagnostic tools will by necessity be compared to standard diagnostics to ensure sensitivity and specificity and will need to be field-tested to guarantee performance in patient care settings, as environmental conditions and end-user application might affect performance. Proven assays, if affordable, promise to improve care in resource-limited settings where undifferentiated febrile illness is the norm and where gaps or delays in diagnosis, targeted care, and infection control contribute to infectious disease mortality and spread. For example, TB results in an estimated 1.3 million deaths annually, the leading cause from a single infectious agent, and most deaths could be prevented with early diagnosis and treatment (14). Assays might be expanded to provide insight into pathogen resistance patterns to guide antimicrobial therapy, molecular correlates of pathogen viability to guide infection control, and compatibility with additional specimen types such as stool, respiratory secretions, and cerebrospinal fluid, that when integrated with clinical judgment might for example, differentiate etiology of enteritis, pneumonia, and meningitis. Future work will expand upon the range of diagnostic applications for infectious and noninfectious diseases in the clinic, laboratory, and field where assay accuracy, reliability, simplicity, speed, flexibility, and cost will determine the scope of impact. ∎

REFERENCES AND NOTES

1. M. J. Broadhurst, T. J. G. Brooks, N. R. Pollock, *Clin. Microbiol. Rev.* **29**, 773 (2016).
2. A. C. Komor, A. H. Badran, D. R. Liu, *Cell* **168**, 20 (2017).
3. J. S. Chen *et al.*, *Science* **360**, 436 (2018).
4. C. Myhrvold *et al.*, *Science* **360**, 444 (2018).
5. J. S. Gootenberg *et al.*, *Science* **360**, 439 (2018)
6. F. Hille *et al.*, *Cell* **172**, 1239 (2018).
7. O. Piepenburg *et al.*, *PLOS Biol.* **4**, e204 (2006).
8. apps.who.int/iris/bitstream/handle/10665/94830/ 9789241548694_eng.pdf?sequence=1
9. J. S. Gootenberg *et al.*, *Science* **356**, 438 (2017).
10. C. I. Paules, A. S. Fauci, *N. Engl. J. Med.* **376**, 1397 (2017).
11. I. U. Mysorekar, M. S. Diamond, *N. Engl. J. Med.* **375**, 481 (2016).
12. T. Wardlaw *et al.*, *Lancet* **368**, 1048 (2006).
13. J. Dorward, P. K. Drain, N. Garrett, *Lancet* **5**, e8 (2018).
14. www.who.int/tb/publications/global_report/Exec_ Summary_13Nov2017.pdf?ua=1

GRAPHIC: V. ALTOUNIAN/SCIENCE

Günter Blobel (1936–2018)

Passionate cell biologist and inspirational mentor

By **Peter Walter**[1] and **David Anderson**[2]

On 18 February, we lost our Ph.D. adviser, Günter Blobel, who succumbed to a battle with cancer at age 81. For both of us, it felt akin to losing a close family member. Typical of graduate students, we arrived in his lab unformed and malleable, and we eagerly absorbed not only information, logic, and experimental technique but also Günter's scientific mind-set. Perhaps the most important thing that Günter taught us is that science is not an ordinary job. He lived science as a passion, filled with emotions and purpose, with personality and intuitive preconceptions (that is, hypotheses), and with a keen eye on aesthetics and beauty in approach and execution.

Günter taught us to distinguish experiments that should be done from those that could be done; he taught us to cherish the paradox over the obvious next thing. Importantly, Günter excelled at standing up firmly for one's convictions in the face of controversy. In fact, he thrived on it; lab members used to speak of him searching for new windmills at which to tilt. If scientists can be categorized as the scholar, the technologist, or the knight errant, Günter was certainly the latter, chasing biological holy grails like the hero in his beloved Wagnerian opera *Parsifal*. His energy and relentless drive infected everyone in the lab, stimulating our creativity and impelling us to work at a fever pitch.

Günter was born and raised in a village in Silesia in eastern Germany. Fleeing from the advancing Russian army and the perils of World War II, his family resettled in western Germany. He earned a medical degree from the University of Tübingen in 1960. Frustrated by the prevailing lack of knowledge of normal cell function, he decided to focus on fundamental research problems at the University of Wisconsin, where he started the work on protein targeting that would occupy him for 55 years. After earning his Ph.D. in oncology at the University of Wisconsin in 1967, he joined The Rockefeller University in New York, where he moved up the faculty ranks and remained for the rest of his career.

[1]University of California, San Francisco, CA, USA.
[2]California Institute of Technology, Pasadena, CA, USA.
Email: peter@walterlab.ucsf.edu; wuwei@caltech.edu

Günter's contributions to cell biology are immense. His career was punctuated by numerous discoveries, collectively establishing the rules that determine internal cell organization. In the early 1970s, he established the principle that topologically complex biological phenomena, such as protein sorting and translocation across membranes, can be recapitulated in cell extracts and then systematically dissected using biochemical approaches. Günter, along with biologists David Sabatini and Bernhard Dobberstein,
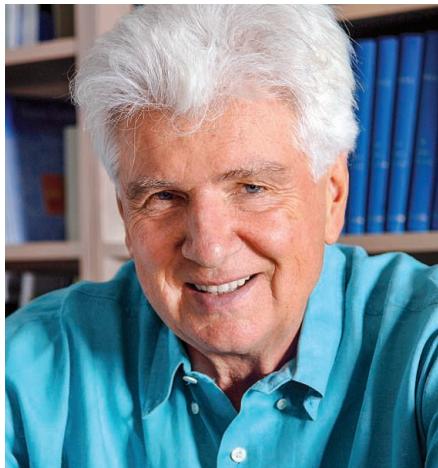
experimentally established that the convoluted membrane network of the endoplasmic reticulum can be isolated from broken cells as tiny vesicles (microsomes). When added to a test tube containing ribosomes and other protein synthesis machinery, these microsomes can miraculously transfer nascent secretory proteins across an otherwise protein-impermeant lipid bilayer. This system led to the discovery of the guiding role of signal sequences in polypeptide chains that determine the ultimate destination of proteins in the cell, as well as their folding in the membrane. It was the reconstitution of complex cellular events in the test tube that won Günter the Nobel Prize in 1999.

Günter was unfaltering in his conviction that proteins moved across membranes through a protein-conducting channel. Sidestepping arguments to the contrary—including those of preeminent biophysicists at the time—his unconditional adherence to the "signal hypothesis" shaped our thinking profoundly and drove the research in the lab. Ironically, it was others (cell biologists Randy Schekman and Tom Rapoport in particular) who identified the channel molecularly and unambiguously proved the principal idea to be true.

The Blobel lab was anarchic, guided by a passion comparable to that of true explorers unconstrained by economic realities. The lab was constantly on the brink of financial collapse, full of decrepit instrumentation, and requiring annual bailouts by the grudgingly acquiescent, if visionary, administration of The Rockefeller University. A constellation of strong individual characters, each driven by his or her own ambitions, worked together (mostly) as an amazingly cohesive unit. Günter's vision firmly united a potpourri of experimental systems. During our time in the lab, students and postdocs pursued protein translocation into every conceivable compartment of the cell. The research encompassed the fate of simple soluble and complex multispanning membrane proteins, traveling near and far. The breadth of the approaches and Günter's reliance on everyone's self-motivation and passion gave each individual sufficient room to pursue their own paths with considerable independence, while guided by his unerring judgment for distinguishing questions that would lead to fundamental conceptual breakthroughs, rather than simply to diversionary details. His generosity in letting us take our projects with us when we left the lab was remarkable.

Günter's experimental fearlessness manifested itself in many other ways. Although he trained in medicine, he drifted continuously over his career toward the power of mechanistic insights. Long after we had left the lab, he fell in love with structural approaches, realizing the richness that molecular detail can provide to our understanding of biology. Günter's longest quest, lasting well into his final days, was to decipher the intricacies of the nuclear pore. The current textbook picture of nuclear pore complexes describes a static assembly of a protein ring. Günter posited instead that the outer ring of nuclear pore complexes would be dynamic, expanding and contracting with the demands of the cargo traversing the pore. The future will tell if these later convictions, currently judged as unorthodox by many, will turn out to be just as insightful as his earlier ones.

For Günter, ideas proved a powerful driving force. Although not all of his intuitions proved correct, the ones that were on the mark revolutionized cell biology. For us, and many others who had the opportunity to work with him, his spirit will remain a powerful example of the value of hypotheses driving discovery, the importance of deep personal engagement, and the search for beauty and truth. ∎

GENOMICS

Advancing the ethics of paleogenomics

Ancestral remains should be regarded not as "artifacts" but as human relatives who deserve respect

By **Jessica Bardill, Alyssa C. Bader, Nanibaa' A. Garrison, Deborah A. Bolnick, Jennifer A. Raff, Alexa Walker, Ripan S. Malhi,** *and* **the Summer internship for INdigenous peoples in Genomics (SING) Consortium**

Recent scientific developments have drawn renewed attention to the complex relationships among Indigenous peoples, the scientific community, settler colonial governments, and ancient human remains (*1, 2*). Increasingly, DNA testing of ancestral remains uncovered in the Americas is being used in disputes over these remains (*3*). However, articulations of ethical principles and practices in paleogenomics have not kept pace (*4*), even as results of these studies can have negative consequences, undermining or complicating community claims in treaty, repatriation, territorial, or other legal cases. Paleogenomic narratives may also misconstrue or contradict community histories, potentially harming community or individual identities. Paleogenomic data can reveal information about descendant communities that may be stigmatizing, such as genetic susceptibilities to disease. Given the potential consequences for Indigenous communities, it is critical that paleogenomic researchers consider their ethical obligations more carefully than in the past.

As Indigenous scientists and bioethicists, and allied non-Indigenous scientists, we offer needed Indigenous perspectives on ethical best practices in paleogenomic research involving the remains of our ancestors (where "ancestors" refers to all pre-European–contact individuals in the Americas as well as postcontact deceased Indigenous individuals

All affiliations and members of the SING Consortium are listed in the supplementary materials.
Email: jessica.bardill@concordia.ca; malhi@illinois.edu

from infants to elders). Currently, little legal structure or ethical guidance is available to help researchers determine ethical best practices for paleogenomic studies. U.S. Institutional Review Boards (IRBs) oversee human subjects research and protect research participants following the principles of the Belmont Report (*5*)—respect for autonomy, beneficence, and justice. However, because only living individuals are considered "human subjects," the remains of ancestors in the United States are designated legally and scientifically as "artifacts" and fall under the purview of the Native American Graves Protection and Repatriation Act (NAGPRA) (*6*) and the National Museum of the American Indian Act (*7*) (for remains held by the Smithsonian). Both laws provide guidance for consultation with federally recognized tribal nations in the context of determining the cultural affiliation of remains for repatriation. However, neither outlines research best practices nor requires consultation for research involving remains deemed culturally "unaffiliated." Further, aside from museums, ancestors uncovered on private lands or outside the United States do not fall under these laws. States like Hawai'i have additional laws regarding ancestors uncovered on private lands but do not provide ethical guidelines for research.

With such uneven ethical terrain, community consultation and engagement practices have varied widely. Many studies have proceeded with little to no engagement. For example, a paleogenomic study of ancestors from Chaco Canyon, New Mexico (*8*), was recently published without tribal consultation because the remains had been deemed culturally unaffiliated. However, many Southwestern tribes have traditional knowledge and oral histories linking them to Chaco Canyon (*9*). Scientific studies of the Ancient One (Kennewick Man) (*3*) were also conducted

against the wishes of the Columbia Plateau tribes who sought repatriation of their ancestor. The results of paleogenomic analysis supported the tribes' efforts to repatriate the Ancient One, but the 20-year delay in reburial and the destructive analysis required were painful and disruptive for many community members.

COMMUNITY-BASED PRACTICES

To minimize harms in the future, we recommend that ancestral remains be regarded not as "artifacts" but as human relatives who deserve respect in research. As the deceased cannot give consent, present-day communities should be consulted. In this context, attention to place is paramount, as both the Indigenous peoples who reside nearby and those with ancestral ties to the region may wish to speak for the ancestor(s). Communities today are also affected by studies of ancestors, so engagement ensures that their concerns will be considered.

By adopting this approach, mutually beneficial relationships can develop between researchers and communities, leading to more robust science and productive collaborations. For example, coauthors R.S.M. and A.C.B. have partnerships with a First Nations community in British Columbia. The community is primarily interested in using paleogenomic studies to identify genetic links between living community members and ancestors in the region to corroborate oral histories and archaeological evidence of residing on the Northwest Coast for thousands of years. This genetic evidence of continuous residence may be more likely to be accepted by the Canadian government as support for treaty rights (*10*).

The Canadian Tri-Council Policy Statement (TCPS2) (*11*) governing human research ethics establishes protections for living and recently deceased individuals but not indi-

viduals who lived hundreds or thousands of years ago. Thus, to ensure ethical and mutually beneficial paleogenomic research, the partners signed an agreement outlining expectations of the researchers and community before samples were collected for analysis from ancestors housed at the Museum of History in Gatineau, Quebec. Research team members visit the community regularly to communicate with research participants, elders, and First Nations government representatives. They review research goals and discuss results and language for presentations, manuscripts, and press releases. Two First Nations members also participated in the SING workshop (https://sing.igb.illinois. edu) to learn about the uses and limitations of paleogenomics as well as ethical, legal, and social considerations. This collaboration has yielded insights about the community's history from population genetic analyses of ancestral skeletal remains and metagenomic analyses of diet from ancestors' dental calculus. Elders and community members inform the inferences made from paleogenomic data, helping to distinguish signals of ancestral diet from genome database biases (12).

RECOMMENDATIONS

If scientists lack knowledge of Indigenous cultures and concerns, unintended cultural harm, mistrust, and/or weakened political authority may occur for communities participating in or affected by paleogenomic studies (9, 13). Community engagement helps ensure that Indigenous peoples have a voice in this research. Our aim here is not to advocate a "one size fits all" approach but to provide ethical guidance applicable to many contexts and communities.

Ideally, researchers should hold discussions with Indigenous communities before research involving ancestors begins, to address the interests and concerns of the community (e.g., incorporating community research questions or using less destructive sampling methods). This should occur regardless of whether ancestors were uncovered on public or private lands, or considered "culturally unidentified" under NAGPRA. This term reflects processes that have obscured cultural connections of ancestral remains, including historical removals of Indigenous peoples from their homelands, archaeological collection practices that disinterred individuals en masse, and settler-colonial practices of labeling Indigenous peoples in ways that merged distinct peoples or misnamed them. Often, sufficient information (e.g., geographic location, items buried with ancestors) exists to identify potential communities for consul-

tation. Many Indigenous communities are intimately connected with the land where they reside, caring for both the land and ancestors held therein, even if they are not direct biological descendants. Geographically proximate communities are therefore appropriate to contact to begin engagement. If current research lacks community input, engagement should start immediately.

Recent technological advances have also enabled paleogenomic studies of DNA from dental calculus, hair, coprolites, and even soil, providing alternatives to destructive analysis of the bones and teeth of ancestors. However, community engagement is still needed in these contexts. Indigenous perspectives on the sacredness of materials from the body and earth should be considered, and paleogenomic studies of these materials can have social, political, and legal consequences for Indigenous communities.

To aid the process of community engagement, we offer these guiding questions for paleogenomic researchers to consider:

1. In the absence of known descendant or culturally affiliated communities, which Indigenous peoples, tied to land where ancestors were buried, will be consulted?
2. Who is the appropriate community body (e.g., tribal council, tribal IRB, elders) or representative (e.g., tribal president, historic preservation officer) to initiate discussions with about paleogenomic analyses?
3. What are potential ethical pitfalls of this research or harms that could affect the community? What cultural concerns of the community, such as destruction of ancestral remains, need to be considered?
4. How will the community benefit from the paleogenomic research?
5. How will the community provide input on study design and interpretation of results? How frequently does the community wish to be contacted during the project?
6. When community members participate directly in the project (e.g., as advisers or laboratory technicians), will they coauthor research publications and presentations? How do communities and individuals wish to be recognized in research products?
7. What happens after the project ends? Who will have access to the data generated? How will remaining samples from ancestors be handled, stored, returned, or reburied?

Because Indigenous communities have diverse practices and views on genomics, the nature and structure of engagement will vary. Although it may not always be obvi-

ous how to proceed if different potentially linked communities hold differing views, we believe engaging with Indigenous communities should be as integral to the research process as hypothesis development.

SHIFTING THE STATUS QUO

The ethical practices proposed here will help paleogenomics avoid perpetuating the long history of unethical and exploitative scientific research with Indigenous communities, both in the United States and Canada as well as abroad (14). Increased community engagement will produce stronger scientific interpretations and improve relationships between scientists and Indigenous peoples, particularly as the number of Indigenous scientists grows. Currently, several initiatives are focused on building capacity for genomic research in Indigenous communities, such as the Genetic Education for Native Americans program that aims to increase communities' genetic literacy (15) and the SING workshop that teaches laboratory and computational skills while facilitating discussions of culturally appropriate uses of genomics. Programs like these help dissolve barriers and distrust. Ultimately, community engagement and capacity building will produce more robust, ethical paleogenomic research. ■

REFERENCES AND NOTES

1. E. Callaway, Nature 522, 404 (2015).
2. Editorial, Nature 533, 7 (2016).
3. M. Rasmussen et al., Nature 523, 455 (2015).
4. T. Turner, Ed., Biological Anthropology and Ethics (SUNY Press, Albany, 2005).
5. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, "The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research," Department of Health, Education and Welfare, 78-0012 (1978).
6. Native American Graves Protection and Repatriation Act, Public law 101-601, 16 November 1990.
7. National Museum of the American Indian Act, Public law 101-185, 28 November 1989.
8. D. J. Kennett et al., Nature 8, 14115 (2017).
9. K. G. Claw et al., Human Biology, 2017; muse.jhu.edu/article/677135.
10. A. Martindale, in Rethinking Colonial Pasts Through Archaeology, N. Ferris, R. Harrison, M. V. Wilcox, Eds. (Oxford Univ. Press, 2014), pp. 397–422.
11. Canadian Institutes of Health Research Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2), December 2014.
12. See supplementary materials for additional examples.
13. R. R. McInnes, Am. J. Hum. Genet. 88, 254 (2011).
14. N. A. Garrison, Cases of how tribes are relating to genetics research, National Congress of American Indians (2011).
15. L. Burhansstipanov, L. Bemis, M. Dignan, F. Dukepoo, Genetics 158, 941 (2001).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6387/384/suppl/DC1

10.1126/science.aaq1131

U.S. Marines conduct a foot patrol with the Modular Advanced Armed Robotic System on 9 July 2016.

BOOKS *et al.*

ARTIFICIAL INTELLIGENCE

The automated battlefield

A sober treatise on the future of warfare warns of the perils of autonomous robotic combatants

By **George Lucas**

Sooner than you may think, robotic swarms will intercept incoming missiles at hypersonic speed, while dueling cyberattacks and countermeasures transpire at nearly the speed of light. Such strikes and counterstrikes will quickly overwhelm the capacities of human combatants to respond.

Fiscal constraints and decreasing human resources, together with the promise of enhanced precision, effective risk management, and force multiplication, are driving Pentagon planners (and their counterparts in other nations) to aggressively pursue battlefield automation. In *Army of None*, Paul Scharre offers an authoritative and sobering perspective on the automated battlefields that will very soon come to characterize military conflict, predicting that autonomous robots, many fully armed and capable of independent targeting decisions, will inevitably come to rule the waves, as well as prevailing on the ground, in the air, and especially in space.

The logic behind such developments seems inescapable. Why should human pilots fly risky missions in hostile territory when remotely piloted, and increasingly autonomous, air frames can conduct the same missions at a fraction of the cost and often with greater endurance, efficiency, and precision? Why should combat supply convoys use human drivers in zones of conflict when autonomous vehicles can move the same supplies without risk of harm? Robotic combatants and logistical support troops, moreover, require no pensions or veterans' benefits—a draft by their human counterparts on future Department of Defense fiscal resources that, if unchecked, will very soon come to consume the entire annual defense budget.

The drive to automate seems relentless and unstoppable, and the headlong rush toward endowing automated combat weapons platforms and logistical support devices with the ability to operate without continuous human oversight or intervention seems the only way of realizing the economies of risk reduction and force multiplication that automation promises. Yet such considerations do not fully address the range of underlying policy questions at stake.

Scharre reminds readers of the compelling example of Soviet Army Lieutenant Colonel Stanislav Petrov, who, in September 1983, recognized that an automated surveillance system alert of an incoming U.S. nuclear missile strike was, in fact, a grave system malfunction and thus very likely averted a nuclear holocaust. Despite subsequent advances in artificial intelligence capabilities, none is sufficient to equip an automated weapons-delivery platform to make the kinds of discerning judgments that humans have made and continue to make routinely in such circumstances.
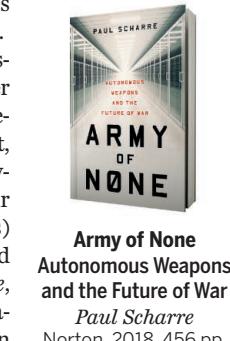
Machines lack the concerns, empathy, and self-consciousness that factor into normal human situational awareness. Thus, despite numerous advantages, Scharre marshals his engagingly detailed accounts of these weapons systems to argue that a policy of blindly arming and deploying autonomous robots in armed conflict would constitute an error fraught with peril.

Debates over the future of military robotics have raged for more than a decade among the delegates to the United Nations' periodic conventions on Certain Conventional Weapons (CCW) in Geneva, while members of ICRAC (International Committee for Robot Arms Control), including eminent Irish roboticist Noel Sharkey, have lobbied strenuously against the further development or deployment of such weapons.

Indeed, Wendell Wallach, a technology ethicist and robotics expert at Yale University, proposed several years ago that any attempts to wed full autonomy or general (deep learning) artificial intelligence with lethal weaponry ought to be preemptively condemned as constituting a means of warfare that is *malum in se*—wrong or evil in itself. By contrast, computer scientist and military roboticist Ronald Arkin proposed addressing such concerns with design modifications that would mimic emotions such as guilt and empathy and thus provide an "ethical governor" for emergent robotic behaviors.

To be sure, a kind of generalized technology anxiety with respect to military robotics is nothing new. P. W. Singer's *Wired for War* covered much of the same well-worn ground in 2009 (*1*). But Scharre brings these previous discussions up to date from a perspective forged in combat as a former Army Ranger and subsequently as a senior policy analyst in the Department of Defense. His is, thus, to a large extent, an insider's perspective on the current and likely future course of development of these revolutionary and disruptive military technologies. As such, it is all the more compelling for its clarity and exhaustive detail—and hence well worth reading. ∎

Army of None
Autonomous Weapons and the Future of War
Paul Scharre
Norton, 2018. 456 pp.

REFERENCE

1. P. W. Singer. *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (Penguin, London, 2009).

10.1126/science.aas9074

The reviewer is the emeritus distinguished chair in ethics, U.S. Naval Academy, Annapolis, MD 21402, USA, and the author of Ethics and Cyber Warfare *(Oxford University Press, New York, 2017). Email: george.r.lucas.jr@gmail.com*

PUBLIC HEALTH

Incentivizing responsible antibiotic use

In the fight against antimicrobial resistance, confronting economic challenges is key

By **Ramanan Laxminarayan**

Ever since the advent of antibiotics, scientists and clinicians have warned of the potential for widespread antibiotic resistance. Indeed, the first in vitro study of resistance to penicillin was published in 1940, 2 years before the first patient was even treated with the drug. In the ensuing decades, experts and the media continued to warn about an impending crisis of resistance but were largely ignored by the public and policy-makers.

Public opinion changed when resistance became clinically relevant. In the mid-1990s, methicillin-resistant *Staphylococcus aureus* (MRSA), a pathogen that had not been encountered outside health care settings, became common in patients who had not been hospitalized. New antibiotics were developed, but with the subsequent discovery of organisms that were resistant to carbapenem—a drug of last resort—now there were patients with infections for whom nothing could be done.

The U.S. Centers for Disease Control and Prevention warned of "nightmare bacteria," and global leaders started talking about the problem. Yet the problem of resistance lacked an effective global spokesperson.

In stepped Dame Sally Davies, the chief medical officer of England. Having warned about antimicrobial resistance (AMR) in England in a 2011 report, Davies set off to make sure that other countries recognized the problem. In 2014, the U.K. government and the Wellcome Trust commissioned a review on AMR. Jim O'Neill, a former chief economist of Goldman Sachs, was recruited to lead the initiative.

O'Neill's review followed a long line of reports that sought to explore the intersection of AMR with infection control, vaccines, diagnostics, and new drug development. But it went beyond earlier reports by examining the incentives that prevent free markets

Superbugs
An Arms Race
Against Bacteria
William Hall,
Anthony McDonnell,
and Jim O'Neill
Harvard University
Press, 2018. 246 pp.

from achieving an efficient solution for the development and use of antimicrobials.

The review had its weaknesses: The estimates of economic burden ($100 trillion) and health burden (10 million deaths by 2050), for example, are not realistic and assume zero change in technology or policies over the next three decades. Nevertheless, and perhaps because of the dire nature of these estimates, it brought the issue of AMR to a broader audience and led to a meeting of the United Nations General Assembly in September 2016.

O'Neill's new book, *Superbugs: An Arms Race Against Bacteria*, written in collaboration with Anthony McDonnell, who served as head of economic research for the review, and William Hall, who served as senior policy adviser, aims to bring the issue to a lay audience. In their discussion of incentives for infection control, antibiotic use, and new antibiotic development, the authors are careful to sidestep arguments made by the infectious disease community, which were not grounded in an understanding of the economics of the pharmaceutical industry. The book also de-emphasizes the report's problematic assessments of health burden. What remains is an immensely readable description of the challenges that encourage overuse of antibiotics and discourage new drug development.

The authors propose a $4 billion annual investment in tackling AMR and a tax on companies not investing in antibiotics R&D to support market-entry rewards for companies that are. However, there is greater focus on spending on new antimicrobials than on conserving the effectiveness of the drugs we already have or on changing practices that encourage the overuse of antibiotics.

The focus of the book is distinctly on global AMR issues as they relate to high-income countries. For instance, there is little attention devoted to the issue of lack of access to antibiotics, which kills more people than does antibiotic resistance in low- and middle-income countries.

Antibiotic resistance has much in common with climate change, in that actions in any single country have the potential to affect the rest of the world. No matter where the next strain of multidrug-resistant *S. aureus* arises, it will become a problem in rich and poor countries alike. And, as with climate change, a key problem is the lack of incentives—for individuals, organizations, and countries—to preserve a global common resource.

The metaphor of an "arms race" against bacteria is outdated. One could argue that the idea that bacteria are our enemy is what got us to the problem of antibiotic overuse in the first place. That said, *Superbugs* is a worthy exposition of the challenges we will have to surmount to incentivize more responsible antibiotic use until we discover new ways of dealing with infections. ∎

10.1126/science.aat1630

The reviewer is the founder and director of the Center for Disease Dynamics, Economics, and Policy in Washington, DC, and a senior research scholar at the Princeton Environmental Institute, Princeton University, Princeton, NJ 08544, USA. Email: ramanan@cddep.org

PODCAST

The Truth About Animals
Stoned Sloths, Lovelorn Hippos,
and Other Tales from the Wild
Side of Wildlife
Lucy Cooke
Basic Books, 2018. 345 pp.

The medieval belief that beavers will chew off their own testicles to escape a hunter may seem laughable now, but many other mischaracterizations of animals—the "lazy" sloth, the hyena as simpering scavenger—remain commonplace. This week on the *Science* podcast, Lucy Cooke discusses some of our strangest misconceptions.

Anthropomorphizing prevents a true understanding of animals.

10.1126/science.aat7146

LETTERS

Edited by **Jennifer Sills**

Bee conservation: Key role of managed bees

In their Perspective "Conserving honey bees does not help wildlife" (26 January, p. 392), J. Geldmann and J. P. González-Varo argue that because managed honey bees are an agricultural animal, their crop pollination does not fit the definition of an ecosystem service. This distinction, the authors suggest, is a key step to wild pollinator conservation. This argument highlights a fundamental misinterpretation of the ecology of ecosystem services: Services are delivered to beneficiaries through ecological processes and interactions, not by organisms alone. Geldmann and González-Varo have confounded the service (i.e., food production from insect pollination) with the organisms involved in the interactions underlying that service.

We disagree with the assumption that managed animals cannot be involved in delivering ecosystem services. Managed animals are recognized in current ecosystem service classifications (1) as vital contributors to ecosystem services delivery, both directly (through food and fiber products such as meat, milk, and wool) and indirectly (through interactions such as pollination and pest control). The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) pollination assessment explicitly recognizes that both wild and managed pollinators have "globally significant" roles in crop pollination (2, 3). This includes non-*Apis* managed pollinators, such as bumble bees (*Bombus* spp.), stingless bees (*Meliponini*), and solitary bees (2). The relationship between pollinator diversity and crop pollination services depends on a suite of ecological and environmental variables, including floral traits, landscape context, weather conditions, and on-farm management (4–6).

The concept of ecosystem services is not about humans passively receiving benefits from "wild" nature. Rather, it encourages mindful management and interaction with surrounding ecosystems that sustain natural processes and human well-being synergistically. Wild pollinator conservation will indeed benefit from more research and public communication about the differences and interactions between managed and wild pollinators. However, ignoring the global contribution of managed pollinators to ecosystem services will not facilitate wild pollinator conservation. Instead, it disregards a vital component of ecosystem services necessary to feed an increasingly populous planet.

Manu E. Saunders,[1,2]* **Tobias J. Smith,**[1] **Romina Rader**[1]

[1]School of Environmental and Rural Science, University of New England (UNE), Armidale, NSW, Australia. [2]UNE Business School, UNE, Armidale, NSW, Australia.
*Corresponding author.
Email: manu.saunders@une.edu.au

REFERENCES

1. R. Haines-Young, M. B. Potschin, Common International Classification of Ecosystem Services (CICES) V5.1 (2018); www.cices.eu.
2. IPBES, "The assessment report on pollinators, pollination, and food production," S. G. Potts et al., Eds. (2016); www.ipbes.net/assessment-reports-0.
3. L. V. Dicks et al., Science **354**, 975 (2016).
4. L. A. Garibaldi et al., Science **339**, 1608 (2013).
5. R. Rader et al., Proc. Natl. Acad. Sci. U.S.A. **113**, 146 (2016).
6. P. Balvanera et al., Science **291**, 2047 (2001).

10.1126/science.aat1535

Bee conservation: Inclusive solutions

In their Perspective "Conserving honey bees does not help wildlife" (26 January, p. 392), J. Geldmann and J. P. González-Varo point out that promoting managed honey bees does not help wild pollinators.

Policies regarding managed bees, such as this bumble bee (*Bombus* spp.), affect wild pollinators as well.

We agree that, at high densities, honey bees can adversely affect wild pollinator populations. However, focusing only on the negative aspects of their interactions may be counterproductive for both wild and managed pollinators.

Countries such as the Netherlands (1) have increasingly restricted honey bee access to protected areas based on incomplete evidence for negative impacts on wild pollinators and plants (2, 3). Such restrictions are mostly symbolic acts, given that honey bees can forage up to 10 km from their hive and continue to use resources within protected areas even when hives remain outside (4). However, the regulations fuel tensions between beekeepers and conservationists.

A more productive approach would be to promote the suite of pollinators—both wild and managed—that provide pollination services to crops and wild plants (5). A united front of beekeepers and conservation organizations, together representing millions of citizens, is more likely to succeed in driving policy changes and public awareness than different sectors advocating either wild or managed species. New generations of initiatives to promote pollinators, such as the Dutch Bee Strategy (6), the English National Pollinator Strategy (7), and the International Pollinator Initiative (8), all use this inclusive approach. Moreover, all of these initiatives include the agricultural and environmental sectors, as well as the private sectors, because only solutions that are supported by all parties can deliver sustainable results.

Whether considering food security, national economies, or nature conservation, we must safeguard both wild and managed pollinators. Arguing that one group is more important than another overlooks the key

global challenges and opportunities that wider society needs to address.

David Kleijn,[1]* Koos Biesmeijer,[2,3] Yoko L. Dupont,[4] Anders Nielsen,[5] Simon G. Potts,[6] Josef Settele[7]

[1]Plant Ecology and Nature Conservation Group, Wageningen University, Wageningen, Netherlands. [2]Naturalis Biodiversity Center, 2300 RA Leiden, Netherlands. [3]Institute for Environmental Sciences Leiden University, 2300 RA Leiden, Netherlands. [4]Department of Bioscience, Aarhus University, Denmark. [5]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway. [6]Centre for Agri-Environmental Research, School of Agriculture, Policy, and Development, Reading University, Reading RG6 6AR, UK. [7]Helmholtz Centre for Environmental Research–UFZ, Department of Community Ecology, 06120 Halle, Germany.
*Corresponding author. Email: david.kleijn@wur.nl

REFERENCES

1. E. Van der Spek, *Entomol. Ber.* **72**, 103 (2012) [in Dutch].
2. R. E. Mallinger et al., *PLOS ONE* **12**, 32 (2017).
3. D. R. Paini, *Austral. Ecol.* **29**, 399 (2004).
4. M. Beekman, F. L. W. Ratnieks, *Funct. Ecol.* **14**, 490 (2000).
5. IPBES, "The assessment report on pollinators, pollination, and food production," S. G. Potts et al., Eds. (2016); www.ipbes.net/assessment-reports-0.
6. Government of the Netherlands, "NL pollinator strategy—bed & breakfast for bees" (2018); www.government.nl/documents/reports/2018/02/02/nl-pollinator-strategy-bed--breakfast-for-bees.
7. U.K. Department for Environment, Food, and Rural Affairs, "The National Pollinator Strategy: For bees and other pollinators in England" (2014); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/409431/pb14221-national-pollinators-strategy.pdf.
8. Convention on Biological Diversity, "Decision adopted by the Conference of the Parties to the Convention on Biological Diversity XIII/15" (2016); www.cbd.int/doc/decisions/cop-13/cop-13-dec-15-en.pdf.

Response

Saunders et al. argue that honey bees play a significant role in crop pollination and that managed species, in general, can deliver ecosystem services. We agree. In our Perspective, referencing the same source as Saunders et al. (1), we unequivocally state the importance of managed pollinators for food production and emphasize the significant role of honey bees for global food security. We also agree that managed animals can deliver ecosystem services, from grazing cattle maintaining meadows to chicken feathers used in cultural costumes.

However, we disagree with Saunders et al.'s assertion that all ecosystems and all types of services should be classified as ecosystem services. Saunders et al. advocate the Common International Classification of Ecosystem Services' definition, developed based on the Millennium Ecosystem Assessment (2), which lacks a qualifying definition of "ecosystem," meaning it could be interpreted to include any ecosystem, no matter how artificial, as well as any service, no matter how commercial. Even fossil fuels, the product of ecosystems that existed millions of years ago, could be considered an ecosystem service under the auspice of their definition. However, we question whether this definition is useful, generally accepted, or in line with the founding ideas, which emphasized the need to complement economic metrics, such as the gross domestic product (GDP), for services that were not easily captured by the market (3, 4).

We advocate a definition where the services are provided by more natural, native, or wild elements, which are more likely to deliver biodiversity conservation benefits. Pest control by insectivorous birds and bats (5) cannot be equated with *Pyrethrum*-derived pesticides, and crop pollination by wild animals should not be equated with pollination by managed bees. In both cases, even if naturally derived, the pesticides and the managed bees are externalities to the local ecosystems with the only aim of improving crop yield.

Finally, we do not suggest removing the ecosystem service tag from crop pollination by managed honey bees as a key step for pollinator conservation, but rather for increasing public understanding of the difference between managed and wild pollinators. Key steps for wild pollinator conservation should focus on expanding and protecting natural areas that wild pollinators rely on and minimizing the effects of agricultural intensification (e.g., pesticides and fertilizers) in these areas. Furthermore, regulating managed honey bees in areas of importance to wild pollinators and increasing the availability and diversity of noncrop food sources (i.e., native wild flowers) in the more cultivated landscapes will help address the decline in wild pollinators.

In their Letter, Kleijn et al. feel that we only focus on the negative aspects of honey bees, which we find surprising. We clearly state that managed honey bees are a necessary agricultural tool for improving crop yield; that they serve as a "canary in a coal mine" because pressures affecting them are also affecting wild pollinators; and that honey bees have been important in raising awareness for conservation issues (6).

We agree with Kleijn et al. that inclusive solutions are important, if not essential, for the success of conservation strategies. However, inclusiveness should not be interpreted as permissiveness. We highlight two important reasons for site-specific regulations of managed pollinators. First, beekeeping extracts pollen and nectar from the environment, which are resources needed by wild pollinators. Cane and Tepedino (7) recently estimated that a 40-hive apiary located on natural habitats for 3 months collects the pollen equivalent of 4 million wild bees. We accept that extractive activities (logging, cattle grazing, and even hunting) are allowed and can be sustainable within certain protected areas. However, the impacts of such activities are normally assessed, and the activities regulated accordingly, which is rarely the case for beekeeping. Second, we must distinguish between the native and non-native range of the honey bee. Within the native range (Europe and Africa), restricting hive numbers at low densities in protected areas could mirror the past densities of wild honey bees. However, in their non-native range, any density of honey bees is unnatural, yet hive numbers in protected areas keep growing (7, 8).

Kleijn et al. claim that restricting beekeeping in protected areas is a symbolic act because honey bees can forage up to 10 km from hives located outside. We disagree. The hives present in the landscape determine the density of foraging honey bees (9). Thus, restricting hive numbers will at least keep honey bee densities lower. Moreover, the long foraging distances invoked by Kleijn et al. are unlikely: Mean foraging distances are usually ~1 km from the hive (10, 11) and the probability of foraging sharply declines beyond 1 km (12). Accordingly, the detrimental effects of honey bees on native bumble bees decreased markedly along 1200 m while moving away from apiaries (13).

Inclusive solutions bringing together different societal sectors should be compatible with pollinator conservation, and this requires case-specific regulations (14). Advocating a precautionary principle in our protected and vulnerable landscapes, where the need for crop pollination is negligible, is not the same as widely banning beekeeping.

Juan P. González-Varo and Jonas Geldmann

Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. Email: jpgvaro@outlook.com (J.P.G.-V.); jg794@cam.ac.uk (J.G.)

REFERENCES

1. IPBES, "The assessment report on pollinators, pollination, and food production," S. G. Potts et al., Eds. (2016); www.ipbes.net/assessment-reports-0.
2. Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: Synthesis* (Island Press, 2005).
3. G. Daily, *Nature's Services: Societal Dependence on Natural Ecosystems* (Island Press, 1997).
4. J. Boyd, S Banzhaf, *Ecol. Econ.* **63**, 616 (2007).
5. B. Maas et al., *Ecol. Lett.* **16**, 1480 (2013).
6. M. Allsopp, R. Tirado, P. Johnston, D. Santillo, P. Lemmens, *Plan Bee—Living Without Pesticides* (Greenpeace International, Amsterdan, 2014).
7. J. H. Cane, V. J. Tepedino, *Conserv. Lett.* **10**, 205 (2017).
8. F. Hancock, *The Dark Side of NZ's Honey Bee* (Newsroom, Auckland, 2018).
9. J. P. González-Varo, M. Vilà, *Biol. Conserv.* **212**, 376 (2017).
10. N. Danner et al., *Ecol. Appl.* **26**, 1920 (2016).
11. I. Steffan-Dewenter, A. Kuhn, *Proc. R. Soc. B Biol. Sci.* **270**, 569 (2003).
12. M. Couvillon et al., *Curr. Biol.* **24**, 1212 (2014).
13. D. Thomson, *Ecology* **85**, 458 (2004).
14. European Commission, Plants, Neonicotinoids (2018); https://ec.europa.eu/food/plant/pesticides/approval_active_substances/approval_renewal/neonicotinoids_en.

Rwandan President Paul Kagame challenged Africa's science and technology community to come together to build an innovation ecosystem.

Africa cultivates innovation to boost global reach

The Next Einstein Forum's Global Gathering 2018 explores opportunities, identifies challenges

By **Anne Q. Hoy**

Tolullah Oni, a public health physician, researcher and epidemiologist, recently traced her life journey—birthplace in Lagos, Nigeria; education in London; and professional career in South Africa—to underscore the power of inspiration, the value of education, and the importance of mentors to the vitality of Africa's scientific community, particularly for women and the continent's growing demographic of young people.

The story of Kenyan research scientist Rose Mutiso strikes similar notes. Educated at Ivy League institutions in the United States, she returned to the continent with a Ph.D. in materials science and engineering and two undergraduate degrees in engineering in hand. Wasting no time to begin to fulfill her longtime ambition to improve the lives of Africans, Mutiso cofounded and now heads the Mawazo Institute, a Nairobi-based nonprofit research institute that supports African female Ph.D. candidates whose research is focused on reversing the continent's development challenges.

Oni and Mutiso were among more than 1,600 participants from 91 countries in Africa and beyond who convened last month in Kigali, Rwanda, at the Next Einstein Forum Global Gathering 2018, where African scientists highlighted their research and participants explored how to accelerate the continent's science, technology, and innovation capacity; expand precision health care; respond to climate impacts on energy, agriculture, and economic growth; and accelerate the adoption of digital technologies to connect more African countries with each other and the rest of the world. The Forum featured a Nobel laureate, two African presidents, and representatives of the American Association for the Advancement of Science and other leading scientific societies, industries, governments, and private philanthropic organizations from across the globe.

Oni was selected as a fellow of the Next Einstein Forum, a competitive fellowship that provides recipients with global opportunities to advance their careers. She also is cochair of the Global Young Academy, an international organization dedicated to cultivating the next generation of researchers on the continent and those living abroad, and a fellow of a World Economic Forum council on the future of health and health care. She has participated in AAAS training sessions and events sponsored by its Center for Science Diplomacy.

Last year, Mutiso was named the Next Einstein Forum's Kenyan ambassador, an initiative that selects a network of working scientists under the age of 42 to represent each of Africa's 54 nations and participate in a range of public engagement activities to share news about the continent's science, technology, and innovation advances. She also is a senior fellow of the Energy for Growth Hub Initiative, for which she conducts independent research on energy sector topics. Earlier in her career, after completing graduate studies in the United States, she served as a AAAS Science & Technology Policy fellow, a position that placed her in a U.S. Senate office and at the Energy Department, where she delved into issues such as renewable energy and energy access in sub-Saharan Africa and Asia, seeing firsthand the nexus between science and public policy.

The life trajectories of Oni and Mutiso open a window into the advances taking place in Africa as the continent becomes "a generator of knowledge, innovation, creativity, and technology, rather than being solely an adapter of trends produced elsewhere in the world," as described in a white paper prepared by the African Academy of Sciences' Sustainable Development Goals Center for Africa and Carnegie Mellon University Africa. Their stories and experiences also point to continuing obstacles.

As gifted and ambitious African scholars, they followed their curiosity and absorbed the world around them from an early age. Educated abroad at top-shelf institutions, they were exposed to inspiring teachers and professors and found access to advanced laboratories and time to discover their passion. They participated

in scientific conferences with powerful networking opportunities. "This is all part of the international research community, and if you are not in it, you are out," said Mutiso.

Africa faces the world's largest shortage of primary and secondary school teachers at the very time the population of school-age students is on track to grow at the world's largest pace. In addition, 32.3% of sub-Saharan Africa's children, adolescents, and youth do not attend school, the world's lowest educational participation rate, according to the UNESCO Institute for Statistics.

Despite the work many African nations have devoted to elevating the continent's global standing in science and technology, significant challenges remain. Recent World Bank statistics show the continent contributes 1.3% to the world's total scientific research output and its research and development investments rank among the world's lowest. Scientific laboratories at African universities also are often ill-equipped, Next Einstein Forum participants said.

Creating an environment in Africa for science, engineering, and innovation to flourish begins with an educational system that is focused on science, technology, engineering, and mathematics from the elementary level through college and graduate school, a necessity repeatedly cited by multiple Forum participants.

"For too long Africa has allowed itself to be left behind, but that is starting to change as we see with the important work on display at this forum," Rwandan President Paul Kagame said in a keynote address on the Forum's opening day.

"We need to start with the basics," added Albert G. Zeufack, the World Bank's chief of economics for Africa, during a Forum session. "The learning crisis must be addressed if we are to generate the next Einstein in Africa."

Oni said Africa's young scientists need interdisciplinary and collaborative efforts to thrive.

African countries are forging agreements with other African nations to develop homegrown science and technology solutions that address specific African problems, an approach that several philanthropic groups, foundations, scientific societies, and international organizations across the world are supporting. Africa's leading scientific countries participating in the Forum each pledged to increase their national research and development budgets 1% by 2020 and 3% by 2025, according to a draft 2018 Kigali Declaration reached at the Forum—levels higher than those agreed to by participants 2 years ago.

The drive to transform the continent's scientific profile also has seen the emergence of African-run scientific research institutions, from the African Academy of Sciences' Alliance for Accelerating Excellence in Science in Africa to the African Institute for Mathematical Sciences, which established the Next Einstein Forum in partnership with the German-based Robert Bosch Stiftung (Foundation) in 2013.

Africa's growing consumer base is being met by more than 314 active technology hubs in 93 cities and 42 African countries, spurring innovations in every sector from education and health to agriculture and energy, according to the white paper.

Kagame also singled out gender disparities among Africa's scientific community as a problem that must be addressed. "As Africa catches up to the rest of the world, we cannot afford to leave our women and girls out of the equation," he said.

Attracting and preparing women and the continent's growing population of young people for careers in science, technology, and innovation was a goal repeatedly cited at the Forum. Outreach to women and the continent's youth demographic is essential to fostering the "knowledge acquisition" necessary to propel innovation in Africa, participants said.

New avenues for African women and young scientists need to be paved to expand Africa's scientific reach, said Oni during a presentation focused on bridging the gender gap in Africa's science and technology. "We need to fix the system. We need to think about the conditions in the pipeline itself. We need to do more than just plug the holes," she said. "We should be invested, more invested in new kinds of partnerships that are really finding and nurturing these hidden talents that are often stifled."

One such avenue unveiled at the Forum was the launch of a peer-reviewed, multidisciplinary, open-access journal dedicated to research by African scientists. *Scientific African*, a quarterly journal to be published by Elsevier, will focus on relevant African topics and seek to expand opportunities to showcase African research. It is slated to begin publishing later this year.

A closing day session that featured a panel of global scientific leaders, including Rush Holt, chief executive officer and executive publisher of the *Science* family of journals, examined the state of science and challenges confronting the scientific community across the world, a discussion that identified situations that the African research community may want to avoid.

The impact of the world's rise in nationalist movements along with a sometimes-negligent attitude toward evidence pose risks for the scientific enterprise and stir public distrust in science, said panel participants.

These political and social developments increase the need for nongovernmental scientific organizations such as AAAS to continue to engage and inform the public about evidence-based realities and demonstrate the many ways the scientific enterprise improves lives and contributes to global well-being, Holt said.

He pointed, for instance, to some of AAAS's many programs—science communication workshops, media fellowships that place graduate students in news outlets, and international journalism awards to highlight leading science writing—designed to implement AAAS's mission to "advance science for the benefit of all people."

As publisher of the *Science* family of journals, Holt said, AAAS makes scientific papers available at reduced rates to scientists in low- and-middle income nations and makes it possible for those who have papers published in the *Science* journals to post their manuscripts on PubMed or place them on their own repositories, in recognition of their need to share and cite scientific research.

Turning to the Next Einstein Forum's objective to help Africa build institutions and opportunities to produce the next Albert Einstein, Holt said preserving public trust in science and the freedom of scientists to pursue their ideas are paramount to global progress.

"Political, academic, and scientific freedom are not separable," Holt said. "For science to thrive, young scientists must find confidence, freedom, and sustained support. You have to have organizations that are monitoring the situation, continuing to speak out and advocating for them."

"Those who want them to succeed have to insist that science is not a luxury for nations in Africa or elsewhere."

RESEARCH

Cryo-ET elucidates flagellar motor mechanisms
Lin and Nicastro, p. 396

PSYCHOLOGY
Cultural differences in Starbucks

For thousands of years, people in northern China grew wheat, whereas people in southern China farmed rice. The latter required shared labor and coordinated irrigation. Rice cultivation thus fostered an interdependent culture distinct from the more individualistic wheat-growing culture. Talhelm *et al.* wanted to understand whether traits related to rice versus wheat farming still influence culture today. They observed Starbucks customers in six Chinese cities. Customers in rice-growing regions were less likely to sit alone and squeezed themselves through narrow aisles, whereas customers in wheat-growing regions were more likely to move a chair blocking their way. Thus, it seems that historical cultural differences related to farming may continue to influence people's behavior. —PJB

Sci. Adv. 10.1126/sciadv.aap8469 (2018).

A Starbucks cafe in Shenzhen, China

QUANTUM ENTANGLEMENT
Splitting the entanglement

When particles in a quantum mechanical system are entangled, a measurement performed on one part of the system can affect the results of the same type of measurement performed on another part—even if these subsystems are physically separated. Kunkel *et al.*, Fadel *et al.*, and Lange *et al.* achieved this so-called distributed entanglement in a particularly challenging setting: an ensemble of many cold atoms (see the Perspective by Cavalcanti). In all three studies, the entanglement was first created within an atomic cloud, which was then allowed to expand. Local measurements on the different, spatially separated parts of the cloud confirmed that the entanglement survived the expansion. —JS

Science, this issue p. 413, p. 409, p. 416; see also p. 376

POLYMERS
Recycle, recycle, recycle

Some polymers, such as polyethylene terephthalate in soft drink bottles, can be depolymerized back to the starting monomers. This makes it possible to repolymerize true virgin material for repeated use. Zhu *et al.* developed a polymer based on a five-membered ring cyclic monomer derived from γ-butyrolactone that could be produced at ambient temperature and mild conditions (see the Perspective by Sardon and Dove). The high-molecular-weight polymer exhibited high crystallinity and thermal stability. However, at hot enough conditions, or at lower temperatures in the presence of a zinc chloride catalyst, the polymer could be returned to its starting monomers and thus recycled into new material. —MSL

Science, this issue p. 398; see also p. 380

BIOTECHNOLOGY
Taking CRISPR technology further

CRISPR techniques are allowing the development of technologies for nucleic acid detection (see the Perspective by Chertow). Taking advantages of the distinctive enzymatic properties of CRISPR enzymes, Gootenberg *et al.* developed an improved nucleic acid detection technology for multiplexed quantitative and highly sensitive detection, combined with lateral flow for visual readout. Myhrvold *et al.* added a sample preparation protocol to create a field-deployable viral diagnostic platform for rapid detection of specific strains of pathogens in clinical samples. Cas12a (also known as Cpf1), a type V CRISPR protein, cleaves double-stranded DNA and has been adapted for genome editing. Chen *et al.* discovered that Cas12a also processes single-stranded DNA threading activity. A technology platform based on this activity detected human papillomavirus in patient samples with high sensitivity. —SYM

Science, this issue p. 439, p. 444, p. 436; see also p. 381

METABOLISM
When beige is not boring

Shifting energy-storing white adipocytes or their progenitors into energy-burning beige adipocytes could be a strategy to combat obesity. Working in mice, Babaei *et al.* found that transient inflammation induced adipocyte progenitors to differentiate into beige adipocytes (see the Focus by Sun *et al.*). Unfortunately, mitophagy causes beige adipocytes to revert to white adipocytes in the absence of "browning" stimuli. Lu *et al.* found that a browning stimulus prevented the mitochondrial recruitment of the mitophagy factor Parkin and that Parkin-deficient mice maintained beige fat upon stimulus withdrawal (see the Focus by Sarraf and Youle). —WW

Sci. Signal. **11**, eaap8526, eaai7838; see also eaat1082, eaat3192 (2018).

OCEAN HYPOXIA
Haunted by the past

Reducing the extent of hypoxia in the Gulf of Mexico will not be as easy as reducing agricultural nitrogen use. Van Meter *et al.* report that so much nitrogen from runoff has accumulated in the Mississippi River basin that, even if future agricultural nitrogen inputs are eliminated, it will still take 30 years to realize the 60% decrease in load needed to reduce eutrophication in the Gulf. This legacy effect means that a dramatic shift in land-use practices, which may not be compatible with current levels of agricultural production, will be needed to control hypoxia in the Gulf of Mexico. —HJS

Science, this issue p. 427

"Legacy" nitrogen in river runoff into the Gulf of Mexico continues to cause hypoxia.

NEUROSCIENCE
Memories are stored in synapses

Memory formation is thought to change the strength of synaptic connections between neurons. However, direct measurements between neurons that participate in a learning process are difficult to obtain. Choi *et al.* developed the "dual-eGRASP" technique to identify synaptic connections between hippocampal CA3 and CA1 pyramidal cells. This method could label two different sets of synapses so that their convergence on the same dendrites would be quantified. After contextual fear conditioning in mice, the number and size of spines were increased on CA1 engram cells receiving input from CA3 engram cells. —PRS

Science, this issue p. 430

REGENERATION
A recipe for regeneration

Unlike humans, planarian flatworms can regenerate certain tissues. During regeneration, existing tissues remodel, and undifferentiated and progenitor cells convert into specialized cell types at specified locations. Atabay *et al.* examined planarian eye regeneration (see the Perspective by Tanaka). Surgical and transplantation experiments revealed three properties governing regenerative progenitor behavior: cell self-organization, an extrinsic migratory target for progenitors, and a broad progenitor-specification zone. Predictions from this model enabled generation of animals with multiple stable eyes. —BAP

Science, this issue p. 404; see also p. 374

IN OTHER JOURNALS
Edited by **Caroline Ash** *and* **Jesse Smith**

SCHOLARLY PUBLISHING
Lessons learned from the *JACS* Challenge

How do we gauge the importance of a scientific publication? Citation counts are, by their nature, crowdsourced, but their inherent meaning is somewhat unclear. To explore this question in more depth, Borchardt *et al.* asked chemists in a survey to look back at a 10-year-old issue of the *Journal of the American Chemical Society* (*JACS*) and predict, without checking, which three papers had been most highly cited. Respondents were also asked which papers they construed as most important and which they would share with other chemists or the public more broadly. Citations not only proved rather hard to predict, but also correlated poorly with the papers chosen to share. —JSY

PLOS ONE **13**, e0194903 (2018).

DIABETES
Staving off diabetes

Childhood obesity is a growing epidemic associated with increased risk of health issues later in life. Bjerregaard *et al.* asked whether, and at what age, shedding of excess pounds by overweight kids influenced their chance of developing type 2 diabetes as adults. The authors examined more than 62,000 Danish men for whom height and weight data were available from childhood into early adulthood. Individuals who were overweight throughout adolescence were four times as likely to develop adult type 2 diabetes. However, males who lost excess weight by age 13 cut their diabetes risk to that of those who were never overweight. Thus, maintaining a healthy body mass index throughout adolescence should reduce future chances of type 2 diabetes in men. —PNK

N. Engl. J. Med. 10.1056/NEJMoa171323.1 (2018).

HEMATOPOIESIS
Gut bugs encourage hematopoietic recovery

Antibacterial measures are vital to the success of many clinical interventions. But increasing evidence shows that we need to be more discriminating in our

mitochondrial activity and protein quality-control mechanisms. —PAK

NEURODEVELOPMENT
The brain as a work in progress

Although the brain grows little in volume after childhood, small structural changes, such as myelination and synapse pruning, continue into adulthood. Using multiecho functional magnetic resonance imaging to study subjects ranging in age from 8 to 46 years, Kundu *et al.* found that the functional organization of the brain shifts. Localized networks that characterize youth meld into larger and more functionally distinct networks with maturity. Not all parts of the brain change at the same rate; some regions are more dynamic. These regions correlate with aspects of cognition, such as the ability to monitor one's performance, to estimate others' intentions, or to develop a sense of self—all skills that are works in progress during adolescence. —PJH

Cognitive skills, such as those relating to executive function and coordination, continue to develop into adulthood.

STELLAR ACTIVITY
A brief flash from our closest neighbor

Proxima Centauri, a small red dwarf, is the closest star to us other than the Sun. It hosts an exoplanet that may be about Earth's mass and might lie in the habitable zone, but many red dwarfs have frequent stellar flares that could strip away a planet's atmosphere. MacGregor *et al.* investigated millimeter-wavelength observations of Proxima Centauri, discovering that they contain a bright flare lasting less than a minute. Such short-duration millimeter-wavelength flaring has not previously been recognized and casts doubt on previous claims of dust lanes and asteroid belts in the system. Its effect on habitability is unknown. —KTS

efforts at bacterial extirpation. Profound immunodeficiency is a major challenge for postoperative bone marrow transplantation (BMT) patients. Considering that the gut microbiota is important for immune training during early life, Staffas *et al.* examined whether microbes in the intestine play a role in hematopoietic recovery after BMT. A major reduction in lymphocytes and neutrophils occurred in antibiotic-treated mice after surgery, accompanied by a loss of visceral fat. Hematopoietic stem and progenitor cells were less affected. Experiments with different antibiotics showed that ampicillin was associated with the poorest recovery and greatest fat loss because it caused near-total ablation of the microbiota. —BAP

MEDICINE
A cool model for biomedical research

Hibernation is a fascinating physiological adaptation. One important question is which cellular signaling pathways allow hibernating mammals to enter and exit a cold-tolerant state without damage to cells and organs. Understanding this could conceivably lead to interventions that prolong the shelf life of donor organs before transplant. Ou *et al.* studied these pathways by establishing induced pluripotent stem cells (iPSCs) from a hibernating ground squirrel. Neurons derived from the iPSCs retained cold-resistant features, including microtubule stability. Comparison of ground squirrel and human iPSC-derived neurons revealed that differences in cold tolerance result in part from species differences in druggable signaling pathways that govern

Neural stem cells in thirteen-lined ground squirrels are geared for hibernation.

MICROBIOME
A safe haven for the small

The menu for gut microbes is decided by their host and is not always healthy. Linoleic acid, a component of soybean oil, is toxic to beneficial lactobacilli in the laboratory, but its role in shaping the microbiome is less clear. Di Rienzi *et al.* found that cultured lactobacilli developed strategies to resist the toxic fatty acid and that lactobacilli from mice fed a high-fat diet were more likely to be resistant than those fed a low-fat diet. However, in strains isolated from mice treated with linoleic acid, the resistant lactobacilli were not enriched. Gut bacteria may therefore be shielded from the harmful effects of linoleic acid by the local environment within the gut. —MAF

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

CILIA AND FLAGELLA
Switching how to make flagella beat

Motile cilia and flagella are hairlike cellular appendages that power the movement of individual cells or liquid across tissues, as exemplified by the cilia found in airways. The question of how they move in rhythmic oscillations has puzzled scientists for centuries. Lin and Nicastro used cryo–electron tomography (cryo-ET) to visualize the activity states of individual dynein motors with respect to their locations within beating flagella. They observed an asymmetric distribution of dynein activity and the switching of conformations of dyneins and their regulators between opposite sides of active flagella. The results confirm the switching aspect of the prevailing "switch-point" hypothesis but change the view with respect to how dynein activities are coordinated to drive flagellar motility. —SMH

Science, this issue p. 396

BIOMECHANICS
Hop, skip, jump, or massive leap

In biological and engineered systems, an inherent trade-off exists between the force and velocity that can be delivered by a muscle, spring, or combination of the two. However, one can amplify the maximum throwing power of an arm by storing the energy in a bow or sling shot with a latch mechanism for sudden release. Ilton *et al.* used modeling to explore the performance of motor-driven versus spring-latch systems in engineering and biology across size scales. They found a range of general principles that are common to animals, plants, fungi, and machines that use elastic structures to maximize kinetic energy. —MSL

Science, this issue p. 397

ORGANIC CHEMISTRY
Light and acid steer a radical addition

So-called Minisci reactions have been used for decades in pharmaceutical and agrochemical synthesis to make carbon-carbon bonds. The reactions link carbon radicals to the carbon centers adjacent to nitrogen in pyridine rings. Proctor *et al.* devised a method to steer these reactions to just one of two possible mirror-image products. To make the radicals, they prepared derivatives of widely available amino acids and then activated them with an iridium photocatalyst. At the same time, a chiral phosphoric acid catalyst was used to activate the pyridine and bias the reaction geometry. —JSY

Science, this issue p. 419

BIOPHYSICS
Watching proteins' weight

Careful measurements of light scattering can provide information on individual macromolecules and complexes. Young *et al.* used a light-scattering approach for accurate mass determination of proteins as small as 20 kDa (see the Perspective by Lee and Klenerman). Movies of protein complex association and dissociation were analyzed to extract biophysical parameters from single molecules and assemblies without labeling. Using this approach, the authors determined in vitro kinetics of fibril and aggregate growth and association constants for a complex protein-glycoprotein assembly. —MAF

Science, this issue p. 423;
see also p. 378

IMMUNOLOGY
Immunometabolism as therapeutic target

Dimethyl fumarate (DMF) is an immunomodulatory compound used to treat multiple sclerosis and psoriasis whose mechanisms of action remain only partially understood. Kornberg *et al.* found that DMF and its metabolite, monomethyl fumarate, succinate the glycolytic enzyme GAPDH (see the Perspective by Matsushita and Pearce). After DMF treatment, GAPDH was inactivated, and aerobic glycolysis was down-regulated in both myeloid and lymphoid cells. This resulted in down-modulated immune responses because inflammatory immune-cell subsets require aerobic glycolysis. Thus, metabolism can serve as a viable therapeutic target in autoimmune disease. —STS

Science, this issue p. 449;
see also p. 377

DIAGNOSTICS
A fluid transition into the field

Many point-of-care diagnostics rely on lateral flow assays or microfluidics; however, these methods generally cannot test multiple samples simultaneously. Ng *et al.* optimized inkjet-printed digital microfluidic cartridges and a portable control system to perform serological immunoassays in remote settings. Digital microfluidics use electrostatic forces to mix and separate reagents and samples in small droplets of fluids. The system measured immunoglobulin G (IgG) antibodies for measles and rubella in human blood samples obtained from adults and children on site in a refugee camp in Kenya. Four samples could be tested simultaneously, although digital microfluidic IgG detection was less sensitive and specific than laboratory-based ELISA (enzyme-linked immunosorbent assay) testing of matched serum samples. The emergence of this field-compatible technology brings with it tools for advancing global health. —CC

Sci. Transl. Med. **10**, aar6076 (2018).

THYMUS
Fueling T cell proliferation

Previous studies on BRCA1-associated protein-1 (BAP1) have documented its importance in suppressing the development of myeloid leukemia. BAP1 is a deubiquitinase (DUB) that acts on histone H2A monoubiquitinated at Lys[119] (H2AK119ub), a chromatin modification associated with gene repression. Arenzana *et al.* report that BAP1 is essential for the development of T cells in the thymus and for promoting peripheral T cell proliferation. Deletion of BAP1 impaired expression of genes associated with cell cycle progression in thymocytes and peripheral T cells. In both cases, the effect of BAP1 deletion was dependent on the DUB activity of BAP1, calling for a closer examination of the role of H2AK119ub in T cell development and differentiation. —AB

Sci. Immunol. **3**, eaal1953 (2018).

RESEARCH ARTICLE SUMMARY

CILIA AND FLAGELLA

Asymmetric distribution and spatial switching of dynein activity generates ciliary motility

Jianfeng Lin and Daniela Nicastro*

INTRODUCTION: Motile cilia and flagella are highly conserved, hairlike appendages of eukaryotic cells that propel the movement of cells or fluids. They play important roles in the normal development and health of many species, including humans. Flagellar beating is driven by the coordinated activities of multiple dynein isoforms that must be spatially and temporally regulated. Although the prevailing "switch-point" hypothesis posits that flagellar motility results from periodic switching of spatially restricted, asymmetrical activation of dyneins, no direct evidence has been reported, and how the thousands of dyneins inside a flagellum work together to generate flagellar motility remains elusive.

RATIONALE: Here we rapidly froze swimming sea urchin sperm cells and used cryo–electron tomography (cryo-ET) to image their beating flagella. Subtomogram averaging and classification analyses allowed us to identify and visualize the different activity states of individual dyneins and their regulators in situ. These conformational states were then mapped to their locations along the sinusoidal wave of the beating flagellum, for example, in relation to principal bend, reverse bend, or straight regions between bends. The results allowed us to elucidate the distinct roles played by various dyneins and to propose a model for the mechanism that underlies ciliary and flagellar motility.

RESULTS: The native three-dimensional structures of flagellar complexes were determined in situ with resolutions sufficient for identifying different activity states. Dyneins of immotile control flagella were found to be in post–power stroke conformations (unprimed, inactive states). By contrast, in beating flagella, most dyneins were in pre–power stroke conformations (primed, active states), with only a few dyneins in intermediate conformations.

Moreover, for all outer dyneins, the intermediate and inactive conformations were only found in bent regions and were clustered on one side of the flagellum in a bend direction–dependent manner. For inner dyneins, certain isoforms (dyneins I1, a, d, and g) showed similar bend direction–dependent distribution patterns in bent regions of flagella, whereas the distribution patterns of other isoforms (dyneins b, c, and e) lacked obvious correlations with bending direction.

Our results revealed three key tenets that are important for generating flagellar motility: (i) The asymmetric distribution of dynein activity on opposite sides of the flagellum results in unidirectional bending, and (ii) the switching of dynein conformations between opposite sides causes the undulating waveform of beating flagella, both of which directly confirmed the switching aspect of the previously proposed switch-point hypothesis. (iii) In contrast to predictions, however, the findings also suggested the paradigm-shifting model that dyneins are active by default and that the asymmetry of dynein activity is driven by spatially restricted inhibition rather than activation of dyneins on alternating sides of the flagellum. This "switch-inhibition" mechanism was further supported by our analyses of a regulation-deficient *Chlamydomonas* mutant, which revealed that dyneins consumed adenosine triphosphate (ATP) and adopted pre–power stroke conformations, even though flagella were paralyzed.

CONCLUSION: Our comprehensive structural analysis combined with biochemical investigations provides an enhanced understanding of the distinct roles played by various dyneins and regulatory complexes in the motility of cilia and flagella and suggests critical modifications to previous hypotheses regarding robust molecular mechanisms underlying flagellar motility. Our study demonstrates that comparative cellular cryo-ET studies provide the conceptual framework and experimental tools to better understand molecular mechanisms and cellular functions. ∎

Asymmetric dynein activity underlies beating of cilia and flagella. Cryo-ET was used to image the active flagellum of swimming sea urchin sperm cells. Different activity states of the motility-driving dynein motors were identified. Magnified views show active (right) and inactive intermediate states (left). The distribution patterns of dynein conformations along the undulating waveform suggest a switch-inhibition mechanism for ciliary and flagellar motility.

ILLUSTRATION: XVIVO SCIENTIFIC ANIMATION

**RESEARCH ARTICLE**

CILIA AND FLAGELLA

Asymmetric distribution and spatial switching of dynein activity generates ciliary motility

Jianfeng Lin[1,2] and Daniela Nicastro[1,2]*

Motile cilia and flagella are essential, highly conserved organelles, and their motility is driven by the coordinated activities of multiple dynein isoforms. The prevailing "switch-point" hypothesis posits that dyneins are asymmetrically activated to drive flagellar bending. To test this model, we applied cryo–electron tomography to visualize activity states of individual dyneins relative to their locations along beating flagella of sea urchin sperm cells. As predicted, bending was generated by the asymmetric distribution of dynein activity on opposite sides of the flagellum. However, contrary to predictions, most dyneins were in their active state, and the smaller population of conformationally inactive dyneins switched flagellar sides relative to the bending direction. Thus, our data suggest a "switch-inhibition" mechanism in which force imbalance is generated by inhibiting, rather than activating, dyneins on alternating sides of the flagellum.

Motile cilia and flagella (*1*) are microtubule-based organelles that propel the movement of cells or fluids. They play essential roles in nearly all eukaryotes; for example, in humans, their motility is critical for embryonic development and organ function, and defects in ciliary assembly or motility lead to a range of diseases termed ciliopathies [reviewed in (*2*)]. From protozoa to humans, motile cilia share a highly conserved 9 + 2 axonemal structure, with nine doublet microtubules (DMTs, consisting of A- and B-tubules) surrounding a central pair complex (CPC) that contains two singlet microtubules with attached projections (Fig. 1A and fig. S1) [reviewed in (*3*)]. Along the DMT length, tens of thousands of dynein motors, regulators, and other axonemal complexes are spatially organized into 96-nm-long repeat units (Fig. 1, B and C, and fig. S1). Ciliary motility is driven by the coordinated activities of multiple dynein isoforms that must be spatially and temporally regulated (*4*, *5*).

Dyneins are multisubunit enzymes that hydrolyze adenosine triphosphate (ATP) to transport cargo toward the minus end of microtubules (*5*–*7*). Their major subunits are one to three large dynein heavy chains (HCs) and several smaller intermediate and light chains that associate with the N-terminal tail of the HC, assisting in oligomerization and cargo binding (Fig. 1D) (*5*). The motor unit of the dynein HC can be divided into four parts (Fig. 1D): the ATP-hydrolyzing head (a

ring of six AAA+ domains), the linker (lever arm between head and tail), the coiled-coil stalk, and the microtubule-binding domain (MTBD) at the stalk's tip (*5*, *7*). Typically, cells encode a single dynein HC isoform for the homodimeric cytoplasmic dynein that fulfills different roles in a plethora of cellular processes, from retrograde transport to mitosis. By contrast, ciliary motility requires more than a dozen different axonemal dynein HC isoforms (*8*). Usually, each axonemal repeat contains 11 dynein complexes that attach permanently with their tails to the A-tubule and form two distinct rows of "arms" along the length of each DMT (Fig. 1C and fig. S1B). One row consists of heterodimeric (α- and β-HC; for example, in sea urchin and human) or heterotrimeric (for example, in protozoa) outer dynein arms (ODAs). In some species, the ODAs of one specialized DMT (usually DMT1 or DMT5) are replaced by nonmotile, cross-linking structures, such as the o-SUB5-6 bridge on DMT5 of sea urchin sperm flagella (Fig. 1A) (*9*). The second row consists of the inner dynein arms (IDAs), which include the heterodimeric (1α- and 1β-HC) I1 dynein and six dyneins (a to e, and g) that have a single HC each (Fig. 1C and fig. S1B). Thus, each repeat contains 16 or 20 dynein motor heads (fig. S1B). Previous in vitro biochemical studies have demonstrated markedly different molecular motor properties (e.g., velocity, processivity, regulatory function) for individual axonemal dyneins (*10*). However, how the various activities of these multiple dyneins are spatially and temporally coordinated to affect ciliary motility remains unknown.

To generate ciliary bending, the dyneins from DMT *n* walk toward the minus end of the neighboring DMT *n* + 1 in an ATP-dependent manner, causing sliding between these adjacent DMTs (fig. S2). In intact cilia, the extent of sliding

is restricted by the basal body and interdoublet linkers (for example, the nexin link), converting the interdoublet sliding into axonemal bending in the direction of DMT *n* + 1 (fig. S2) (*11*, *12*). The undulatory motion of cilia and flagella requires bending in alternate directions. The prevailing "switch-point" hypothesis posits that, during ciliary beating, spatially restricted dynein activity switches regularly by means of alternate activation of dyneins on opposite sides of the axoneme (*13*). This model is supported by several lines of indirect evidence (*14*–*16*) derived from genetic (for example, phenotypes of motility mutants) and biochemical (for example, in vitro assays, DMT extrusion) approaches, as well as conventional electron microscopy with limited resolution and in silico simulations. However, to date, the activity states of individual dyneins in the context of intact beating cilia, and how these contribute to the mechanics of motility, have not been directly demonstrated. It is also largely unknown how the many dynein isoforms are coordinated and how switching of the bending direction is regulated.

To directly test the switch-point hypothesis and define the structural basis for the spatially and temporally coordinated activity of multiple dyneins during ciliary motility, we rapidly froze swimming sea urchin sperm cells and applied cryo–electron tomography (cryo-ET), subtomogram averaging, and particle classification to identify the activity state of individual dyneins. These activity states were then mapped to their locations in the axoneme and in relation to bent or straight regions along the beating flagellum, allowing a direct correlation between structure and function. Our higher-resolution structural analyses elucidate the distinct roles played by various dyneins and regulators to ensure their coordinated activity for robust ciliary beating.

Cryo-ET allows the visualization of the activity states of individual dynein motors along beating flagella

Swimming sea urchin sperm cells with actively beating flagella (active flagella; fig. S3 and movie S1) were rapidly frozen to trap the dynein motors and regulators in their current conformational states (*17*). The natively preserved flagella were then imaged by cryo-ET along their entire length except for the very base of the flagellum, because the ice close to the sperm head was too thick for electrons to penetrate under low-dose conditions. After calculating the three-dimensional (3D) reconstruction of the flagella (Fig. 1, E to G), subtomograms containing dyneins and regulators were extracted, aligned, and averaged (fig. S3 and table S1) (*18*). Then, structural classification (*19*) was performed to separate the different activity states of the dyneins into class averages on the basis of changes in the overall molecule shape and the relative position of major domains within and between molecules (Fig. 2, fig. S3, and table S2).

Dyneins generate force through an ATP-consuming cycle of pre– and post–power stroke conformational changes that cause relative motion

[1]Departments of Cell Biology and Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2]Department of Biology, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454, USA.
*Corresponding author. Email: daniela.nicastro@utsouthwestern.edu

especially between dynein's linker and head domains, resulting in a power stroke that pulls the tail-bound DMT (cargo) toward the minus end of the neighboring DMT (track) [reviewed in (7)]. Extensive previous studies of isolated dynein have related dynein's mechanochemical and stepping cycle to changes in its structure (7), including the 3D atomic structures in different conformational states. This allowed us to interpret the activity state of the class-averaged dynein structures observed here with high confidence (Fig. 2, A to C'). We identified several distinct conformations for all axonemal dynein isoforms (table S2). For example, it has been shown that a small angle between the dynein linker and

stalk domains is representative of dynein's "inactive" post–power stroke conformation (7) (Fig. 2, A, A', F, and G). By contrast, a large angle is indicative of a "primed" ("spring-loaded") pre–power stroke conformation (7) (Fig. 2, B to C', K, and L). In axonemes, we never observed the autoinhibited "phi particle" that was recently described for cytoplasmic dynein (20). The linker of the phi dimer is in a primed position, but the second HC is rotated 180°; thus, the stalk would point toward the microtubule plus end, which could likely not form between properly assembled axonemal dyneins because of geometric constraints. The previous structure-function studies allowed us to identify the activity state of indi-

vidual dynein motors relative to their positions within the axoneme and along the beating flagellum.

Asymmetric and bend direction–dependent distribution of the activity of outer dyneins

To confirm our structure and function assignments, we first analyzed immotile flagella that were inactivated by using either the adenosine triphosphatase (ATPase) inhibitor erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA) (immotile-inhibited; Fig. 1E, fig. S3A, and movie S2) or ATP-depleted, demembranated flagella (immotile-demembranated; Fig. 1F and fig. S3A). The
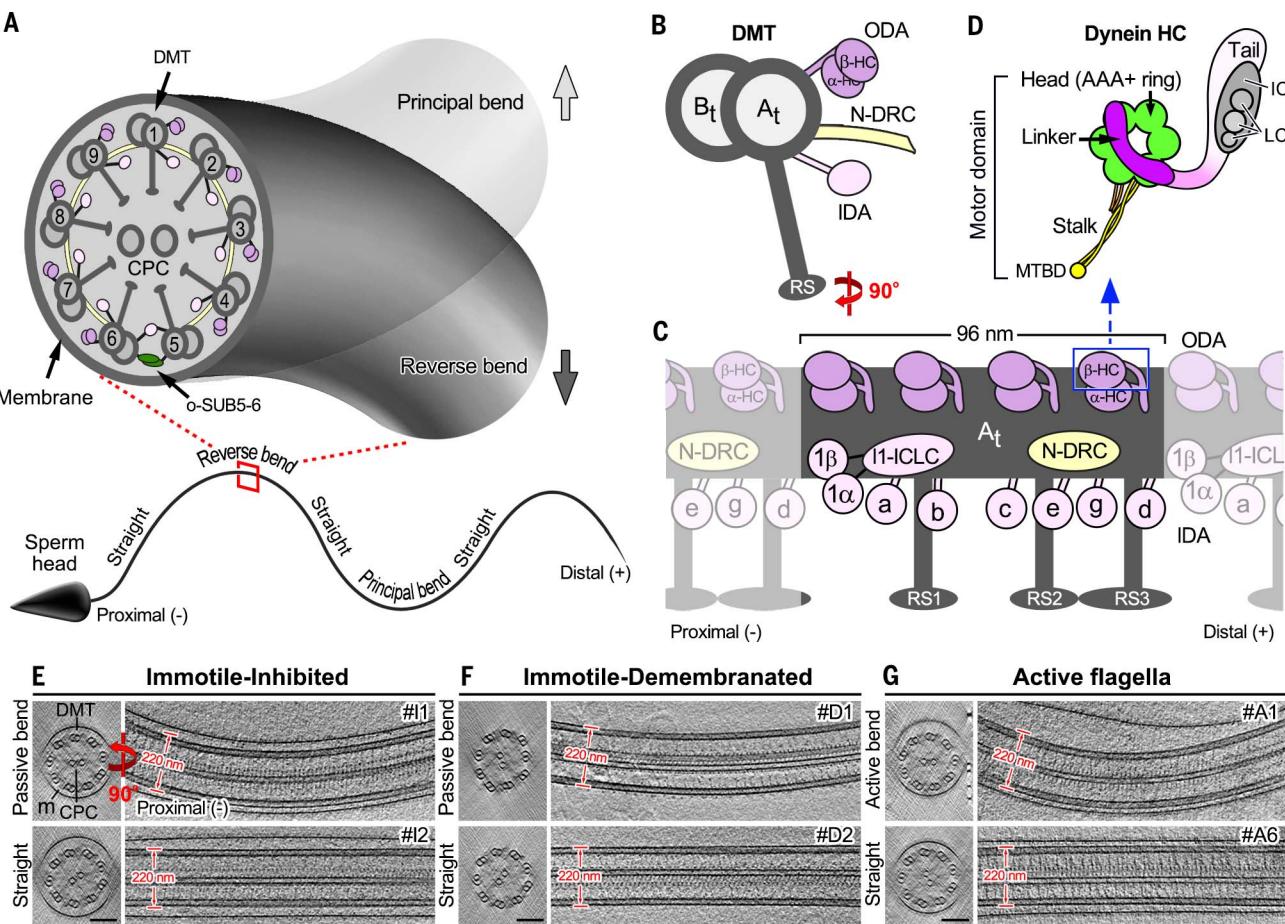
Fig. 1. 3D visualization of native sea urchin sperm flagella using cryo-ET.
(**A**) Diagrams of a sea urchin sperm (bottom) and zoom-in of the flagellum (top). The flagellar bending directions toward DMT1 (principal bend) or the o-SUB5-6 bridge (reverse bend) are indicated (note that the nomenclature of bend direction in relation to DMT numbers is not consistent among species). The o-SUB5-6 bridges on DMT5 of sea urchin sperm flagella replace the ODAs. Microtubule polarity is indicated by "+" and "−" ends. (**B** and **C**) Cross-sectional (B) and longitudinal (C) views of a DMT, which consists of 96-nm-long repeat units. Each repeat contains four ODAs with two heavy chains (α- and β-HC), the I1 dynein with two HCs (1α and 1β) and an intermediate–light chain complex (I1-ICLC), and six single-headed inner dyneins (IDAs a to e, and g). At, A-tubule; Bt, B-tubule; RS, radial spoke.

(**D**) Schematic of domain organization of a dynein HC [magnified view of region outlined in blue in (C)] with associated ICs and LCs that bind to the N-terminal tail domain of the HC and facilitate oligomerization as well as cargo binding (in cilia the cargo is the DMT A-tubule). (**E** to **G**) Tomographic slices of representative bent and straight regions from sea urchin sperm flagella in the following states, as viewed in cross sections and longitudinal sections: intact inhibited immotile flagella (E), demembranated immotile (ATP-free) axonemes (F), and intact active flagella (G). Even in the controls with an immotile flagella and/or axoneme, we occasionally observed bends [(E) and (F), top]. However, these are likely passive bends induced by outside forces (for example, liquid motion during sample preparation). m, membrane. Scale bar in (E) to (G) is 100 nm.

Fig. 2. The outer dyneins exhibit distinct conformations in active sea urchin sperm flagella. (**A** to **C′**) Tomographic slices of ODAs show the three predominant activity states without [(A) to (C)] and with [(A′) to (C′)] schematic model superimposed: the post–power stroke conformation (post) [(A) and (A′)] in immotile flagella and the pre–power stroke states pre-I [(B) and (B′)] and pre-II [(C) and (C′)] in active flagella. Note that the stalk (yellow arrow) with MTBD is microtubule bound in both the post (A) and the primed pre-II (C) states and detached in the primed pre-I state (B) (*17*). (**D**) Spatial resolution of the class averages shown in (A) to (C), as determined by Fourier shell correlation (FSC = 0.5). (**E**) Transitions between dynein conformations. Interpretation of the nucleotide states is based on (*51*). P$_i$, inorganic phosphate. (**F** and **G**) 3D isosurface renderings of the class average post conformation that has both dynein heads in the post–power stroke state. Domain coloring is as shown in the legend in (J). The double-headed arrow in (F) highlights the small gap between the dynein heads and tails. (**H** and **I**) Longitudinal [(H), left] and cross-sectional (I) tomographic slices of the two class averages (ODA-post and o-SUB5-6) in inhibited immotile flagella and a simplified diagram [(H), top right] of the two dynein head positions. The percentage of subtomograms included in each average is indicated. (**J**) 3D isosurface rendering of the averaged axonemal repeat of active flagella, as viewed in cross section. The location of the tomographic slices shown in (A) to (C), (H), and (M) and the viewing direction of (F) and (K) are indicated by their corresponding letters. (**K** and **L**) 3D isosurface renderings of the class average pre conformation that has both dynein heads in the pre–power stroke state. The double-headed arrow in (K) highlights the large gap between the dynein heads and tails. (**M** and **N**) Longitudinal [(M), left] and cross-sectional (N) tomographic slices and simplified diagrams [(M), right] show changes in the dynein-head position of the four averaged ODA classes (pre, IM3, IM2, and IM1) and o-SUB5-6 identified in active flagella. The dynein heads are color coded depending on their conformations. Dotted lines indicate the relative positions of neighboring DMTs, and red arrowheads highlight the positional shift of DMT *n* + 1 in intermediate states IM1 to IM3. Scale bars are 10 nm in (A) to (C′), (H), and (M) and 20 nm in (I) and (N).

occasionally observed bends in these immotile samples (Fig. 1, E and F) were most likely passive bends induced by external force (for example, liquid flow) during sample freezing. As expected for immotile flagella and axonemes, all outer dynein motors (ODA α- and β-HC) were in the tightly microtubule-bound, inactive post–power stroke conformation (Fig. 2, A, A′, and F to I, and fig. S4, post), regardless of whether the flagella

were straight or passively bent (Fig. 3, A and B, and fig. S5).

The classification algorithm also automatically identified a class containing the o-SUB5-6 bridge structures on DMT5 of inhibited (and active) flagella (bottom rows in Fig. 2, H, I, M, and N). As previously reported, the o-SUB5-6 bridge structures cross-link DMT5 and DMT6 in sea urchin sperm flagella (*9*) and likely form a fairly rigid plane

parallel to the two CPC microtubules that are also rigidly cross-linked by a specialized bridge structure (*21*). Together, these two complexes are thought to restrict the flagellar bending direction perpendicular to the planar arrangement of DMT5 and DMT6 and the CPC microtubules, respectively, to impede flagellar twisting or out-of-plane beating (*10*, *15*).

The switch-point hypothesis and prevailing view of flagellar motility predict that active and
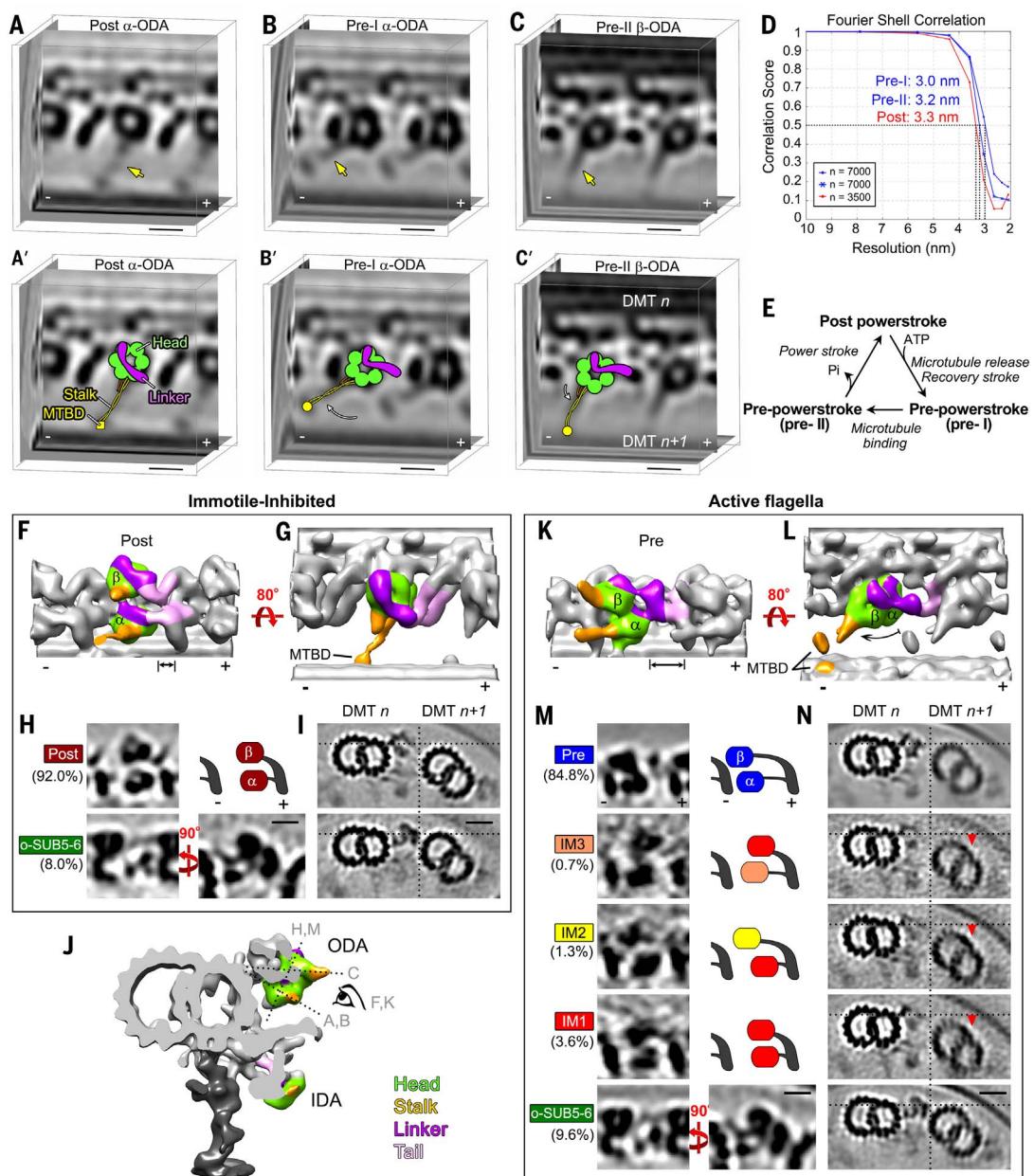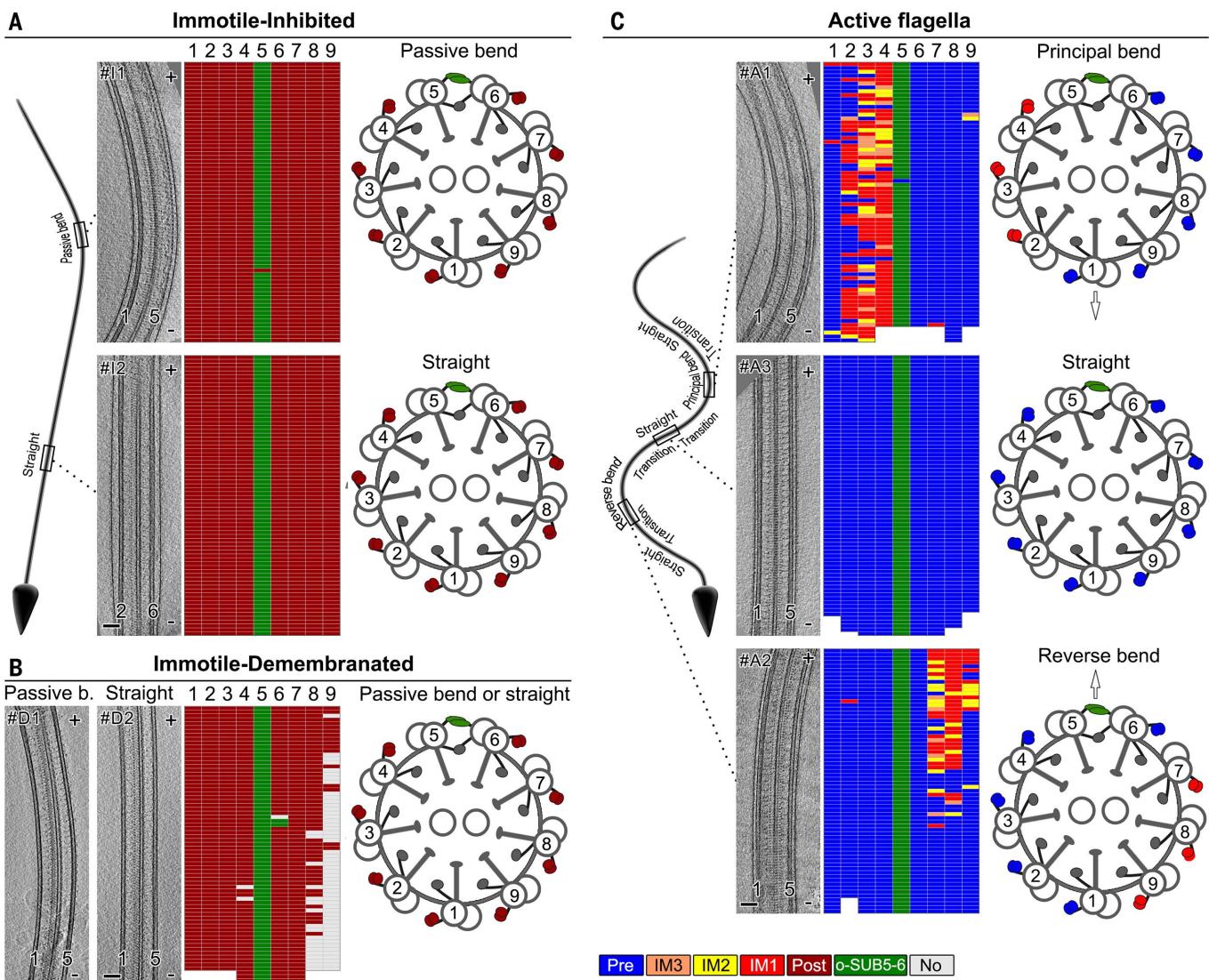
Fig. 3. The distinct conformations of outer dyneins in active sea urchin sperm flagella correlate with bend direction. (**A** to **C**) Distributions of ODA conformations in different regions of individual immotile-inhibited (A), immotile-demembranated (B), and active (C) flagella. For each flagellum, the following is shown: a longitudinal tomographic slice (left), the distribution pattern of ODA conformations (middle), and a diagram of an axonemal cross section (right) (arrows indicate bend directions). In the distribution patterns, the ODA repeats on the nine DMTs (1 to 9) are schematically shown as individual grids; the grid color represents the conformational state of each ODA repeat as shown in the legend. Diagrams of an actively swimming sperm cell (C) and an immotile-inhibited sperm cell (A) were depicted to indicate specific functional regions of the flagellum. Scale bars, 100 nm.

primed dynein states will be found spatially restricted to only a few DMTs on one side of curved regions of the flagellum (*13*). Surprisingly, however, the predominant class of outer dyneins throughout intact flagella had both motor domains in active pre–power stroke conformations (Fig. 2, B to C′ and K to N, pre; and movie S3), in which the dynein heads are closer to the minus end of the DMT relative to the post–power stroke conformation (Fig. 2, compare K with F) and the linkers are in the primed position (Fig. 2, compare L with G). In addition, α- and β-ODA are in the microtubule-detached pre-I and the microtubule-bound pre-II states, respectively (Fig. 2, B to C′).

The remaining three outer dynein classes in active flagella showed intermediate (IM) confor-

mations (IM1 to IM3) that were either post–power stroke–like (Fig. 2M, IM1) or had the dynein head located somewhere between the post– and the pre–power stroke positions (Fig. 2M, IM2 and IM3). In the averages of intermediate conformations, the stalk density was blurred out and thus not clearly visible. This is typically observed when structures that exhibit positional heterogeneity are averaged and suggests that the stalks and MTBDs in these intermediate states were detached from or only weakly bound to the DMT, rather than tightly bound to the DMT. On the basis of the observed structural features, it is likely that the intermediate conformations (IM1 to IM3) of the outer dyneins neither generate force nor resist interdoublet sliding.

Intriguingly, the intermediate conformations were spatially restricted to one side of only bent regions of active flagella. Specifically, depending on the bend direction, they were found either on DMTs 2 to 4 in principal bend regions, or on DMTs 7 to 9 in reverse bends (Fig. 3C). Our direct visualization of the spatial organization of dynein conformation supports two key predictions of the switch-point hypothesis (*13*): (i) Active and inactive states of the outer dyneins are indeed asymmetrically distributed on opposite sides of beating flagella, allowing the actively walking dyneins to drive unidirectional bending, and (ii) to alternate the bending direction during beating, the sides on which the outer dyneins generate force switch. However, in contrast to previous

predictions, in straight regions of active flagella, all outer dyneins were uniformly in primed pre–power stroke conformations (Fig. 3C). Even in bent regions of the flagella, the outer dyneins adopted their pre–power stroke states on at least five of the eight DMTs with ODAs. By contrast, the population of conformationally inactive dyneins was spatially restricted to two to three DMTs and switched flagellar sides relative to the bending direction (Fig. 3C). These data strongly suggest that, at steady state, most outer dyneins readily bind ATP—which is omnipresent in the ciliary matrix—and adopt active, pre–power stroke states within their mechanochemical cycle. This interpretation is consistent with previous in vitro results, including (i) that most isolated dyneins can undergo mechanochemical cycles upon addition of just ATP and Mg^{2+} (22) and (ii) that ATP-

deprived axonemes that have been treated with protease to cleave interdoublet linkers require only the addition of ATP to extrude DMT and undergo axoneme disintegration through linear interdoublet sliding (12, 23). Thus, when the outer dyneins on all DMTs are in their active states, the flagellum is straight, that is, the forces generated on opposite sides of the flagellum are balanced and no net bending force is generated. Spatially clustered inactive dyneins disrupt this balance, allowing dyneins on the opposite side of the flagellum to drive motility.

Inner dynein isoforms also show asymmetric and bend direction–dependent distribution of activity

Similar to the results for outer dyneins, inner dyneins were predominantly in the inactive post–

power stroke state in immotile controls (Figs. 4B and 5A and fig. S5; post) and in the primed pre–power stroke conformations in active flagella (Figs. 4, B and C, and 5, B and G, pre; and movie S4). In addition, for four of the inner dynein isoforms—dyneins a, g, d (Fig. 4C), and I1 dynein (Fig. 5G)—the distribution of the smaller population of intermediate conformations (IM1 and IM2) was highly correlated with the bend direction, that is, they were found either on DMTs 1 to 4 in principal bend regions or on DMTs 6 to 9 in reverse bends of active flagella. This suggests that these inner dyneins contribute to flagellar motility through the same conformational switching mechanisms as seen for the outer dyneins.

In contrast to the outer dyneins, however, we found overall more conformational heterogeneity and specializations among the inner dynein
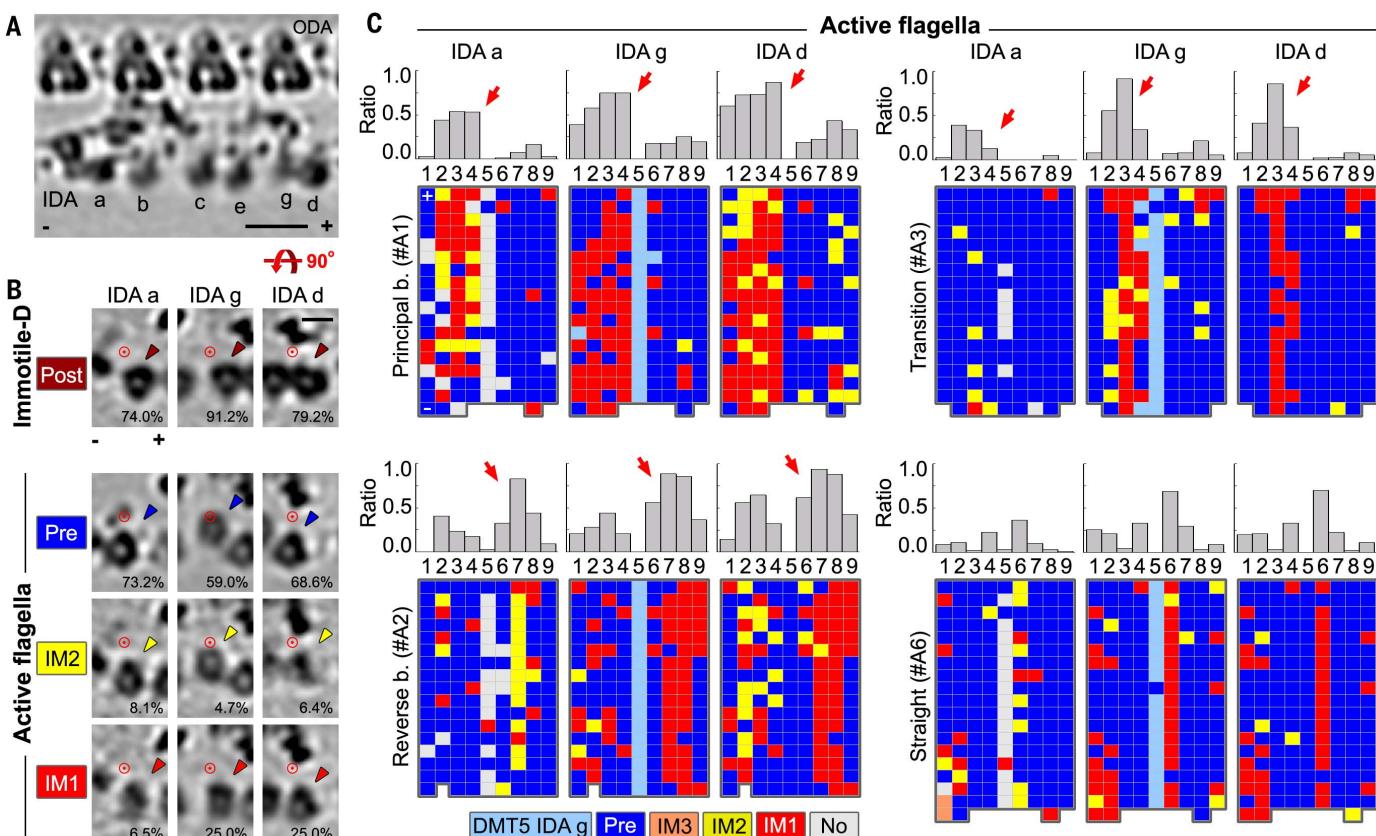
Fig. 4. The inner dyneins a, g, and d exhibit distinct conformations that correlate with bend direction in active sea urchin sperm flagella. (**A**) A longitudinal tomographic slice of the averaged 96-nm repeat of active flagella shows the arrangement of IDAs a to e, and g. (**B**) Tomographic slices of class averages of IDAs a, g, and d in immotile-demembranated axonemes (immotile-D) and active flagella. The percentage of subtomograms included in each class average is indicated. The identified inner dynein conformations are post–power stroke (post, dark red), pre–power stroke (pre, blue), and the intermediates IM2 (yellow) and IM1 (red). Arrowheads point at the particular dynein heads. The red circles mark identical locations in each column to allow better correlation of positional changes of the IDA heads. (**C**) Distributions of the activity states of IDAs a, g, and d in four functional regions of the flagellar wave of active flagella. For each inner dynein isoform, an averaged histogram (top) and color-coded

distribution pattern (bottom) of a representative flagellum are shown. The histograms depict the ratios of intermediate states (IM1 and IM2) among all repeats of each DMT. For the reverse bend (b), principal bend, transition, and straight regions, two, four, two, and five tomograms were included, respectively. Note that mildly bent flagella were mostly excluded from the histograms because of some ambiguity in assigning them to specific functional regions. The red arrows indicate the clustering of IM1 and IM2 conformations for IDAs a, g, and d in a bend direction–specific manner. For the distribution patterns, the conformations of the IDAs on the nine DMTs are schematically shown as individual grids; the grid colors represent the conformation of each inner dynein according to the color scheme in (B). In addition, a unique conformation of IDA g on DMT5 and the empty position of occasionally missing inner dyneins are indicated in light blue and gray, respectively. Scale bars are 20 nm in (A) and 10 nm in (B).
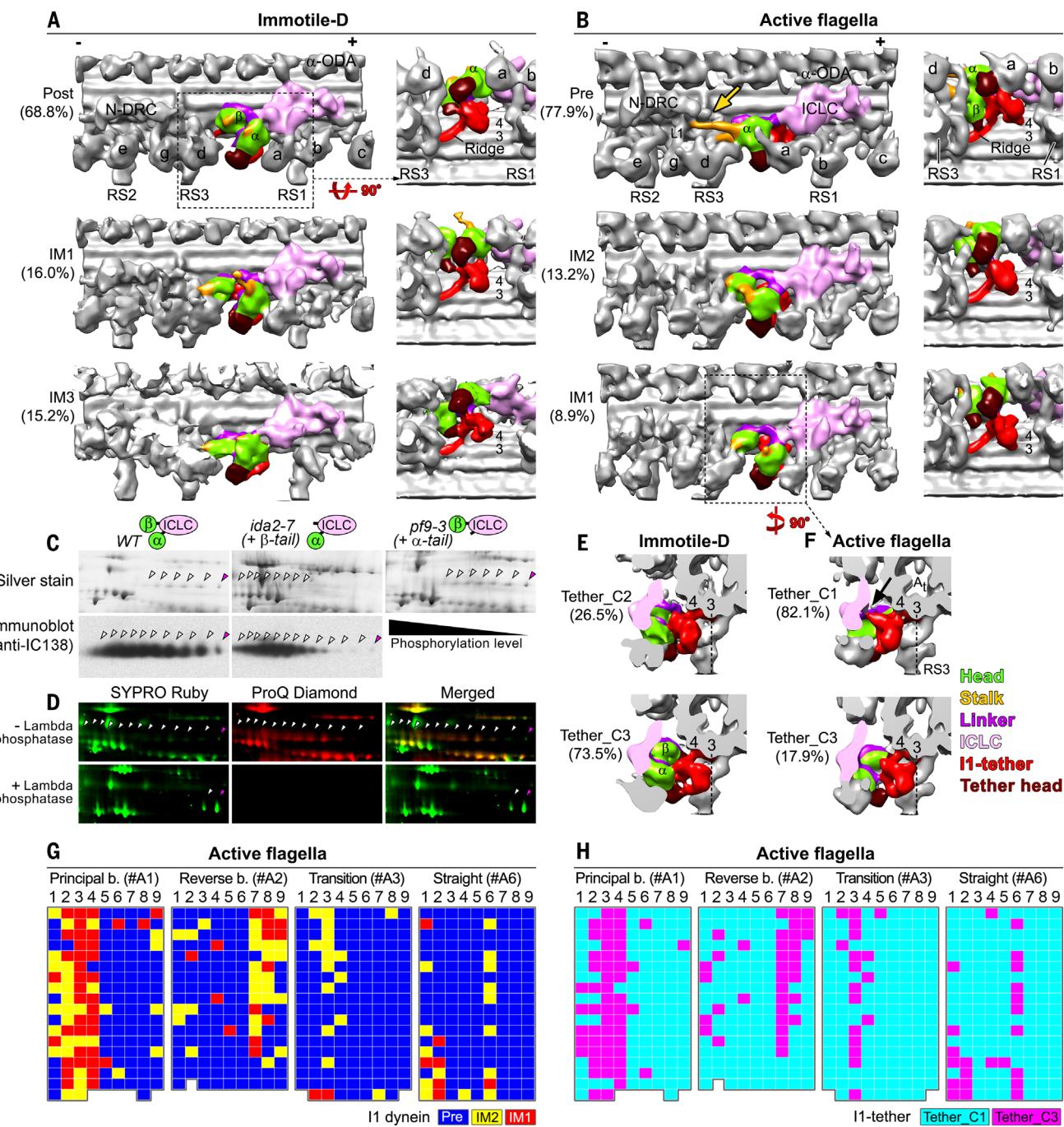
Fig. 5. The I1 dynein and I1-tether exhibit distinct and bend-correlated conformations in active flagella. (**A** and **B**) 3D isosurface renderings of the I1 dynein in the post–power stroke; intermediate IM1, IM2, and IM3; and pre–power stroke conformations in immotile-demembranated axonemes (A) and active flagella (B). Domain coloring is as shown in the legend in (F). Percentage of included subtomograms per class is indicated. a to e, and g indicate IDAs. The yellow arrow in (B) indicates the stalk of the pre–power stroke 1α-dynein. (**C**) 2D electrophoresis gels (top) and 2D immunoblots (bottom) of axonemal proteins show an increase in the phosphorylation of IC138 in *Chlamydomonas ida2-7*(+β-tail); note the higher abundance of highly phosphorylated isoforms on the acidic (left) side of the gels relative to isoforms with lower levels of phosphorylation. In WT and the *pf9-3*(+α-tail) mutant, IC138 exhibits a more evenly distributed string of spots (white and magenta arrowheads indicate the phosphorylated and nonphosphorylated isoforms of IC138, respectively).

(**D**) Identification of phosphorylated IC138 isoforms in WT *Chlamydomonas* axonemes by 2D electrophoresis analysis. The six images show the spot pattern in the absence (top row) or presence (bottom row) of lambda phosphatase. Phosphorylated isoforms (white arrowheads) were detected by both total protein stain (SYPRO Ruby) and phosphoprotein stain (ProQ Diamond), whereas nonphosphorylated isoforms (magenta arrowheads) were only detected by total protein stain. (**E** and **F**) 3D isosurface renderings show the identified conformations of the I1-tether in immotile-demembranated (E) and active (F) flagella. Note the apparent interaction between the I1-tether (red) and I1-ICLC complex (light pink) in the tether_C1 conformation in active flagella (black arrow). Dashed lines serve as a reference to highlight the positional changes of the I1-tether. (**G** and **H**) Distributions of conformations of I1 dynein (G) and I1-tether (H) in different regions of active flagella. In (A), (B), and (E) to (H), flagella are from sea urchin sperm; in (C) and (D), flagella are from *Chlamydomonas*.

isoforms than the outer dyneins. This suggests that the fundamental roles of inner and outer dyneins during ciliary motility are distinct. For some, but not all, inner dynein isoforms, we observed some small clusters of intermediate conformations that were also in straight regions immediately neighboring significant bends, here called "transition" regions (Figs. 4C and 5G). We conclude that individual inner dynein isoforms likely play different roles in ciliary motility, including initiating flagellar bending in transition regions, preceding the conformational switching of the outer dyneins (Fig. 3C). This interpretation is consistent with genetic evidence from *Chlamydomonas* that suggests that inner dyneins are critical for waveform determination, whereas outer dyneins determine the power and speed of the bending. Thus, mutant flagella lacking inner dyneins display substantially reduced bend angles (*24*), which could be a result of defective bend initiation by inner dyneins a, g, d, and I1. By contrast, *Chlamydomonas* mutant flagella lacking outer dyneins are motile without considerable changes in waveform, but they exhibit reduced beat frequency (*25*). The conformational distributions of IDAs b, c, and e lacked obvious correlation with the bending direction, making these dynein isoforms likely not essential for bend initiation. However, they could still play roles in regulating flagellar waveform.

Potential regulatory mechanisms for conformational switching

Our study provides direct visualization of the asymmetric distribution of dynein activity and conformational switching between opposite sides of a beating flagellum. However, the nature of the switching signal(s) is still unknown. Four major models have been proposed for how the regular switching of dynein activity might be regulated, including (i) the distributor model, which proposes that a series of enzymatic and mechanical interactions between regulatory complexes [e.g., CPC, radial spokes, nexin-dynein regulatory complex (N-DRC), and I1 dynein] modulate dynein activity (*23*, *26*); (ii) the geometric clutch model, which posits that bending-induced distortions of the axoneme change the spacing between DMTs, acting as a "clutch" to disengage dyneins from their DMT tracks so that they can no longer generate force (or impede interdoublet sliding) (*27*); (iii) the sliding-control model, which postulates collective dynein behavior and a positive feedback mechanism in which the activity of dyneins leads to load-accelerated dissociation of dynein motors, meaning that the force per dynein head decreases as sliding velocity increases (*4*, *28*); and (iv) the curvature-control model, which hypothesizes that dynein's activity is regulated by local axonemal curvature, in that, when the axoneme is bent to a sufficient curvature, it triggers the activation of a different set of dyneins (*29*).

Our data reveal that not only dyneins but also major regulatory complexes display conformational switching in active flagella, which supports the distributor model (*26*). For example, both the regulatory I1 dynein and its associated I1-tether

(*30*) undergo substantial conformational changes (Fig. 5, B and F) that correlate with the bend direction (Fig. 5, G and H), similar to IDAs a, g, and d.

We also observed previously undescribed attachments and dissociations between different regulators or regulators and dyneins, respectively. These conformational changes of regulators were spatially restricted in active flagella (for example, Fig. 5H), similar to specific activity states of the dyneins, and thus may represent the structural basis for transduction of regulatory signals. For example, the following four coordinated and bend direction–specific "connections" could allow for transient signal transduction between three major regulatory complexes: (i) the stalk of the pre–power stroke 1α-dynein (yellow arrow in Fig. 5B, pre, on left) connected to the proximally located nexin linker (part of the N-DRC); (ii) the stalk of pre–power stroke 1β-dynein (yellow in Fig. 5B, pre, on left) also projected proximally, but seemed to connect to IDA g and/or the base of radial spoke RS3; (iii) the head of pre–power stroke 1β-dynein (green in Fig. 5B, pre, on right) connected through the I1-tether and tether ridge (red in Fig. 5B, pre, on right) to the base of radial spoke RS3 (Fig. 5B, pre, on right, and movie S5); and (iv) the I1-tether in tether-class 1 (red in Fig. 5F, tether_C1) connected to the intermediate chain–light chain complex (ICLC, pink) of I1 dynein (Fig. 5F, arrow). The ICLC is known to be important for I1 docking and the regulatory function of I1 through its key regulatory component, the intermediate chain IC138, which is thought to be a "phospho-switch" (*31*). These spatially restricted states of regulatory complexes and their interactions could be the structural basis for the conformational switching mechanisms seen for the dyneins during flagellar motility.

To further probe the possible functional importance of the I1 dynein motor domains for the regulation of switching of dynein activity, we tested the effect of the deletion of each of the I1 motor domains on the activity state of the regulatory subunit IC138. Knockout of either of the two I1 dynein HCs in *Chlamydomonas reinhardtii* results in mutant flagella (for example, mutants *ida2-7* and *pf9-3*) that lack the entire I1 dynein, including all ICLC subunits. However, transformation of these I1 mutants with just the tail fragment of the missing HC leads to partial rescue of the structural defects, in that all I1 subunits are present, except for the motor domain of the HC, which is still missing; specifically, mutant *pf9-3*(+α-tail) lacks the 1α-head (*32*) and mutant *ida2-7*(+β-tail) lacks the 1β-head (*33*) (Fig. 5C, cartoons). We found that IC138 is hyperphosphorylated in the mutant missing the 1β-HC motor domain [*ida2-7*(+β-tail)] but not in the mutant missing the 1α-HC motor domain [*pf9-3*(+α-tail)] (Fig. 5, C and D). Hyperphosphorylated IC138 was previously described as an "off" state that results in severe motility defects and the failure to regulate microtubule sliding in vitro (*31*, *34*). This highlights the importance of the 1β-HC motor domain for the regulation of ciliary motility.

Cross sections of active flagella show that the axoneme diameter and relative position of DMTs

remain mostly the same in both bent and straight flagellar regions (Fig. 1G). However, between neighboring DMTs where the outer dyneins were in intermediate conformations in spatially restricted regions of active flagella, a small change was observed in the relative position of the DMTs, that is, here the neighboring doublet (DMT $n + 1$) was located slightly lower (toward the axoneme center) (Fig. 2N, red arrows). This subtle, bend direction–dependent change in the relative positions of neighboring DMTs might be sufficient to disengage or weaken dynein's interaction with the neighboring DMT, either by increasing interdoublet distance or by decreasing the binding affinity of dynein to the DMT as a result of a changed angle between the dynein stalk and microtubule interface. Thus, our results are also consistent with the geometric clutch model (*27*).

In "paralyzed-flagella" mutants, the outer dyneins are active and in a tug-of-war

Our observation that, at steady state, most inner and outer dyneins in active flagella were in pre–power stroke conformations strongly suggests that axonemal dyneins are ready to bind ATP and undergo their mechanochemical cycle. So far, dyneins in paralyzed mutant flagella were assumed to be in their inactive post–power stroke state, as shown for axonemes deprived of ATP. Our data, however, suggest that dyneins in paralyzed mutants with defective regulators should still exhibit (mostly) primed pre– rather than post–power stroke conformations as long as ATP is present. Paralyzed flagella would therefore be immotile not because of a lack of dynein activity but because the tug-of-war between dyneins on opposite sides of the flagella results in balanced forces, similar to straight regions of active flagella.

The protein composition (*35*) and structure of axonemes are highly conserved across species (fig. S1B). For example, the loss of subunit DRC1 from the N-DRCs or of subunit RSP1 from the radial spokes results in similar motility defects in the green algae *Chlamydomonas* as in human primary ciliary dyskinesia patients (*36*, *37*). In addition, sea urchin sperm and *Chlamydomonas* flagella show comparable ATP-consumption rates during beating (*38*), suggesting that the molecular mechanism of flagellar motility is universally conserved. Therefore, we tested our hypothesis about the activity states of dyneins in paralyzed flagella by applying cryo-ET to readily available *Chlamydomonas pf19* flagella, which are paralyzed because of the absence of the CPC, a major regulatory complex of ciliary motility (*23*, *26*).

Indeed, in *pf19* flagella, the outer dyneins were predominantly in pre–power stroke conformations (Fig. 6, A to D), similar to the outer dyneins in active wild-type (WT) *Chlamydomonas* flagella (Fig. 6, E to H) and active sea urchin sperm flagella (Fig. 2, K to M). By contrast, the outer dyneins in control flagella—that is, ATP-free isolated *Chlamydomonas* WT flagella and pseudo-WT (pWT) axonemes—were in the post–power stroke conformation (Fig. 6, I to P). Most paralyzed flagellar mutants, including *pf19*, can undergo dynein-driven sliding disintegration, that is, linear DMT
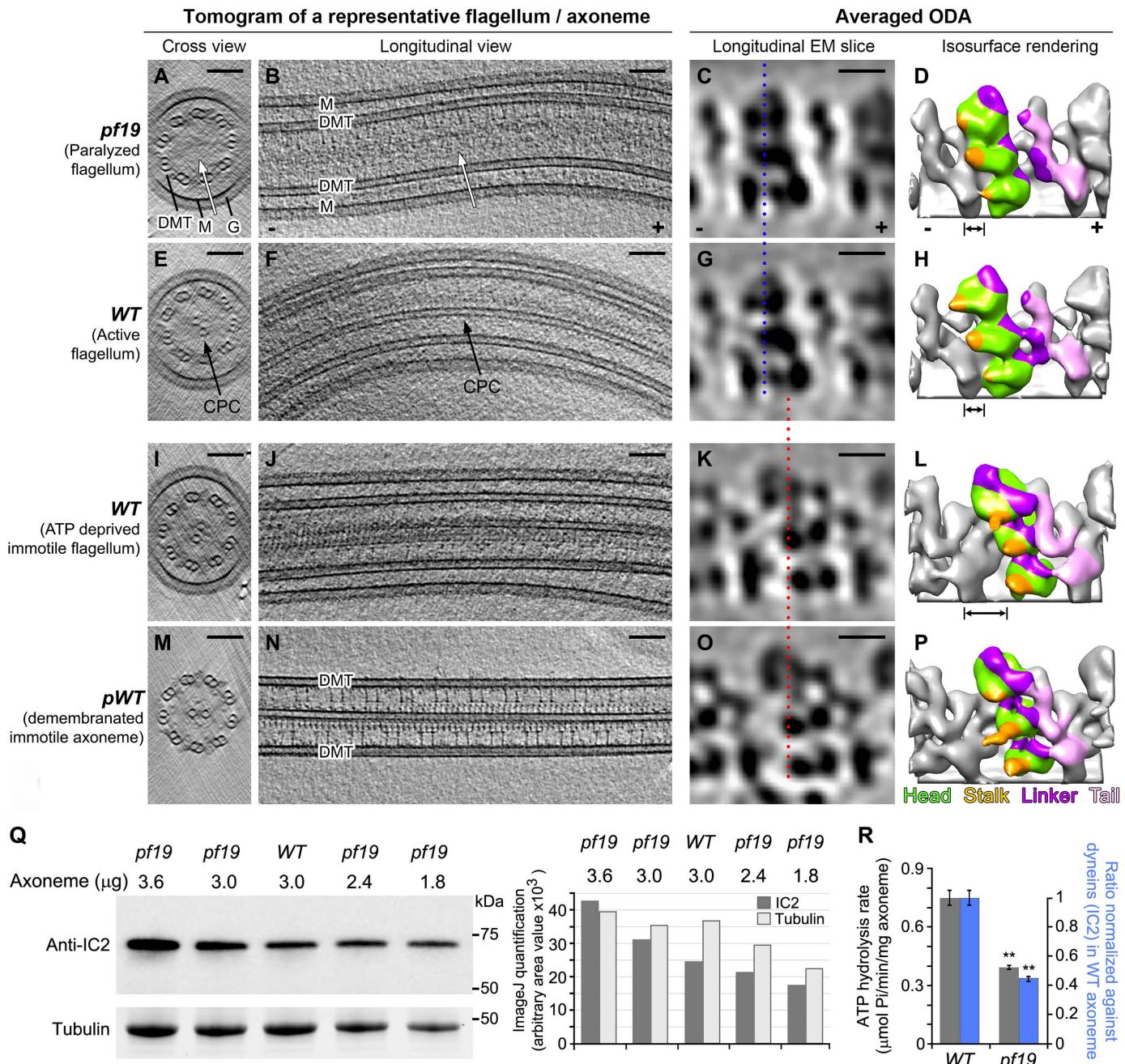
Fig. 6. Outer dyneins in paralyzed *pf19* mutant flagella from *Chlamydomonas* are in pre–power stroke conformations.

(**A** to **P**) Representative tomograms and averaged ODAs in the following *Chlamydomonas* samples: paralyzed flagella of intact *pf19* cells (white arrow indicates missing CPC) [(A) to (D)], active flagella from intact WT cells [(E) to (H)], detached immotile WT flagella without ATP [(I) to (L)], and immotile-demembranated axonemes from rescued pWT strain (*pf2-4; PF2:: GFP*) [(M) to (P)]. Note that the intact flagellar membrane (indicated by M) is surrounded by a dense glycocalyx (indicated by G) layer, resulting in thicker samples and noisier tomograms compared to demembranated axonemes in (M) and (N). The averaged outer dyneins in both paralyzed *pf19* and active WT flagella were predominantly in their pre–power stroke conformations, with the three dynein motor heads (green) and stalks (orange) closer to the microtubule minus end [(C), (D), (G), and (H)] as compared to the post–power stroke conformation observed in immotile detached and demembranated flagella [(K),

(L), (O), and (P)]. The blue and red dotted lines along the proximal edge of the bottom dynein head [in (C), (G), (K), and (O)] and the double-headed arrows [in (D), (H), and (L)] facilitate comparison of the head positions. (**Q**) Comparison of the relative abundance of the ODA intermediate chain IC2 in WT and *pf19* axoneme samples by immunoblot analysis (left) and densitometry quantification of the bands (right). The abundance of IC2 in *pf19* axonemes was 1.19 ± 0.03 times that in WT axonemes due to the missing CPC proteins in *pf19*. Tubulin is a loading control (but note that the ratio between tubulin and total protein mass is slightly changed in the mutant). (**R**) ATP hydrolysis rate of WT and *pf19* axonemes (bar chart shows means $\pm$ SD, $n = 3$; $**P < 0.01$; Student's t test). For the same amount of axonemes, the ATPase activity of *pf19* is 0.53 ± 0.02 that of WT (gray bars), and, after accounting for the amount of dyneins and normalizing against WT, the ATPase activity of *pf19* is significantly lower, 0.44 ± 0.02 that of WT (blue bars). Scale bars are 100 nm in (A), (B), (E), (F), (I), (J), (M), and (N) and 10 nm in (C), (G), (K), and (O).

extrusion, after proteolytic cleavage of interdoublet linkers and addition of ATP (*23*). This confirms that the pre–power stroke outer dyneins in *pf19* observed here require only ATP, rather than additional activating signals, to induce interdoublet sliding.

The observed prevalence of pre–power stroke dyneins in intact beating sea urchin sperm flagella, and in WT and paralyzed mutant *Chlamydomonas* flagella (Figs. 2 to 6), can be partially explained by previous kinetic studies that have indicated that, in the dynein ATPase cycle, the power stroke and release of adenosine diphosphate (ADP) (that is, the prestroke-to-poststroke transition) are the rate-limiting steps, whereas the recovery stroke (poststroke-to-prestroke transition) proceeds rapidly, about 1000 times faster (*39*). However, other mechanisms preventing the primed dyneins from completing their power stroke could also contribute. For example, previous optical trap experiments demonstrated that dynein's detachment rate decreases with increasing opposing load, that is, dynein binds tighter to microtubules when a force is applied on the dynein tail that is directed toward the plus end of the microtubule (*40*, *41*). In flagella, the activity of dyneins on opposite sides of the axoneme would cause opposing directions of interdoublet sliding, which would be perceived as a load increase (a force exerted on the dynein tails directed toward the microtubule plus end). Thus, simultaneously active dyneins on opposite sides of the axoneme could cause tighter microtubule binding or even catch bond of these dyneins. The tension applied by means of the tail domain and allosteric interhead communication between neighboring motor domains of the dimeric ODAs could prevent the primed dyneins from completing their power stroke (*7*).

The prevalence of pre–power stroke dyneins in beating and paralyzed mutant flagella (Figs. 2 to 6), together with the dynein distribution patterns that agree well with the prevailing notion of bend generation (fig. S2D) and the switch-point hypothesis (*13*), strongly suggests that, at steady state, dyneins in cilia are in an active state by default and ready to undergo a mechanochemical cycle, without requiring specific activating signals, as long as ATP is present. These direct structural findings are consistent with previous biochemical studies, such as DMT extrusion from protease-treated WT and mutant axonemes, and motility assays of isolated dyneins (*12*, *22*, *23*). It therefore follows that, for flagellar bending to occur, proper regulatory signals are required to reduce dynein activity on one side of the flagellum to generate the force imbalance required to bend the flagellum. In the absence of such inhibitory signals, the dynein forces generated

on opposite sides are balanced, resulting in "rigor" paralyzed flagella.

Although the outer dyneins in *pf19* flagella were predominantly in pre–power stroke conformations, we hypothesized that the tug-of-war between opposite sides of the flagellum could prevent the dynein power stroke and keep many dyneins in their primed pre-II states, thus preventing futile ATP hydrolysis. This is supported by our comparative measurement of ATP consumption by isolated axonemes from

Movie 1. Switch-inhibition mechanism. The animation summarizes the switch-inhibition mechanism of ciliary and flagellar beating, as suggested by our cryo-ET study of active sea urchin sperm flagella. Note that the depicted durations of different conformational states of dyneins do not reflect the native kinetics of the molecules.

WT and *Chlamydomonas pf19* flagella; the ATP hydrolysis rate of *pf19* axonemes is less than half (44%) that of WT axonemes (Fig. 6, Q and R).

"Switch-inhibition" model

Our direct analysis of the spatial regulation of dynein activity revealed three tenets that are important for dynein-driven flagellar motility: (i) The asymmetric distribution of dynein activity results in unidirectional bending, (ii) the switching of dynein conformations between opposite sides of the flagellum causes the undulating waveform of beating flagella, and (iii) at steady state, dyneins are in the active pre–power stroke states, and the asymmetry of dynein activity is most likely caused by selectively inhibiting dyneins on one side of the flagellum. Given that bends are usually initiated at the flagellar base and travel over time to the tip, the spatially restricted and cooperative conformational changes seen along frozen flagella likely also represent the temporal sequence of structural changes during wave formation. By correlating structural changes with their locations in the flagellar bend, we are able to propose the following switch-inhibition model, in which ciliary and flagellar motility is generated by inhibiting, rather than activating, dyneins on alternating sides of the flagellum (Fig. 7 and Movie 1):

1) Straight regions: Under physiological conditions (ATP present) and with the regulators

in their off state or the inhibitory signal still below a threshold, the dynein forces on opposite sides of the flagellum are counterbalanced, keeping the flagellum straight (Figs. 7, third cross section from the top, and 3C, straight). Without proper function of regulators that can "break" the force balance, such as what occurs in the *pf19* mutant, flagella appear paralyzed and stiff.

2) Bend initiation: Regulatory signals lead first to the inhibition of inner dyneins (for example, dyneins a, g, d, and I1) either on DMTs 2 to 4 (for principal bend) or DMTs 7 to 9 (for reverse bend) (Figs. 7, fourth cross section from the top, and 4C and 5G, transition). When enough inner dyneins are inhibited on one side, dyneins on the opposite side can begin to exert a net force sufficiently strong enough to initiate a mild bend in the flagellum.

3) Bending: The positional changes of DMTs caused by the mild bend initiation and/or transduction of the inhibitory signal downstream to the outer dyneins on the inhibited side of the axoneme allow for a rapid increase of net bending force generated by the active dyneins on the opposite side of the flagellum. This causes further interdoublet sliding, which is restricted by the basal body and molecular interdoublet links, causing a full bend of the flagellum in one direction, that is, the principal or the reverse bend (Figs. 7, first and fifth cross sections, and 3C, principal and reverse bend regions of active flagella).

After the inhibitory signals are turned off by as-yet-unknown trigger(s), dynein activity on the previously inhibited side recovers, causing the flagellum to straighten (Figs. 7, second cross section from the top, and 4C and 5G, transition) and the bending cycle restarts. By switching the side of dynein inhibition in regular time intervals, the flagellum alternates the direction of bending, resulting in the typical planar waveform of sea urchin sperm flagella.

Switch inhibition is a robust mechanism

Although the original switch-point hypothesis (*13*) did not specify the regulatory mechanism that underlies the asymmetric distribution of dynein activity, data interpretation in the field focused mainly on active sliding and dynein activation. Our switch-inhibition model provides a comprehensive molecular mechanism for flagellar beat generation that appears more robust than the previously assumed switch-activation model. Flagella beat over a long time with fairly consistent waveform, amplitude, and beat frequency of 50 to 60 Hz, that is, the bending direction switches rapidly every ~10 ms. However, as with a switch-activation mechanism, the regulators would have to activate dyneins in a spatiotemporally coordinated manner with high efficiency,

and the bend amplitude would directly depend on the number of synchronously activated dyneins.

By contrast, a switch-inhibition mechanism could happen in an "all-or-none" fashion. Indeed, the spatially restricted inhibition of just a fraction of the dyneins would be sufficient to cause cooperative microtubule release of all dyneins on the inhibited flagellar side, allowing the predominantly active dyneins on the opposite side to rapidly drive bending. The advantage of the all-or-none principle is that the strength of a response can be independent of the strength of the input signal. As long as the inhibitory signal exceeds a specific threshold (for example, a minimum number of inhibited dyneins), the flagellum would bend consistently with 100% wave amplitude, resulting in the relatively constant and sustained beating of cilia and flagella. All-or-none mechanisms are often observed in cellular processes that require rapid switching and/or preservation of signal strength, such as transduction of action potentials and heart-muscle contraction (42, 43). Our findings will guide future interpretations and shift the field's focus to molecular mechanisms underpinning the spatio-temporally coordinated inactivation of dyneins.

Materials and methods
Sample preparation

Spawning of male adult sea urchins (*Strongylocentrotus purpuratus* purchased from Monterey Abalone, Monterey, CA) was induced by the injection of 1 to 2 ml of 0.5 M KCl into the perivisceral cavity. Sperm samples were collected and a small aliquot was diluted in artificial seawater (360 mM NaCl, 50 mM MgCl₂, 10 mM CaCl₂, 10 mM KCl, and 30 mM HEPES, pH 8.0) to examine the motility by light microscopy using the differential interference contrast (DIC) mode of a Marianas spinning disk confocal system (3I, Denver, CO) consisting of a Zeiss Axio Observer Z1 microscope (Carl Zeiss, Jena, Germany) equipped with a Yokogawa CSU-X1 spinning disk confocal head (Yokogawa, Tokyo, Japan) and a QuantEM 512SC EMCCD camera (Photometrics, Tucson, AZ). All harvested sperm cells were motile (movie S1), and the samples were then divided to prepare three different types of samples: (part A) active flagella, (part B) ATPase-inhibited immotile flagella (immotile-I), and (part C) demembranated immotile axonemes (immotile-D). Part A was diluted in artificial seawater and rapidly frozen (as described below). Part B was diluted in artificial seawater containing the ATPase inhibitor erythro-9-(2-hydroxy-3-nonyl)adenine hydrochloride (EHNA hydrochloride, 2 mM; Santa Cruz Biotechnology) (44); after incubation for 5 min, we confirmed by light microscopy that the sperm were completely immotile (movie S2) and then rapidly froze the sample. Part C was diluted in demembranation buffer (30 mM HEPES pH 8.0, 150 mM KCl, 4 mM MgCl₂, 0.5 mM EGTA, and 0.1% Triton X-100) to remove the flagellar membrane (but axonemes remained attached to the cell body). After incubation for 1 min, the sperm were collected by centrifugation at 1000g, resuspended in demembranation buffer (but without Triton X-100), and rapidly frozen.
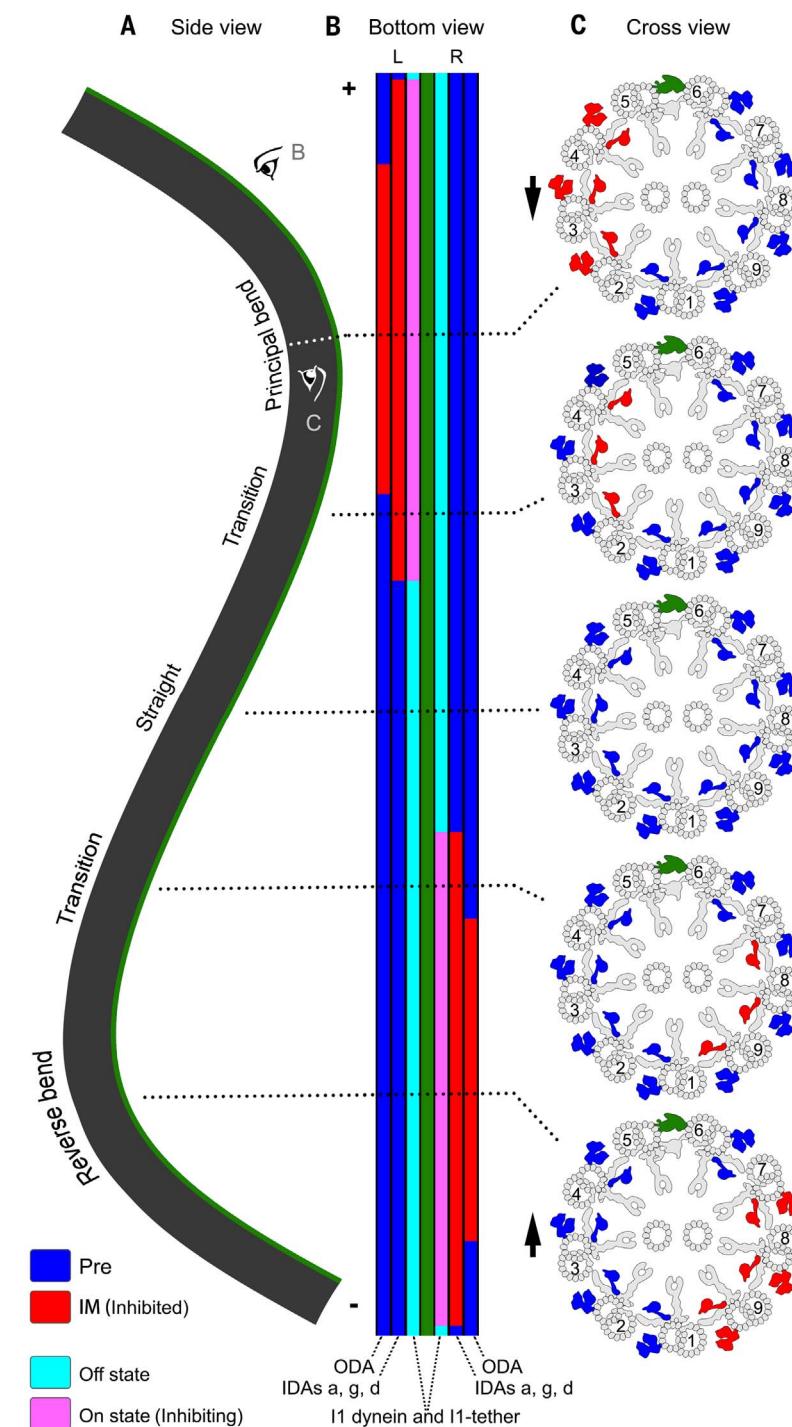
Fig. 7. Schematic model of the switch-inhibition mechanism of ciliary and flagellar motility. (**A** to **C**) Summary illustration of a sinusoidal wave of an active flagellum that contains reverse bend, transition, straight, and principal bend regions. The schematic shows the flagellum in different views: longitudinal views from its "left" side (where DMTs 2 to 4 are located) (A), as viewed from the bottom passing through the o-SUB5-6 bridge and DMT1 (L and R indicate the left and right sides of the flagellum) (B), and cross-sectional views from the proximal end (C). The view directions of (B) and (C) are indicated in (A) by their corresponding letters. In (B) and (C), the distributions of the different activity states of the dyneins and the I1-tether (as indicated below the bottom view) are shown in the different functional regions of the flagellum. The different states are indicated by distinct colors (as specified in the color legend at the bottom left). Note that the flagellar wave travels from the proximal base (minus end) to the distal tip (plus end) of the flagellum; this means that the functional state shown, for example, in the fourth cross section from the top (transition and bend initiation) precedes the state shown in the fifth cross section (reverse bend) in time.

Chlamydomonas reinhardtii strains used in this study include WT (wild-type strain: CC-125, 137c mt+), pWT (a pseudo-WT strain: *pf2-4; PF2::GFP*) (*9*), a CPC-lacking mutant *pf19* (strain: cc-1037 mt+) (*23*), and two I1 dynein mutant strains: *pf9-3*(+α-tail) (strain: *pf9-3; pf9-3::G41a*) and *ida2-7*(+β-tail) (strain: *ida2-7; ida2-7::pCAP1*). The *pf9-3*(+α-tail) and *ida2-7*(+β-tail) strains lack the motor domains of 1α- and 1β-HC, respectively (*32, 33*). *Chlamydomonas* cells were grown in liquid Tris acetate-phosphate medium at room temperature with a light:dark cycle of 16:8 hours. Axonemes were isolated from *Chlamydomonas reinhardtii* cells as previously described (*30*). Briefly, flagella were detached from the cells using the pH-shock method and purified by two centrifugation steps over 20% sucrose cushions. Purified flagella were demembranated with 1% IGEPAL CA-630 (Sigma-Aldrich, St. Louis, MO), and axonemes were collected by centrifugation at 10,000g for 10 min. Except for WT and *pf19* axonemes that were used for an ATPase assay and pWT axonemes that were used for cryo-ET analysis, the axoneme pellet was directly dissolved in two-dimensional electrophoresis (2DE) lysis buffer [7 M urea, 2 M thiourea, 4% (wt/wt) CHAPS, 65 mM DTT, and 2% (vol/vol) IPG buffer (pH 3-10NL; GE Healthcare)] by vigorously stirring for 0.5 hours. Cell debris and insoluble material were removed by centrifugation at 45,000g for 1 hour. The supernatant was aliquoted and stored at −70°C until subsequent analysis.

For the phosphorylation analysis of IC138, the protein samples dissolved in 2DE lysis buffer were precipitated using the 2-D Clean-Up Kit (GE Healthcare) and resuspended in Milli-Q water to approximately 4 mg/ml. Phosphatase treatment using Lambda Protein Phosphatase was performed as previously described (*45*). Briefly, two 85-μl aliquots of the protein sample were mixed with 10 μl of 10% SDS and were vigorously vortexed for 20 s, followed by the addition of 695 μl of Milli-Q water, 100 μl of 10 mM MnCl$_2$, 100 μl of 10× Lambda Protein Phosphatase buffer (New England Biolabs), and 10 μl of protease inhibitor cocktail (P9599, Sigma-Aldrich). To one of the two aliquots, 1200 units of Lambda Protein Phosphatase (New England Biolabs) was added, and both aliquots were then incubated overnight at 30°C. Phosphatase-treated and untreated protein samples were precipitated using acetone (−20°C), resuspended in 2DE lysis buffer, further purified using the 2-D Clean-Up Kit (GE Healthcare), and resuspended in 2DE lysis buffer to a final concentration of approximately 4 mg/ml.

Cryo-sample preparation and cryo-ET

Cryo-samples were prepared, imaged by cryo-ET, and processed as previously described (*17*). Briefly, Quantifoil holey carbon grids (Quantifoil Micro Tools GmbH, Germany) were glow discharged, coated with 10-nm gold (Sigma-Aldrich), and loaded onto a homemade plunge-freezing device. Three microliters of sample [that is, actively swimming sea urchin sperm cells, *Chlamydomonas* WT cells, immotile sea urchin sperm cells (demembranated or treated with EHNA), paralyzed

Chlamydomonas pf19 cells, isolated *Chlamydomonas* WT flagella or pWT axonemes, respectively] and 1 μl of fivefold concentrated BSA coated 10-nm colloidal gold solution were applied to the grid, blotted with filter paper for 1.5 to 2.5 s, and immediately plunge-frozen in liquid ethane. Vitrified specimens were transferred into a Tecnai F30 transmission electron microscope (FEI, Hillsboro, OR) with a cryo-holder (Gatan, Pleasanton, CA). Flagella and/or axonemes that appeared well preserved by EM inspection were imaged at 300 keV, with −6- or −8-μm defocus, under low-dose conditions and using an energy filter in zero-loss mode (Gatan, Pleasanton, CA) (20-eV slit width). Tilt series were recorded while stepwise rotating the sample from about −65° to +65° with 1.5° to 2.5° increments using the microscope control software SerialEM (*46*). The cumulative electron dose per tilt series was limited to approximately 100 e/Å^2. All images were digitally recorded on a 2k×2k charge-coupled device camera (Gatan, Pleasanton, CA) at a nominal magnification of 13,500, resulting in a pixel size of approximately 1 nm.

Image processing

The tilt series images were reconstructed into 3D tomograms by weighted back projection using the IMOD software package (*47*). Some tomograms were previously utilized for the analysis of axonemal dyneins (*17*). Only tomograms of intact and noncompressed flagella and/or axonemes were used for further data analysis. To enhance the signal-to-noise ratio and improve the resolution, subtomograms that contained the 96-nm axonemal repeat units along the doublet microtubules (volume size: 110 nm by 84 nm by 80 nm) or that contained individual ODAs (volume size: 56 nm by 56 nm by 56 nm) were extracted from the raw tomograms, aligned, and averaged (including missing wedge compensation) using the PEET program to obtain subtomogram averages (*18*). To identify distinct conformations of various axonemal structures, classification analyses were performed on the aligned subtomograms using a clustering approach (principal component analysis) built into the PEET program (*19*). Before classification, appropriate masks were applied to focus the classification on structures of interest. Subtomograms that contained the structure of interest with an identical conformation were grouped into a class, and were averaged to generate a class average. The automatic classification into different conformational states of axonemal complexes was performed by an automated algorithm without prior knowledge about from which functional region of a flagellar wave individual subtomograms were extracted. Using the information provided by the classification analysis, we then mapped the conformational state of each subtomogram and/or repeat back onto the respective location in the raw tomograms. The numbers of tomograms and subtomograms analyzed by classification are summarized in table S1. The resolution of the resulting averages was estimated in a (30 nm)^3 subvolume in the center of the structure of interest using the Fourier shell

correlation method with a criterion of 0.5 (*48*). The structures were visualized as 2D tomographic slices and 3D isosurface renderings using IMOD (*47*) and UCSF Chimera (*49*), respectively.

Electrophoresis and phosphorylation analysis

2DE analysis was performed as previously described (*45*). Briefly, *Chlamydomonas* axonemal proteins (70 μg) were separated in the first dimension on 13-cm immobilized pH 3-10NL IPG strips (GE Healthcare) for 24 kVh, followed by 10% SDS-PAGE for the second dimension. All samples (two to seven samples for each strain) were run in at least duplicate to confirm reproducibility. To visualize total proteins, the gels were stained with silver nitrate; to visualize phosphoproteins, the gels were stained with Pro-Q Diamond Phosphoprotein Gel Stain (Thermo Fisher Scientific) according to the manufacturer's instructions. After image acquisition using a Typhoon 9410 Variable Mode Imager (GE Healthcare), the Pro-Q Diamond-stained gels were poststained with SYPRO Ruby Protein Gel Stain (Thermo Fisher Scientific) to detect total protein.

Immunoblot

For 2D immunoblot analysis of IC138, 35 μg of total axonemal proteins was separated by 2DE with 7-cm immobilized pH 3-10NL IPG strips (GE Healthcare) for the first dimension and 10% SDS-PAGE gels for the second dimension. Immunoblot analysis was performed using a polyclonal IC138 antibody (1:10,000) (*31*). Signals were visualized using the ECL detection system (Bio-rad).

For immunoblot analysis of ODA IC2 protein, the axonemal proteins were resolved by SDS-PAGE on Any kD TGX Strain-Free™ protein gels (Bio-rad Cat# 4568123). After visualizing by ChemiDoc™ Touch Imaging System (Bio-Rad), proteins were blotted onto PVDF membranes (Bio-Rad) and probed with anti-IC2 monoclonal antibody (Sigma-Aldrich, D6168; 1:5000). Signals were visualized using the ECL detection system (Bio-rad), and quantified by ImageJ software.

ATPase assay and analysis of relative ATP consumption

The rate of phosphate release by ATP hydrolysis of axonemes was measured in bulk as described before (*50*). Briefly, isolated axonemes were washed and resuspended in HMEEK buffer (30 mM Hepes, pH 7.4, 5 mM MgSO$_4$, 1 mM EGTA, 0.1 mM EDTA, and 25 mM KCl). 150 μl of axonemes (10.3 μg) were incubated with 1.3 μl of 50 mM ATP (resulting in 0.43 mM ATP in solution) for 1 min. The ATP hydrolysis was stopped by adding 150 μl of 12% SDS solution. Color was developed by incubation with 300 μl of a 1:1 solution of 6% ascorbic acid in 1 N HCl and 1% (NH$_4$)$_6$Mo$_7$O$_{24}$.4H$_2$O in 12% SDS for 10 min and was stopped by adding 450 μl of 2% sodium citrate, 2% NaAsO$_2$, and 2% acetic acid followed by 20 min incubation at room temperature. The concentrations of the released phosphate were calculated from the color absorbance at 850 nm. The relative abundance of the ODA intermediate

chain protein IC2 in WT and *pf19* axoneme was used to correlate total protein amount to number of dynein molecules (more dyneins per total protein in *pf19* because of lack of CPC). The densitometry quantification of the IC2 Western blot bands was performed by ImageJ software (Fig. 6Q). On the basis of analyses of three independent samples, the abundance of IC2 in *pf19* axoneme was 1.19 ± 0.03 times higher than that in WT axonemes (Fig. 6Q). For the same amount of axonemes, the ATPase activity of *pf19* is 0.53 ± 0.02 of that of WT. As the abundance of dyneins in *pf19* axonemes is 1.19 times that in WT axonemes, the ATPase activity of *pf19* is 0.44 ± 0.02 of that of WT after accounting for the amount of dyneins and normalizing against WT (Fig. 6R). The data were analyzed by Student's *t* test.

REFERENCES AND NOTES

1. Cilia and flagella share the same conserved structure and differ only in waveform and/or length; therefore, the terms "cilia" and "flagella" are often used interchangeably, including in this paper.
2. H. M. Mitchison, E. M. Valente, Motile and non-motile cilia in human pathology: From function to phenotypes. *J. Pathol.* **241**, 294–309 (2017). doi: 10.1016/j.jmb.2012.05.040; pmid: 22683354
3. N. Mizuno, M. Taschner, B. D. Engel, E. Lorentzen, Structural studies of ciliary components. *J. Mol. Biol.* **422**, 163–180 (2012). doi: 10.1016/j.jmb.2012.05.040; pmid: 22683354
4. T. J. Mitchison, H. M. Mitchison, Cell biology: How cilia beat. *Nature* **463**, 308–309 (2010). doi: 10.1038/463308a; pmid: 20090745
5. M. Kikkawa, Big steps toward understanding dynein. *J. Cell Biol.* **202**, 15–23 (2013). doi: 10.1083/jcb.201304099; pmid: 23836927
6. I. R. Gibbons, A. J. Rowe, Dynein: A protein with adenosine triphosphatase activity from cilia. *Science* **149**, 424–426 (1965). doi: 10.1126/science.149.3682.424; pmid: 17809406
7. H. Schmidt, A. P. Carter, Review: Structure and mechanism of the dynein motor ATPase. *Biopolymers* **105**, 557–567 (2016). doi: 10.1002/bip.22856; pmid: 27062277
8. P. Höök, R. B. Vallee, The dynein family at a glance. *J. Cell Sci.* **119**, 4369–4371 (2006). doi: 10.1242/jcs.03176; pmid: 17074830
9. J. Lin, T. Heuser, K. Song, X. Fu, D. Nicastro, One of the nine doublet microtubules of eukaryotic flagella exhibits unique and partially conserved structures. *PLOS ONE* **7**, e46494 (2012). doi: 10.1371/journal.pone.0046494; pmid: 23071579
10. R. Kamiya, T. Yagi, Functional diversity of axonemal dyneins as assessed by in vitro and in vivo motility assays of *Chlamydomonas* mutants. *Zool. Sci.* **31**, 633–644 (2014). doi: 10.2108/zs140066; pmid: 25284382
11. W. S. Sale, P. Satir, Direction of active sliding of microtubules in *Tetrahymena* cilia. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2045–2049 (1977). doi: 10.1073/pnas.74.5.2045; pmid: 266725
12. K. E. Summers, I. R. Gibbons, Adenosine triphosphate-induced sliding of tubules in trypsin-treated flagella of sea-urchin sperm. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 3092–3096 (1971). doi: 10.1073/pnas.68.12.3092; pmid: 5289252
13. J. Wais-Steider, P. Satir, Effect of vanadate on gill cilia: Switching mechanism in ciliary beat. *J. Supramol. Struct.* **11**, 339–347 (1979). doi: 10.1002/jss.400110309; pmid: 120905
14. S. Hayashi, C. Shingyoji, Mechanism of flagellar oscillation-bending-induced switching of dynein activity in elastase-treated axonemes of sea urchin sperm. *J. Cell Sci.* **121**, 2833–2843 (2008). doi: 10.1242/jcs.031195; pmid: 18682495
15. W. S. Sale, The axonemal axis and Ca²⁺-induced asymmetry of active microtubule sliding in sea urchin sperm tails. *J. Cell Biol.* **102**, 2042–2052 (1986). doi: 10.1083/jcb.102.6.2042; pmid: 2940250
16. S. L. Tamm, S. Tamm, Alternate patterns of doublet microtubule sliding in ATP-disintegrated macrocilia of the ctenophore *Beroë*. *J. Cell Biol.* **99**, 1364–1371 (1984). doi: 10.1083/jcb.99.4.1364; pmid: 6480696
17. J. Lin, K. Okada, M. Raytchev, M. C. Smith, D. Nicastro, Structural mechanism of the dynein power stroke. *Nat. Cell Biol.* **16**, 479–485 (2014). doi: 10.1038/ncb2939; pmid: 24727830

18. D. Nicastro et al., The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944–948 (2006). doi: 10.1126/science.1128618; pmid: 16917055
19. J. M. Heumann, A. Hoenger, D. N. Mastronarde, Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J. Struct. Biol.* **175**, 288–299 (2011). doi: 10.1016/j.jsb.2011.05.011; pmid: 21616153
20. K. Zhang et al., Cryo-EM reveals how human cytoplasmic dynein is auto-inhibited and activated. *Cell* **169**, 1303–1314.e18 (2017). doi: 10.1016/j.cell.2017.05.025; pmid: 28602352
21. B. I. Carbajal-González et al., Conserved structural motifs in the central pair complex of eukaryotic flagella. *Cytoskeleton (Hoboken)* **70**, 101–120 (2013). doi: 10.1002/cm.21094; pmid: 23281266
22. T. Kon, T. Mogami, R. Ohkura, M. Nishiura, K. Sutoh, ATP hydrolysis cycle-dependent tail motions in cytoplasmic dynein. *Nat. Struct. Mol. Biol.* **12**, 513–519 (2005). doi: 10.1038/nsmb930; pmid: 15880123
23. G. B. Witman, J. Plummer, G. Sander, *Chlamydomonas* flagellar mutants lacking radial spokes and central tubules. Structure, composition, and function of specific axonemal components. *J. Cell Biol.* **76**, 729–747 (1978). doi: 10.1083/jcb.76.3.729; pmid: 632325
24. C. J. Brokaw, R. Kamiya, Bending patterns of *Chlamydomonas* flagella: IV. Mutants with defects in inner and outer dynein arms indicate differences in dynein arm function. *Cell Motil. Cytoskeleton* **8**, 68–75 (1987). doi: 10.1002/cm.970080110; pmid: 2958145
25. D. R. Mitchell, J. L. Rosenbaum, A motile *Chlamydomonas* flagellar mutant that lacks outer dynein arms. *J. Cell Biol.* **100**, 1228–1234 (1985). doi: 10.1083/jcb.100.4.1228; pmid: 3156867
26. E. F. Smith, P. Yang, The radial spokes and central apparatus: Mechano-chemical transducers that regulate flagellar motility. *Cell Motil. Cytoskeleton* **57**, 8–17 (2004). doi: 10.1002/cm.10155; pmid: 14648553
27. C. B. Lindemann, A "geometric clutch" hypothesis to explain oscillations of the axoneme of cilia and flagella. *J. Theor. Biol.* **168**, 175–189 (1994). doi: 10.1006/jtbi.1994.1097
28. I. H. Riedel-Kruse, A. Hilfinger, J. Howard, F. Jülicher, How molecular motors shape the flagellar beat. *HFSP J.* **1**, 192–208 (2007). doi: 10.2976/1.2773861; pmid: 19404446
29. P. Sartori, V. F. Geyer, A. Scholich, F. Jülicher, J. Howard, Dynamic curvature regulation accounts for the symmetric and asymmetric beats of *Chlamydomonas* flagella. *eLife* **5**, e13258 (2016). doi: 10.7554/eLife.13258; pmid: 27166516
30. T. Heuser et al., Cryoelectron tomography reveals doublet-specific structures and unique interactions in the I1 dynein. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2067–E2076 (2012). doi: 10.1073/pnas.1120690109; pmid: 22733763
31. R. Bower et al., IC138 defines a subdomain at the base of the I1 dynein that regulates microtubule sliding and flagellar motility. *Mol. Biol. Cell* **20**, 3055–3063 (2009). doi: 10.1091/mbc.E09-04-0277; pmid: 19420135
32. S. H. Myster, J. A. Knott, K. M. Wysocki, E. O'Toole, M. E. Porter, Domains in the 1α dynein heavy chain required for inner arm assembly and flagellar motility in Chlamydomonas. *J. Cell Biol.* **146**, 801–818 (1999). doi: 10.1083/jcb.146.4.801; pmid: 10459015
33. C. A. Perrone, S. H. Myster, R. Bower, E. T. O'Toole, M. E. Porter, Insights into the structural organization of the I1 inner arm dynein from a domain analysis of the 1β dynein heavy chain. *Mol. Biol. Cell* **11**, 2297–2313 (2000). doi: 10.1091/mbc.11.7.2297; pmid: 10888669
34. S. Toba et al., Distinct roles of 1α and 1β heavy chains of the inner arm dynein I1 of *Chlamydomonas* flagella. *Mol. Biol. Cell* **22**, 342–353 (2011). doi: 10.1091/mbc.E10-10-0806; pmid: 21148301
35. G. J. Pazour, N. Agrin, J. Leszyk, G. B. Witman, Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170**, 103–113 (2005). doi: 10.1083/jcb.200504008; pmid: 15998802
36. M. Wirschell et al., The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. *Nat. Genet.* **45**, 262–268 (2013). doi: 10.1038/ng.2533; pmid: 23354437
37. J. Lin et al., Cryo-electron tomography reveals ciliary defects underlying human RSPH1 primary ciliary dyskinesia. *Nat. Commun.* **5**, 5727 (2014). doi: 10.1038/ncomms6727; pmid: 25473808
38. D. T. Chen, M. Heymann, S. Fraden, D. Nicastro, Z. Dogic, ATP consumption of eukaryotic flagella measured at a single-cell level. *Biophys. J.* **109**, 2562–2573 (2015). doi: 10.1016/j.bpj.2015.11.003; pmid: 26682814
39. T. Mogami, T. Kon, K. Ito, K. Sutoh, Kinetic characterization of tail swing steps in the ATPase cycle of *Dictyostelium* cytoplasmic dynein. *J. Biol. Chem.* **282**, 21639–21644 (2007). doi: 10.1074/jbc.M701914200; pmid: 17548361

40. B. J. Reddy et al., Load-induced enhancement of Dynein force production by LIS1-NudE in vivo and in vitro. *Nat. Commun.* **7**, 12259 (2016). doi: 10.1038/ncomms12259; pmid: 27489054
41. C. Leidel, R. A. Longoria, F. M. Gutierrez, G. T. Shubeita, Measuring molecular motor forces in vivo: Implications for tug-of-war models of bidirectional transport. *Biophys. J.* **103**, 492–500 (2012). doi: 10.1016/j.bpj.2012.06.038; pmid: 22947865
42. M. Wacke, G. Thiel, Electrically triggered all-or-none Ca²⁺-liberation during action potential in the giant alga *Chara*. *J. Gen. Physiol.* **118**, 11–22 (2001). doi: 10.1085/jgp.118.1.11; pmid: 11429441
43. J. Weiss, Y. Goldman, M. Morad, Electromechanical properties of the single cell-layered heart of tunicate *Boltenia ovifera* (sea potato). *J. Gen. Physiol.* **68**, 503–518 (1976). doi: 10.1085/jgp.68.5.503; pmid: 993769
44. P. Bouchard, S. M. Penningroth, A. Cheung, C. Gagnon, C. W. Bardin, erythro-9-[3-(2-Hydroxynonyl)]adenine is an inhibitor of sperm motility that blocks dynein ATPase and protein carboxylmethylase activities. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1033–1036 (1981). doi: 10.1073/pnas.78.2.1033; pmid: 6453342
45. J. Lin et al., Building blocks of the nexin-dynein regulatory complex in *Chlamydomonas* flagella. *J. Biol. Chem.* **286**, 29175–29191 (2011). doi: 10.1074/jbc.M111.241760; pmid: 21700706
46. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). doi: 10.1016/j.jsb.2005.07.007; pmid: 16182563
47. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996). doi: 10.1006/jsbi.1996.0013; pmid: 8742726
48. G. Harauz, M. Van Heel, Exact filters for general geometry three dimensional reconstruction. *Optik (Stuttg.)* **73**, 146–156 (1986).
49. E. F. Pettersen et al., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
50. S. Chifflet, A. Torriglia, R. Chiesa, S. Tolosa, A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: Application to lens ATPases. *Anal. Biochem.* **168**, 1–4 (1988). doi: 10.1016/0003-2697(88)90002-4; pmid: 2834977
51. A. J. Roberts, T. Kon, P. J. Knight, K. Sutoh, S. A. Burgess, Functions and mechanics of dynein motor proteins. *Nat. Rev. Mol. Cell Biol.* **14**, 713–726 (2013). doi: 10.1038/nrm3667; pmid: 24064538

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6387/eaar1968/suppl/DC1
Figs. S1 to S5
Tables S1 and S2
References (*52–54*)
Movies S1 to S5

13 October 2017; accepted 20 February 2018
10.1126/science.aar1968

RESEARCH ARTICLE SUMMARY

BIOMECHANICS

The principles of cascading power limits in small, fast biological and engineered systems

Mark Ilton, M. Saad Bhamla,* Xiaotian Ma,* Suzanne M. Cox,* Leah L. Fitchett, Yongjin Kim, Je-sung Koh, Deepak Krishnamurthy, Chi-Yun Kuo, Fatma Zeynep Temel, Alfred J. Crosby, Manu Prakash, Gregory P. Sutton, Robert J. Wood, Emanuel Azizi,† Sarah Bergbreiter,† S. N. Patek†‡

INTRODUCTION: Mechanical power, whether for launched missiles or running humans, is limited by the universal, physical trade-off between force and velocity. However, many biological systems use power-amplifying mechanisms that enable unmatched accelerations in challenging environments and across a wide range of size scales. How these mechanisms actually enhance power output remains unclear. Power-amplified biological systems are of particular interest because they achieve a trio of combined capabilities that exceed current engineering performance: (i) high accelerations that (ii) can be continuously fueled through metabolic processes and (iii) are used repeatedly with minimal performance degradation throughout the life of the organism. Although engineers have struggled to design lightweight and long-lasting devices that can deliver high power output, biological systems have been performing such feats for millions of years and using these systems for a myriad of functions.

Size range of power amplified systems
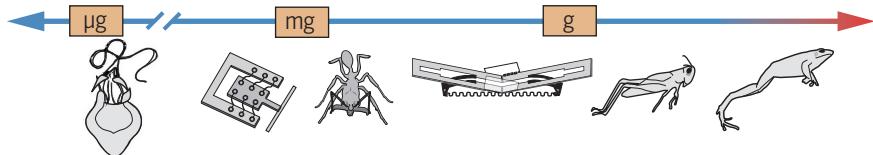
Scaling of power delivery
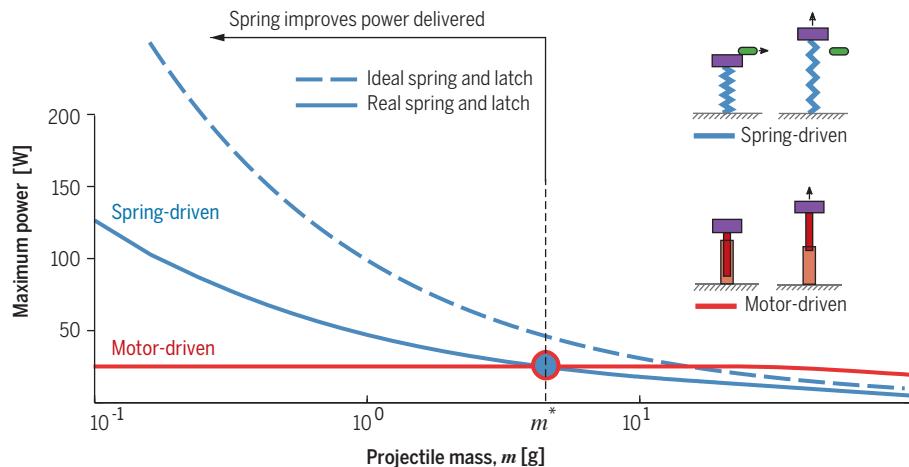
Power-amplified biological and synthetic systems use spring elements to drive motion over a range of size scales. Mathematical modeling reveals a cascade of power limits and mass-dependent transitions in power delivery that arise from the integration of motors, springs, and latches to actuate movement. Variation of these components creates synergistic effects relevant to the analysis and synthesis of diverse power-amplified systems.

RATIONALE: Through a mathematical analysis that is equally applicable to biological and synthetic systems, we investigate how power enhancement emerges through the dynamic coupling of motors, springs, latches, and projectiles and relate the findings to data on existing biological and engineered systems. The model incorporates nonideal behavior of spring and latch systems in a scalable framework using both dimensional and dimensionless approaches.

RESULTS: Motors, springs, and latches all experience force-velocity trade-offs, and their integration exemplifies the cascading effects of power limits. Springs circumvent motor power limits when projectile mass is small and the motor's force-velocity dynamics limit performance. However, springs also exhibit force-velocity trade-offs when their mass, mechanical properties, and time dependence are incorporated. Latches dynamically modulate spring power through variation in latch shape and velocity. Motor-driven and spring-driven movements are distinct in their transitions across performance (power, maximum velocity, and duration), which are largely dictated by projectile mass. When analyzed as a single, integrated system, the necessity for tuning and inherent tunability are evident. Simply increasing the force output of a motor does not enhance performance; the spring and latch capacities must also shift. Simply decreasing the size of the system also does not enhance performance; spring energy storage falls off at smaller scales due to the effects of materials, stiffness, and geometry. With this mathematical foundation of scaling and integration, we apply a new lens to patterns of biological scaling limits and propose new design principles for integrated and tuned systems.

CONCLUSION: Our model reveals a foundational framework for the scaling, synthetic design, and evolutionary diversification of power-amplified systems. The model enables a straightforward approach to analyzing biological systems, encourages a rich design space and functionality for synthetic systems, and highlights a compelling need for the integrative analysis of spring and latch dynamics in both synthetic and biological systems. ∎

RESEARCH ARTICLE

BIOMECHANICS

The principles of cascading power limits in small, fast biological and engineered systems

Mark Ilton,[1] M. Saad Bhamla,[2]*† Xiaotian Ma,[3]* Suzanne M. Cox,[4]* Leah L. Fitchett,[4] Yongjin Kim,[1] Je-sung Koh,[5]‡ Deepak Krishnamurthy,[2] Chi-Yun Kuo,[4]§ Fatma Zeynep Temel,[5] Alfred J. Crosby,[1] Manu Prakash,[2] Gregory P. Sutton,[6] Robert J. Wood,[5] Emanuel Azizi,[7]|| Sarah Bergbreiter,[3]|| S. N. Patek[4]||¶

Mechanical power limitations emerge from the physical trade-off between force and velocity. Many biological systems incorporate power-enhancing mechanisms enabling extraordinary accelerations at small sizes. We establish how power enhancement emerges through the dynamic coupling of motors, springs, and latches and reveal how each displays its own force-velocity behavior. We mathematically demonstrate a tunable performance space for spring-actuated movement that is applicable to biological and synthetic systems. Incorporating nonideal spring behavior and parameterizing latch dynamics allows the identification of critical transitions in mass and trade-offs in spring scaling, both of which offer explanations for long-observed scaling patterns in biological systems. This analysis defines the cascading challenges of power enhancement, explores their emergent effects in biological and engineered systems, and charts a pathway for higher-level analysis and synthesis of power-amplified systems.

Certain organisms are renowned for their ability to circumvent the force-velocity trade-off of muscle motors through mechanisms of power amplification that greatly reduce the amount of time required to perform a given amount of work (force × velocity = work/time = power) (*1*–*3*). Numerous studies have explored muscular power output and the underlying force-velocity trade-offs (*4*, *5*), as well as the enhanced power output achieved using springs and latches (*6*–*11*); however, these studies have yet to fully explain the limits of mass-specific power output [power density (W/kg)] in these organisms. For example, in many of these systems, the motor does work to store energy in a spring, and then the spring solely actuates the motion. The spring serves

as the actuator for the system and, therefore, must operate under its own mass-specific power limits analogous to the motor; however, the power densities of elastic materials are largely unknown for biological and synthetic systems.

We investigate how nonidealized components of power-amplified systems, such as springs and latches, enhance and mediate power output. We apply a modeling approach that is grounded in a simplified power amplification system (Fig. 1), which delineates the limits, integration, and scaling of these systems. Our goal is to formulate the fundamental interactions of mechanical power amplification that apply to both engineered and biological systems. The model is used to successively investigate the role of motors, springs, and latches in mechanical power amplification by addressing four central questions: (i) Under which inertial-loading conditions is a projectile best launched by a motor versus a spring? (ii) What are the implications of spring force-velocity behavior for the generation of power amplification? (iii) How do latches mediate the dynamics of energy release? and (iv) What are the general principles of integrated tuning for power-amplified systems? As we answer each guiding question, we place the findings into an interdisciplinary context and explain trends and limits in existing systems.

A foundational lesson from archery

Archery illustrates the benefits and limits of power amplification that emerge from the integration of a motor, spring, and latch. Arm muscles serve

as motors that put energy into the bow (the spring). Fingers act as latches that resist the release of the elastic energy stored in the bow and determine the timing of energy delivery from the bow to the arrow. The arrow is launched solely by the bow's stored elastic energy. The addition of a bow has two distinct benefits. First, the bow decouples the arm muscles from the arrow's launch, such that the muscles need not contract quickly. Instead, the muscles contract slowly and forcefully to load the stiff spring. Second, the bow can launch the lightweight arrow without the inertial load of the arm, resulting in higher velocity, kinetic energy, and acceleration than if the arrow had been thrown. Even so, although it is sensible to launch a lightweight arrow with a bow, we would not reach for the same bow to launch a heavy stone. The projectile's size should determine whether the bow yields higher launching kinetic energy than possible from the arm alone (*12*). To optimize the system, the properties of the elastic bow should be tuned to the force capacity of the arm muscles and the fingers' ability to restrain the bow's release. Just as in archery, the hallmarks of biological power amplification are the spatial and temporal decoupling of motor and spring; the use of slow and forceful muscle contractions; the tuning of muscle, spring, and latch for maximum work capacity; and the reduction of inertial load (*6*, *7*, *13*).

The lessons from archery and their manifestation in biology set up a series of questions that we address sequentially, beginning with the major components of power amplification—motor, spring, and latch—and concluding with an analysis of their integration. We formulate a mathematical approach that is equally applicable to biological and synthetic systems. To achieve a general framework, we model the system in the simplest possible terms that enable dimensional (presented here) and dimensionless (see supplementary materials) analyses that define a parameter space for investigation and synthesis across systems and scales.

Launching with a motor versus a spring

Our mathematical model—consisting of a motor, spring, and latch—simulates the launch of projectiles of varying mass, m (see supplementary materials for details). A linear force-velocity relation for the motor is held constant during the simulation (Fig. 1A). We solve the dynamical equations of motion for a projectile launched either by the motor alone (Fig. 1B) or by a spring that is preloaded by that same motor (Fig. 1C). In this section, we focus on the question, Under which inertial-loading conditions is a projectile best launched by a motor versus a spring?

The simulation highlights a key transition between small masses, which can achieve high velocities through the addition of elastic elements, and large masses, which are constrained by their inertial loads (Fig. 2). The addition of a spring is beneficial under two scenarios, when (i) takeoff velocity of the projectile is limited by motor velocity and (ii) projectile mass is small. For motor-only actuation, if projectile mass, m, is low, then

[1]Department of Polymer Science and Engineering, University of Massachusetts Amherst, Amherst, MA 01003, USA. [2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [3]Department of Mechanical Engineering and Institute for Systems Research, University of Maryland, College Park, College Park, MD 20742, USA. [4]Department of Biology, Duke University, Durham, NC 27708, USA. [5]School of Engineering and Applied Sciences and Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA. [6]School of Biological Sciences, University of Bristol, Bristol BS8 1TH, UK. [7]Ecology and Evolutionary Biology, University of California Irvine, Irvine, CA, USA.
*These authors contributed equally to this work. †Present address: School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. ‡Present address: Department of Mechanical Engineering, Ajou University, Suwon, 16499, South Korea. §Present address: Division of Evolutionary Biology, Ludwig Maximilian University of Munich, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. ||These authors contributed equally to this work.
¶Corresponding author. Email: snp2@duke.edu (S.N.P.)

takeoff velocity, v_{to}, is limited by the speed of the motor, as seen in the asymptotic approach to v_{max} as m approaches zero. When projectile mass is high, takeoff velocity is limited by the inertia of the projectile. For spring-driven actuation, takeoff velocity is not limited by the motor when the projectile mass is low, and the resulting takeoff velocity is much larger than in the motor-driven case. These findings are consistent with previous analyses of motor-spring systems (12).

Building on these classic findings, we investigate the effects of varying mass on kinematic performance. We find that the transition to effective spring actuation depends on the focal performance metric. If the goal is to maximize takeoff velocity, the transition occurs at a smaller projectile mass than if the focal metric is takeoff duration or maximum power output (Fig. 2).

The relationship between force capacity of a motor and stiffness of a spring determines the amount of stored elastic energy (12, 14). This is important for biological systems because motor force capacity varies as a function of length scale. Biological motors used for power amplification maximize force development at the cost of loading velocity and do so at exclusively small inertial loads (Tables 1 and 2). Invertebrates such as trap-jaw ants and mantis shrimp use long sarcomeres (the contractile units of muscle) to generate slow but forceful contractions to load the elastic elements that subsequently power their extremely rapid movement (15, 16). Plants such as Venus flytraps, aquatic bladderworts, and fern sporangia use slow and forceful nonmuscular hydraulic movements (or nastic movements) coupled with elastic instabilities to achieve rapid motion (17–20). For many power-amplified systems, loading energy into the spring often takes orders of magnitude longer than releasing the energy (13), exemplifying the trade-off of generating work (force × distance) through large forces and small displacements at the cost of velocity.

Force-velocity trade-offs of spring-driven systems are ubiquitous in engineered systems. Most engineered systems use electromagnetic motors with transmissions that generate large torques that slowly load spring elements (21–31). For small systems, a shape memory alloy (SMA) is chosen for its high specific force and linear actuation (32–34). This high specific force comes with low velocity: Heating and cooling the material above and below its phase transition temperature is time-consuming. Engineered power-amplified systems often propel larger inertial loads than biological systems (Tables 1 and 2). Still, some engineered jumping systems (Table 2) effectively use combustion, powerful pneumatic actuators, or high-power electromagnetic actuators that actuate directly without the use of power amplification (29, 30, 35–40).

Power of springs

Although the focus on power amplification has historically revolved around the force-velocity trade-off of motors, the dynamic behavior of springs delivers this power to the payload in spring-mass systems (12). Springs have typically been characterized in terms of force-displacement curves and defined by geometrical and material properties, while still leaving their potential for power enhancement unresolved (41, 42). Furthermore, springs are typically assumed to operate as ideal structures without internal mass or dissipation, such that the energy put into a spring is equal to the energy leaving the spring. By this logic, a spring driving no load mass would accelerate infinitely quickly. Here, we examine how realistic spring actuators limit or enable power amplification. Building on the mathematical model of the previous section, in this section we ask, What are the implications of spring force-velocity behavior for the generation of power amplification? Given that springs have traditionally been assumed to be ideal, the force-velocity

behavior and influence of the material and geometric properties that generate a given stiffness had not been individuated. Therefore, we model kinematic performance of nonideal springs to examine force-velocity behavior in the context of the spring's inertia, stiffness, material properties, and geometry.

Spring inertia leads to a force-velocity trade-off for the spring-driven system (Fig. 3), even for the case of the relatively light spring ($m_s = 0.1$ g) used in the previous section. The net force acting on the projectile depends on position and velocity for both the motor-driven (Fig. 3A) and spring-driven (Fig. 3B) systems, which suggests that force-displacement-velocity trade-offs of springs should be taken into consideration alongside force-velocity trade-offs for motor-driven systems. The force-displacement characteristics of the spring-driven system are set by the spring stiffness, k. Given a fixed maximum force, F_{max}, and range-of-motion, d, for the motor, the spring stiffness determines the amount of elastic energy that the motor can store in the spring. Output of the spring-driven system is therefore sensitive to the stiffness of the spring, which ultimately depends on both the material properties and geometry of the spring.

To illustrate the effect of changing material and geometric properties of the spring, we perform a simulation of the spring-driven system in which the spring is assumed to be a uniform rod, storing and releasing elastic energy under uniaxial compression. The maximum kinetic energy of the spring-driven projectile depends on both cross-sectional area and elastic modulus of the spring material (Fig. 4A) and is strongly dependent on how close the resulting spring stiffness is to the optimal stiffness, k_{opt}. We define k_{opt} as the stiffest spring that can still be loaded a maximum distance d with the same motor. Compared to the $F(x,v)$ relationship for the spring at optimal stiffness (Fig. 3B), using a

Fig. 1. We mathematically model the factors influencing the power output of a projectile driven by either a motor or a spring. We apply a linear force-velocity trade-off that operates in the same range of force and velocity as a biological motor (muscle) (see Box 1, Glossary, for variable definitions). (**A**) This linear force-velocity trade-off approximates the output of a biological system and exhibits a maximum force (F_{max}) of 20 N and a maximum velocity (v_{max}) of 5 m/s. (**B**) Using the force-velocity relationship in (A), the motor (red) directly launches a projectile (purple) with velocity, v, equal to motor velocity (v_{mot}). At the instant that the projectile leaves the motor, the projectile's velocity is defined as its takeoff velocity (v_{to}). The duration from initiation ($t = 0$) of projectile movement to its launch is defined as launch duration (Δt_{to}). The projectile's displacement, x, is defined such that $x(t = 0) = 0$. (**C**) The same motor loads a spring (blue) that solely launches the projectile at the velocity of the spring ($v = v_{sp}$). A latch (green) moving at velocity (v_L) controls the timing and release of elastic potential energy.

$$F = F_{max}\left(1 - \frac{v}{v_{max}}\right)$$
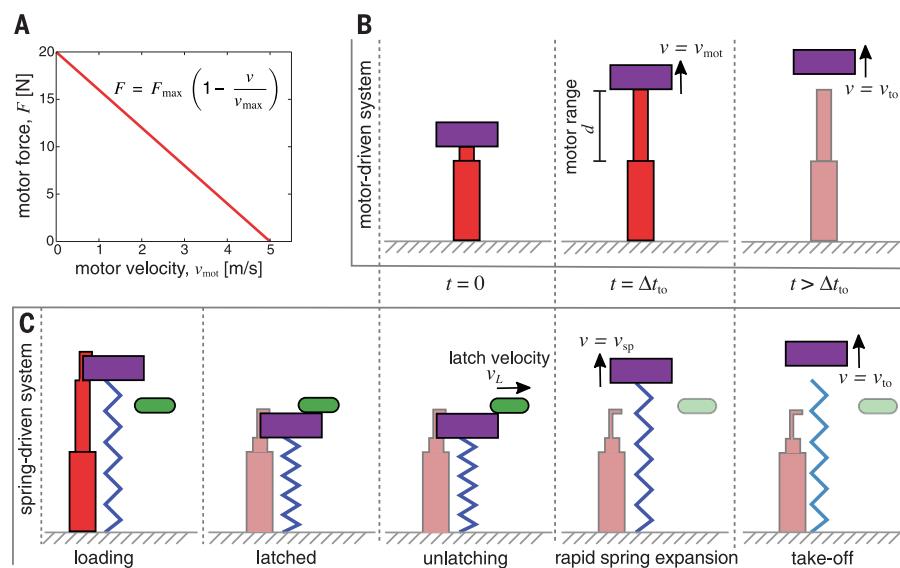
Box 1. Glossary

σ_y: Yield strength of the spring material.
A: Cross-sectional area of the spring material.
d: Motor range of motion.
E: Young's modulus of spring material.
$F(x,v)$: Force-displacement-velocity relationship. The net force acting on the projectile as a function of its displacement and velocity.
F_{max}: Maximum motor force.
k: Hookean spring constant. Defined by the length, cross-sectional area, and Young's modulus of the spring as $k = EA/L$.
k_{opt}: Optimum Hookean spring constant. Defined by the motor properties as $k_{opt} = F_{max}/d$.
KE: Kinetic energy.
KE_{max}: Maximum kinetic energy reached by the projectile. Defined by the projectile mass and takeoff velocity as $KE_{max} = \frac{1}{2} m v_{to}^2$.
L: Equilibrium length of the spring.
m: Projectile mass.
m_s: Mass of the spring. Defined by the density, length, and cross-sectional area of the spring material as $m_s = \rho L A$.
P_{max}: Maximum power delivered to the projectile during its acceleration.
PA: Power amplification. Defined as the ratio of the maximum power output of the spring-driven to the maximum power output of the motor-driven system driving a given projectile.
PE: Potential energy.
R: Radius of curvature of the latch edge.
SMA: Shape memory alloy.
Δt_{to}: Projectile takeoff duration. Elapsed time from the start of motion until the moment the force acting on the projectile falls to zero.
v: Projectile velocity.
v_L: Velocity at which the latch is removed from blocking the projectile.
v_{max}: Maximum motor velocity.
v_{mot}: Motor velocity.
v_{sp}: Velocity of spring during projectile launch.
v_{to}: Projectile takeoff velocity.
x: Projectile displacement.

Fig. 2. Projectile launch dynamics diverge in motor-driven versus spring-driven mechanisms when tested across a range of projectiles. Using the models in Fig. 1, we simulate the launch of projectiles of varying mass, m (0.1 to 100 g), and calculate four metrics relevant to the kinematics of projectile launching: takeoff velocity (v_{to}), takeoff duration (Δt_{to}), maximal power output (P_{max}), and maximum kinetic energy (KE_{max}). These performance metrics illuminate key transitions across the projectile sizes. At larger sizes, the motor-driven system performs moderately better than the spring-driven system. At smaller sizes, the spring-driven projectiles experience substantial enhancements in kinetic energy, maximal power, and takeoff velocity compared with motor-driven projectiles.

spring that is either too compliant (Fig. 4B) or too stiff (Fig. 4C) reduces projectile velocity. Stiffness is a composite property that has both geometric and materials contributions. Therefore, if a spring material with a low modulus is used, a larger cross-sectional area of the spring is required to reach k_{opt}. For the same material density, this requires a heavier spring and increases the inertial contribution of the spring, thus reducing the final kinetic energy achieved by the projectile (upper left of Fig. 4A). On the other hand, a spring material with a high modulus requires a small cross-sectional spring area to achieve the optimal stiffness. Small cross-sectional areas lead to high stress and the potential for

yielding and failure of the spring material, reducing the kinetic energy of the projectile (lower right of Fig. 4A).

Force-velocity trade-offs therefore guide both motors and springs. Spring-driven motion can be analyzed and synthesized in terms of a net force acting on an inertial load as a function of both its displacement and velocity, $F(x,v)$. Spring power capacity is determined by the inertia of the driven mass, spring materials, and materials failure properties. When the driven mass is much greater than the mass of the spring, the recoil rate of the spring becomes insignificant. On the other hand, springs produce the greatest accelerations while driving the smallest masses, such that spring limitations cannot be ignored. Limits on power amplification depend on how

fast springs can recoil and the forces they generate during recoil.

Spring force-velocity behavior may explain spring diversity in biological systems. The small sizes and high accelerations of the most extreme biological systems necessarily constrain the geometries and material properties of their springs. With decreasing size, springs must be proportionally more compliant and geometrically arranged to minimize increasing losses of spring recoil that accompany decreasing size. Varying spring properties in the model would shift the log-log space of Fig. 4A, but the slope would remain consistent. With a more severe force-velocity trade-off at smaller size-scales, the plateau of maximum kinetic energy in Fig. 4A would move toward the material failure regime. This may explain why the larger organisms in Table 1 (e.g., frogs) store energy in long collagen tendons (Young's modulus 0.5 to 1.5 GPa) (43, 44), whereas at smaller scales, energy stores are made out of increasingly stiffer materials. Grasshoppers combine resilin (Young's modulus 1 MPa) and chitinous cuticle (Young's modulus 1 to 40 GPa) to form a composite structure, whereas much smaller froghoppers exclusively use chitinous cuticle, and some plant pollen ejection systems use cellulose (Young's modulus 25 to 150 GPa) (45). Few studies have expressly analyzed spring diversity within and across organisms, yet recent analyses point to multiple optima for spring stiffness depending on the temporal limitations set by the animal's behavioral use of the system (46).

Using a system repeatedly and repeatably likely involves other material constraints (Table 2). Single-use systems can strain materials to failure, enabling enhanced projection, such as in plant pollen ejection. Likewise, materials available to particular organisms may limit their projection mechanisms. Cellulose, for example, has a much higher Young's modulus than collagen, making cellulose more appropriate to accelerate very small projectiles. Conversely, this may explain why vertebrates, which are limited to using collagen with its far lower Young's modulus, only minimally use elastic structures for direct actuation. By contrast, spring-based actuation is extremely widespread in animals, such as insects, that have chitin in their materials toolbox.

Material and geometric properties of biological springs are diverse (Table 2), yet it is not understood how their architecture influences elastic energy storage capacity, recoil rate, and internal energy dissipation. Arthropod springs are built of multiple materials (47–50), such as composites of structured stiff, high-energy-density materials bound within softer, resilient matrices (45, 51–54). The locust's leg combines the protein resilin with stiff cuticle to generate a flexible yet stiff spring (48); resilin provides resilience, while the chitin nanofibers provide the extensional stiffness required for high energy density. Biological springs are typically monolithic, with distributed and integrated flexible and stiff regions. For example, mantis shrimp raptorial appendages dynamically flex both dense, stiff regions and thin, flexible regions during spring loading and release (55).

Some engineered springs have been carefully designed to store the maximum amount of energy, given motor or size constraints. They make use of various geometries and materials, including shaped polymer-fiber composites, SMA, molded elastomers, torsional and linear springs, and steel wires and ribbons (Table 2). In one jumping robot, to maximize stored energy density when driven by an electromagnetic motor, a tapered conical cross section of elastomer equalizes shear stress throughout the spring (56, 57). The springs of compound bows and jumping robots (27) use specific force-displacement profiles to delay peak acceleration. These systems typically incorporate a nonlinear spring within a more complex mechanism. Some jumping robots use SMA for both actuation and energy storage to reduce size and weight of the integrated system (32, 33), decreasing the inertial component of their force-velocity trade-off.

Experimental analyses of the force-velocity dynamics of springs are needed. Most dynamic materials tests operate at known strain rates and measure resistive force (58–66). These tests do not reveal how the actual loading and unloading conditions influence recoil dynamics of the spring. Measuring material properties at rates comparable to the natural unloading rate of the spring presents considerable challenges (67), yet it is essential to understand how materials properties of springs influence their recoil rate when driving a small mass. Free recoil of rubber and elastic bands have revealed the kinetics of compliant materials with simple geometries (68–74). Visualization of material deformation for strain rate calculation is viable for these large deformations of compliant materials but difficult in biological springs undergoing small-scale motions at high rates. Measurements of the maximum rates of spring release under variable loading conditions could address questions such as, Which

springs convert potential to kinetic energy with minimal internal dissipation? What properties influence recoil time? Does a bending spring recoil faster than one in uniaxial extension?

Dynamics of latch removal

For latches, a short release time yields power amplification, and a sufficiently long duration can completely eliminate power amplification or even result in power attenuation. The central influence of latches in power amplification extends beyond the notion of the latch as a switch or a simple mechanism for energy release. Indeed, latches mediate the time, space, and rate over which potential energy is converted to kinetic energy (Fig. 5A). The latch's force capacity determines the maximum amount of stored energy, given the capacity of the underlying spring. The latch's shape and movement should therefore influence the rate of the spring's delivery of force and velocity to the projectile. Thus, we ask, How do latches mediate the dynamics of energy release? We focus on latches that are simple mechanical structures with adjustable geometry and dynamics.

In this simulation, we incorporate latch removal velocity, shape, force capacity, and mass to determine how these factors influence projectile kinematics. We find that the duration and kinematic profile of latch release are influenced by the latch's shape (Fig. 5B). The latch's shape is altered by increasing the radius of curvature of its edges, which reduces the speed of latch removal and allows the spring to release energy before the latch is fully removed. Consequently, the projectile's kinetic energy declines, the release point shifts (there is more stored energy left in the spring when the latch force goes to zero for a smaller radius latch), and release velocity decreases (constrained by the latch velocity and shape). The projectile's maximum kinetic energy decreases as the latch deviates further

from an ideal latch (Fig. 6) in terms of geometry and latch removal kinematics (larger corner radius of curvature and slower latch removal velocity). Latch properties thus dramatically alter the $F(x,v)$ landscape of the spring-driven system (Fig. 6).

Even with many latch designs and patents in engineering, and the diverse evolutionary history of latching and control in biology (Table 2) (75, 76), the relevant metrics and dynamics of latch performance are still in need of basic characterization and analysis. Latches span a range of physical forces, such as osmotic transitions, contact, and phase changes (76–78) (Tables 1 and 2). Biological latches use contact forces, geometric instability, pressure transitions, and cohesive forces (Table 2). Some trap-jaw ants (Myrmicinae) remove a physical block to release their fast-rotating mandibles (79). Venus flytraps use turgor pressure to alter their leaf curvature, such that energy is suddenly released when leaf geometry changes from convex to concave (80). Fern sporangia resist energy release with water pressure; sudden cavitation of the water triggers spore release (17). Similar in principle to the ferns, snapping shrimp use water cohesion to enable energy storage in the system until sufficient tension is generated and the cohesive forces are overcome (81). Ballistospores of the jelly fungus are launched at the instant when two water surfaces coalesce and release surface tension energy (82). Grasshoppers use a lever arm system that generates a positive feedback loop to trigger their jumps (83).

Latches are necessary for controlling the release of considerable elastic energy over vanishingly short durations. Several kinds of insects—fleas, froghoppers, and leafhoppers—rely on well-tuned relationships between latch linkages and springs to minimize jerk. In these systems, linkages increase mechanical advantage of the spring at approximately the same rate that spring force decreases, resulting in an approximately

Fig. 3. Motor-driven and spring-driven systems exhibit distinct force-displacement-velocity behavior.
Color contours indicate the net force of the motor or spring on the projectile, and black lines depict three examples of different projectile masses with their trajectories through x-v space (phase space). Takeoff occurs when the projectiles reach the maximum displacement of the motor or spring at x = d = 5 mm (vertical white line). (**A**) Motor-driven projectiles, regardless of size, are constrained to takeoff velocities below v_{max} = 5 m/s (see motor force-velocity constraints in Fig. 1A). (**B**) Spring-driven projectiles, by contrast, encompass roughly a sixfold greater range of launch velocities than motor-driven projectiles. The spring's force and velocity delivery are determined by a linear, Hookean relationship, but with inertia of the spring included. Therefore, instead of vertical force contour lines in phase space, which would represent a Hookean spring, the force contour lines are curved to reflect the inertial effects of the spring's mass on its force-velocity behavior. The star symbol indicates that optimal spring stiffness is used in the simulation (see Fig. 4) and that the projectile is released using a latch with 0.2 mm radius of curvature (R) and removal velocity (v_L) of 5 m/s (see Fig. 5).
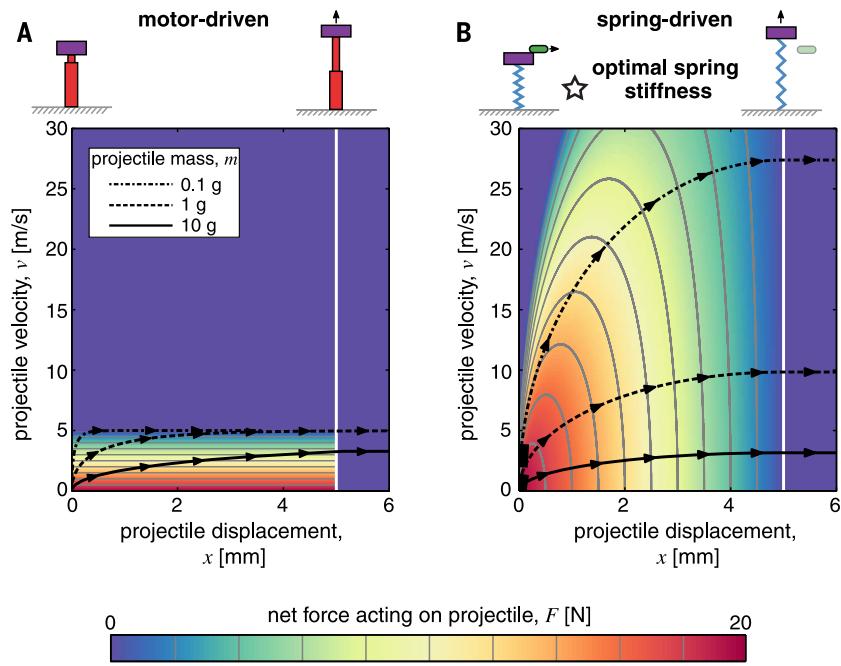
Fig. 4. Spring properties influence the force-velocity profile of projectile launching. (**A**) The spring modulus and cross-sectional area affects the maximum kinetic energy of the projectile (KE_{max}). Using the same motor and latch behavior from the previous simulations, the Young's modulus (E) and cross-sectional area (A) of the spring are varied, while keeping a constant spring length ($L = 10$ mm), density ($\rho = 10$ kg/m^3), and yield strength ($\sigma_y = 10$ MPa). Spring stiffness is calculated as $k = EA/L$. Spring cross-sectional area is a proxy for geometric stiffness. Spring modulus serves as a proxy for material stiffness. Each symbol on the KE_{max} heat map corresponds to the projectile simulations in (B), (C), and Fig. 3. (**B**) A more compliant spring (represented by the circle symbol) yields lower force but similar projectile velocity when compared with (**C**) projectiles launched by a stiffer spring (square symbol). The optimal spring stiffness for achieving maximal projectile velocity is depicted in Fig. 3 (star symbol).

Fig. 5. Latches determine the timing, magnitude and rate of energy release. (**A**) As the spring is loaded, potential energy (PE) (dashed line) in the spring increases. The latch holds the loaded spring in place. Once the latch starts to release, PE decreases and kinetic energy (KE) (solid line) increases until takeoff when the projectile is completely free of the latch. (**B**) Projectile KE is affected by latch shape and rate of removal. Three simulations apply different values for the latch corner radius (R) and latch removal velocity (v_L).

constant acceleration (*49*, *84*). The ability of higher Young's modulus materials to enable energy storage at small scales is likely tied to the origin of latches that are formed through contacting surfaces. Evolutionary origins of latches likely accompanied high-capacity energy storage and the need for tuning of energy release.

Engineered systems also employ a wide variety of latches, including disengagement of physical structures based on contact forces and geometric rearrangements (Table 2). One of the oldest examples of a contact-based force latch is a cam mechanism by Leonardo da Vinci (*85*). This device uses intermittent contact to store potential energy in a slowly rising hammer that generates a large impact when released. The common mousetrap uses a contact latch (physical contact between two structures) that releases the spring-loaded trap upon disengagement (*86*). The small (7 g) jumping robot from Ecole Polytechnique Fédérale de Lausanne (EPFL) is a more modern design that uses the cam mechanism as a contact latch and releases the energy stored in torsional springs; this enables efficient jumping with small payloads through rough terrains (*21*). Inspired by latches in jumping insects, the water strider jumping robot uses a torque reversal catapult (TRC) mechanism to maintain a force profile that enables takeoff on both terrestrial and water surfaces (*32*). A Venus flytrap–inspired system uses bistable composite plates: SMAs slowly push the plates toward an unstable configuration, resulting in a rapid transition between the two stable states (*34*).

Integrated tuning and output of power-amplified systems

Power amplification emerges from the dynamic integration of motors, springs, and latches, yet it has most often been understood as a simple ratio of gross power output relative to power of the underlying motor. In biological systems, it is calculated as the system's power output relative to the maximum power output of the underlying muscle or, in a more conservative approach for systems in which muscle power output has not yet been quantified, relative to the highest measured power output of any biological system. In engineering, power amplification is typically referenced as the enhanced power density of a system. Not only does this approach black-box dynamics of these systems, it also misses the rich array of performance outputs that can be achieved and tuned through integration of these components. Therefore, we ask, What are the general principles of integrated tuning for power-amplified systems?

To achieve the greatest power amplification, spring properties must be optimized and matched to the motor, and the motor's properties must be shifted away from high power output and toward high force output. Spring-driven performance increases with motor force or motor range of motion but does not depend on maximum motor velocity. Therefore, increasing F_{max} or d (even at the expense of v_{max}) improves performance. However, if other elements (e.g., spring or latch) are

fixed, increasing motor force or range of motion does not necessarily lead to improved performance. The system must be fully integrated to confer the benefits of increasing the motor's capabilities.

We mathematically explore integration by allowing the motor's maximum force output to increase (Fig. 7). If spring stiffness is tuned to its optimal value as a function of F_{max}, then maximum kinetic energy of a spring-driven projectile increases monotonically as a function of F_{max} (Fig. 7A, solid blue curve), and power amplification intensifies (Fig. 7B, dashed green curve). On the other hand, if spring stiffness is fixed to the optimal stiffness for $F_{max} = 20$ N, and, of equal importance, if the motor has no increased range of motion or the position of the latch is fixed, then increasing the motor's maximum force capacity above 20 N does not result in any additional kinetic energy delivered to the projectile (Fig. 7A, dashed red curve), and power amplification is diminished for $F_{max} = 30$ N (Fig. 7B, dash-dotted orange curve). In other words, although a motor with a larger force capacity has the potential capability to produce more energy, unless the spring can store that energy over the limited range of motion of the motor, the motor will not reach its maximum force capacity and the overall performance of the system will not improve. This coupling is important not only between the spring and motor but also between all elements of the system, including the latch.

Even though the model is composed of few components, defined in the simplest terms, and with a modest array of nonideal behaviors, the outputs venture into a richer space than expected from previous research. Integration of these components not only determines the array of outputs and transition points, it also enables analysis of failure and scaling limits as more parameters are incorporated. A particular strength of power-amplified systems, especially ones with spring actuation, is that components are often spatially and temporally separated and thereby enable a modular approach to both synthesis and analysis.

When our modeling results are placed in the context of the diversity of biological systems and the limited array of engineered systems, it is clear that integration is central to achieving a rich performance space. The components of biological systems necessarily evolved together, yet engineering design often struggles to achieve similarly tuned integration. Only a few studies and systems have been examined in the context of tuning and integration to achieve power amplification. The mantis shrimp's motor, spring, and projectile coevolved as integrated components ([87]–[91]), such that species producing the greatest accelerations evolved springs with greater work capacity and muscles with force-modified architecture at the expense of contraction velocity. They evolved with varying degrees of integration (i.e., correlated change among components): More tightly integrated components are associated with greater acceleration and spring work

but with slower accumulation of change over evolutionary time ([89]). This trade-off between integration and evolvability is relevant to engineering design, given that retaining flexibility or modularity of design for different goals may bear a cost in terms of integration and performance.

Integrated design approaches offer a key pathway for achieving improved dynamics and scaling in engineered systems. The tuning of engineered jumpers has been examined, including a galago-inspired jumping robot ([56]). In this case, various configurations of motors, springs, and linkages were compared using a vertical jumping agility metric (a combination of jump height and frequency). With the motor and spring in series, which enabled rapid leg repositioning for multiple jumps, a linkage was used to modulate the power delivery from the spring to the ground. Even though the spring was not included in this particular optimization [other studies have incorporated spring optimization analyses; see ([57], [92])], the galago study analyzed trade-offs between motor power density and linkage design for series elastic jumpers. Navigation of the rich design space for engineered systems may be particularly challenging; however, advances in additive manufacturing alleviate this challenge,
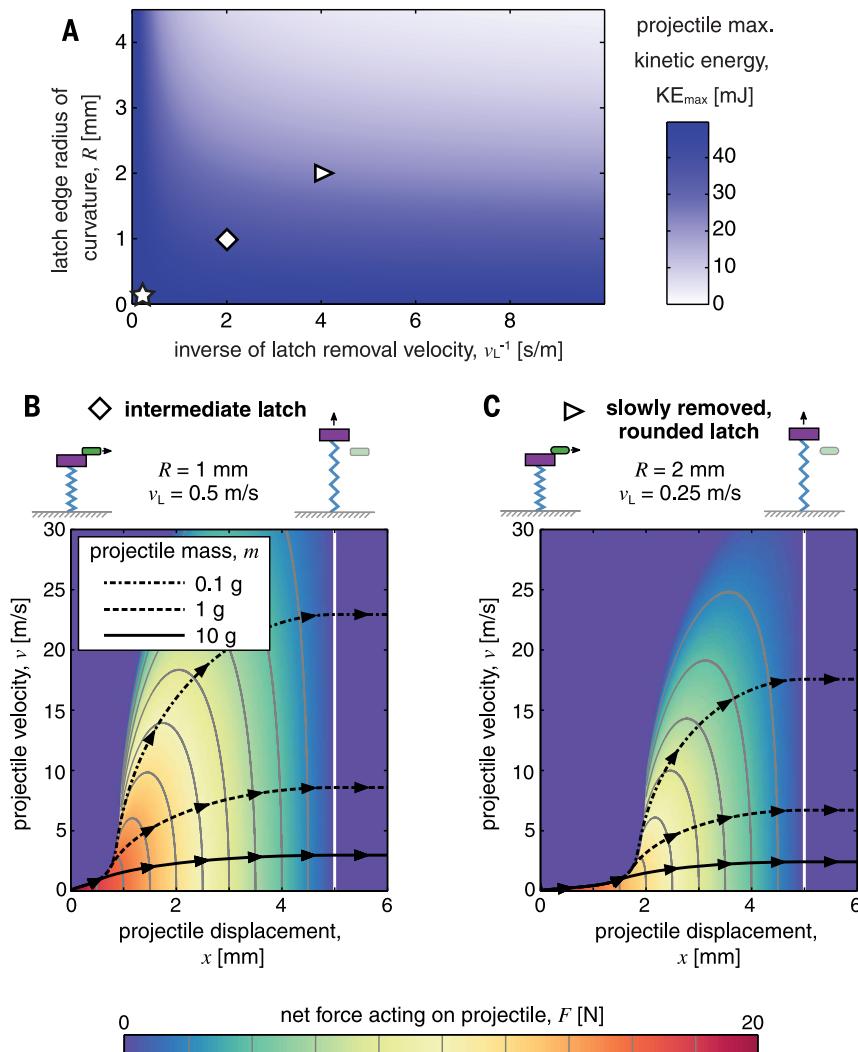
Fig. 6. Latch geometry and latch removal kinematics influence projectile kinetic energy and the force-velocity profile of spring-driven launches. (**A**) The maximum kinetic energy of the projectile (KE_{max}) decreases as the latch's radius of curvature (R) (see Fig. 5) is increased (star, $R = 0.2$ mm; diamond, $R = 1$ mm; triangle, $R = 2$ mm) and latch removal speed (v_L) is decreased (increased inverse latch removal speed, $1/v_L$). The star symbol refers to the simulation depicted in Fig. 3B, to which both optimal latch and spring dynamics were applied. (**B**) This simulation applied the 1 mm radius of curvature latch at 0.5 m/s latch removal speed (diamond). In this case, force on the projectile was preserved and launch velocity remained high for small projectiles. (**C**) Incorporation of a large radius of curvature (2 mm) and slower removal velocity (0.25 m/s) caused a substantial reduction in force development during launching, which primarily affected the kinematics of smaller projectiles.

where, as in self-assembly, the cost of complexity is dramatically reduced (93).

Applications

A natural next question is whether this model can be applied to even broader challenges, such as establishing design principles underlying biological diversity. Scaling of power-amplified biological systems has generated centuries of hypotheses from Borelli to Vogel (94). The most widely accepted scaling explanation is that jump performance (kinetic energy) scales with body mass and muscle volume. However, as pointed out by Vogel (95, 96), this scaling pattern persists in organisms that lack muscles, implying that muscle-volumetric scaling is not sufficient to explain enhanced performance at small sizes and declining performance at larger sizes. Our findings suggest a more complete explanation for the

scaling of power-amplified systems in biology that incorporates a lower size limit, an upper size limit, and an approach to addressing "optimal" size (Fig. 8).

The upper size limit of biological power-amplified systems has garnered the most attention: Why are power-amplified systems exclusively small (Fig. 8A)? Our analysis of muscle-spring dynamics answers that question (Fig. 2): Above a certain size, a spring does not enhance kinematic output and an organism is best served by using muscle. However, the inverse question is rarely asked: Why does kinematic performance fall off at smaller sizes in jumping insects (Fig. 8B)? Our analysis suggests that a trade-off between Young's modulus and size may reduce performance at smaller scales (Fig. 4A). To actuate motion with a spring and achieve high performance at small scales, the Young's modulus must be large. However, as Young's modulus increases

and spring cross-sectional area decreases, the failure limit of the material will be reached and thereby sets a lower size limit to spring actuation.

Upper and lower limits to muscle-based power amplification may explain a conundrum in the scaling of jumping insects: The smallest insects and the largest insects jump more slowly than midsized insects (Fig. 8B). Fleas (0.5-mg body mass) have a low takeoff velocity relative to other spring-loaded jumpers (97, 98), whereas the best-recorded jumpers have body mass on the order of 10s of milligrams [e.g., froghoppers: 5 to 6 m/s, 20 to 50 mg (99); pygmy mole crickets: 5 to 6 m/s, 10 mg (98, 100)]. Given the constraints of their materials toolbox, and following the inherent trade-off between Young's modulus and cross-sectional area (Fig. 4A), insect springs are also likely to decrease in performance with decreasing size. In other words, at a threshold

Table 1. Fast biological systems are used for a greater diversity of functions and operate at smaller size scales than engineered systems. Kinematics of a representative sample of fast biological and engineered movements are arranged by characteristic length scale. Characteristic length represents projectile length (e.g., stinging needle of nematocysts and magnet from chameleon coil gun), accelerated part length (e.g., combustion-based jumping systems, jumping insects, and mandibles of trap-jaw ants), or the leg length for jumping robots, given that legs are the dominant length in these systems. Duration represents acceleration time. Distance represents maximum horizontal or vertical distance traveled by a projectile, jumping system, or maximum rotation of striking and trapping systems. Biological movements operate in air (frog, grasshopper, Venus flytrap, trap-jaw ant, fungi, and chameleon) or water (mantis shrimp, *Hydra*, and bladderworts). All listed engineered systems are designed for operation in air, although the water strider robot jumps at the air-water interface. Char., characteristic; Max., maximum; Accel., acceleration. *Estimated or calculated values from figures or data reported. **Velocity and acceleration of fluid displaced by bladderwort trap.

Movement	System	Char. length (m)	Mass (kg)	Duration (s)	Distance (m)	Max. speed (m/s)	Accel. (m/s²)	References
Biological systems								
Nematocyst discharge	*Hydra*	2.4×10^{-5}*	2.3×10^{-12}	1×10^{-6}*	1.3×10^{-5}	3.7×10^{1}	5.4×10^{7}	(103, 106)
Ballistospore ejection	Fungi	4.2×10^{-6}	3.7×10^{-13}*	1×10^{-5}*	4.0×10^{-4}*	1.6	1.2×10^{5}	(107)
Pollen ejection	Bunchberry dogwood stamen	8.0×10^{-4}	4.0×10^{-7}	4.0×10^{-4}	1.2×10^{-3}	7.5	2.4×10^{4}	(108, 109)
Mandible strike	Trap-jaw ant	1.4×10^{-3}	1.5×10^{-7}	6×10^{-5}	–	6.4×10^{1}	1×10^{6}	(110)
Jump	Plant louse	1.9×10^{-3}	7.0×10^{-7}	4×10^{-4}	1.5×10^{-3}	2.5	6.3×10^{3}	(111)
Suction trap	Aquatic bladderworts	2.0×10^{-3}	–	1×10^{-3}	–	1.5**	6.0×10^{3}**	(18)
Jump	Froghopper	6.1×10^{-3}	1.2×10^{-5}	8.8×10^{-4}	7.0×10^{-1}	4.7	5.4×10^{3}	(99)
Snap buckling	Venus flytrap	1.0×10^{-2}	–	1.0×10^{-1}	–	1.0×10^{-2}	–	(20)
Appendage strike	Mantis shrimp	8.2×10^{-3}	9.2×10^{-3}	2.7×10^{-4}	–	3.1×10^{1}	2.5×10^{5}	(112)
Jump	Frog	4.1×10^{-2}	8.8×10^{-3}	5.4×10^{-2}	2.2	4.5	1.4×10^{2}	(113)
Tongue projection	Chameleon	2.0×10^{-2}*	8.7×10^{-5}	6×10^{-3}*	1.2×10^{-1}	5.3	2.6×10^{3}	(114, 115)
Jump	Locust	5.0×10^{-2}	1.8×10^{-3}	3×10^{-2}*	9.5×10^{-1}	3.2	1.8×10^{2}	(116)
Engineered systems								
Jump	Micro elastomer jumper	4.0×10^{-3}	8.0×10^{-6}	–	3.2×10^{-1}	3.0	–	(117)
Jump	Energetic silicon jumper	7.0×10^{-3}	3.1×10^{-4}	–	8.0×10^{-2}	1.3	–	(37)
Projection	Chameleon tongue–inspired sys.	8.0×10^{-3}	7.7×10^{-4}	8.0×10^{-3}	1.6×10^{-1}	5.4	9.2×10^{2}	(118)
Jump	Flea-inspired robot	3.0×10^{-2}	2.3×10^{-3}	8.0×10^{-3}	1.2	7.0	8.8×10^{2}*	(119)
Jump	Steerable MSU jumper	3.3×10^{-2}	2.4×10^{-2}	–	9.0×10^{-1}	4.3	-	(120)
Jump	Water strider–inspired robot	5.0×10^{-2}	6.8×10^{-5}	2×10^{-2}*	1.4×10^{-1}	1.6	1.4×10^{2}	(32)
Jump	EPFL 7 g robot	5.0×10^{-2}*	7.0×10^{-3}	1.5×10^{-2}	1.4	6.0	4.0×10^{2}*	(21)
Jump	JPL hopper (2nd gen.)	1.0×10^{-1}*	1.3	–	2.0	–	–	(27)
Strike, catch	High-speed fingered hand	1.1×10^{-1}	1.0×10^{-1}*	2×10^{-2}	180°	4.5	2×10^{2}*	(30)
Jump	Locust-inspired robot	1.4×10^{-1}	2.3×10^{-2}	2×10^{-2}	3.7	9.0	4×10^{2}*	(31)
Catch	Flytrap-inspired robot	1.5×10^{-1}	–	4.1×10^{-1}	90°	–	–	(34)
Jump	Galago-inspired "Salto" robot	1.5×10^{-1}	1.0×10^{-1}	–	1.0	–	–	(56)
Jump	Sand flea–inspired robot	1.5×10^{-1}	5.0	–	1.0×10^{1}	–	–	(39)
Jump	Sandia Mars hopper	2.0×10^{-1}*	5.0×10^{-1}	–	1.8	–	–	(121, 122)
Jump	Soft combustion robot	3.0×10^{-1}*	9.7×10^{-1}	–	7.6×10^{-1}	–	–	(93)
Jump	Bipedal jumper "Mowgli"	6.0×10^{-1}*	3.0	2.5×10^{-1}	5.0×10^{-1}	–	–	(38)

specific to their elastic system, the smallest jumpers cannot load their springs without encountering failure. Other biological systems at the scale of fleas that achieve greater kinematic performance should therefore be using either different materials (i.e., cellulose) or different mechanisms that circumvent material limits at small scales. (e.g., the surface tension catapult of fungal ballistospores).

Our simple model can be parameterized to pinpoint limits and optima that are system-specific and can be extended to engineered systems. Performance limits and optimal mass have been found for small engineered jumpers based on environmental drag; similar methods can be extended to motor-spring dynamics and material limits (101). Even so, compared to the impressive diversity of motion and scales in biological systems (which are naturally fully inclusive of power, control, and actuation), engineered systems occupy a relatively narrow range of dynamics,

scales, and uses (Table 2 and Fig. 8A) (35, 36), and our analyses suggest a far broader range of potential capabilities for synthetic systems.

Limitations

By adjusting the scaling of our archery example, the multifaceted challenges of energy dissipation become apparent. Energy dissipation begins with heat produced by the actively contracting arm muscles and continues through material dissipation of the contracting and releasing spring, frictional or hysteresis losses of the latching mechanism, and interaction between the arrow, gravity, and its fluid environment while moving toward its target. The environment surrounding the arm and bow especially influences the amount of dissipation, especially in the context of scaling. For an arrow flying through air at high Reynolds numbers [Re, a dimensionless measure of the relative importance of inertial and viscous forces in a flow (102)], drag forces

scale with the second power of the arrow velocity and cause external dissipation. However, if the arrow is scaled down to the size of many small power-amplified systems, drag forces increase linearly with velocity (low Re limit), thus changing the scaling of the external dissipation (102).

If we move the elastic bow from air to liquid—the internal environment of most biological springs—dissipation increases substantially. Small Re is relevant for microscale systems, such as the puncturing needles of nematocysts (103): Projectile dynamics are dominated by viscous effects, inertia is of no consequence, and the dynamics are counterintuitive (104). Additional complexity occurs when the ambient fluid is non-Newtonian or complex [i.e., not characterized by a single, constant parameter such as viscosity (105)]. The rheology of the materials constituting the spring and latch can assume greater importance at small scales. Our

Table 2. Fast movement is achieved through integration of diverse motor, spring, and latch components, with some systems operating repeatedly and others self-destructing after one use. Repeatable systems (R) function many times, whereas single-shot systems operate only once (NR). Latch mechanisms include contact (physical contact between two structures), fluidic (mediated by microscopic and macroscopic fluid properties), and geometric (dependent on changes in forces, moment arms, and elastic instabilities due to geometrical configura-

tions). A final distinction is whether the system can repeat the motion without external manipulations (R-i) or external manipulation is required to prepare the system to fire again (R-e). Any repeatable biological system must be able to internally reset the system. However, many engineered systems still require an external device (or person) to reconfigure the system to the correct condition to fire again. DC, direct current; MA, mechanical advantage; SMA, shape memory alloy; TRC, torque reversal catapult; Repeat., repeatability.

Movement	System	Work input	Energy storage	Latch	Repeat.	References
Biological systems						
Appendage strike	Mantis shrimp	Muscle contraction	Exoskeleton	Contact	R-i	(123)
Claw closure	Snapping shrimp	Muscle contraction	Exoskeleton	Fluidic (cohesion)	R-i	(81, 124)
Jump	Frog	Muscle contraction	Plantaris tendon	Geometric	R-i	(125, 126)
Jump	Grasshoppers, locusts	Muscle contraction	Resilin and chitin	Geometric	R-i	(47, 48)
Leaf closure	Venus flytrap	Turgor pressure	Cell wall	Geometric (instability)	R-i	(20)
Mandible closure	Trap-jaw ant	Muscle contraction	Exoskeleton	Contact	R-i	(79, 110)
Nematocyst discharge	*Hydra*	Osmotic gradient	Cell membrane	Cohesion	NR	(103)
Spore ejection	Basidiomycota fungi	Water condensation	Surface tension	Fluidic (coalescence)	NR	(107, 127)
Spore ejection	Fern sporangium	Dehydration	Annulus wall	Fluidic (pressure)	NR	(17)
Stalk contraction	Vorticella	Ionic gradient	Spasmoneme protein	Unknown	R-i	(128)
Tongue projection	Chameleon	Muscle contraction	Collagen sheaths	Contact	R-i	(129, 130)
Water suction	Bladderworts	Osmotic gradient	Trap wall	Geometric (instability)	R-i	(18)
Engineered systems						
Catch	Flytrap-inspired robot	SMA contraction	Bistable composite	Geometric (instability)	R-i	(34)
Jump	Micro elastomer jumper	External force	Elastomer tension	Contact	R-e	(117)
Jump	EPFL 7 g robot	DC motor	Steel torsion	Contact	R-i	(21)
Jump	Steerable MSU jumper	DC motor	Steel torsion	Geometric	R-i	(120)
Jump	JPL hopper (2nd gen.)	DC motor	Steel torsion	Contact	R-i	(27)
Jump	Locust-inspired robot	DC motor	Steel torsion	Contact	R-i	(31)
Jump	Galago-inspired "Salto" robot	DC motor	Elastomer torsion	Geometric (MA)	R-i	(56)
Jump	Water strider–inspired robot	SMA contraction	SMA sheet + cantilever	Geometric (TRC)	R-e	(32)
Jump	Flea-inspired robot	SMA contraction	SMA coil	Geometric (TRC)	R-e	(119)
Jump	Energetic silicon jumper	Directly powered by chemical reaction			NR	(37)
Jump	Soft combustion robot	Directly powered by chemical reaction			R-e	(93)
Jump	Sandia Mars hopper	Directly powered by chemical reaction			R-i	(121, 122)
Jump	Bipedal jumper "Mowgli"	Directly powered by pneumatic actuation			R-i	(38)
Projection	Chameleon tongue–inspired sys.	Directly powered by electromagnetic coilgun			R-i	(118)
Strike	Mousetrap	External force	Steel torsion	Contact	R-e	(86)
Strike, catch	High-speed fingered hand	Directly powered by DC motor			R-i	(30)
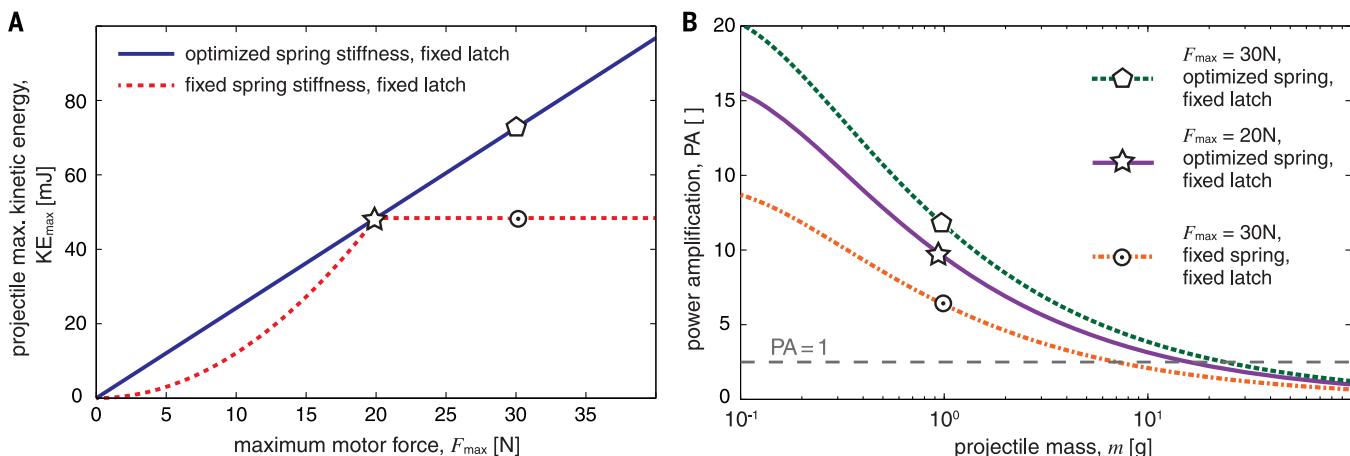
Fig. 7. Power amplification and kinetic energy represent the effects of strategic tuning of springs, motors, and latches. (**A**) The dependence of the maximum kinetic energy of a projectile on the motor force capacity is determined by whether the spring stiffness is optimized as a function of motor force (blue solid curve) or is fixed (red dashed curve). This simulation is applied to a $m = 1$ g projectile. We simulate a motor with expanded force capacity compared with previous simulations (Fig. 1A), while keeping the motor range of motion fixed. Spring properties are optimized using the process illustrated in Fig. 4A. (**B**) Power amplification, PA (the ratio of maximum power delivered to the projectile from the spring-driven or motor-driven system), is strongly influenced by both the tuning of spring properties to motor properties and the projectile mass, particularly at smaller sizes. We simulate power amplification using earlier spring configurations (A) and again increase the force capacity of the motor. Tuning spring properties to motor force capacity enhances power amplification beyond what is achieved with a fixed spring stiffness, exemplifying the need to integrate and tune the motor, spring, and projectile load when attempting to maximize the projectile kinematics.
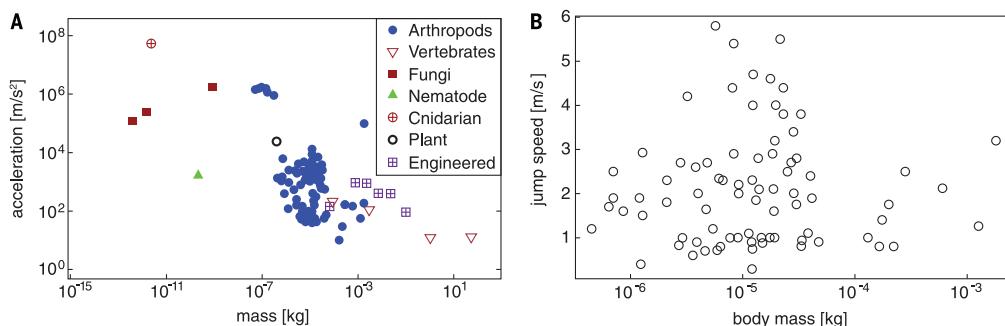
Fig. 8. Power-amplified biological systems span orders of magnitude in acceleration and mass, yet smaller sizes may not always confer benefits in terms of speed. (**A**) The most extreme accelerations are found in fungi and the stinging cells of cnidarians, whereas the top accelerations in the arthropods are performed by trap-jaw ants and mantis shrimp. Engineered systems exceed some vertebrates but have yet to extend into the dynamics and mass scales of most power-amplified biological systems. (**B**) Even though broad comparisons of acceleration and mass demonstrate increasing acceleration at smaller masses, a closer look at the jumping velocities of insects reveals peak jumping speeds at an intermediate size. Engineering data are from Table 1. Biological data are listed in table S3.

mathematical model can be improved by including environment-system interactions and incorporating a more detailed analysis of the friction and mechanical interactions during unlatching. In addition, the role of time-dependent visco-elastic materials, which are ubiquitous in nature and engineering, should be incorporated into the motor, spring, and latch. Extension of our current framework to include dissipative elements will further illuminate design principles that very small-scale power-amplification systems encounter and successfully resolve.

Conclusions

Our results explore and resolve the intriguing intersection of force-velocity dynamics, size, and materials in the diversity and performance of biological and engineered power amplification.

We discovered that $F(x,v)$ of a spring-driven system is a whole-system property of the combined motor, spring, latch, and projectile. As demonstrated through a parameterized latch model, latch design has a substantial impact on performance. Optimal spring design requires careful tuning of spring stiffness to motor properties, which can be achieved through geometric or materials stiffness. Ultimately, the spring's stiffness combined with its force-velocity trade-off and material failure properties set an optimal size-scale of the spring: A large spring is slowed down by its own force-velocity trade-off, whereas a small spring is more likely to fail. Individual components of power-amplified systems and their interactions are essential to dynamic outputs and offer a greater potential for kinematic performance than has been previously recognized.

This study demonstrates the rich potential for understanding, analyzing, and designing effective, integrated dynamics of power-amplified systems. Analysis of power-amplified systems provides an important opportunity to establish fundamental principles of actuators, materials, and latches in their own right and in the context of their dynamic interactions. Researchers across the fields of mathematics, engineering, and biology are well poised to resolve these new and classic challenges through advances in high-speed imaging, materials testing and synthesis, integrated engineering design systems, and new biological discoveries.

Methods

The comparative biological data set was constructed based on an exhaustive literature search of fast movements in biology. This initial data set

consisted of 278 species-level data points about fast organismal movement. This data set was then reduced to include only species for which the mass of the accelerated component was reported and then reduced again to include only one entry per species, such that the final data set of 104 species was compiled from the most recent or most high-quality kinematic data set. When available, the maximum reported acceleration and velocity were included in the data set and, if a range was not provided, we instead included the average reported values. If maximum acceleration and average acceleration were not reported, acceleration was estimated from the ratio of maximum speed to duration in Table 1.

A mathematical model was constructed using a linear elastic spring and a geometric latch model described in the supplementary text (along with tables S1 and S2 and figs. S1 to S12). For a given set of motor, spring, latch, and load mass parameters, the latch release time was numerically calculated using MATLAB. From the latch release time, the kinematic variables (force-displacement-velocity relationship, takeoff velocity, maximum power, projectile takeoff duration, and maximum kinetic energy) were calculated using the equations in sections S1 to S6 of the supplementary materials.

REFERENCES AND NOTES

1. P. Aerts, Vertical jumping in *Galago senegalensis*: The quest for an obligate mechanical power amplifier. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **353**, 1607–1620 (1998). doi: 10.1098/rstb.1998.0313

2. M. M. Peplowski, R. L. Marsh, Work and power output in the hindlimb muscles of Cuban tree frogs *Osteopilus septentrionalis* during jumping. *J. Exp. Biol.* **200**, 2861–2870 (1997). pmid: 9344973

3. M. Burrows, Biomechanics: Froghopper insects leap to new heights. *Nature* **424**, 509 (2003). doi: 10.1038/424509a; pmid: 12891345

4. O. Barth, Harmonic piezodrive—miniaturized servo motor. *Mechatronics* **10**, 545–554 (2000). doi: 10.1016/S0957-4158(99)00062-8

5. L. C. Rome, S. L. Lindstedt, The quest for speed: Muscles built for high-frequency contractions. *Physiology* **13**, 261–268 (1998). doi: 10.1152/physiologyonline.1998.13.6.261; pmid: 11390801

6. T. J. Roberts, E. Azizi, Flexible mechanisms: The diverse roles of biological springs in vertebrate movement. *J. Exp. Biol.* **214**, 353–361 (2011). doi: 10.1242/jeb.038588; pmid: 21228194

7. S. N. Patek, D. M. Dudek, M. V. Rosario, From bouncy legs to poisoned arrows: Elastic movements in invertebrates. *J. Exp. Biol.* **214**, 1973–1980 (2011). doi: 10.1242/jeb.038596; pmid: 21613512

8. W. Gronenberg, Fast actions in small animals: Springs and click mechanisms. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **178**, 727–734 (1996). doi: 10.1007/BF00225821

9. R. McN. Alexander, H. C. Bennet-Clark, Storage of elastic strain energy in muscle and other tissues. *Nature* **265**, 114–117 (1977). doi: 10.1038/265114a0; pmid: 834252

10. R. M. Alexander, *Elastic Mechanisms in Animal Movement* (Cambridge Univ. Press, 1988).

11. D. P. Ferris, M. Louie, C. T. Farley, Running in the real world: Adjusting leg stiffness for different surfaces. *Proc. R. Soc. London Ser. B* **265**, 989–994 (1998). doi: 10.1098/rspb.1998.0388; pmid: 9675909

12. A. Galantis, R. C. Woledge, The theoretical limits to the power output of a muscle-tendon complex with inertial and gravitational loads. *Proc. R. Soc. London Ser. B* **270**, 1493–1498 (2003). doi: 10.1098/rspb.2003.2403; pmid: 12965015

13. S. N. Patek, The most powerful movements in biology. *Am. Sci.* **103**, 330–337 (2015). doi: 10.1511/2015.116.330

14. D. Cofer, G. Cymbalyuk, W. J. Heitler, D. H. Edwards, Neuromechanical simulation of the locust jump. *J. Exp. Biol.* **213**, 1060–1068 (2010). doi: 10.1242/jeb.034678; pmid: 20228342

15. W. Gronenberg, B. Ehmer, The mandible mechanism of the ant genus *Anochetus* (Hymenoptera, Formicidae) and the possible evolution of trap-jaws. *Zoology* **99**, 153–162 (1996).

16. M. M. Blanco, S. N. Patek, Muscle trade-offs in a power-amplified prey capture system. *Evolution* **68**, 1399–1414 (2014). doi: 10.1111/evo.12365; pmid: 24475749

17. X. Noblin *et al.*, The fern sporangium: A unique catapult. *Science* **335**, 1322 (2012). doi: 10.1126/science.1215985; pmid: 22422975

18. O. Vincent *et al.*, Ultra-fast underwater suction traps. *Proc. R. Soc. London Ser. B* **278**, 2909–2914 (2011). doi: 10.1098/rspb.2010.2292; pmid: 21325323

19. J. M. Skotheim, L. Mahadevan, Physical limits and design principles for plant and fungal movements. *Science* **308**, 1308–1310 (2005). doi: 10.1126/science.1107976; pmid: 15919993

20. Y. Forterre, J. M. Skotheim, J. Dumais, L. Mahadevan, How the Venus flytrap snaps. *Nature* **433**, 421–425 (2005). doi: 10.1038/nature03185; pmid: 15674293

21. M. Kovac, M. Fuchs, A. Guignard, J.-C. Zufferey, D. Floreano, in *IEEE International Conference on Robotics and Automation* (IEEE, 2008), pp. 373–378.

22. F. Li *et al.*, Jumping like an insect: Design and dynamic optimization of a jumping mini robot based on bio-mimetic inspiration. *Mechatronics* **22**, 167–176 (2012). doi: 10.1016/j.mechatronics.2012.01.001

23. J. Zhao, in *IEEE International Conference on Robotics and Automation* (IEEE, 2011), pp. 4614–4619.

24. B. G. A. Lambrecht, A. D. Horchler, R. D. Quinn, in *IEEE International Conference on Robotics and Automation* (IEEE, 2005), pp. 1240–1245.

25. S. A. Stoeter, P. E. Rybski, M. Gini, N. Papanikolopoulos, in *IEEE Int. Conf. Intell. Robot. Syst.* **1**, 721–726 (2002).

26. N. Fukamachi, H. Mochiyama, in *IEEE International Conference on Advanced Intelligent Mechatronics* (IEEE, 2015) pp. 1102–1107.

27. J. Burdick, P. Fiorini, Minimalist jumping robots for celestial exploration. *Int. J. Robot. Res.* **22**, 653–674 (2003). doi: 10.1177/02783649030227013

28. M. Kaneko, M. Higashimori, in *Automation Congress* (IEEE, 2004), pp. 117–122.

29. A. M. Johnson, D. E. Koditschek, in *IEEE International Conference on Robotics and Automation* (IEEE, 2013), pp. 2568–2575.

30. A. Namiki, Y. Imai, M. Ishikawa, M. Kaneko, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2003), pp. 2666–2671.

31. V. Zaitsev *et al.*, A locust-inspired miniature jumping robot. *Bioinspir. Biomim.* **10**, 066012 (2015). doi: 10.1088/1748-3190/10/6/066012; pmid: 26602094

32. J. S. Koh *et al.*, Jumping on water: Surface tension-dominated jumping of water striders and robotic insects. *Science* **349**, 517–521 (2015). doi: 10.1126/science.aab1637; pmid: 26228144

33. J. S. Koh, S. Jung, R. J. Wood, K. Cho, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2013), pp. 3796–3801.

34. S. W. Kim *et al.*, Flytrap-inspired robot using structurally integrated actuation based on bistability and a developable surface. *Bioinspir. Biomim.* **9**, 036004 (2014). doi: 10.1088/1748-3182/9/3/036004; pmid: 24615620

35. W. Lindsay, D. Teasdale, V. Milanovic, K. Pister, C. Fernandez-Pello, in *IEEE International Conference on Micro Electro Mechanical Systems* (IEEE, 2001), pp. 606–610.

36. S. J. Apperson *et al.*, Characterization of nanothermite material for solid-fuel microthruster applications. *J. Propuls. Power* **25**, 1086–1091 (2009). doi: 10.2514/1.43206

37. W. A. Churaman, L. J. Currano, C. J. Morris, J. E. Rajkowski, S. Bergbreiter, The first launch of an autonomous thrust-driven microrobot using nanoporous energetic silicon. *J. Microelectromech. Syst.* **21**, 198–205 (2012). doi: 10.1109/JMEMS.2011.2174414

38. R. Niiyama, A. Nagakubo, in *IEEE International Conference on Robotics and Automation* (IEEE, 2007), pp. 2546–2551.

39. Boston dynamics Inc., Sandflea, https://www.bostondynamics.com/sandflea.

40. H. Tsukagoshi, M. Sasaki, A. Kitagawa, T. Tanaka, in *IEEE International Conference on Robotics and Automation* (IEEE, 2005), pp. 1276–1283.

41. F. E. Zajac, Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control. *Crit. Rev. Biomed. Eng.* **17**, 359–411 (1989). pmid: 2676342

42. T. J. Roberts, R. L. Marsh, Probing the limits to muscle-powered accelerations: Lessons from jumping bullfrogs. *J. Exp. Biol.* **206**, 2567–2580 (2003). doi: 10.1242/jeb.00452; pmid: 12819264

43. R. F. Ker, Dynamic tensile properties of the plantaris tendon of sheep (*Ovis aries*). *J. Exp. Biol.* **93**, 283–302 (1981). pmid: 7288354

44. R. F. Ker, R. McN. Alexander, M. B. Bennett, Why are mammalian tendons so thick? *J. Zool.* **216**, 309–324 (1988). doi: 10.1111/j.1469-7998.1988.tb02432.x

45. J. F. V. Vincent, U. G. K. Wegst, Design and mechanical properties of insect cuticle. *Arthropod Struct. Dev.* **33**, 187–199 (2004). doi: 10.1016/j.asd.2004.05.006; pmid: 18089034

46. M. V. Rosario, G. P. Sutton, S. N. Patek, G. S. Sawicki, Muscle-spring dynamics in time-limited, elastic movements. *Proc. R. Soc. London Ser. B* **283**, 20161561 (2016). doi: 10.1098/rspb.2016.1561; pmid: 27629031

47. M. Burrows, S. R. Shaw, G. P. Sutton, Resilin and chitinous cuticle form a composite structure for energy storage in jumping by froghopper insects. *BMC Biol.* **6**, 41 (2008). doi: 10.1186/1741-7007-6-41; pmid: 18826572

48. M. Burrows, G. P. Sutton, Locusts use a composite of resilin and hard cuticle as an energy store for jumping and kicking. *J. Exp. Biol.* **215**, 3501–3512 (2012). doi: 10.1242/jeb.071993; pmid: 22693029

49. G. P. Sutton, M. Burrows, Biomechanics of jumping in the flea. *J. Exp. Biol.* **214**, 836–847 (2011). doi: 10.1242/jeb.052399; pmid: 21307071

50. S. N. Patek, M. V. Rosario, J. R. A. Taylor, Comparative spring mechanics in mantis shrimp. *J. Exp. Biol.* **216**, 1317–1329 (2013). doi: 10.1242/jeb.078998; pmid: 23239886

51. J. Gosline *et al.*, Elastic proteins: Biological roles and mechanical properties. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357**, 121–132 (2002). doi: 10.1098/rstb.2001.1022; pmid: 11911769

52. D. Raabe, C. Sachs, P. Romano, The crustacean exoskeleton as an example of a structurally and mechanically graded biological nanocomposite material. *Acta Mater.* **53**, 4281–4292 (2005). doi: 10.1016/j.actamat.2005.05.027

53. J. F. V. Vincent, Arthropod cuticle: A natural composite shell system. *Compos., Part A Appl. Sci. Manuf.* **33**, 1311–1315 (2002). doi: 10.1016/S1359-835X(02)00167-7

54. D. Klocke, H. Schmitz, Water as a major modulator of the mechanical properties of insect cuticle. *Acta Biomater.* **7**, 2935–2942 (2011). doi: 10.1016/j.actbio.2011.04.004; pmid: 21515418

55. M. V. Rosario, S. N. Patek, Multilevel analysis of elastic morphology: The mantis shrimp's spring. *J. Morphol.* **276**, 1123–1135 (2015). doi: 10.1002/jmor.20398; pmid: 26195244

56. D. W. Haldane, M. M. Plecnik, J. K. Yim, R. S. Fearing, Robotic vertical jumping agility via series-elastic power modulation. *Sci. Robot.* **1**, eaag2048 (2016). doi: 10.1126/scirobotics.aag2048

57. D. Rollinson, S. Ford, B. Brown, H. Choset, in *Proceedings of the ASME 2013 Dynamic Systems and Control Conference.* (ASME, 2013) p. v001T08A002.

58. B. Hopkinson, A method of measuring the pressure produced in the detonation of high explosives or by the impact of bullets. *Phil. Trans. R. Soc. A* **213**, 437–456 (1914). doi: 10.1098/rsta.1914.0010

59. K. W. Hillier, H. Kolsky, An investigation of the dynamic elastic properties of some high polymers. *Proc. Phys. Soc. London. Sect. B* **62**, 111–121 (1949). doi: 10.1088/0370-1301/62/2/304

60. B. Justusson, M. Pankow, C. Heinrich, M. Rudolph, A. M. Waas, Use of a shock tube to determine the bi-axial yield of an aluminum alloy under high rates. *Int. J. Impact Eng.* **58**, 55–65 (2013). doi: 10.1016/j.ijimpeng.2013.01.012

61. J. Yi, M. C. Boyce, G. F. Lee, E. Balizer, Large deformation rate-dependent stress–strain behavior of polyurea and polyurethanes. *Polymer* **47**, 319–329 (2006). doi: 10.1016/j.polymer.2005.10.107

62. G. H. Staab, A. Gilat, A direct-tension split Hopkinson bar for high strain-rate testing. *Exp. Mech.* **31**, 232–235 (1991). doi: 10.1007/BF02326065

63. M. Hudspeth *et al.*, High speed synchrotron x-ray phase contrast imaging of dynamic material response to split Hopkinson bar loading. *Rev. Sci. Instrum.* **84**, 025102 (2013). doi: 10.1063/1.4789780; pmid: 23464246

64. A. D. Mulliken, M. C. Boyce, Mechanics of the rate-dependent elastic–plastic deformation of glassy polymers from low to high strain rates. *Int. J. Solids Struct.* **43**, 1331–1356 (2006). doi: 10.1016/j.ijsolstr.2005.04.016

65. C. C. Chen, J. Y. Chueh, H. Tseng, H. M. Huang, S. Y. Lee, Preparation and characterization of biodegradable PLA polymeric blends. *Biomaterials* **24**, 1167–1173 (2003). doi: 10.1016/S0142-9612(02)00466-0; pmid: 12527257

66. B. Wetzel, P. Rosso, F. Haupert, K. Friedrich, Epoxy nanocomposites: Fracture and toughening mechanisms. *Eng. Fract. Mech.* **73**, 2375–2398 (2006). doi: 10.1016/j.engfracmech.2006.05.018

67. C. R. Siviour, J. L. Jordan, High strain rate mechanics of polymers: A review. *J. Dyn. Behav. Mat.* **2**, 15–32 (2016). doi: 10.1007/s40870-016-0052-8

68. R. B. Bogoslovov, C. M. Roland, Viscoelastic effects on the free retraction of rubber. *J. Appl. Phys.* **102**, 063531 (2007). doi: 10.1063/1.2784018

69. P. H. Mott et al., Comparison of the transient stress-strain response of rubber to its linear dynamic behavior. *J. Polym. Sci., B, Polym. Phys.* **49**, 1195–1202 (2011). doi: 10.1002/polb.22292

70. C. C. Lawrence, G. J. Lake, A. G. Thomas, The deformation and fracture of balloons. *Int. J. Non-linear Mech.* **68**, 59–65 (2015). doi: 10.1016/j.ijnonlinmec.2014.08.009

71. J. G. Niemczua, "On the response of rubbers at high strain rates" (Tech. Rep., Sandia National Laboratories, Albuquerque, NM, and Livermore, CA 2010).

72. R. Vermorel, N. Vandenberghe, E. Villermaux, Rubber band recoil. *Proc. R. Soc. A* **463**, 641–658 (2007). doi: 10.1098/rspa.2006.1781

73. L. B. Tunnicliffe, A. G. Thomas, J. J. C. Busfield, The free retraction of natural rubber: A momentum-based model. *Polym. Test.* **47**, 36–41 (2015). doi: 10.1016/j.polymertesting.2015.07.012

74. C. M. Roland, Mechanical behavior of rubber at high strain rates. *Rubber Chem. Technol.* **79**, 429–459 (2006). doi: 10.5254/1.3547945

75. K. Kagaya, S. N. Patek, Feed-forward motor control of ultrafast, ballistic movements. *J. Exp. Biol.* **219**, 319–333 (2016). doi: 10.1242/jeb.130518; pmid: 26643091

76. A. Sakes et al., Shooting mechanisms in nature: A systematic review. *PLOS ONE* **11**, e0158277 (2016). doi: 10.1371/journal.pone.0158277; pmid: 27454125

77. M. S. Rodgers, J. J. Allen, K. D. Meeks, B. D. Jensen, S. L. Miller, in *Proceedings of SPIE* (1999), pp. 212–222.

78. R. E. Fischell, L. Wilson, Spacecraft application of subliming materials. *J. Spacecr. Rockets* **2**, 376–379 (1965). doi: 10.2514/3.28187

79. W. Gronenberg, J. Tautz, B. Hölldobler, Fast trap jaws and giant neurons in the ant odontomachus. *Science* **262**, 561–563 (1993). doi: 10.1126/science.262.5133.561; pmid: 17733239

80. Y. Forterre, Slow, fast and furious: Understanding the physics of plant movements. *J. Exp. Bot.* **64**, 4745–4760 (2013). doi: 10.1093/jxb/ert230; pmid: 23913956

81. R. Ritzmann, Snapping behavior of the shrimp *Alpheus californiensis*. *Science* **181**, 459–460 (1973). doi: 10.1126/science.181.4098.459; pmid: 17793337

82. F. Liu et al., Asymmetric drop coalescence launches fungal ballistospores with directionality. *J. R. Soc. Interface* **14**, 20170083 (2017). doi: 10.1098/rsif.2017.0083; pmid: 28747394

83. W. J. Heitler, The locust jump. *J. Comp. Physiol.* **89**, 93–104 (1974). doi: 10.1007/BF00696166

84. G. Bonsignori et al., The green leafhopper, *Cicadella viridis* (Hemiptera, Auchenorrhyncha, Cicadellidae), jumps with near-constant acceleration. *J. Exp. Biol.* **216**, 1270–1279 (2013). doi: 10.1242/jeb.076083; pmid: 23487271

85. F. C. Moon, *The Machines of Leonardo Da Vinci and Franz Reuleaux: Kinematics of Machines from the Renaissance to the 20th Century* (Springer Science & Business Media, 2007).

86. J. M. Keep, Animal trap, Patent, U.S. Patent Office (1879).

87. M. M. Muñoz, P. S. L. Anderson, S. N. Patek, Mechanical sensitivity and the dynamics of evolutionary rate shifts in biomechanical systems. *Proc. R. Soc. London Ser. B* **284**, 20162325 (2017). doi: 10.1098/rspb.2016.2325; pmid: 28100817

88. T. Claverie, E. Chan, S. N. Patek, Modularity and scaling in fast movements: Power amplification in mantis shrimp. *Evolution* **65**, 443–461 (2011). doi: 10.1111/j.1558-5646.2010.01133.x; pmid: 20840593

89. T. Claverie, S. N. Patek, Modularity and rates of evolutionary change in a power-amplified prey capture system. *Evolution* **67**, 3191–3207 (2013). doi: 10.1111/evo.12185; pmid: 24152002

90. P. S. L. Anderson, D. C. Smith, S. N. Patek, Competing influences on morphological modularity in biomechanical

systems: A case study in mantis shrimp. *Evol. Dev.* **18**, 171–181 (2016). doi: 10.1111/ede.12190; pmid: 27161948

91. P. S. L. Anderson, S. N. Patek, Mechanical sensitivity reveals evolutionary dynamics of mechanical systems. *Proc. R. Soc. London Ser. B* **282**, 20143088 (2015). doi: 10.1098/rspb.2014.3088; pmid: 25716791

92. J. Aguilar, A. Lesov, K. Wiesenfeld, D. I. Goldman, Lift-off dynamics in a simple jumping robot. *Phys. Rev. Lett.* **109**, 174301 (2012). doi: 10.1103/PhysRevLett.109.174301; pmid: 23215192

93. N. W. Bartlett et al., A 3D-printed, functionally graded soft robot powered by combustion. *Science* **349**, 161–165 (2015). doi: 10.1126/science.aab0129; pmid: 26160940

94. S. Vogel, *Glimpses of Creatures in Their Physical Worlds* (Princeton Univ. Press, Princeton 2009).

95. S. Vogel, Living in a physical world III. Getting up to speed. *J. Biosci.* **30**, 303–312 (2005). doi: 10.1007/BF02703667; pmid: 16052068

96. S. Vogel, Living in a physical world II. The bio-ballistics of small projectiles. *J. Biosci.* **30**, 167–175 (2005). doi: 10.1007/BF02703696; pmid: 15886452

97. H. C. Bennet-Clark, E. C. Lucey, The jump of the flea: A study of the energetics and a model of the mechanism. *J. Exp. Biol.* **47**, 59–67 (1967). pmid: 6058981

98. M. Burrows, M. D. Picker, Jumping mechanisms and performance of pygmy mole crickets (Orthoptera, Tridactylidae). *J. Exp. Biol.* **213**, 2386–2398 (2010). doi: 10.1242/jeb.042192; pmid: 20581268

99. M. Burrows, Jumping performance of froghopper insects. *J. Exp. Biol.* **209**, 4607–4621 (2006). doi: 10.1242/jeb.02539; pmid: 17114396

100. M. Burrows, G. P. Sutton, Pygmy mole crickets jump from water. *Curr. Biol.* **22**, R990–R991 (2012). doi: 10.1016/j.cub.2012.10.045; pmid: 23218011

101. S. E. Bergbreiter, thesis, University of California, Berkeley (2007).

102. L. G. Leal, *Advanced Transport Phenomena: Fluid Mechanics and Convective Transport Processes*, vol. 7 (Cambridge Univ. Press, 2007).

103. T. Nüchter, M. Benoit, U. Engel, S. Ozbek, T. W. Holstein, Nanosecond-scale kinetics of nematocyst discharge. *Curr. Biol.* **16**, R316–R318 (2006). doi: 10.1016/j.cub.2006.03.089; pmid: 16682335

104. E. Purcell, Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977). doi: 10.1119/1.10903

105. S. Spagnolie, *Complex Fluids in Biological Systems: Experiment, Theory, and Computation* (Springer, New York 2014).

106. A. W. Koch et al., Spinalin, a new glycine- and histidine-rich protein in spines of Hydra nematocysts. *J. Cell Sci.* **111**, 1545–1554 (1998). pmid: 9580562

107. A. Pringle, S. N. Patek, M. Fischer, J. Stolze, N. P. Money, The captured launch of a ballistospore. *Mycologia* **97**, 866–871 (2005). doi: 10.1080/15572536.2006.11832777; pmid: 16457355

108. J. Edwards, D. Whitaker, S. Klionsky, M. J. Laskowski, Botany: A record-breaking pollen catapult. *Nature* **435**, 164 (2005). doi: 10.1038/435164a; pmid: 15889081

109. D. L. Whitaker, L. A. Webster, J. Edwards, The biomechanics of *Cornus canadensis* stamens are ideal for catapulting pollen vertically. *Funct. Ecol.* **21**, 219–225 (2007). doi: 10.1111/j.1365-2435.2007.01249.x

110. S. N. Patek, J. E. Baio, B. L. Fisher, A. V. Suarez, Multifunctionality and mechanical origins: Ballistic jaw propulsion in trap-jaw ants. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12787–12792 (2006). doi: 10.1073/pnas.0604290103; pmid: 16924120

111. M. Burrows, Jumping mechanisms in jumping plant lice (Hemiptera, Sternorrhyncha, Psyllidae). *J. Exp. Biol.* **215**, 3612–3621 (2012). doi: 10.1242/jeb.074682; pmid: 22771753

112. M. J. McHenry et al., The comparative hydrodynamics of rapid rotation by predatory appendages. *J. Exp. Biol.* **219**, 3399–3411 (2016). doi: 10.1242/jeb.140590; pmid: 27807217

113. R. S. James, R. S. Wilson, Explosive jumping: Extreme morphological and physiological specializations of Australian rocket frogs (*Litoria nasuta*). *Physiol. Biochem. Zool.* **81**, 176–185 (2008). doi: 10.1086/525290; pmid: 18190283

114. C. V. Anderson, T. Sheridan, S. M. Deban, Scaling of the ballistic tongue apparatus in chameleons. *J. Morphol.* **273**, 1214–1226 (2012). doi: 10.1002/jmor.20053; pmid: 22730103

115. C. V. Anderson, Off like a shot: Scaling of ballistic tongue projection reveals extremely high performance in small

chameleons. *Sci. Rep.* **6**, 18625 (2016). doi: 10.1038/srep18625; pmid: 26725508

116. H. C. Bennet-Clark, The energetics of the jump of the locust *Schistocerca gregaria*. *J. Exp. Biol.* **63**, 53–83 (1975). pmid: 1159370

117. A. P. Gerratt, S. Bergbreiter, Incorporating compliant elastomers for jumping locomotion in microrobots. *Smart Mater. Struct.* **22**, 014010 (2013). doi: 10.1088/0964-1726/22/1/014010

118. A. Debray, Manipulators inspired by the tongue of the chameleon. *Bioinspir. Biomim.* **6**, 026002 (2011). doi: 10.1088/1748-3182/6/2/026002; pmid: 21422504

119. J. S. Koh, S. P. Jung, M. Noh, S. W. Kim, K. J. Cho, in *IEEE International Conference on Robotics and Automation* (IEEE, 2013), pp. 26–31.

120. J. Zhao et al., MSU Jumper: A single-motor-actuated miniature steerable jumping robot. *IEEE Trans. Robot.* **29**, 602–614 (2013). doi: 10.1109/TRO.2013.2249371

121. P. Weiss, Hop… Hop… Hopbots! *Sci. News* **159**, 88–91 (2001). doi: 10.2307/3981566

122. J. German, Sandia-developed intelligent software agents challenge electronic intruders. *Sandia Lab News* **52**(10), LN05-19-00 (19 May 2000); http://www.sandia.gov/LabNews/LN05-19-00/software_story.html.

123. S. N. Patek, B. N. Nowroozi, J. E. Baio, R. L. Caldwell, A. P. Summers, Linkage mechanics and power amplification of the mantis shrimp's strike. *J. Exp. Biol.* **210**, 3677–3688 (2007). doi: 10.1242/jeb.006486; pmid: 17921168

124. M. Versluis, B. Schmitz, A. von der Heydt, D. Lohse, How snapping shrimp snap: Through cavitating bubbles. *Science* **289**, 2114–2117 (2000). doi: 10.1126/science.289.5487.2114; pmid: 11000111

125. H. C. Astley, T. J. Roberts, Evidence for a vertebrate catapult: Elastic energy storage in the plantaris tendon during frog jumping. *Biol. Lett.* **8**, 386–389 (2012). doi: 10.1098/rsbl.2011.0982; pmid: 22090204

126. H. C. Astley, T. J. Roberts, The mechanics of elastic loading and recoil in anuran jumping. *J. Exp. Biol.* **217**, 4372–4378 (2014). doi: 10.1242/jeb.110296; pmid: 25520385

127. X. Noblin, S. Yang, J. Dumais, Surface tension propulsion of fungal spores. *J. Exp. Biol.* **212**, 2835–2843 (2009). doi: 10.1242/jeb.029975; pmid: 19684219

128. L. Mahadevan, P. Matsudaira, Motility powered by supramolecular springs and ratchets. *Science* **288**, 95–99 (2000). doi: 10.1126/science.288.5463.95; pmid: 10753126

129. J. H. de Groot, J. L. van Leeuwen, Evidence for an elastic projection mechanism in the chameleon tongue. *Proc. R. Soc. London Ser. B* **271**, 761–770 (2004). doi: 10.1098/rspb.2003.2637; pmid: 15209111

130. U. K. Müller, S. Kranenbarg, Power at the tip of the tongue. *Science* **304**, 217–219 (2004). doi: 10.1126/science.1097894; pmid: 15073361

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

POLYMERS

A synthetic polymer system with repeatable chemical recyclability

Jian-Bo Zhu, Eli M. Watson, Jing Tang, Eugene Y.-X. Chen*

The development of chemically recyclable polymers offers a solution to the end-of-use issue of polymeric materials and provides a closed-loop approach toward a circular materials economy. However, polymers that can be easily and selectively depolymerized back to monomers typically require low-temperature polymerization methods and also lack physical properties and mechanical strengths required for practical uses. We introduce a polymer system based on γ-butyrolactone (GBL) with a trans-ring fusion at the α and β positions. Such trans-ring fusion renders the commonly considered as nonpolymerizable GBL ring readily polymerizable at room temperature under solvent-free conditions to yield a high–molecular weight polymer. The polymer has enhanced thermostability and can be repeatedly and quantitatively recycled back to its monomer by thermolysis or chemolysis. Mixing of the two enantiomers of the polymer generates a highly crystalline supramolecular stereocomplex.

V arious approaches (*1–5*) have been pursued to address the unsustainable annual generation and disposal of several hundred million metric tons of synthetic polymers, with the goal of a circular plastics economy (*6*). The use of renewable resources as feedstock materials (*7, 8*) generally does not address materials' end-of-use problems. The development of biodegradable polymers (*9*) for biological recycling (*10*) also provides a partial solution but fails to recover valuable building block chemicals. Degraded materials, especially those that only partially degrade, can also cause unintended environmental consequences. Mechanical reprocessing (*11*) tends to degrade the quality of the polymers. In contrast, chemical recycling (*12, 13*) can allow for recovery of the precursor building block chemicals via depolymerization or creative reuse or repurposing through the generation of value-added materials (*14–16*). With specifically designed monomers, reaction conditions can be used to select the direction of the monomer-polymer equilibrium or the closed-loop chemical cycle, with low temperatures and bulk or high monomer concentrations favoring polymerization and high temperatures or dilution triggering depolymerization. Several classes of recently designed recyclable polymers operate under this thermodynamic principle, such as poly[2-(2-hydroxyethoxybenzoate)] (*17, 18*), poly(β-methyl-δ-valerolactone) (*19*), and a polycarbonate (PC) derived from copolymerization of CO_2 with a *meso*-epoxide (*20*). Poly[2-(2-hydroxyethoxybenzoate)] exhibited relatively low glass (~27°C), melting (~69°C), and decomposition (~146°C) temperatures; the thermostability of the PC was also

Department of Chemistry, Colorado State University, Fort Collins, CO 80523-1872, USA.
*Corresponding author. Email: eugene.chen@colostate.edu

limited (below 260°C), and its depolymerization underwent decarboxylation.

However, the chemical recycling approach still faces challenges, including the selectivity involved in chemical recycling processes and circular monomer-polymer-monomer cycles, as well as trade-offs between polymers' depolymerizability and properties. A notable example for depolymerization selectivity is biodegradable poly(L-lactide) [P(L-LA)], which produces a mixture of many products upon thermolysis (*21*) or a mixture of LA stereoisomers and cyclic oligomers upon chemolysis with an Sn catalyst (*22*), thus requiring substantial separation and purification before the recovered L-LA can be reused. Polymers with a low ceiling temperature (T_c) (*23, 24*) are readily depolymerizable under mild conditions, but they typically do not have robust enough physical and mechanical properties to be useful for most common applications. For example, poly(γ-butyrolactone) (PGBL), synthesized via catalyzed ring-opening polymerization (ROP) of the renewable, nonstrained, thermodynamically highly stable five-membered γ-butyrolactone (GBL) (*25, 26*), can be selectively and quantitatively depolymerized back to GBL upon heating of the bulk material at 260° or 300°C, depending on PGBL topology (*27, 28*). However, the synthesis of PGBL requires energy-intensive, industrially undesirable low-temperature conditions (typically –40°C), and PGBL exhibits limited thermostability and crystallinity, with a low melting transition temperature (T_m) of ~60°C. Another example of a completely recyclable polymer was produced through the chemoselective ROP of bioderived α-methylene-γ-butyrolactone; however, not only was a low temperature (–60°C) required for the polymer synthesis, but the resulting polymer was also a noncrystalline amorphous material (*29*).

Room temperature, solvent-free polymerization to high–molecular weight polymers

To design monomer and polymer structures that can deliver desired properties, it is advantageous to keep the highly stable, five-membered GBL core so that the complete chemical recyclability of the designed polymers can be preserved (for both thermodynamic and kinetic reasons). We reasoned that the ring strain of the parent (nonstrained five-membered GBL), or the thermodynamic polymerizability, can be tuned via suitable substituents and substitution patterns on the GBL ring. *trans*-Hexahydro-2(3H)-benzofuranone [i.e., 4,5-*trans* six-membered ring–fused GBL (4,5-T6GBL)] was found to be polymerizable even at 40°C by typical anionic initiators but not by a coordination polymerization catalyst such as tin(II) octoate, whereas the cis-fused isomer is completely inert toward ROP (*30*). However, the resulting product was reported to be only an oligomer, with a number-average molecular weight (M_n) up to only 6.2 kg/mol [by gel permeation chromatography (GPC)] or 2.6 kg/mol [by nuclear magnetic resonance (NMR)]. We hypothesized that removing the substituent at the 5 (or γ) position of the GBL ring could not only further enhance the thermodynamic polymerizability (by increasing the ring strain) and rate of polymerization (by releasing the steric pressure at the ester –$OC_\alpha H$), thus affording useful high–molecular weight polymers in short time periods, but also render polymers with possibly high crystallinity assisted by ordered secondary structures and/or $OC_\alpha H$–O=C hydrogen bonds (*31–33*) between polyester chains. Guided by these hypotheses, we arrived at 3,4-T6GBL (**M1**), where the cyclohexyl ring is trans-fused to GBL at the α and β positions and the γ position is left unsubstituted to enhance polymerizability, reaction rates, and H bonding. This monomer can be prepared on relatively large scales from commercially available *trans*-1,2-cyclohexanecarboxylic acid anhydride (*34*).

The polymerizability of **M1** was probed via measuring the thermodynamics of its ROP with a discrete molecular catalyst, yttrium complex **Y1** (*35*) (Fig. 1A), which is known to be effective for the ROP of the parent GBL (*27*), revealing standard-state thermodynamic parameters of $\Delta H°_p$ (change in enthalpy of polymerization) = –20 kJ mol^{-1} and $\Delta S°_p$ (change in entropy of polymerization) = –72 J mol^{-1} K^{-1} (figs. S9 and S10). The T_c was calculated to be 0°, 62°, or 88°C for an initial **M1** concentration ([**M1**]$_0$) of 1.0, 5.0, or 8.2 (bulk) M, respectively. These data showed that **M1** exhibits much higher thermodynamic polymerizability than GBL, as indicated by a much larger negative change in enthalpy and a substantially elevated T_c: $\Delta H°_p$ = –20 kJ mol^{-1} versus –5.4 kJ mol^{-1} and T_c = 0°C (1.0 M) versus –136°C (1.0 M) for the ROP of **M1** and GBL, respectively. Accordingly, we chose the solvent-free, bulk condition to perform the polymerization at room temperature, as summarized in table S1. Quantitative conversion of **M1** was achieved even with common anionic initiators such as potassium *tert*-butoxide and TBD (1,5,7-triazabicyclo[4.4.0]dec-5-ene), but

the product was not that of a polymer; instead, **M1** was isomerized under the basic conditions to its cis isomer, which is nonpolymerizable, as verified by its independent synthesis and subsequent polymerization surveys with different catalysts and conditions (table S2). To overcome this isomerization issue, we used a coordinative-insertion ROP catalyst, La[N(SiMe_3)_2]_3 (where Me is methyl) (**La1**), as La is more earth abundant and less expensive within the lanthanide series (36) and was demonstrated to be effective for the ROP of GBL (27). **La1** exhibited high selectivity and activity toward the ROP to afford poly-**M1** [P(**M1**)], achieving greater than 80% conversion with 0.2 or 0.1 mole % (mol %) catalyst loading. For example, with a monomer (M)/catalyst (cat)/initiator (I) ratio of [**M1**]/[**La1**]/[Ph_2CHCH_2OH] (where Ph is phenyl) = 500/1/3, the ROP occurred rapidly to reach 73% **M1** conversion in under 1 min. At 82% conversion, the isolated P(**M1**) was measured by a GPC instrument equipped with multi (18)–angle light scattering and differential refractive index detectors to have a medium M_n of 21.0 kg/mol and an extremely low dispersity index ($Đ$) value of 1.01; this measured M_n is close to the calculated M_n of 19.4 kg/mol, thus indicating a high initiation efficiency of 92%. Lowering the catalyst loading to 0.1 mol % (1000/1/3 ratio) still achieved a relatively high conversion of 84%, producing P(**M1**) with M_n = 46.0 kg/mol and $Đ$ = 1.01. To further increase the molecular weight of the resulting P(**M1**), the M/cat/I ratio was increased to 5000/1/3 with a catalyst loading of 0.02 mol %, affording P(**M1**) a further enhanced M_n (67.9 kg/mol) and a still low $Đ$ (1.01) at 45% conversion. The absolute molecular weight measured by GPC

(M_n = 21.0 kg/mol, $Đ$ = 1.01) (table S1, run 2) was close to both the M_n of 19.4 kg/mol calculated on the basis of the [**M1**]/[I] ratio (table S1) and the M_n of 19.9 kg/mol calculated on the basis of the chain ends of the polymer characterized by NMR (figs. S11 and S12), the latter of which also revealed a linear structure {linear P(**M1**) [*l*-P(**M1**)]} and showed high end-group fidelity. The linear structure of the P(**M1**) produced by **La1** with ROH (where R is Ph_2CHCH_2O) was further confirmed by matrix-assisted laser desorption/ionization–time-of-flight mass spectroscopy (MALDI-TOF MS) of a low–molecular weight sample. Specifically, the MS spectrum (fig. S13) consisted of only one series of molecular ion peaks, with the spacing between the two neighboring molecular ion peaks corresponding to the exact molar mass of the repeat unit, **M1** [mass/charge ratio (m/z) = 140.18], as shown by the slope (140.14) of the linear plot of m/z values (y axis) versus the number of **M1** repeat units (x axis). The intercept of the plot, 221.8, represents the total mass of chain ends plus the mass of Na⁺ [M_{end} = 198.3 (Ph_2CHCH_2O/H) g/mol + 23.0 (Na⁺) g/mol], corresponding to linear structure Ph_2CHCH_2O–[**M1**]_n–H.

With the demonstrated ability of **La1** to form PGBL with a cyclic structure when no initiating ROH is added (27), we explored the possibility of producing cyclic polymer *c*-P(**M1**) by using **La1** alone. The polymerizations with different catalyst loadings (1.0 to 0.05 mol % **La1**) were rapid and achieved relatively high monomer conversions, from 79 to 84%. Unlike the controlled polymerization with **La1** and 3ROH, where the [**M1**]/[**La1**] ratio could determine the M_n while maintain-

ing a low $Đ$ value of 1.01 for all the *l*-P(**M1**) materials produced (table S1), the polymerizations with **La1** alone afforded polymers with similar M_n values in a narrow range of 73.0 to 85.5 kg/mol coupled with higher $Đ$ values from 1.34 to 1.48, despite a 20-fold change in the [**M1**]/[**La1**] ratio and a 7-fold change in the reaction scale (table S1). These observations indicated that possible cyclization reactions occurred under the neat and room temperature conditions used once a certain chain length was reached, which would create a cyclic polymer structure, *c*-P(**M1**). Consistent with this proposed scenario, no end groups were detected from the NMR spectra of the polymers produced by using **La1** alone (figs. S14 to S16). Analysis of a low–molecular weight sample by MALDI-TOF MS (fig. S17) also revealed no end groups (the linear plot of m/z values versus the number of **M1** repeat units gave an intercept of 23, corresponding to the mass of Na⁺). The cyclic structure of *c*-P(**M1**) was further confirmed by GPC analysis with triple detection by a light-scattering detector, a refractometer, and a viscometer. A double-logarithm (Mark-Houwink-Sakurada) plot of intrinsic viscosity [η] versus absolute weight-average molecular weight (M_w) in the lower molecular weight regime (Fig. 1B) showed a lower intrinsic viscosity for *c*-P(**M1**), due to a smaller hydrodynamic volume, than for *l*-P(**M1**). The [η]_cyclic/[η]_linear ratio was found to be 0.7, which is in agreement with the theoretically predicted value for this ratio and the experimentally observed value for other cyclic polymers (37). Expanding to an optically active cyclic polymer structure, we synthesized two chiral polymers with similar M_w values from the ROP of (R)-**M1** with
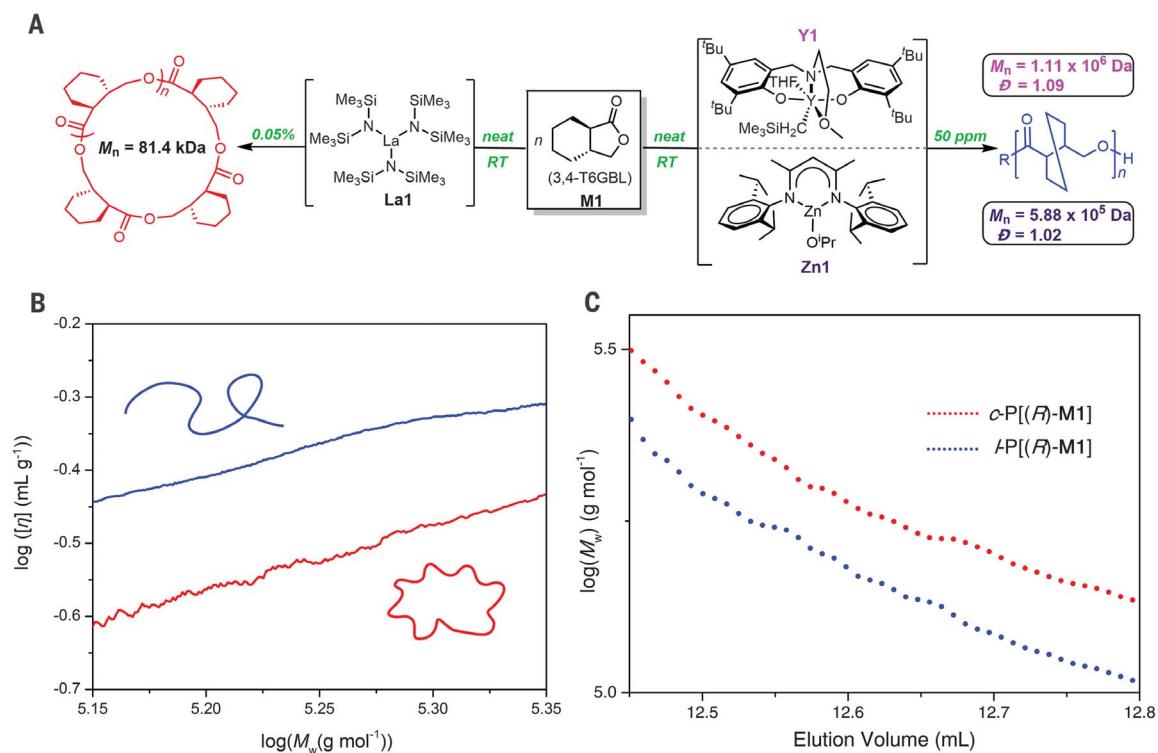

Fig. 1. Structures, intrinsic viscosity, and elution behavior. (**A**) Structures of the monomer, catalysts, and the resulting linear and cyclic polymers. RT, room temperature; *t*-Bu, *tert*-butyl; THF, tetrahydrofuran; ⁱPr, isopropyl. (**B**) Double-logarithm (Mark-Houwink-Sakurada) plots of intrinsic viscosity [η] versus absolute M_w of linear (blue) and cyclic (red) P(**M1**) produced by **La1** with ROH and **La1** alone, respectively. (**C**) Logarithm plots of M_w versus the elution volume of optically active linear and cyclic P[(R)-**M1**] produced by **La1** with ROH and **La1** alone, respectively.

Fig. 2. Thermostability and chemical recyclability of P(M1). (**A**) TGA curves for *c*-P(**M1**) obtained with [**M1**]/[**La1**] = 2000/1 (red) and *l*-P(**M1**) obtained with [**M1**]/[**La1**]/[Ph$_2$CHCH$_2$OH] = 500/1/3 (blue) and a comparative example for the linear PGBL obtained with [GBL]/[**La1**]/[Ph$_2$CHCH$_2$OH] = 400/1/3 (black). (**B**) Overlays of DTG curves for *l*-P(**M1**) obtained with **Y1** (orange), P[(*R*)-**M1**] with **Y1** (blue), P[(*S*)-**M1**] with **Y1** (red), and a 1:1 P[(*R*)-**M1**]–P[(*S*)-**M1**] blend (green) and a comparative example for linear PGBL (black). (**C**) Overlays of ^{1}H NMR spectra (25°C, CDCl$_3$, with residual solvent peaks at 7.26 and 1.56 ppm for CHCl$_3$ and H$_2$O, respectively): (a) *c*-P(**M1**); (b) the colorless liquid product recovered after depolymerization (toluene, 2 mol % ZnCl$_2$, 120°C, 24 hours); (c) clean-starting **M1** for comparison.
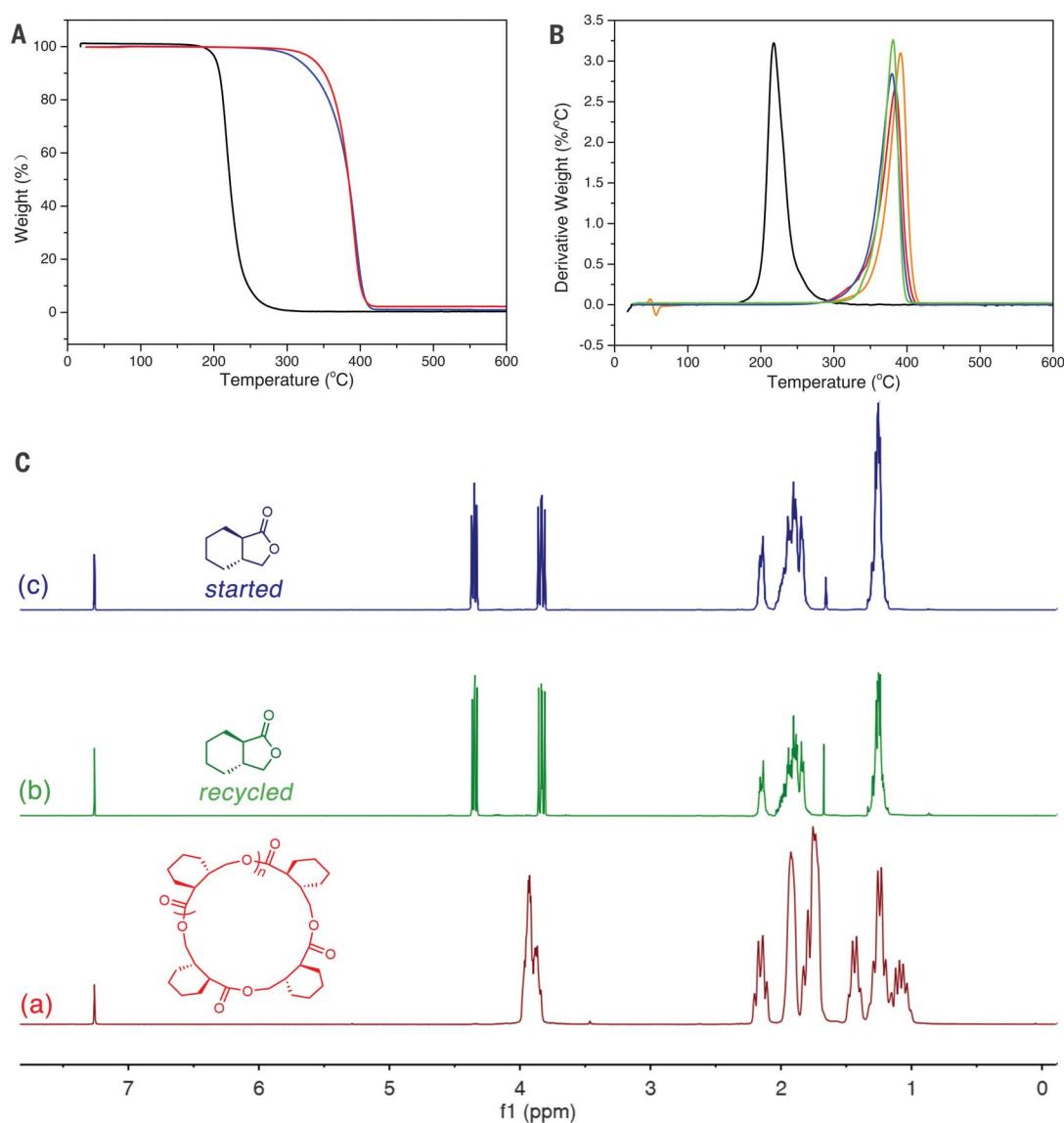
La1 and ROH and with **La1** alone. The logarithm plots of M_w versus the elution volume revealed that the chiral polymer P[(*R*)-**M1**] obtained in the absence of ROH was eluted later than the polymer obtained in the presence of ROH (Fig. 1C). By analogy to achiral P(**M1**), it can be likewise assigned to a cyclic structure because its hydrodynamic volume is smaller than that of its linear analog.

The synthesis of pure cyclic polymers with appreciable (medium to high) molecular weights, which are critical for polymer topology–property relationship studies, still presents a challenge for many types of polymers. This problem has led to substantial interest in developing synthetic methodologies for cyclic polymers (*38, 39*), such as the ring-opening metathesis polymerization route to cyclic polyethylene (*40*) and the *N*-heterocyclic carbene–mediated zwitterionic polymerization route to cyclic polylactide (*41*) and poly(α-peptoid)s (*42*). The cleaner formation of *c*-P(**M1**) than of cyclic PGBL (*27*), as demonstrated by the NMR, MS spectra, and [η]$_{cyclic}$/[η]$_{linear}$ ratio, is noteworthy.

As the molecular weight of the cyclic polymer is limited by the propensity to cyclize once a certain chain length is reached during the polymerization, we explored other catalyst-initiator systems that could further increase the molecular weight of *l*-P(**M1**). In this context, we arrived at discrete yttrium complex **Y1** supported by the tetradentate amino-bisphenolate ligand bearing a pendant ether group, which has been shown to be highly efficient in the ROP of lactide and lactones (*43, 44*). With **Y1** as the catalyst and catalyst loading as low as 50 parts per million (ppm), we achieved relatively high **M1** conversions of 80 to 91% (table S1). Thus, in a 10-g polymerization, a high–molecular weight P(**M1**) with $M_n = 1.11 \times 10^6$ g/mol (*Đ* = 1.09) was readily produced under neat and room temperature conditions with only 50 ppm **Y1**. Furthermore, the molecular weight of the resulting polymer could be readily controlled by the addition of the ROH (Ph$_2$CHCH$_2$OH) initiator; thus, with the [**M1**]/[**Y1**] ratio of 10000/1 held constant, the equiva-

lent of ROH added (relative to **Y1**) was varied from 25 to 50 to 100, affording *l*-P(**M1**) with correspondingly reduced M_n values from 49.2 kg/mol (*Đ* = 1.01) to 18.4 kg/mol (*Đ* = 1.01) to 11.6 kg/mol (*Đ* = 1.01) (table S1). These M_n values were close to the M_n values calculated on the basis of the [**M1**]/[ROH] ratio and **M1** conversion data, thus demonstrating a high initiation efficiency of >96%. Overall, **Y1** brings about "immortal" ROP (*45, 46*) of **M1** in the presence of ROH, producing well-defined *l*-P(**M1**) materials with near ideal dispersity (1.01) in a catalytic fashion (affording up to 100 polymer chains per **Y1**).

An environmentally more benign and earth-abundant zinc catalyst, 2,6-diisopropylphenyl-substituted β-diiminate zinc isopropoxide complex [(BDI)ZnO^{i}Pr] (**Zn1**) (*47*), was also examined for the ROP of **M1** at room temperature under neat conditions. The results, summarized in table S1, showed that **Zn1** is also highly effective for this polymerization. For example, with catalyst loading of 0.02 mol %, the ROP achieved 82% conversion
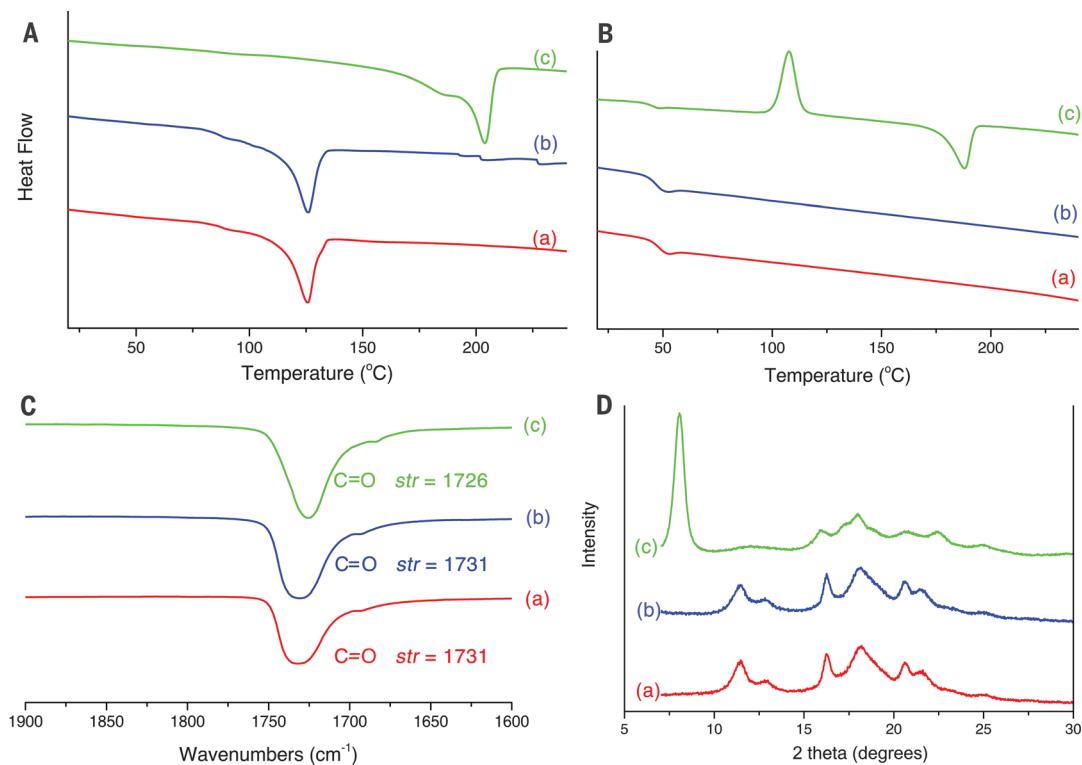
Fig. 3. Thermal transitions and spectroscopic properties. (**A**) DSC first-heating-scan curves (10°C/min). (**B**) DSC second-heating-scan curves [5°C/min for (a) and (b) or 1°C/min for (c), after first cooling at 10°C/min]. (**C**) Overlays of FTIR spectra in the carbonyl stretching region. str, stretching frequency. (**D**) Powder XRD profiles. In all cases, l-P(**M1**) polymers were prepared with [**M1**]/[**Y1**] = 1000/1: l-P[(S)-**M1**] (a, red), l-P[(R)-**M1**] (b, blue), and 1:1 l-P[(R)-**M1**]–l-P[(S)-**M1**] blend or sc-P(**M1**) (c, green).

in 3 hours, producing P(**M1**) with a high M_n of 307 kg/mol and an extremely narrow $Đ$ of 1.01. Comparing these data with those obtained with **Y1**—12 hours, 87% conversion, M_n = 363 kg/mol, and $Đ$ = 1.15—under the same conditions shows that, under this set of conditions (room temperature, neat conditions, 0.02 mol % catalyst), the Zn catalyst performed better than the Y catalyst in terms of its higher polymerization activity and lower polymer dispersity.

Thermostability, chemical recyclability, and circular M1-P(M1)-M1 cycle

The thermostability of both linear and cyclic P(**M1**) materials was examined by thermogravimetric analysis (TGA) in terms of onset decomposition temperature T_d (defined by the temperature of 5% weight loss) and maximum decomposition temperature T_{max} [defined by the peak value in the relative derivative thermogravimetry (DTG)]. Comparing TGA curves of the cyclic and linear polymers revealed that c-P(**M1**) produced with **La1** alone had a noticeably higher T_d (337°C) than l-P(**M1**) obtained with [**M1**]/[**La1**]/ [Ph$_2$CHCH$_2$OH] = 500/1/3, with T_d = 316°C (Fig. 2A), although the T_{max} values were similar (390° versus 394°C) (figs. S22 and S23). The T_d and T_{max} of l-P(**M1**) were 115° and 176°C higher, respectively, than those of the linear PGBL (obtained with [GBL]/[**La1**]/[Ph$_2$CHCH$_2$OH] = 400/ 1/3), which had a T_d of 201°C and a T_{max} of 218°C. The l-P(**M1**) produced by **Y1** ([**M1**]/[**Y1**] = 2000) was even more thermally robust, with a high T_d of 342°C and a T_{max} of 391°C (Fig. 2B). The T_d (344°C) of the physical blend of 1:1 P[(R)-**M1**]– P[(S)-**M1**] was found to be about 17 to 21°C

higher than those of the respective enantiomeric polymers (figs. S25 to S29), whereas the T_{max} values varied only slightly (Fig. 2B).

The chemical recyclability of P(**M1**) materials was examined by both thermolysis at high temperatures and chemolysis in the presence of a catalyst at milder temperatures. An l-P(**M1**) sample (M_n = 46.0 kg/mol, $Đ$ = 1.01) prepared with [**M1**]/[**La1**]/[Ph$_2$CHCH$_2$OH] = 1000/1/3 was heated in a sealed tube at ≥300°C for 1 hour; gravimetric and NMR analyses (figs. S30 and S31) of the resulting colorless liquid showed that monomer **M1** was recovered in a pure state at a quantitative yield. Likewise, heating a c-P(**M1**) sample (M_n = 82.0 kg/mol, $Đ$ = 1.43) prepared with [**M1**]/[**La1**] = 100/1 at ≥300°C for 24 hours also afforded the recycled monomer in a pure state (figs. S32 and S33) at a quantitative yield. To reduce the energy input in the recycling process, the chemical recyclability of P(**M1**) was also investigated by chemolysis with a catalytic amount of a simple metal salt (ZnCl$_2$) at 120°C. Thus, subjecting either l-P(**M1**) (figs. S34 and S35) or c-P(**M1**) (Fig. 2C and figs. S36 and S37) to the above-listed mild chemolysis conditions also demonstrated the full chemical recyclability.

The circular monomer-polymer-monomer cycle was examined through three consecutive polymerization-depolymerization cycles on a multigram scale. Thus, pure **M1** was first polymerized by **Zn1** to well-defined P(**M1**) ($Đ$ = 1.01) after achieving 85% conversion, which is the typical conversion achieved under the ambient temperature and neat conditions used (the unreacted monomer can be recovered). The isolated and purified P(**M1**) was then subjected to

chemolysis in the presence of a simple metal salt (ZnCl$_2$, 2 mol %) at 180°C under vacuum pressure (0.01 torr); the collected colorless liquid was confirmed to be pure **M1** by ^{1}H NMR analysis (figs. S38 to S44). The recovered monomer (97% isolated yield) was repolymerized directly without further purification by **Zn1** to produce well-defined P(**M1**) ($Đ$ = 1.02), achieving the same conversion (85%). This process was repeated three times, and the mass balance of the regenerated polymer product and the recovered monomer was tracked over the three consecutive polymerization-depolymerization cycles, showing essentially quantitative recovery of pure **M1** (96 to 97% isolated yield) after each cycle. This recovered **M1** can be directly repolymerized without a decrease in the subsequent monomer conversion and polymer quality.

Physical blending to yield highly crystalline stereocomplexed material

Physical blending of enantiomers of certain chiral polyesters in a stoichiometric ratio offers a powerful strategy to generate crystalline stereocomplexed (sc) materials that often exhibit much enhanced materials properties, such as increased T_m and crystallization rate, compared with their constituent enantiomers (48–52). In this context, a 1:1 physical blend of enantiomeric isotactic polymers (table S3 and figs. S48 to S53) derived from enantiomeric monomers, either linear enantiomers l-P[(R)-**M1**] and l-P[(S)-**M1**] produced with [**M1**]/[**Y1**] = 1000/1 or cyclic enantiomers c-P[(R)-**M1**] and c-P[(S)-**M1**] produced with [**M1**]/[**La1**] = 500/1, showed substantially different thermal properties and solubility as well

Fig. 4. Mechanical and rheological properties.
(**A**) Overlay of storage modulus E' and loss modulus E'' for sc-P(**M1**) measured by DMA (tension film mode, 0.05% strain, 1 Hz, 3°C min^{-1}). (**B**) Stress-strain curves for rac-P(**M1**) and sc-P(**M1**) measured by tensile testing (5.0 mm/min, room temperature, with the break point indicated by ×). (**C**) Rheology master curve (dynamic storage modulus G' and loss modulus G'' versus angular frequency ω) for rac-P(**M1**), reported as a time-temperature superposition curve at reference temperature 215°C. (**D**) Dynamic shear viscosity of rac-P(**M1**) as a function of the shear rate measured at 215°C.
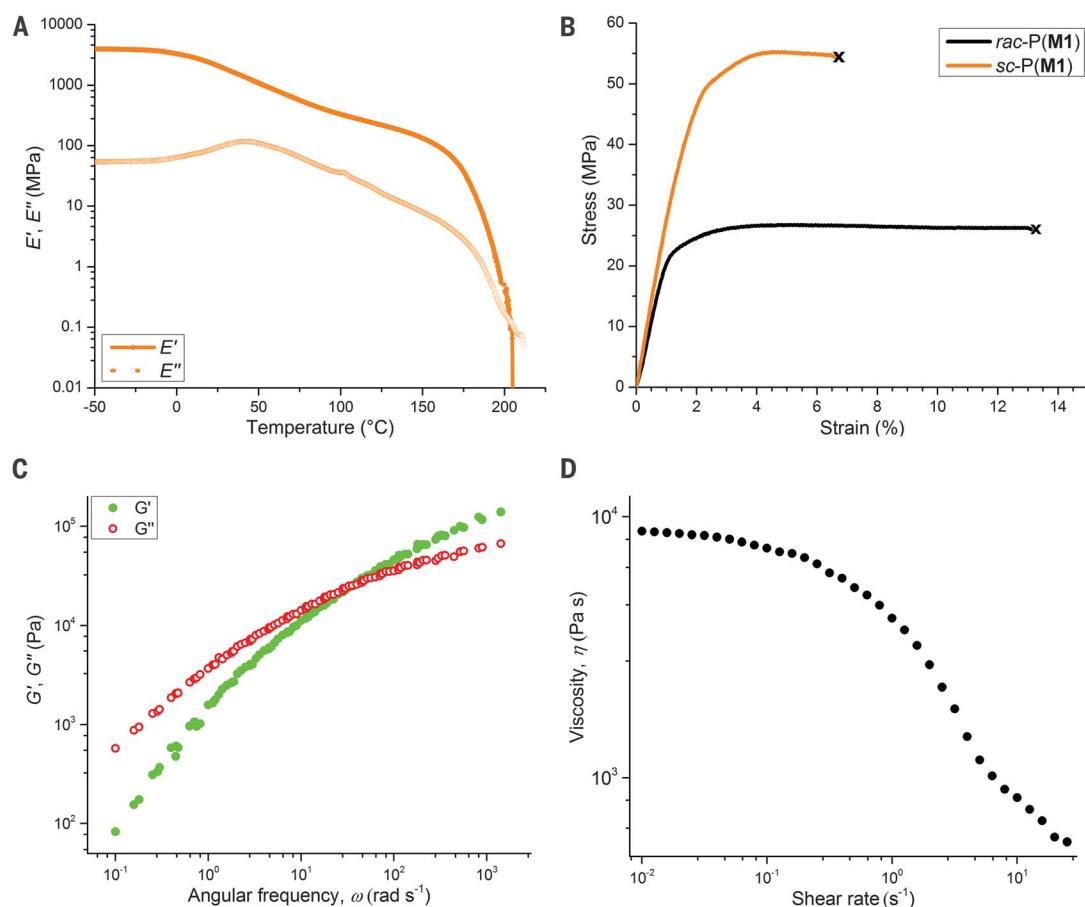
as noticeably different spectroscopic features from the parent enantiomers. Differential scanning calorimetry (DSC) curves for the first heating scans (Fig. 3A) of the linear enantiomeric polymers, previously crystallized from CHCl$_3$, displayed a crystalline peak with T_m = 126°C (heat of fusion ΔH_m = 28 to 32 J/g), but the 1:1 blend produced much higher melting and heat of fusion values, with T_m = 203°C and ΔH_m = 53 J/g. More notably, on the second heating scans (after cooling at 10°C/min) only the physical blend continued to show a melting peak of T_m = 188°C (T_c = 108°C), whereas the enantiomeric polymers became amorphous, displaying only a glass transition temperature T_g of ~49°C (Fig. 3B). These results indicate that these enantiomeric polymers have relatively low crystallization rates and that stereocomplexation in the blend markedly enhanced not only the crystallinity but also the crystallization rate. Linear enantiomeric polymers produced by the [**La1**]-3[Ph$_2$CHCH$_2$OH] system displayed more or less the same DSC curves on the first and second heating scans (figs. S55 and S56). Furthermore, a comparison of DSC curves for c-P(**M1**) showed the same trend: The first heating scans revealed a T_m of 127°C (ΔH_m = 34 to 37 J/g) for the enantiopure polymers but a much higher T_m of 198°C and ΔH_m of 61 J/g for the c-P[(R)-**M1**]–c-P[(S)-**M1**] blend (figs. S65 and S66), and the second heating scans showed a melting peak

of T_m = 188°C (T_c = 122°C) only for the blend (fig. S66).

Overlays of Fourier transform infrared (FTIR) spectra in the carbonyl stretching region (Fig. 3C and fig. S70) revealed the red shift of the C=O stretching frequency ($v_{C=O}$) for the blend of the two linear enantiomeric polymers to a wave number 5 cm^{-1} lower than that for the parent enantiomers. Likewise, a red shift of 7 cm^{-1} was also observed for the blend of the two cyclic enantiomers relative to the parent enantiomers (figs. S71 and S72). These results are consistent with the hypothesis that the blend forms a stereocomplex, sc-P(**M1**), assisted by the weak to moderate OC$_\alpha$H–O=C hydrogen bonds. Powder x-ray diffraction (XRD) profiles (Fig. 3D) of enantiomeric l-P[(R)-**M1**] [(a) in Fig. 3D] and l-P[(S)-**M1**] (prepared using [**M1**]/[**Y1**] = 1000/1), as well as their 1:1 l-P[(R)-**M1**]–l-P[(S)-**M1**] blend or sc-P(**M1**), all crystallized from CHCl$_3$, revealed substantially different crystalline diffraction patterns of the blend in comparison with the parent enantiomers. Chiefly, whereas the two enantiomeric polymers showed identical patterns (consisting of four major diffraction signals at 11.5°, 16.2°, 18.2°, and 20.6°, along with three minor peaks at 12.8°, 21.6°, and 25.2°), the blend exhibited a new, intense diffraction peak at 8.1° [d spacing (the spacing between adjacent planes) = 1.1 nm] and was also devoid of the two signals (major,

11.5°, and minor, 12.8°) present in the enantiomers, which is attributable to the formation of the stereocomplex. Although P[(R)-**M1**] or P[(S)-**M1**] is readily soluble in common polar organic solvents such as CHCl$_3$, CH$_2$Cl$_2$, and N,N'-dimethylformamide (DMF), sc-P(**M1**) is only partially soluble in CHCl$_3$ and insoluble in CH$_2$Cl$_2$, tetrahydrofuran, and DMF. Overall, the above corroborative evidence showed that a nanocrystalline stereocomplex formed between the two enantiomeric P(**M1**) polymers and that such a stereocomplex exhibited markedly enhanced crystallinity, crystallization rate, and solvent resistance over those of the parent enantiomers.

In contrast, mixing of enantiomeric polymers P[(+)-4,5-T6GBL] and P[(−)-4,5-T6GBL] (table S3) in a 1:1 stoichiometric ratio, followed by crystallization, yielded a physical blend exhibiting thermal properties and spectroscopic characteristics essentially identical to those of either the starting enantiomeric polymer or the racemic polymer P[(±)-4,5-T6GBL]. DSC curves (fig. S45) for the two enantiomeric polymers and their physical blend displayed the same features of an amorphous material, with a T_g of ~72°C. FTIR spectra (figs. S46 and S47) also revealed the same absorption features, with an identical $v_{C=O}$ stretching frequency of 1725 cm^{-1} observed for all of the three polymers (two enantiomeric polymers and their 1:1 blend). Overall, the above collective

evidence showed that no stereocomplexation occurred upon mixing of these two enantiomeric polymers, and thus no enhancement of properties occurred through blending.

Mechanical and rheological properties

The thermomechanical properties of the amorphous polymer derived from racemic **M1** [rac-P(**M1**)] prepared with **Y1** (M_n = 875 kg/mol, Đ = 1.07) (table S1, run 19) and semicrystalline stereocomplex sc-P(**M1**) prepared with **Y1** (table S3, runs 4 and 8) were examined by dynamic mechanical analysis (DMA) in a tension film mode. The thermomechanical spectra of sc-P(**M1**) (Fig. 4A) and rac-P(**M1**) (fig. S83) show that, at room temperature (the glassy state), both sc-P(**M1**) and rac-P(**M1**) exhibited high storage modulus (E') values, although E' (1.58 ± 0.44 GPa) of sc-P(**M1**) was somewhat higher than that (1.47 ± 0.25 GPa) of rac-P(**M1**). However, after the glass transition region with similar T_g values [83 to 90°C, as defined by the peak maxima of tan δ, the loss modulus/storage modulus ratio (E''/E')] (fig. S84), E' of rac-P(**M1**) dropped by more than three orders of magnitude and then quickly went to the viscous flow state. In contrast, E' of sc-P(**M1**) decreased only by approximately one order of magnitude after T_g, and the material still maintained a high E' in the rubbery plateau until reaching a flow temperature of ~180°C, characteristic of a semicrystalline material having a high T_m (186°C by DSC).

Tensile testing of dog-bone–shaped specimens of rac-P(**M1**) and sc-P(**M1**) yielded stress-strain curves (Fig. 4B), revealing that semicrystalline sc-P(**M1**) exhibited a much higher ultimate tensile strength (σ_B = 54.7 ± 4.0 MPa) and Young's modulus (E = 2.72 ± 0.25 GPa) than amorphous rac-P(**M1**) (σ_B = 26.2 ± 3.2 MPa, E = 1.85 ± 0.30 GPa). As a glassy material, sc-P(**M1**) displayed an elongation at break (ε_B = 6.5 ± 1.2%), and the ε_B value for rac-P(**M1**) was approximately doubled, with ε_B = 13.1 ± 3.5%. Overall, the key thermal and mechanical properties of the crystalline P(**M1**) (T_g ~ 50°C, T_m ~ 188°C, T_d ~ 340°C; σ_B ~ 55 MPa, E ~ 2.7 GPa, ε_B ~ 7%) compare well to those of typical crystalline P(L-LA) materials (T_g ~ 54° to 59°C, T_m ~ 159° to 178°C, T_d ~ 235° to 255°C; σ_B ~ 28 to 50 MPa, E ~ 1.2 to 3.0 GPa, ε_B ~ 2 to 6%) (53).

The angular frequency (ω) dependencies of the dynamic storage or elastic modulus (G') and loss or viscous modulus (G'') of rac-P(**M1**) and sc-P(**M1**) were characterized at six different temperatures (165°, 175°, 185°, 195°, 205°, and 215°C) in the linear viscoelastic regime (1.0% strain) established by the strain sweeps at 215°C (figs. S85 and S86). The data obtained from frequency sweep experiments at each temperature were compiled to generate a master curve reported as a time-temperature superposition curve at reference temperature 215°C [Fig. 4C for rac-P(**M1**) and fig. S87 for sc-P(**M1**)]. A G' and G'' crossover point where G' becomes larger than G'', indicating the transition from the terminal (viscous) to the rubbery (elastic) region, was seen for both rac-P(**M1**) and sc-P(**M1**). The crossover

frequencies measured at 165°, 175°, 185°, 195°, 205°, and 215°C for rac-P(**M1**) were found to be 0.43, 0.75, 1.21, 1.95, 3.18, and 4.89 Hz, corresponding to relaxation times of 2.32, 1.33, 0.83, 0.51, 0.31, and 0.21 s, respectively. The crossover frequencies observed for sc-P(**M1**) were more than six times higher at each of the same temperatures, thus giving rise to much shorter relaxation times, from 0.34 s (165°C) to 0.03 s (215°C). For the high–molecular weight rac-P(**M1**) (M_n = 875 kg/mol), the melt processability was preliminarily tested by examining the dynamic melt viscosity as a function of the shear rate measured at 215°C (Fig. 4D), showing that shear thinning started to develop at a low shear rate of ~0.1 s^{-1} and became pronounced at ~1 s^{-1}.

Summary

This work introduces a solution to three key challenges facing the development of chemically recyclable polymers: selectivity in depolymerization, trade-offs between polymers' depolymerizability and their properties and performance, and a circular monomer-polymer-monomer cycle. The results showed that, with judiciously designed monomer and polymer structures, it is possible to create chemically recyclable polymers that exhibit quantitative recyclability and useful materials properties and that the polymer synthesis and recycling processes can be performed under ambient or industrially relevant conditions.

REFERENCES AND NOTES

1. X. Zhang, M. Fevre, G. O. Jones, R. M. Waymouth, *Chem. Rev.* **118**, 839–885 (2018).
2. D. K. Schneiderman, M. A. Hillmyer, *Macromolecules* **50**, 3733–3749 (2017).
3. J. M. Garcia, M. L. Robertson, *Science* **358**, 870–872 (2017).
4. J. M. Eagan et al., *Science* **355**, 814–816 (2017).
5. X. Jia, C. Qin, T. Friedberger, Z. Guan, Z. Huang, *Sci. Adv.* **2**, e1501591 (2016).
6. World Economic Forum, Ellen MacArthur Foundation, and McKinsey & Company, "The new plastics economy: Rethinking the future of plastics" (Ellen MacArthur Foundation, 2016); www.ellenmacarthurfoundation.org/publications/the-new-plastics-economy-rethinking-the-future-of-plastics.
7. M. A. Hillmyer, *Science* **358**, 868–870 (2017).
8. Y. Zhu, C. Romain, C. K. Williams, *Nature* **540**, 354–362 (2016).
9. A.-C. Albertsson, M. Hakkarainen, *Science* **358**, 872–873 (2017).
10. S. Yoshida et al., *Science* **351**, 1196–1199 (2016).
11. N. Torres, J. J. Robin, B. Boutevin, *Eur. Polym. J.* **36**, 2075–2080 (2000).
12. M. Hong, E. Y.-X. Chen, *Green Chem.* **19**, 3692–3706 (2017).
13. A. Rahimi, J. M. García, *Nat. Rev. Chem.* **1**, 0046 (2017).
14. J. M. García et al., *Science* **344**, 732–735 (2014).
15. G. O. Jones, A. Yuen, R. J. Wojtecki, J. L. Hedrick, J. M. García, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7722–7726 (2016).
16. K. Fukushima et al., *J. Polym. Sci. A Polym. Chem.* **49**, 1273–1281 (2011).
17. J. P. MacDonald, M. P. Shaver, *Polym. Chem.* **7**, 553–559 (2016).
18. E. Lizundia, V. A. Makwana, A. Larrañaga, J. L. Vilas, M. P. Shaver, *Polym. Chem.* **8**, 3530–3538 (2017).
19. D. K. Schneiderman et al., *ACS Macro Lett.* **5**, 515–518 (2016).
20. Y. Liu, H. Zhou, J. Z. Guo, W. M. Ren, X. B. Lu, *Angew. Chem. Int. Ed.* **56**, 4862–4866 (2017).
21. F.-D. Kopinke, M. Remmler, K. Mackenzie, M. Möder, O. Wachsen, *Polym. Degrad. Stab.* **53**, 329–342 (1996).
22. H. Nishida et al., *Polym. Degrad. Stab.* **81**, 515–523 (2003).
23. C. E. Diesendruck et al., *Nat. Chem.* **6**, 623–628 (2014).
24. J. A. Kaitz, C. E. Diesendruck, J. S. Moore, *J. Am. Chem. Soc.* **135**, 12755–12761 (2013).
25. K. N. Houk, A. Jabbari, H. K. Hall Jr., C. Alemán, *J. Org. Chem.* **73**, 2674–2678 (2008).
26. A. Duda, A. Kowalski, in *Handbook of Ring-Opening Polymerization*, P. Dubois, O. Coulembier, J.-M. Raquez, Eds. (Wiley-VCH, 2009), chap. 1.
27. M. Hong, E. Y.-X. Chen, *Nat. Chem.* **8**, 42–49 (2016).
28. M. Hong, E. Y.-X. Chen, *Angew. Chem. Int. Ed.* **55**, 4188–4193 (2016).
29. X. Tang et al., *J. Am. Chem. Soc.* **138**, 14326–14337 (2016).
30. H. Haba, H. Itabashi, *Polym. J.* **46**, 89–93 (2014).
31. S. M. Anderson, B. K. Mueller, E. J. Lange, A. Senes, *J. Am. Chem. Soc.* **139**, 15774–15783 (2017).
32. J.-R. Sarasua, N. L. Rodriguez, A. L. Arraiza, E. Meaurio, *Macromolecules* **38**, 8362–8371 (2005).
33. R. Vargas, J. Garza, D. A. Dixon, B. P. Hay, *J. Am. Chem. Soc.* **122**, 4750–4755 (2000).
34. P. D. Kennewell, S. S. Matharu, J. B. Taylor, R. Westwood, P. G. Sammes, *J. Chem. Soc. Perkin Trans. 1* **0**, 2563–2570 (1982).
35. A. Amgoune, C. M. Thomas, T. Roisnel, J.-F. Carpentier, *Chemistry* **12**, 169–179 (2005).
36. A. S. Dudnik, V. L. Weidner, A. Motta, M. Delferro, T. J. Marks, *Nat. Chem.* **6**, 1100–1107 (2014).
37. J. A. Semlyen, Ed., *Cyclic Polymers* (Kluwer Academic, ed. 2, 2000).
38. K. Endo, *Adv. Polym. Sci.* **217**, 121–183 (2008).
39. H. A. Brown, R. M. Waymouth, *Acc. Chem. Res.* **46**, 2585–2596 (2013).
40. C. W. Bielawski, D. Benitez, R. H. Grubbs, *Science* **297**, 2041–2044 (2002).
41. D. A. Culkin et al., *Angew. Chem. Int. Ed.* **46**, 2627–2630 (2007).
42. L. Guo, D. Zhang, *J. Am. Chem. Soc.* **131**, 18072–18074 (2009).
43. J.-F. Carpentier, *Organometallics* **34**, 4175–4189 (2015).
44. A. Amgoune, C. M. Thomas, S. Ilinca, T. Roisnel, J.-F. Carpentier, *Angew. Chem. Int. Ed.* **45**, 2782–2784 (2006).
45. T. Aida, S. Inoue, *Acc. Chem. Res.* **29**, 39–48 (1996).
46. Y. Wang, W. Zhao, X. Liu, D. Cui, E. Y.-X. Chen, *Macromolecules* **45**, 6957–6965 (2012).
47. M. Cheng, A. B. Attygalle, E. B. Lobkovsky, G. W. Coates, *J. Am. Chem. Soc.* **121**, 11583–11584 (1999).
48. J. M. Longo, A. M. DiCiccio, G. W. Coates, *J. Am. Chem. Soc.* **136**, 15897–15900 (2014).
49. J. Slager, A. J. Domb, *Adv. Drug Delivery Rev.* **55**, 549–583 (2003).
50. J. Kumaki, T. Kawauchi, K. Okoshi, H. Kusanagi, E. Yashima, *Angew. Chem. Int. Ed.* **46**, 5348–5351 (2007).
51. K. Fukushima, Y. Kimura, *Polym. Int.* **55**, 626–642 (2006).
52. A. M. Liquori, et al., *Nature* **206**, 358–362 (1965).
53. I. Engelberg, J. Kohn, *Biomaterials* **12**, 292–304 (1991).

REGENERATION

Self-organization and progenitor targeting generate stable patterns in planarian regeneration

Kutay Deniz Atabay,[1,2,3] Samuel A. LoCascio,[1,2,3] Thom de Hoog,[1,3] Peter W. Reddien[1,3,4]*

During animal regeneration, cells must organize into discrete and functional systems. We show that self-organization, along with patterning cues, govern progenitor behavior in planarian regeneration. Surgical paradigms allowed the manipulation of planarian eye regeneration in predictable locations and numbers, generating alternative stable neuroanatomical states for wild-type animals with multiple functional ectopic eyes. We used animals with multiple ectopic eyes and eye transplantation to demonstrate that broad progenitor specification, combined with self-organization, allows anatomy maintenance during regeneration. We propose a model for regenerative progenitors involving (i) migratory targeting cues, (ii) self-organization into existing or regenerating eyes, and (iii) a broad zone, associated with coarse progenitor specification, in which eyes can be targeted by progenitors. These three properties help explain how tissues can be organized during regeneration.

T he capacity to regenerate is widespread and variable throughout the animal kingdom. There is much current interest in understanding how regenerative progenitors reform missing organs and tissues. Complex tissue architectures can be formed through an adaptive, nonlinear, and dynamic process called self-organization, described by the emergence of orderly structure through local interactions between the basic elements of a system (*1, 2*). In self-organizing systems, ordered patterns emerge from an initially disordered and noisy environment. Self-organization has been invoked in embryonic development (*3–9*), regeneration (*10, 11*), and the in vitro formation of complex structures such as the optic cup (*12*) and organoids (*13, 14*). To explore the role of self-organization in regeneration, we investigated the ability of planarians to completely regenerate missing organs.

Planarians are flatworms with a remarkable ability to regenerate complex tissues, such as their centralized nervous system. Planarians also continually renew all cells in their bodies via cellular turnover, involving dividing cells called neoblasts, which include pluripotent stem cells (*15*) and diverse specialized neoblasts that are progenitors for different cell types. Specialized neoblasts can be specified in broad spatial domains and produce migratory progenitors that converge and differentiate at defined target locations, such as the eyes (Fig. 1A) and brain (*16–18*).

Transient discordance between anatomy and positional information in regeneration

As a target tissue for studying self-organization in regeneration, we selected planarian eyes. Pla-

narians have two eyes composed of photoreceptor neurons, which project axons ipsilaterally and contralaterally to the brain, and pigmented optic cup cells (Fig. 1A) (*17*). Because the eyes are visible and dispensable for animal viability, they are amenable to extensive manipulation for regeneration studies. Eyes are located at two predictable bilaterally symmetric positions relative to the existing body size and are regenerated and maintained during tissue turnover from *ovo*+ eye-specific progenitor cells (Fig. 1, A and B, and figs. S1A and S2A) (*16, 17, 19*). *ovo*+ progenitors are formed coarsely in the dorsal prepharyngeal region and migrate anteriorly, where they can incorporate into existing eyes or nucleate eyes de novo (fig. S1, A to D) (*17*).

Understanding how progenitors are targeted to a specific location and organize into new tissues is central to understanding regeneration. We realized that progenitors can face conflicting targeting choices during regeneration and that understanding how this dilemma is resolved could explain key regeneration principles. Synthesis of two pieces of information reveals the existence of these conflicting choices. First, restoration of full anatomy and proportions in planarian regeneration involves both blastema formation (new tissue growth at wounds) and substantial changes (called morphallaxis) to the remaining body fragment itself (*20*). Morphallaxis involves new cell production, cell death (*21*), and the remodeling of differentiated tissues (*22*). During morphallaxis, existing organs and major tissues are retained but gradually change their proportions and relative positions (Fig. 1, C and D). Whereas new tissue formation in a blastema occurs within days, morphallaxis can take up to several weeks (depending on degree of proportion resolution needed and nutrient status). For example, after decapitation, the eyes and brain of a head fragment gradually (over weeks) shrink and move anteriorly, while continuously maintaining their form and function.

Second, regenerating correct planarian anatomy depends upon regionally and constitutively

expressed genes called position control genes (PCGs), which are proposed to act in muscle to establish adult positional coordinates (*23, 24*). Inhibition of multiple PCGs (such as *bmp4*, *wntP-2*, and *ndl-3*) by means of RNA interference (RNAi) causes patterning phenotypes, such as ectopic heads, pharynges, or eyes (*23–26*). After amputation, PCG expression domains are rapidly reestablished in muscle (in a matter of 2 to 3 days). For example, within 48 hours after decapitation, a head fragment initiates posterior PCG expression at the wound, and anterior PCG expression domains shift toward the head tip. Initial PCG expression domain regeneration does not require neoblasts (*23*) and precedes new differentiated tissue generation.

PCG expression domain regeneration upon amputation proceeds faster—by days to weeks— than do changes to the scale and position of existing differentiated anatomy during morphallaxis. Therefore, there is a substantial period of mismatch between PCG expression patterns and pattern of underlying anatomy in planarian regeneration (Fig. 1E and fig. S3, A to E). For example, for several days during regeneration of head fragments, eyes and brain are mis-positioned with respect to PCG expression domains but are located in the correct position with respect to remaining differentiated tissues (Fig. 1E and fig. S3, A to E). Eyes slowly shrink and move anteriorly (morphallaxis) to ultimately align anatomy and PCG expression domains (Fig. 1, C and D). Regardless of whether PCGs directly control this process, their expression changes indicate that rapid positional information shifting during regeneration leads to its discordance with anatomy (Fig. 1E and fig. S3, A to E). Because progenitors continuously target the eye during tissue turnover (*17*), we presume that eye progenitors are incorporated into existing eyes during morphallaxis. These factors together suggest the targeting choice dilemma that regenerative progenitors must resolve: Should the cells target their "correct" anatomical location or their "correct" position with respect to positional coordinates during this period of positional information-anatomy discordance in regeneration?

Dynamic positional coordinates guide regenerative progenitor targeting

We refer to the position where progenitors nucleate in de novo organ or tissue formation as the "target zone" (TZ). We assessed targeting decisions of eye progenitors by amputating heads and then unilaterally resecting one eye after 3 days (Fig. 1F). Prior work demonstrated that eye resection does not cause eye progenitor amplification, with eye regeneration occurring as an emergent property of constant progenitor production and progenitor-target equilibrium dynamics (*19*). Therefore, we anticipated that eye progenitors would form new eye cells at equal rates on both sides of these head fragments. Does the regenerating eye nucleate at the original, correct anatomical location, or does it form at the new, correct position with respect to shifting positional coordinates (Fig. 1F)? As is predicted by a model in which positional information guides progenitors and shifts early in regeneration, de novo eye nucleation occurred at

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: reddien@wi.mit.edu

more anterior locations than the position of the remaining eye, generating asymmetric animals (Fig. 2A and fig. S4, A and B). The original and regenerating eyes were both targeted by progenitors, which originated posterior to the eyes, demonstrating that eye progenitors targeted two different locations in these head fragments (Fig. 2B and fig. S4).

We predicted that if positional coordinates were shifted in different directions, then different asymmetric eye configurations would emerge. We tested this prediction by removing the anterior head tip, which shifts positional information posteriorly (fig. S3C), and unilaterally resecting eyes 3 days later. Accordingly, regenerating eyes shifted posteriorly (Fig. 2C). These data

indicate that rapidly changing positional information after injury can define the eye progenitor TZ.

Eyes act as self-organizing centers and can trap migratory eye progenitors

In the experiments above, why did eye progenitors not go to their correct position with respect to positional information on the side with the
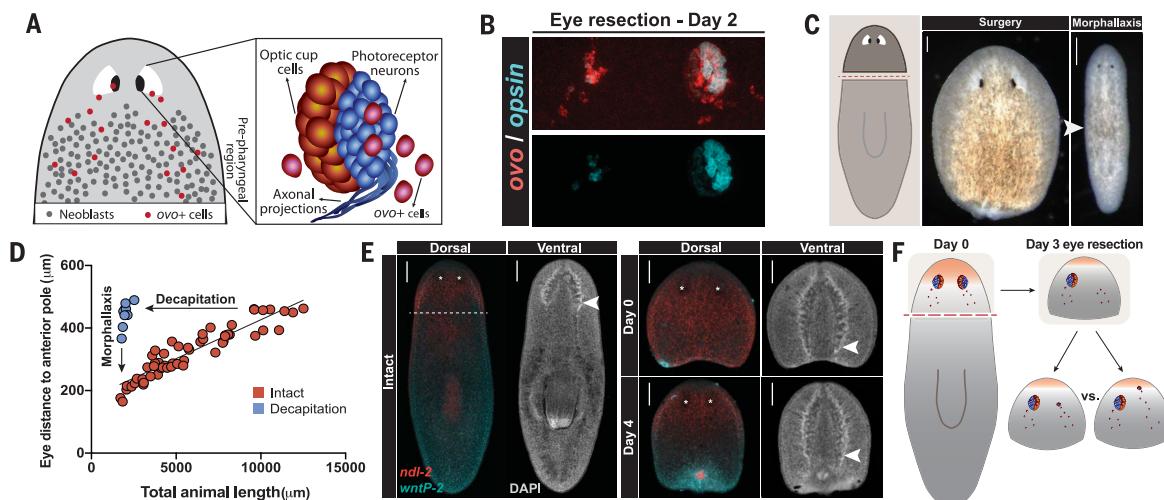
Fig. 1. Discordance between positional coordinates and anatomy during regeneration. (**A**) Eye progenitors emerge from neoblasts in a broad specification zone and migrate anteriorly to form eyes. (**B**) Eye nucleation after unilateral eye resection. (**C**) An amputated head fragment morphallaxed slowly, with eyes moving anteriorly, matching final animal proportions over time (30 days shown). (**D**) Eye positioning occurs relative to existing body size. Head fragments slowly morphallax to resolve

proportions. (**E**) Positional control gene expression in head fragments [*ndl-2* and *wntP-2* fluorescence in situ hybridization (FISH)] rescales before anatomy changes occur. White arrowheads indicate brain posterior. (**F**) Positional information pattern regenerates faster than do anatomical changes. Experimental strategy to reveal a predicted target zone (nucleation target) shift in a decapitated and day 3 eye-resected head fragment. Scale bars, 200 μm.
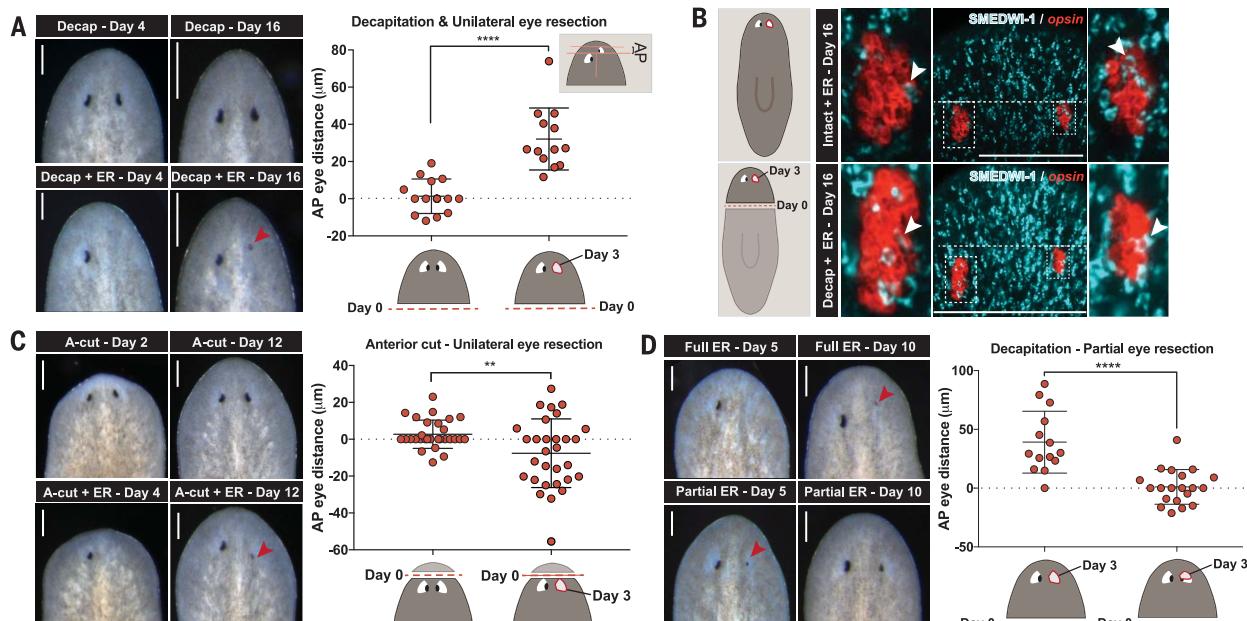
Fig. 2. Planarian eyes act as attractors and renucleate at predictable positions. (**A**) Resected eyes on day 3 after decapitation nucleate more anteriorly. (**B**) Eye progenitors [SMEDWI-1+ (*19*)] target both nonresected eyes and regenerating eyes at different positions in the same animal (white

arrowheads). (**C**) Resected eyes on day 3 after a head-tip cut nucleate more posteriorly. (**D**) Partially resected eyes prevent anterior eye nucleation. ER, eye resection. In (A), (C), and (D), red arrowheads indicate regenerating eyes. Student's *t* test, **$P \le 0.01$, ****$P \le 0.0001$. Scale bars, 200 μm.

remaining eye? We postulate that existing eyes act as stable attractors and "trap" migrating progenitors (Fig. 1A and fig. S5A), preventing them from reaching more anterior positions. In self-organizing biological systems, molecular coupling strategies lead to maintained, functional anatomical architectures (*27*). Self-organizing systems act as attractors in such cases, leading to the stability of shape and function. To test whether existing eyes behave as attractors, we decapitated animals and, after a delay, partially resected eyes. Even a small amount of remaining eye tissue was sufficient to prevent incoming eye progenitors from reaching their new TZ, resulting in eye regeneration at the original eye location (Fig. 2D).

We hypothesized that if eyes act as attractors and have limited or local attractive boundaries, then moving the migration path of incoming eye progenitors outside of such a boundary could cause de novo nucleation of an ectopic third eye (fig. S5B). Parasagittal amputation, lateral to an eye, combined with unilateral eye resection, led to medially shifted eye regeneration (fig. S5C). This raised the possibility that combining decapitation with parasagittal amputation in large animals would allow progenitors to medially evade the attractive boundary of an eye, like light escaping an event horizon. We tested this possibility (also resecting right eyes 3 days after amputations). As predicted, animals formed a third eye, located anteriorly and on the same side (left) as the uninjured eye (Fig. 3, A and B, and fig. S5D). All three eyes were targeted by eye progenitors (Fig. 3, C to E, and fig. S5E), and third eyes extended projections into the visual circuitry (Fig. 3F). In this injury context, eye progenitors

arrived at two different locations on the same side: the original eye and a more anterior location.

We also decapitated and concurrently inflicted parasagittal amputation on three-eyed animals (from Fig. 3A), generating four- and five-eyed animals, with all eyes integrated into visual circuitry (Fig. 3G and fig. S5, G and H). We tested misplaced eye function by examining three-eyed animals (from Fig. 3A); misplaced eyes were alone sufficient to drive negative phototaxis in a light-graded arena (Fig. 3H; fig. S8, A to C; and movies S1 to S4).

Eye progenitors can nucleate in a stable position during dynamic positional information shifting

In the three-eyed animals generated above, we noticed that the third eye (the anterior left eye) (Fig. 3A) was always anterior to the regenerated right eye. This third eye also regenerated later than the right eye did (fig. S5F); we hypothesized that this delay might explain the relative anterior-posterior (AP) eye positions. Specifically, the TZ might have shifted further anteriorly, after right-eye nucleation, by the time the third (medial-left) eye nucleated. This could occur if the uninjured left eye trapped progenitors past the new right eye nucleation time point (fig. S5F). This interpretation predicts that the TZ location moves gradually and continually after amputation; eye position would depend on its time of nucleation after amputation. To test this possibility, we decapitated animals and unilaterally resected right eyes on subsequent days (days 1 to 6). De novo eye nucleation occurred at progressively more anterior locations depending on the time of eye resection after decapitation (Fig. 4A). This indicates that

eye TZ rescaling is a continuous process after injury and that eye regeneration occurs wherever the shifting TZ is located when eye progenitors first nucleate. We further propose that nucleated eyes became fixed in location, growing and incorporating progenitors even when the TZ continued to move away from this position. This self-organizing process would ensure that progenitors do not form a long trail of differentiated cells deposited along a moving target front but as discrete, organized structures, even if the structure forms at a position that is ultimately incorrect.

Consistent with this model, shifting injury timing yielded predictable eye positioning. We resected right eyes on days 1, 3, or 5 after decapitation and left-side parasagittal amputation. As predicted, the later (day 5) right-eye resection group displayed roughly equal positioning of the regenerating right and ectopic left eyes (Fig. 4, B and C). We also observed formation of two eyes on the right side in the day 1 right-eye resection group, leading to four-eyed animals with ectopic, anteriorly shifted left and right eyes (Fig. 4B). The ectopic right eye was always less anterior than the ectopic left eye. This four-eye configuration can be explained with the conceptual model described above: Eye progenitors read their TZ at any particular time point during positional coordinate shifting. After day 1 resection, a new right eye nucleates close to its original location. Because a head fragment is much smaller than the original animal, the theoretically correct eye positions should be more medial on both sides (fig. S6A). Because of coordinate rescaling, migrating progenitors ultimately medially escape the attractive nature of both the left (uninjured) and
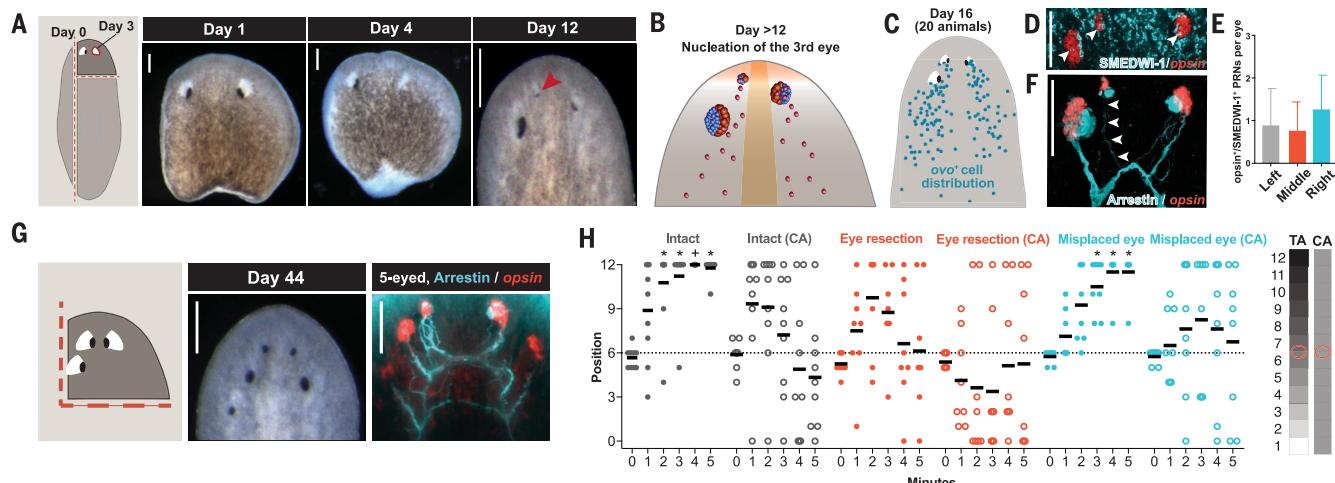
Fig. 3. Generation of an alternative stable neuroanatomical state. (**A** and **B**) Decapitation and parasagittal amputation in large sexual animals plus day 3 unilateral eye resection results in three-eyed animals. Red arrowhead indicates regenerating eye. (**C**) Map of *ovo⁺* cells from 20 three-eyed animals (day 16 after surgery). (**D** and **E**) SMEDWI-1⁺/*opsin*⁺ newly differentiated cells (white arrowheads) were detected in all three eyes (*n* = 144 eyes examined). (**F**) Arrestin immunohistochemistry and *opsin* FISH. Ectopic eyes integrate into visual circuitry. White arrowheads indicate axonal projections.

(**G**) Decapitation and left parasagittal amputation on three-eyed animals resulted in five-eyed animals. (**H**) Behavior analyses: Misplaced eyes drive negative phototaxis. CA, control arena; TA, test arena. Statistical significance: one-sample *t* test comparing each column mean with a hypothetical value of 6 corresponding to chance (*n* = 8 animals per cohort); Bonferroni correction was applied. Scale bars, (A), (D), (E), and (G), 200 μm; (B), (C), and (G), 100 μm. ER, eye resection. Statistical significance: one-way analysis of variance (ANOVA): *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, *****P ≤ 0.0001. NS, not significant.

the newly forming right eye. Because the first-formed right eye is small, progenitors escape its influence before progenitors on the left can escape the influence of the nonresected eye. Therefore, the second-formed right eye will nucleate less anteriorly than the ectopic left eye.

Understanding eye progenitor targeting dynamics allowed even further predictable anatomy changes through simple implementation of injury type and timing. Decapitation and eye resection did not lead to anterior shifting of the brain. By contrast, decapitation combined with sagittal amputation 3 days later caused both the brain and the eye to form more anteriorly on the regenerated side. These animals also generated an anterior third eye on the uninjured side (fig. S7, A and B). These findings suggest that similar self-organizing principles are at play for the brain as well as the eye and that the position of the eye and the brain can be decoupled.

Molecular nature of the target zone

We next explored the molecular attributes of the TZ with PCG RNAi. Medial-lateral (ML) planarian patterning involves *slit* (28). The *slit* medial expression domain is restricted by laterally expressed *wnt5* (29). *wnt5* and *slit* RNAi can affect ML

eye formation (30). After *wnt5* RNAi, serial eye nucleation was observed, progressing laterally as *slit* expression boundaries expanded (Fig. 5, A and B). Unilateral eye resection in *wnt5* RNAi animals, before ectopic eye appearance, resulted in the eye regenerating laterally, indicating that existing eyes can locally influence ectopic lateral eye formation (Fig. 5B). Eye progenitors normally move from posterior to anterior. We therefore posit that unlike the case of AP TZ movement (such as in wild-type head fragments), ML TZ movement allows new, serial eye nucleation without existing eyes "shielding" the new TZ (fig. S9A). Additionally, *wnt5* and *slit* RNAi affected both the ML migration and specification pattern of eye progenitors (fig. S10, A to C). These results implicate a *slit-wnt5* circuit (acting directly or indirectly) as a ML TZ determinant.

notum encodes a broadly conserved Wnt inhibitor (31) expressed near the anterior brain and in the anterior pole and is involved in anterior tissue patterning (25, 31). *notum(RNAi)* animals develop a set of much more anterior and medial eyes under homeostatic conditions (25). We postulated that this anatomical pattern can be explained by using the model described above for experiments in wild-type animals. Specifically,

we postulate that *notum* RNAi leads to progressive anterior TZ movement, but eyes do not appear anteriorly initially because of the attractive influence of remaining eyes. Ultimately, with sufficient TZ anterior movement, arcing medially, eye progenitors could escape existing eyes and nucleate new eyes (similar to those in Fig. 3A). As predicted by this model, unilateral *notum (RNAi)* eye resection, before ectopic anterior eye appearance, resulted in anteriorly shifted eye regeneration. A third eye, anterior to the intact eye, also later formed in these animals, as predicted by the model (Fig. 5C).

A broad targetable zone enables maintenance of anatomy in incorrect locations

We next examined what happens to the extra eyes formed by the surgical manipulations described above. We fed three-eyed wild-type animals (as in Fig. 3A), allowing them to grow and undergo long-term tissue turnover. All eyes remained (Fig. 6A). Therefore, these wild-type animals now stably maintained an alternative and functional anatomical state. All three eyes incorporated new progenitors (early-stage, Fig. 3, C to E; late-stage, fig. S11A). Because eyes were
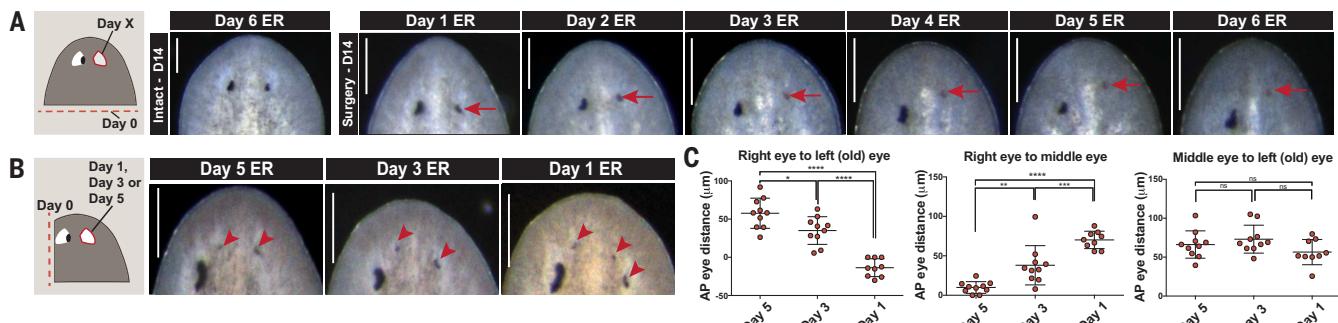
Fig. 4. Eyes nucleate at predictable positions during dynamic positional information shifting. (**A**) Unilateral eye resection time course (days 1 to 6 after decapitation) reveal a continuously shifting TZ. (**B** and **C**) Unilateral eye resections after decapitation and parasagittal amputation. In (A) and (B), red arrows and arrowheads indicate regenerating eyes. Statistical significance: one-way ANOVA: *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, *****$P \leq 0.0001$. NS, not significant. Scale bars, 200 μm.
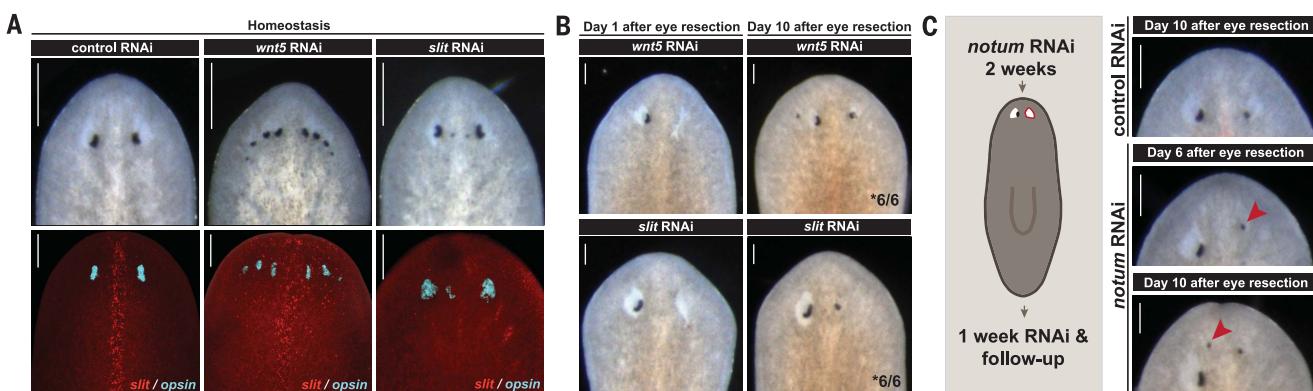
Fig. 5. *slit*, *wnt5*, and *notum* are involved in establishing a TZ. (**A**) *wnt5*, *slit*, and control RNAi in uninjured animals (more than eight double-stranded RNA feedings, more than 4 weeks). FISH: *slit* expression expansion and reduction after *wnt5* and *slit* RNAi, respectively. (**B**) Unilateral eye resections after *wnt5*, *slit*, and control RNAi reveal the local attractive nature of eyes, preventing de novo eye nucleation and a mediolateral TZ shift. (**C**) Unilateral eye resection after *notum* RNAi led to anteriorly shifted eye nucleation without decapitation, implicating *notum* in AP TZ regulation. Red arrowheads indicate regenerating eyes. Asterisks indicate ratio of animals with outcome as shown. Scale bars, 200 μm.
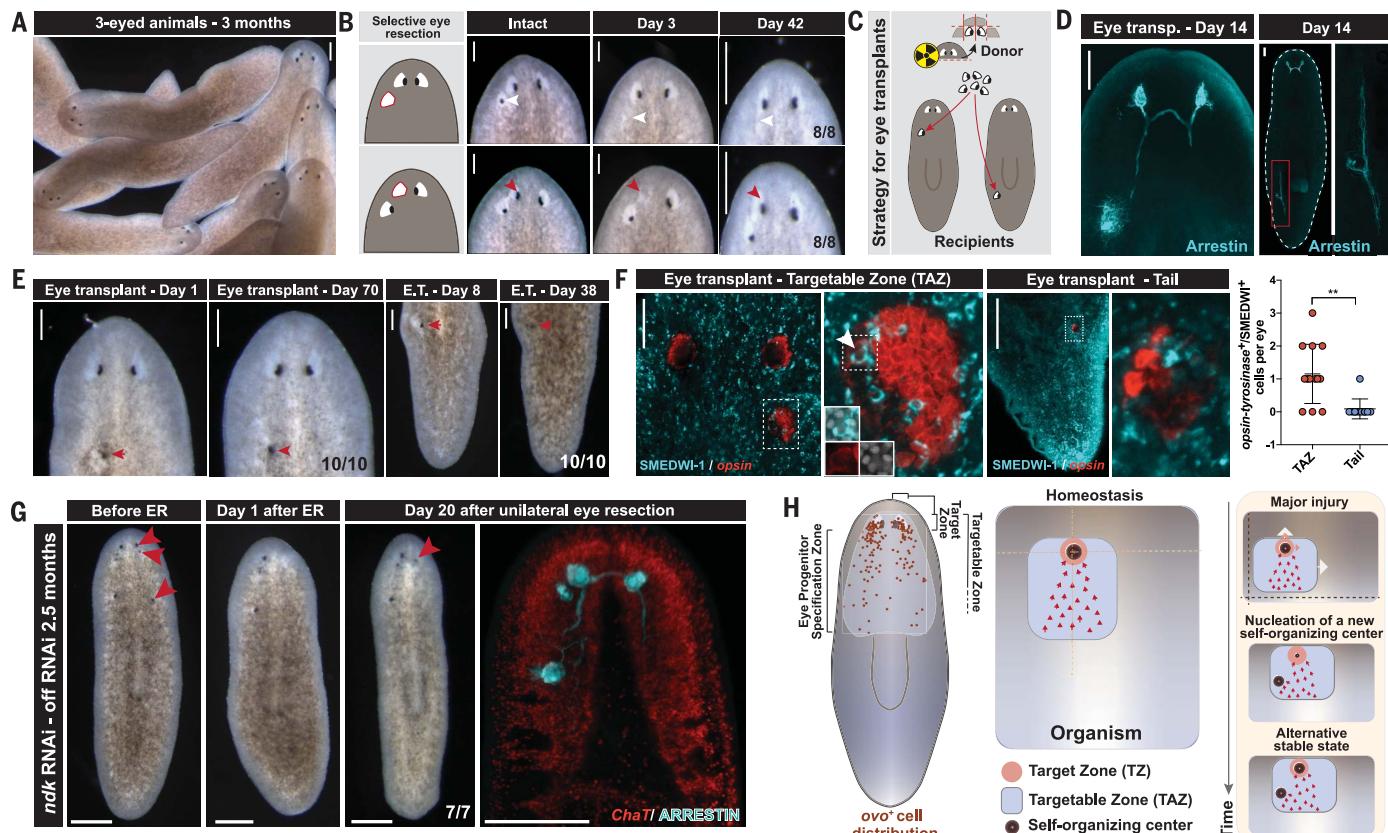
Fig. 6. Ectopic eyes are maintained in a broad TAZ. (**A**) Long-term feeding of wild-type three-eyed animals. All eyes are stably maintained. (**B**) Selective eye resections reveal a TAZ where ectopic eyes can be maintained but not regenerated (white arrowheads). (**C**) Eye transplantation strategy into or outside of the TAZ. Donor animals were irradiated so as to lack progenitors. (**D**) Transplanted eyes sent projections into the visual circuit. (**E**) Transplanted eyes (red arrowheads) were maintained in the TAZ, but not outside of it (tail). (**F**) Only TAZ-transplanted eyes incorporated progenitors (SMEDWI-1$^+$/$opsin^+$ cells, analyzed 12 to 16 days after transplantation) and SMEDWI-1$^+$/$opsin^+$ cell quantification (**$P \leq 0.01$). White arrowhead indicates SMEDWI-1$^+$/$opsin^+$ cell. (**G**) $ndk(RNAi)$ animals, off RNAi, with unilateral eye resection. Only one eye regenerated (in TZ); nonresected eyes remained. Red arrowheads in (B) and (G) indicate resected and regenerating eyes. (**H**) Eye progenitor specification zone by mapping ovo^+ cells in 13 animals and generalized model for progenitors integrating self-organization, a TZ, and a TAZ. Red arrowheads indicate newly specified migratory progenitors and an example for system-level behavior predicted by the model in response to a regenerative challenge. Scale bars, 200 μm.

maintained in their incorrect anatomical positions, we hypothesized that there exists a "targetable zone" (TAZ), where a mispositioned eye can be maintained because of its self-organizing nature. We define the TAZ as the region where regenerative progenitors are capable of going to maintain or regenerate an organ or tissue. The TAZ includes the TZ, but when larger than the TZ, it allows targeting of self-organizing centers in incorrect (non-TZ) locations. TZ and TAZ concepts make testable predictions. We first asked whether eyes would be regenerated in three-eyed animals through selective eye resection. The original left eye in these animals never reached its correct position (the TZ) during morphallaxis, presumably because the second, more anterior left eye occupied this position. Despite being an original, normal eye, we hypothesized that this posterior eye should not regenerate upon resection because, in its absence as an attractor, progenitors should go to the TZ. Indeed, resected posterior eyes did not regenerate, whereas resected anterior eyes (in the TZ) did (Fig. 6B). Supernumerary eyes can rarely appear spontaneously during errors

in asexual planarian reproduction, and consistent with the above data, these supernumerary eyes do not regenerate after removal (*32*).

We postulated that the region where eye progenitors are specified approximates the TAZ. Indeed, mapping ovo^+ eye progenitors from many uninjured animals showed that the eye progenitor specification zone is regional (in the anterior), but coarse spatially (from eyes to near the centrally located pharynx)—much broader than the location of the eye itself (Fig. 6H). We propose that this broad eye progenitor specification zone explains the TAZ: An eye in this region would have access to eye progenitors and, through its self-organizing properties, could maintain itself, allowing alternative anatomical states to be indefinitely maintained.

To further test the TAZ concept, we developed eye transplantation strategies (Fig. 6, C to F, and fig. S11, B to F). Transplanted eyes in the anterior were maintained, whereas eyes transplanted into tails shrank and ultimately disappeared (Fig. 6, E and F, and fig. S11, B and D). Transplanted eyes sent projections toward their targets in the brain

(Fig. 6D). Transplanted eyes also had neighboring cell types (fig. S11C); understanding how eye cells directly or through neighboring cells can attract eye progenitors in the TAZ will be an interesting future direction. After asexual fissioning, transplanted eyes in the posterior were maintained, presumably because positional information resetting placed them in a new TAZ (fig. S11E).

We also used RNAi phenotypes to study the TAZ concept. *nou darake(RNAi)* animals generate ectopic posterior eyes (*33*). After removing these animals from RNAi conditions, all eyes remained. We performed unilateral eye resections after removal from RNAi. On the eye-resected side, only one eye regenerated, occurring at the wild-type eye location (Fig. 6G). On the contralateral side, all ectopic eyes were maintained. Similarly, unilateral eye resections were performed in *wnt5(RNAi)* animals—in this case, under long-term RNAi (fig. S12A). As predicted by TAZ/TZ concepts, only the lateral-most eye regenerated (a lateral-shifted TZ), whereas all eyes were maintained on the nonresected side. SMEDWI-1 labeling of *wnt5(RNAi)* animals removed from

RNAi conditions showed that both medial and lateral eyes were targeted by progenitors (fig. S12, B and C). Similarly, *slit(RNAi)* animals off RNAi also maintained medial eyes (fig. S12D). Last, *notum(RNAi)* animals showed anterior eye regeneration after eye resection (Fig. 5C), yet multiple eye sets were maintained during long-term RNAi (fig. S12E). These findings demonstrate continuous progenitor targeting to eyes in a variety of locations (the TAZ) by using RNAi phenotypes.

Discussion

We propose a model based on findings described above for the properties governing the behavior of migratory, mesenchymal eye progenitors in regeneration (Fig. 6H). The model involves three key components: (i) an extrinsic TZ (different than the anatomical structure itself) that guides regenerative progenitor migration; (ii) self-organization, in which progenitors can be stably incorporated into existing anatomy, even if this is at the "wrong" location; and (iii) a TAZ, involving a much broader progenitor specification zone than the target location for those progenitors. These three properties yield a systems-level process that can explain how new tissues and organs are formed and maintained in noisy and continuously dynamic conditions, such as changing positional coordinates after injury. The coarse progenitor specification zone is essential, allowing progenitors to maintain organs or tissues that remain entirely or partially present after injury, as opposed to duplicating them in new locations. For example, if progenitor specification occurred very locally, near the TZ, reiterated structures could be deposited as this location shifted during positional rescaling (fig. S13B). Without the TAZ and self-organization properties, anatomy would get scrambled as progenitors targeted shifting positional coordinates after injury (fig. S13B). Replacement of completely missing structures is guided by externally acting positional cues that define TZs at the appropriate position for final anatomy. If TZs were not discretely defined, anatomy duplications would also emerge (fig. S13B). This model identifies a set of rules that are used for stabilizing the system and its coherent functional properties. The existence of adaptive TZs and TAZs could be a broadly used strategy by diverse developmental and regenerative processes to form and maintain complex self-organizing modules in appropriate relative positions. These findings explain principles that guide coherent tissue formation and maintenance in dynamic and noisy biological processes such as regeneration.

REFERENCES AND NOTES

1. E. Bonabeau, M. Dorigo, G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems* (Oxford Univ. Press, 1999).
2. S. Camazine *et al.*, *Self-Organization in Biological Systems* (Princeton Univ. Press, 2003).
3. G. Nicolis, I. Prigogine, *Self-Organization in Non-equilibrium Systems. From Dissipative Structures to Order through Fluctuations* (J. Wiley & Sons, 1977).
4. A. M. Turing, *Phys. Today* **B237**, 37–72 (1952).
5. C. D. Tsiairis, A. Aulehla, *Cell* **164**, 656–667 (2016).
6. F. Corson, L. Couturier, H. Rouault, K. Mazouni, F. Schweisguth, *Science* **356**, eaai7407 (2017).
7. S. E. Harrison, B. Sozen, N. Christodoulou, C. Kyprianou, M. Zernicka-Goetz, *Science* **14**, 356 (2017).
8. G. Halder, P. Callaerts, W. J. Gehring, *Science* **267**, 1788–1792 (1995).
9. A. E. Shyer *et al.*, *Science* **357**, 811–815 (2017).
10. D. L. Stocum, D. A. Melton, *J. Exp. Zool.* **201**, 451–461 (1977).
11. T. Stückemann *et al.*, *Dev. Cell* **40**, 248–263.e4 (2017).
12. M. Eiraku *et al.*, *Nature* **472**, 51–56 (2011).
13. M. A. Lancaster *et al.*, *Science* **345**, 374–379 (2013).
14. T. Sato *et al.*, *Nature* **459**, 262–265 (2009).
15. D. E. Wagner, I. E. Wang, P. W. Reddien, *Science* **332**, 811–816 (2011).
16. S. W. Lapan, P. W. Reddien, *PLOS Genet.* **7**, e1002226 (2011).
17. S. W. Lapan, P. W. Reddien, *Cell Reports* **2**, 294–307 (2012).
18. C. E. Adler, C. W. Seidel, S. A. McKinney, A. Sánchez Alvarado, *eLife* **3**, e02238 (2014).
19. S. A. LoCascio, S. W. Lapan, P. W. Reddien, *Dev. Cell* **40**, 381–391.e3 (2017).
20. T. H. Morgan, *Arch. Entwm.* **7**, 364–397 (1898).
21. J. Pellettieri *et al.*, *Dev. Biol.* **338**, 76–85 (2010).
22. D. J. Forsthoefel, A. E. Park, P. A. Newmark, *Dev. Biol.* **356**, 445–459 (2011).
23. J. N. Witchley, M. Mayer, D. E. Wagner, J. H. Owen, P. W. Reddien, *Cell Reports* **4**, 633–641 (2013).
24. M. L. Scimone, L. E. Cote, T. Rogers, P. W. Reddien, *eLife* **5**, 12845 (2016).
25. E. M. Hill, C. P. Petersen, *Development* **142**, 4217–4229 (2015).
26. R. Lander, C. P. Petersen, *eLife* **5**, e12850 (2016).
27. E. F. Keller, *Hist. Stud. Nat. Sci.* **39**, 1–31 (2009).
28. F. Cebrià, T. Guo, J. Jopek, P. A. Newmark, *Dev. Biol.* **307**, 394–406 (2007).
29. K. A. Gurley *et al.*, *Dev. Biol.* **347**, 24–39 (2010).
30. I. M. Oderberg, D. J. Li, M. L. Scimone, M. A. Gaviño, P. W. Reddien, *Curr. Biol.* **27**, 733–742 (2017).
31. C. P. Petersen, P. W. Reddien, *Science* **332**, 852–855 (2011).
32. F. Sakai, K. Agata, H. Orii, K. Watanabe, *Zool. Sci.* **17**, 375–381 (2000).
33. F. Cebrià *et al.*, *Nature* **419**, 620–624 (2002).

REPORT

QUANTUM ENTANGLEMENT

Spatial entanglement patterns and Einstein-Podolsky-Rosen steering in Bose-Einstein condensates

Matteo Fadel, Tilman Zibold, Boris Décamps, Philipp Treutlein*

Many-particle entanglement is a fundamental concept of quantum physics that still presents conceptual challenges. Although nonclassical states of atomic ensembles were used to enhance measurement precision in quantum metrology, the notion of entanglement in these systems was debated because the correlations among the indistinguishable atoms were witnessed by collective measurements only. Here, we use high-resolution imaging to directly measure the spin correlations between spatially separated parts of a spin-squeezed Bose-Einstein condensate. We observe entanglement that is strong enough for Einstein-Podolsky-Rosen steering: We can predict measurement outcomes for noncommuting observables in one spatial region on the basis of corresponding measurements in another region with an inferred uncertainty product below the Heisenberg uncertainty bound. This method could be exploited for entanglement-enhanced imaging of electromagnetic field distributions and quantum information tasks.

Two quantum mechanical degrees of freedom are entangled (nonseparable) if the quantum state of one cannot be described independently of the other. When measurements are performed on both, entanglement results in correlations between the outcomes. Although entanglement can exist between any quantum degrees of freedom, the conflict with classical physics is particularly evident when the correlations are observed between measurement outcomes obtained in spatially separated regions. Einstein, Podolsky, and Rosen (EPR) pointed out (*1*) that if the correlations are sufficiently strong, local measurements in one region, *A*, can apparently change the quantum state in a spatially separated region, *B*, a scenario Schrödinger named "steering" (*2*). The possibility of steering between spatially separated systems implies that quantum theory is in conflict with a local realist description of the world (*3*, *4*). In fact, steering allows an observer in *A* to use her local measurement outcomes to predict the outcomes of noncommuting measurements in *B* with uncertainties below the Heisenberg uncertainty relation for *B*. EPR steering has been extensively explored with optical systems (*4*). Entanglement was observed between spatially separated atomic ensembles (*5–8*) and between individually address-

able atoms in optical lattices (*9*, *10*), but EPR steering has not yet been achieved for more than two atoms (*11*). Demonstrating the EPR paradox with ensembles of massive particles is desirable, as it puts quantum physics to a stringent test in a regime of increasingly macroscopic systems (*4*). Moreover, it opens up perspectives for applications of such systems in quantum metrology and one-sided device-independent quantum information tasks, which exploit EPR steering as a resource (*12*).

Experiments with ultracold atomic ensembles recently made rapid progress, and a variety of nonclassical states can be prepared (*13*). Besides being of fundamental interest, such states find applications in quantum metrology (*14*), where the correlations between the constituent atoms are exploited to reduce the noise in atom interferometric measurements (*15–18*). Because of the large number of atoms involved, it is usually not possible to address and detect the atoms individually. In the case of Bose-Einstein condensates (BECs), it is even impossible in principle: The atoms are identical particles that occupy the same spatial mode. Still, quantum correlations between them can be characterized with the help of witness observables that involve only collective measurements on the entire ensemble (*19*, *20*). This approach has been used to reveal the presence of entanglement (*15*, *21*), EPR correlations (*22*), and even Bell correlations (*23*) in a cloud of atoms. However, these nonclassical correlations have not yet been observed directly by performing measurements on spatially separated subsystems. Moreover, several authors have questioned whether the concept of entanglement in systems of indistinguishable particles is fully legitimate and useful for tasks other than metrology [for a review see (*24*)].

Department of Physics and Swiss Nanoscience Institute, University of Basel, Klingelbergstrasse 82, 4056 Basel, Switzerland
*Corresponding author. Email: philipp.treutlein@unibas.ch
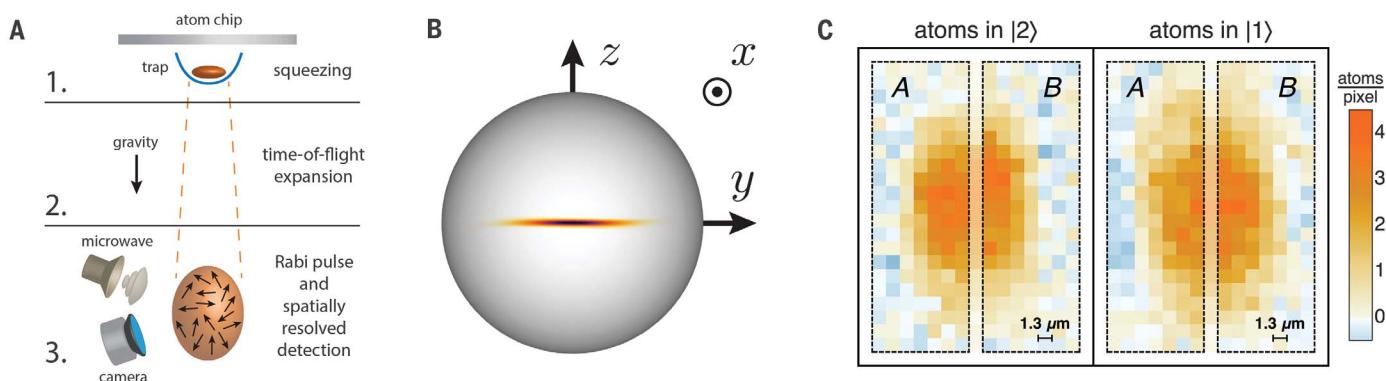
Fig. 1. Extracting entanglement from spatially separated regions of a BEC. (**A**) Experimental sequence. In step 1, we prepare a BEC in a spin-squeezed state on an atom chip. In step 2, the trapping potential is switched off and the BEC expands. In step 3, a Rabi rotation pulse is applied to select the spin component $\hat{S}_{\vec{n}}$ to be measured, followed by recording two high-resolution absorption images of the atomic density distributions in states $|1\rangle$ and $|2\rangle$. (**B**) Illustration of the spin-squeezed state on a sphere (Wigner function, representing the quantum fluctuations of the spin) and definition of the axes $\vec{n}$ used in the measurement of the entanglement and EPR steering criteria. (**C**) Single-shot absorption images of the atomic densities in $|2\rangle$ and $|1\rangle$, showing an example of regions *A* and *B* used to define the collective spins $\hat{S}^A$ and $\hat{S}^B$ entering in the entanglement and EPR steering criteria.

RESEARCH | REPORT

As pointed out in the theoretical work of (24), the presence of entanglement in an ensemble of indistinguishable particles can be unambiguously confirmed by extracting it into spatially separated modes, turning it into a resource for a variety of quantum information tasks. In our experiment, we demonstrate that entanglement can be extracted from spatially separated parts of a spin-squeezed BEC and use the entanglement to demonstrate the EPR paradox with an atomic system.

The quantum degrees of freedom in our experiment are two effective collective spins (13, 25), $\hat{\vec{S}}^A$ and $\hat{\vec{S}}^B$, that describe the internal state of atoms in regions A and B, respectively. Each atom is an effective two-level system with internal states $|1\rangle$ and $|2\rangle$. The component $\hat{S}_z^A = \frac{1}{2\eta_{\text{eff}}^A}(\hat{N}_1^A - \hat{N}_2^A)$ is proportional to half the atom-number difference between the states, evaluated in region A. The normalization by η_{eff}^A (25) takes into account the finite resolution of the imaging system (26). A similar definition holds for $\hat{S}_z^B$. Other spin components can be measured by applying appropriate spin rotations before detection. To detect entanglement, we use the criterion from (27), where it was shown that for all separable states

$$\mathcal{E}_{\text{Ent}} = \frac{4\,\text{Var}\left(g_z\hat{S}_z^A + \hat{S}_z^B\right)\,\text{Var}\left(g_y\hat{S}_y^A + \hat{S}_y^B\right)}{\left(|g_zg_y|\,\left|\langle\hat{S}_x^A\rangle\right| + \left|\langle\hat{S}_x^B\rangle\right|\right)^2} \geq 1 \tag{1}$$

where $\text{Var}(\cdot)$ denotes the variance and g_z, g_y are real parameters that can be optimized to minimize $\mathcal{E}_{\text{Ent}}$. Therefore, $\mathcal{E}_{\text{Ent}} < 1$ is a sufficient condition to certify entanglement (nonseparability) between A and B.

The variances in Eq. 1 quantify the uncertainty with which an observer in A can predict (infer) the outcome of a spin measurement in B on the basis of a corresponding measurement on her own system and are therefore called inferred variances. Because $\hat{S}_z$ and $\hat{S}_y$ do not commute, measuring both inferred variances requires repeated experiments on identically prepared systems. Correlations between A and B improve the prediction of A for such noncommuting measurements performed by B and decrease the product of inferred variances.

EPR steering is demonstrated if the correlations are so strong that the product of the inferred variances falls below the Heisenberg uncertainty bound for system B, i.e., there is a violation of the relation (4)

$$\mathcal{E}_{\text{EPR}}^{A\to B} = \frac{4\,\text{Var}\left(g_z\hat{S}_z^A + \hat{S}_z^B\right)\,\text{Var}\left(g_y\hat{S}_y^A + \hat{S}_y^B\right)}{\left|\langle\hat{S}_x^B\rangle\right|^2} \geq 1 \tag{2}$$

If there are no correlations between A and B, the variances in Eq. 2 are minimized for $g_z = g_y = 0$, for which the spin uncertainty relation for B is recovered. In the presence of a violation of Eq. 2, the observer B must conclude that he is in the paradoxical situation considered by EPR, where the observer in A can predict his measurement results without any classical communication. Note that a violation of Eq. 1 does not imply a violation of Eq. 2, whereas the converse is true. This reflects the fact that entanglement is necessary but not sufficient for EPR steering and that they are inequivalent types of correlations (3, 28). Moreover, the asymmetry between A and B present in Eq. 2 implies that even if A can steer B (denoted $A\to B$), B may not necessarily be able to steer A ($B\to A$), as investi-

gated both theoretically (28–30) and experimentally (31, 32) in optics.

To demonstrate a violation of both Eqs. 1 and 2 with a massive many-particle system, we perform experiments with two-component BECs of $N = 590 \pm 30$ ^{87}Rb atoms, magnetically trapped on an atom chip (33). The two components correspond to the hyperfine states $|F=1, m_F=-1\rangle \equiv |1\rangle$ and $|F=2, m_F=1\rangle \equiv |2\rangle$ and occupy nearly identical spatial modes. They can be described by a collective spin $\hat{\vec{S}}$, referring to the entire BEC.

We prepare the BEC in a spin-squeezed state by controlling atomic collisions with a state-dependent potential, as described in (18, 21, 23). The spin-squeezed state features quantum correlations between the atoms, which reduce fluctuations of $\hat{S}_z$ and increase fluctuations of $\hat{S}_y$ while maintaining a large spin polarization in $\hat{S}_x$ (see Fig. 1B). We obtain typically −3.8(2) dB of spin squeezing according to the Wineland criterion (34). Alternatively, we can prepare the BEC in a coherent spin state, where the atomic internal states are uncorrelated.

To access spatially separated regions in the BEC, we use the sequence illustrated in Fig. 1A. After preparing the state, the atomic cloud is released from the trap and expands during a 2.2-ms time of flight. This expansion is nearly spin-independent (because collisional interactions are very similar for $|1\rangle$ and $|2\rangle$) and leads to a magnification of the atomic cloud. Next, we set the axis $\vec{n}$ of the spin components $\hat{S}_{\vec{n}}^A$ and $\hat{S}_{\vec{n}}^B$ to be measured by applying a Rabi rotation pulse to the entire atomic cloud. Immediately thereafter, we record two high-resolution absorption images (35) of the atomic density distributions in states $|2\rangle$ and $|1\rangle$ by illuminating the atomic cloud twice with a resonant laser beam. The imaging pulses project the spin state and
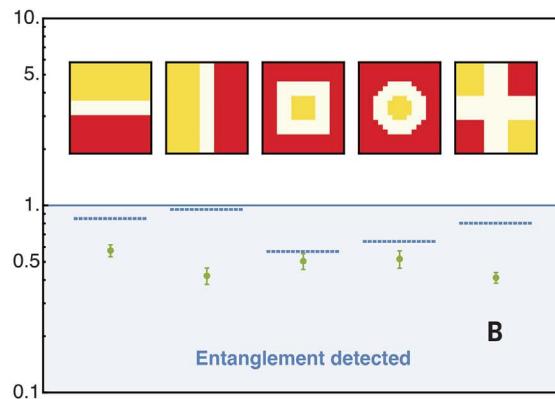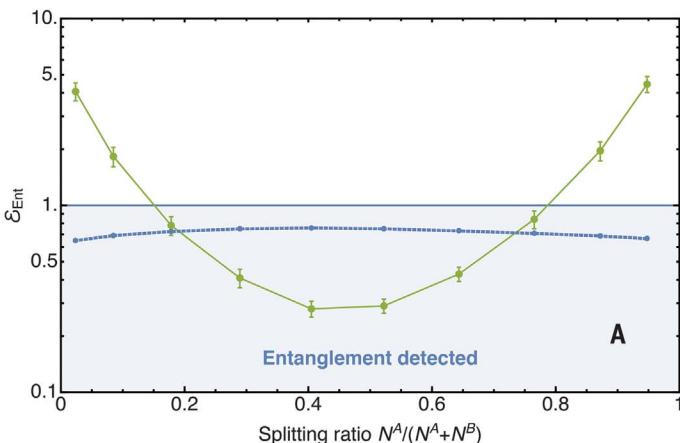

Fig. 2. Spatial entanglement patterns in the atomic cloud.
(**A**) Entanglement criterion Eq. 1 evaluated for a spin-squeezed BEC (green points) for different horizontal positions of the one-pixel gap between regions A and B (see Fig. 1C), corresponding to different splitting ratios $N^A/(N^A + N^B)$. Lines are a guide to the eye, and error bars indicate 1 SEM. The blue points show the maximum violation

(minimum value of ε_{Ent}) that could be explained by detection cross-talk. (**B**) Entanglement between regions of different shapes (A = yellow, B = red) in a spin-squeezed BEC. The pixel pattern used for the analysis is illustrated above the respective data points, and the blue segments show the corresponding maximum violation expected by cross-talk.

Fadel et al., Science **360**, 409–413 (2018)   27 April 2018

2 of 4

Downloaded from http://science.sciencemag.org/ on April 26, 2018
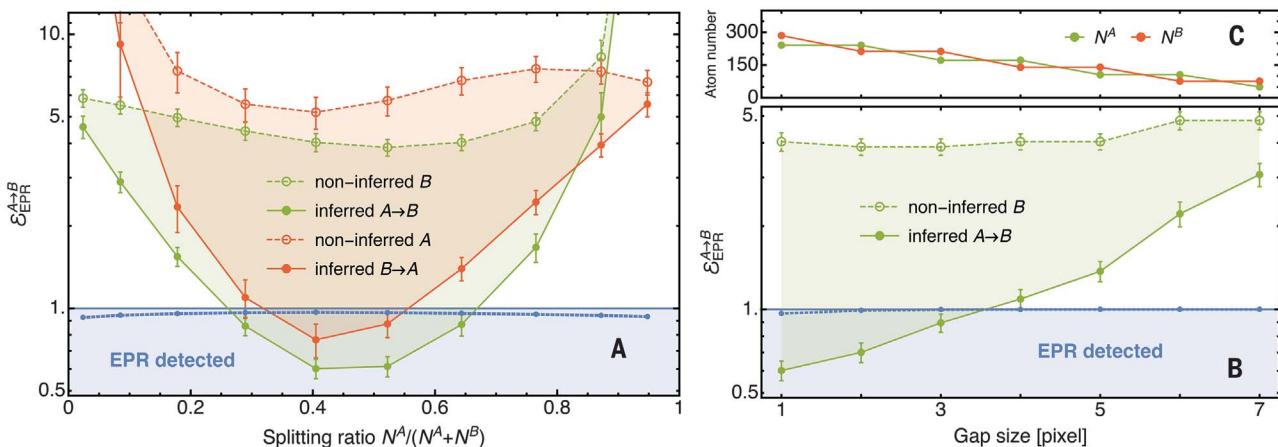
Fig. 3. Observation of Einstein-Podolsky-Rosen steering. (**A**) EPR steering criterion Eq. 2, evaluated for steering $A{\rightarrow}B$ (green filled circles) and $B{\rightarrow}A$ (red filled circles) in a spin-squeezed BEC, for different horizontal positions of the one-pixel gap (see Fig. 1C), corresponding to different splitting ratios $N^A/(N^A + N^B)$. EPR steering is strongest for intermediate splitting ratios. Empty circles: spin uncertainty relation involving the product of noninferred variances in region B (green) and A (red). Lines are a guide to the eye, and the shaded regions are the reduction of the uncertainty product in replacing the noninferred variances with the inferred ones. Blue points: maximum violation that could be explained by detection cross-talk. (**B**) EPR steering $A{\rightarrow}B$ for different widths of the gap in Fig. 1C. The center of the gap is fixed to the position showing maximum EPR steering in Fig. 3A. Even for increased gap size, we find a significant violation of the bound, confirming that the correlations cannot be explained by detection cross-talk between the regions. Lines and shaded regions as in (A). (**C**) Atom number in regions A and B as a function of the gap size.

simultaneously localize the atoms in well-defined positions. Figure 1C shows typical absorption images taken in this way. This experimental sequence is repeated thousands of times, alternating the measurement direction $\vec{n}$ along either x, y, or z.

We now define the two regions, A and B, to be analyzed on all pairs of absorption images (Fig. 1C). Counting the atom numbers N_1^A and N_2^A in region A realizes a single-shot projective measurement of the local spin $\hat{S}_{\vec{n}}^A = \frac{1}{2\eta_{\mathrm{eff}}^A}(\hat{N}_1^A - \hat{N}_2^A)$. The same approach is applied to region B, which yields $\hat{S}_{\vec{n}}^B$. The finite optical resolution and the motion of atoms during the imaging pulses amount to an uncertainty in the atomic position of 1.8 μm (2.5 μm) or 1.4 pixels (2.0 pixels) in the horizontal (vertical) direction. Consequently, spins near the boundary have only partial overlap with the region A or B, which is taken into account by η_{eff} (25). Furthermore, spins overlapping with both A and B lead to detection cross-talk, which we reduce by leaving a gap of 1 pixel between the two regions. The experimentally determined spin variances include a contribution from detection noise, which can be independently measured and subtracted (26).

To detect entanglement between regions A and B, we evaluate Eq. 1 for different positions of the gap, corresponding to different splitting ratios $N^A/(N^A + N^B)$, where $N^A = N_1^A + N_2^A$ and similar for N^B (Fig. 2A, green dots). For a wide range of splitting ratios, we observe a violation of the inequality in Eq. 1 that goes far below the value that could be explained by detection cross-talk (26). This proves that the two local spins $\hat{S}^A$ and $\hat{S}^B$ are entangled. The extracted entanglement derives from the quantum correlations among the indistinguishable atoms in the initial state (24), as the expansion of the cloud, the spin rotation, and detection do not create such correlations.

An intriguing feature of our approach to extract entanglement (24) from a many-body state is that the subsystems can be defined a posteriori on the images. This is in contrast to other experiments where the subsystems are defined by the experimental setup (5–8) or by the source of the state (4, 31). We exploit this feature to detect entanglement between regions A and B patterned in a variety of different shapes (Fig. 2B). That we observe entanglement between all such regions reflects the symmetry of the underlying many-body quantum state: The quantum state of the indistinguishable bosons in the condensate must be symmetric under particle exchange. Consequently, each atom is entangled with all other atoms, and the entanglement extends over the entire atomic cloud. In experiments with atoms in optical lattices, entanglement between different spatial bipartitions was observed by measuring entanglement entropy or concurrence, using systems of up to 10 atoms that were individually addressed (9, 10). By comparison, our experiment reveals entanglement in ensembles of hundreds of atoms by using inequalities that apply in the continuous-variable limit.

The correlations in our system are strong enough to demonstrate an EPR paradox: Fig. 3A shows a measurement of the EPR criterion Eq. 2 for a horizontal splitting of the cloud and different positions of the gap. We observe EPR steering $A{\rightarrow}B$ (green data points) for intermediate splitting ratios. For comparison, we evaluate the spin uncertainty relation $4\,\mathrm{Var}(\hat{S}_z^B)\,\mathrm{Var}(\hat{S}_y^B)/|\langle\hat{S}_x^B\rangle|^2 \geq 1$ for system B, illustrating the reduction of the uncertainty product when replacing the noninferred variances with the inferred ones. As can be seen in

Eq. 2, EPR steering is an asymmetric concept. By relabeling region A as B and vice versa, we can invert the roles of the steering and steered systems. This inverted scenario also shows EPR steering $B{\rightarrow}A$ (red data points in Fig. 3A). The asymmetry between the curves indicates the presence of technical noise in our system (30, 31, 36). For intermediate splitting ratios, we observe two-way steering $A{\leftrightarrow}B$, a prerequisite for observing the even stronger Bell correlations (28). We note that we also observe EPR steering if we do not subtract detection noise from the inferred variances [$\mathcal{E}_{\mathrm{EPR}}^{A{\rightarrow}B} = 0.73(6)$ for a splitting ratio of 0.4], and also for vertical instead of horizontal splitting of the cloud [$\mathcal{E}_{\mathrm{EPR}}^{A{\rightarrow}B} = 0.81(4)$ for a splitting ratio of 0.6].

Finally, we characterize the robustness of the observed EPR steering $A{\rightarrow}B$ to a variation of the gap size. We fix the central position of the gap such that the splitting ratio is 0.40 (the ratio maximizing steering $A{\rightarrow}B$ and $B{\rightarrow}A$ in Fig. 3A) and change the gap width symmetrically with respect to this position (Fig. 3C). We observe that EPR steering vanishes for large widths of the gap, where the size of the steered system is considerably reduced (Fig. 3B).

We have also performed measurements similar to those of Fig. 1 and Fig. 2 with the BEC initially prepared in a coherent spin state, showing no statistically significant violations of Eqs. 1 and 2 beyond detection cross-talk (26). The observed spin noise in each individual region A or B agrees well with projection noise of uncorrelated atoms, confirming our calibration of the imaging system.

Our method can be used for quantum metrology of electromagnetic field patterns. Consider an applied field that shifts the spin components $\hat{S}_y^B$ and $\hat{S}_z^B$ with respect to $\hat{S}_y^A$ and $\hat{S}_z^A$. The EPR entanglement allows one to detect this shift in

the yz plane with an uncertainty characterized by the product of the inferred variances in Eq. 2. The EPR parameter $\mathcal{E}_{\mathrm{EPR}}^{A \to B} < 1$ quantifies by how much this measurement improves over the Heisenberg uncertainty bound for $\hat{S}^B$ and is thus a direct measure of the metrological enhancement provided by the EPR entanglement. Because our imaging method allows us to define the regions A and B a posteriori in a variety of shapes (Fig. 2), a single data set could be used to analyze dipole, quadrupole, and more complex patterns of the applied field. This is different from other field-sensing methods where the pattern is defined by the state preparation (15, 37).

Beyond metrology, EPR steering is a resource for one-sided device-independent quantum information tasks (12). The asymmetry of the steering concept allows tasks such as quantum teleportation, entanglement swapping, or randomness certification to be performed in a situation where one of the involved parties can be trusted but not the other. An interesting perspective in this context is to distribute the correlations over macroscopic distances by splitting the atomic cloud with a double- or multiwell potential, exploiting the full control of BEC wave functions provided by the atom chip (33). Furthermore, our study raises the question of whether Bell correlations could also be observed between spatially separated regions. Although the EPR paradox can be demonstrated with Gaussian states and measurements and identical measurement settings in A and B, a violation of a Bell inequality would require non-Gaussian states or measurements as well as the ability to measure different spin components in the two regions in a single run of the experiment (38). This could be achieved by rotating the collective spins $\hat{S}^A$ and $\hat{S}^B$ independently with on-chip microwave near-fields, followed by atomic fluorescence detection with single-atom reso-

lution. In summary, our results open up a variety of perspectives for quantum science and technology with massive many-body systems.

Complementary to our work, spatially distributed multipartite entanglement was observed in (39) and entanglement of spatially separated modes was observed in (40).

REFERENCES AND NOTES

1. A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.* **47**, 777–780 (1935).
2. E. Schrödinger, M. Born, *Proc. Camb. Philos. Soc.* **31**, 555 (1935).
3. H. M. Wiseman, S. J. Jones, A. C. Doherty, *Phys. Rev. Lett.* **98**, 140402 (2007).
4. M. D. Reid et al., *Rev. Mod. Phys.* **81**, 1727–1751 (2009).
5. B. Julsgaard, A. Kozhekin, E. S. Polzik, *Nature* **413**, 400–403 (2001).
6. C. W. Chou et al., *Nature* **438**, 828–832 (2005).
7. D. N. Matsukevich et al., *Phys. Rev. Lett.* **96**, 030405 (2006).
8. J. Simon, H. Tanji, S. Ghosh, V. Vuletic, *Nat. Phys.* **3**, 765–769 (2007).
9. R. Islam et al., *Nature* **528**, 77–83 (2015).
10. T. Fukuhara et al., *Phys. Rev. Lett.* **115**, 035302 (2015).
11. E. Hagley et al., *Phys. Rev. Lett.* **79**, 1–5 (1997).
12. D. Cavalcanti, P. Skrzypczyk, *Rep. Prog. Phys.* **80**, 024001 (2017).
13. L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, P. Treutlein, arXiv:1609.01609 [quant-ph] (6 September 2016).
14. V. Giovannetti, S. Lloyd, L. Maccone, *Nat. Photonics* **5**, 222–229 (2011).
15. C. Gross, T. Zibold, E. Nicklas, J. Estève, M. K. Oberthaler, *Nature* **464**, 1165–1169 (2010).
16. A. Louchet-Chauvet et al., *New J. Phys.* **12**, 065032 (2010).
17. I. D. Leroux, M. H. Schleier-Smith, V. Vuletić, *Phys. Rev. Lett.* **104**, 250801 (2010).
18. C. F. Ockeloen, R. Schmied, M. F. Riedel, P. Treutlein, *Phys. Rev. Lett.* **111**, 143001 (2013).
19. L. Amico, R. Fazio, A. Osterloh, V. Vedral, *Rev. Mod. Phys.* **80**, 517–576 (2008).
20. O. Gühne, G. Tóth, *Phys. Rep.* **474**, 1–75 (2009).
21. M. F. Riedel et al., *Nature* **464**, 1170–1173 (2010).
22. J. Peise et al., *Nat. Commun.* **6**, 8984 (2015).
23. R. Schmied et al., *Science* **352**, 441–444 (2016).
24. N. Killoran, M. Cramer, M. B. Plenio, *Phys. Rev. Lett.* **112**, 150501 (2014).
25. J. Hu, W. Chen, Z. Vendeiro, H. Zhang, V. Vuletic, *Phys. Rev. A* **92**, 063816 (2015).
26. Supplementary materials are available online.
27. V. Giovannetti, S. Mancini, D. Vitali, P. Tombesi, *Phys. Rev. A* **67**, 022320 (2003).
28. M. T. Quintino et al., *Phys. Rev. A* **92**, 032107 (2015).
29. S. L. W. Midgley, A. J. Ferris, M. K. Olsen, *Phys. Rev. A* **81**, 022101 (2010).
30. Q. Y. He, Q. H. Gong, M. D. Reid, *Phys. Rev. Lett.* **114**, 060402 (2015).
31. V. Händchen et al., *Nat. Photonics* **6**, 596–599 (2012).
32. S. Wollmann, N. Walk, A. J. Bennet, H. M. Wiseman, G. J. Pryde, *Phys. Rev. Lett.* **116**, 160403 (2016).
33. P. Böhi et al., *Nat. Phys.* **5**, 592–597 (2009).
34. D. J. Wineland, J. J. Bollinger, W. M. Itano, D. J. Heinzen, *Phys. Rev. A* **50**, 67–88 (1994).
35. W. Ketterle, D. S. Durfee, D. M. Stamper-Kurn, in *Proceedings of the International School of Physics "Enrico Fermi," Course CXL*, M. Inguscio, S. Stringari, C. E. Wieman, Eds. (IOS Press, Amsterdam, 1999), pp. 67–176.
36. K. Wagner et al., *J. Phys. At. Mol. Opt. Phys.* **47**, 225502 (2014).
37. W. Muessel, H. Strobel, D. Linnemann, D. B. Hume, M. K. Oberthaler, *Phys. Rev. Lett.* **113**, 103004 (2014).
38. N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, *Rev. Mod. Phys.* **86**, 419–478 (2014).
39. P. Kunkel et al., *Science* **360**, 413–416 (2018).
40. K. Lange et al., *Science* **360**, 416–418 (2018).

QUANTUM ENTANGLEMENT

Spatially distributed multipartite entanglement enables EPR steering of atomic clouds

Philipp Kunkel,* Maximilian Prüfer, Helmut Strobel, Daniel Linnemann, Anika Frölian, Thomas Gasenzer, Martin Gärttner, Markus K. Oberthaler

A key resource for distributed quantum-enhanced protocols is entanglement between spatially separated modes. However, the robust generation and detection of entanglement between spatially separated regions of an ultracold atomic system remain a challenge. We used spin mixing in a tightly confined Bose-Einstein condensate to generate an entangled state of indistinguishable particles in a single spatial mode. We show experimentally that this entanglement can be spatially distributed by self-similar expansion of the atomic cloud. We used spatially resolved spin read-out to reveal a particularly strong form of quantum correlations known as Einstein-Podolsky-Rosen (EPR) steering between distinct parts of the expanded cloud. Based on the strength of EPR steering, we constructed a witness, which confirmed genuine 5-partite entanglement.

Quantum mechanics poses a fundamental limit on the simultaneous knowledge of two noncommuting observables, $\hat{Q}_A$ and $\hat{P}_A$. This limit is given by the Heisenberg uncertainty relation for the variances of the observables, $\Delta^2 Q_A \Delta^2 P_A \geq |\langle[\hat{Q}_A, \hat{P}_A]\rangle|^2/4$. At the same time, quantum mechanics allows for nonlocal correlations between two systems, A and B, such that it is possible to infer from the measurement results obtained in system B the outcomes for the same observable in system A more accurately than predicted by the Heisenberg uncertainty constraint for system A (1). This phenomenon was first discussed in the seminal paper by Einstein, Podolsky, and Rosen to show that quantum mechanics is at odds with the assumptions of local realism (2). Schrödinger later termed this steering (3), because in this scenario the state in system A is influenced by the chosen measurement in B (A is steered by B). This is possible only if A and B are strongly entangled, which makes steering a witness for entanglement (4). Originally intended by Einstein, Podolsky, and Rosen to question the completeness of quantum mechanics, entangle-

ment and nonlocality are now regarded as a resource for quantum technologies, such as quantum metrology (5), quantum cryptography (6, 7), and quantum information processing (8).

Pioneering work on spatially distributed entanglement has been reported in pure photonic systems and in hot atomic vapors building on atom-light interaction (9–13). Ultracold atomic gases offer additional possibilities owing to the high coherence of the internal and motional degrees of freedom. In these systems, various schemes for generating nonlocal entanglement between spatially distinct subsystems have been realized using quantum gate operations in optical lattices (14, 15) and long-range interactions in Rydberg systems (16, 17). However, the reliable generation and detection of entanglement in discrete-variable systems so far has been limited to small atom numbers. In the continuous-variable limit considered in this work, the generation of Einstein-Podolsky-Rosen (EPR) entanglement through nonlinear dynamics of spatial multimode systems has been proposed (18–20). Here we present a robust method to spatially distribute entanglement generated

by local interactions (21, 22) in the spin degree of freedom of a Bose-Einstein condensate (BEC) by subsequent expansion of the atomic cloud. This constitutes an explicit experimental implementation of the recently formulated mapping of indistinguishable-particle entanglement in one mode to individually addressable subsystems (23, 24).

Experimentally, we prepared a BEC of $N \approx 11{,}000$ ^{87}Rb atoms in the $F = 1$ hyperfine manifold in the magnetic substate $m_F = 0$ (F, total spin quantum number; m_F, spin projection quantum number). We initiated spin dynamics that coherently populated the states $m_F = \pm 1$ with correlated particle pairs (25), leading to spin-nematic squeezing (26). This led to entanglement shared among all atoms in the condensate. Self-similar expansion for distributing the entanglement was initiated by switching off the longitudinal confinement. The atomic cloud spread in the remaining waveguide potential. After imaging with high optical resolution, we analyzed partitions of the resulting absorption signal to reveal entanglement and EPR steering between the corresponding atomic subsystems (Fig. 1).

As noncommuting observables $\hat{Q}$ and $\hat{P}$, we chose the spin operators $\hat{F}(0)$ and $\hat{F}(\pi/2)$, where $\hat{F}(\phi) = [(\hat{a}_{+1}^\dagger + \hat{a}_{-1}^\dagger)e^{i(\phi-\phi_0)}\hat{a}_0 + \text{h.c.}]/\sqrt{2}$. Here, $\hat{a}_j^\dagger$ is the creation operator for a particle in the spin state $m_F = j$, h.c. denotes the Hermitian conjugate, and ϕ_0 is an offset phase. In the case of negligible populations in $m_F = \pm 1$ compared with the total atom number N, which is fulfilled in the experiment, these operators obey the commutation relation (27)

$$\left[\hat{F}(0), \hat{F}(\pi/2)\right] = 2i\hat{N} \tag{1}$$

The energy difference between the $m_F = \pm 1$ and $m_F = 0$ states caused by the second-order Zeeman shift leads to an evolution of the phase ϕ. By adjusting the hold time before expansion, ϕ was precisely controlled (27).

After the expansion, we applied a resonant radiofrequency pulse corresponding to a $\pi/2$ spin rotation and read out the population difference $N^-(\phi) = N_{+1} - N_{-1}$, which realizes a measurement of $\hat{F}(\phi)$ (27). Because of the commutation relation (Eq. 1), the observed variances of the population differences after spin rotation fulfill the uncertainty relation

$$\frac{\Delta^2 N^-(0)}{N} \frac{\Delta^2 N^-(\pi/2)}{N} \geq 1 \tag{2}$$

This inequality also applies locally to any subsystem where N represents the corresponding subsystem particle number.

We analyzed the absorption images by first partitioning each of them into two halves, A and
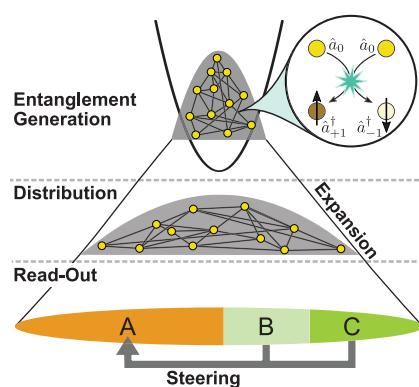
**Fig. 1. Distribution of entanglement.**
In a tightly trapped BEC, entanglement in the spin degree of freedom is generated by local spin-mixing interactions. Switching off the longitudinal confinement leads to a rapid expansion of the atomic cloud, which distributes the entanglement spatially. After local spin measurements with high spatial resolution, we partitioned the detected atomic signal into distinct subsystems. We demonstrated EPR steering between these parts, which evinces the presence of bipartite and even multipartite entanglement.

Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany.
*Corresponding author. Email: steering@matterwave.de

Fig. 2. EPR steering. (**A**) A global change of the phase φ before the measurement allows mapping of the spin observable $\hat{F}(\phi)$ to the read-out direction $\hat{F}(0)$ (inset). We partitioned the atomic signal into two halves and observed that for subsystem A, the fluctuations of $\Delta^2 N_A^-/N_A = \Delta^2(N_{A,+1} - N_{A,-1})/N_A$ are, depending on the value of phase φ, reduced or enhanced compared with the shot-noise limit of a fully separable spin state (dashed line). The solid line is a theoretical prediction based on our experimental parameters (27). At phase φ = 0, fluctuations are reduced, whereas the fluctuations are enhanced at phase φ = π/2. (**B**) The measurement result in subsystem B is used to infer the result in subsystem A (inset), leading to an inference variance $\Delta^2 N_{A|B}^-$. The solid line represents the theoretical prediction. The data in the gray shaded region are used to calculate the EPR steering product $S_{A|B}$. (**C**) We varied the spatial separation (d) between the two subsystems by discarding a fraction η of atomic signal in the middle of the cloud (inset). The red and blue diamonds are the products $S_{A|B} = \Delta^2 N_{A|B}^-(0)\Delta^2 N_{A|B}^-(\pi/2)/N_A^2$ of the inference variances after 60 and 150 ms of spin-mixing time (t_{evo}), respectively. The individual inference variances $\Delta^2 N_{A|B}^-/N_A$ at φ = 0 and φ = π/2 are shown as black triangles and squares, respectively. The steering product remains below the EPR steering bound even if a substantial fraction of the atomic signal is discarded, confirming the spatial distribution of entanglement in our system. The error bars correspond to an estimation of the 1-SD interval.
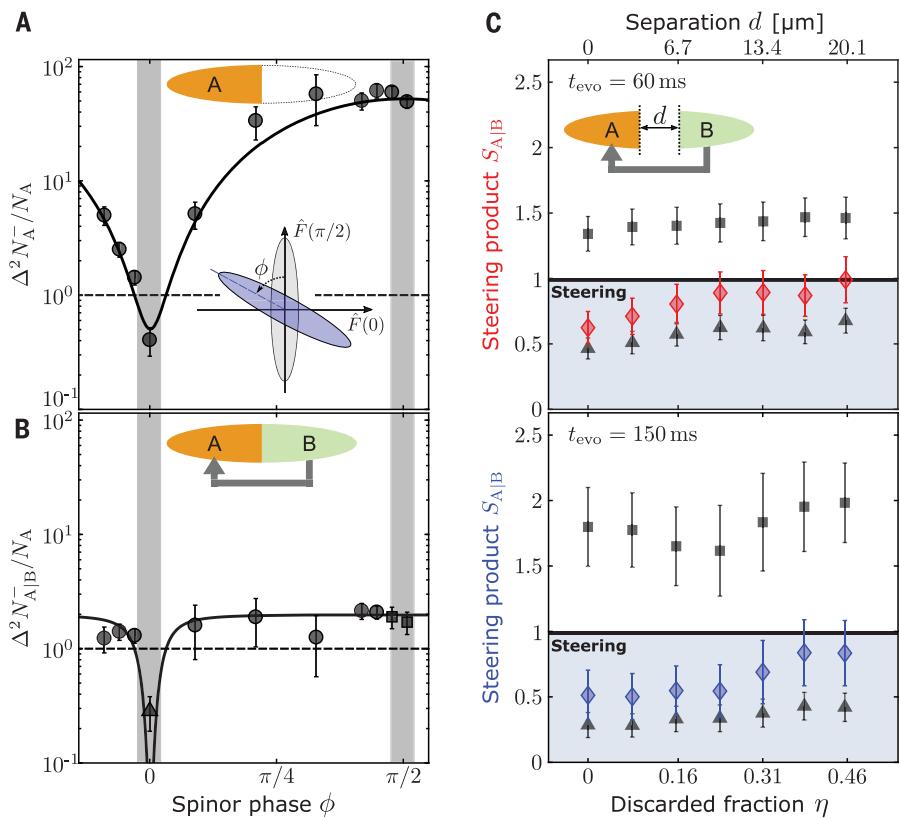
Fig. 3. Three-way EPR steering. By partitioning the absorption signal into three parts of equal length (~20 μm), we show that each of the three subsystems is steered by the other two. For each case, we calculate the steering product $S_{\alpha|\beta\gamma} = \Delta^2 N_{\alpha|\beta\gamma}^-(0)\Delta^2 N_{\alpha|\beta\gamma}^-(\pi/2)/N_\alpha^2$, where $N_{\alpha|\beta\gamma}^-(\phi)$ denotes the optimal inference on the observable $N_\alpha^-(\phi)$ in subsystem α, using the information obtained from the other two subsystems (β, γ). The red (blue) points are the results for 60 ms (150 ms) of spin-mixing time. The black line represents the steering bound. The error bars correspond to an estimation of the 1-SD interval.
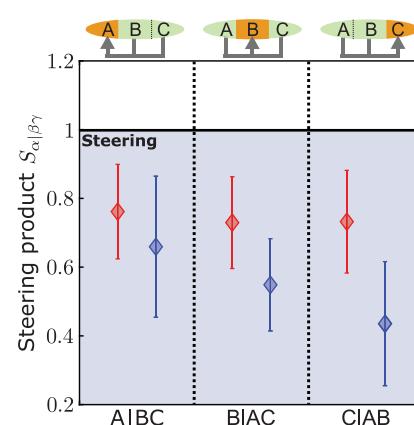
B. In subsystem A, we detected reduced (enhanced) fluctuations of $N_A^-(\phi)$ at phase φ = 0 (φ = π/2) compared with the case of a fully separable initial state (Fig. 2A). The minimum fluctuations are below the separable-state limit (nematic squeezing), and the variance product $\Delta^2 N_A^-(0)\Delta^2 N_A^-(\pi/2)/N_A^2 = 0.41 \cdot 49$ exceeds the uncertainty limit. To reveal EPR steering of A, we demonstrate that a measurement in subsystem B can be used to infer the outcome in A with an accuracy beating the local uncertainty limit $\Delta^2 N_A^-(0)\Delta^2 N_A^-(\pi/2)/N_A^2 \geq 1$ (28).

The outcome in A can be estimated by an arbitrary function of the measurement result in B. We constructed an estimator based on five sub-

divisions of B. The corresponding values of $N_{B,k}^-$ were used to infer the result in A through the linear combination $N_{A,inf}^-(\phi) = \sum_{k=1}^{5} g_k(\phi) N_{B,k}^-(\phi)$. The real numbers $g_k(\phi)$ were chosen to minimize the inference variance

$$\Delta^2 N_{A|B}^-(\phi) = \Delta^2[N_A^-(\phi) - N_{A,inf}^-(\phi)] \quad (3)$$

which is depicted in Fig. 2B. The inference variance quantifies the accuracy with which N_A^- can be inferred by the estimator $N_{A,inf}^-$. To compare the achieved accuracy with the local

uncertainty relation, we evaluated the steering product

$$S_{A|B} = \frac{\Delta^2 N_{A|B}^-(0)}{N_A} \frac{\Delta^2 N_{A|B}^-(\pi/2)}{N_A} \quad (4)$$

$S_{A|B} < 1$ signals EPR steering of A by B. In our experiment, we obtained a value of $S_{A|B} = 0.62 \pm 0.12$ and $S_{A|B} = 0.51 \pm 0.19$ after 60 and 150 ms of spin-mixing dynamics, respectively, verifying bipartite EPR steering in our system. The given errors correspond to the statistical estimation of 1 standard deviation (SD), applying jackknife resampling. For all given variances, the independently characterized photon shot noise contribution to the absorption signal has been subtracted (27). To underscore the spatial separation between the steered subsystems, we discarded a fraction η of the atoms in a region between A and B. Figure 2C shows that EPR steering can be verified up to a discarded fraction of ~30% of the atoms, which corresponds to a separation of ~13 μm between the two subsystems. This is consistent with monogamy of steering (29), which implies that by discarding more than a third of the whole system, no steering between equal partitions of the remaining system is possible.

For indistinguishable particles, one expects that the entanglement is uniformly distributed over the whole system. We illustrated this by
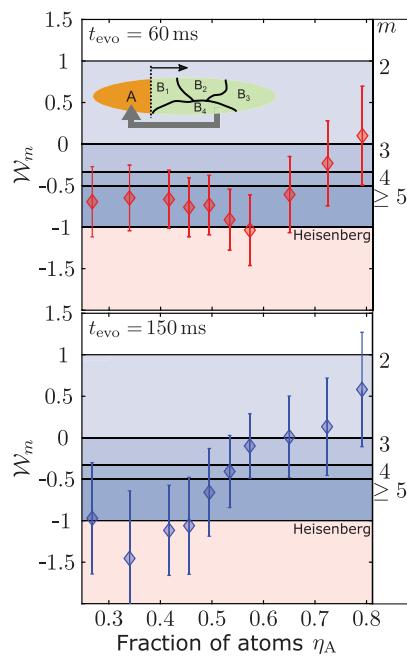
Fig. 4. Genuine multipartite entanglement.
In the bipartite steering scenario, the possible inference of subsystem B on subsystem A is used to reveal genuine multipartite entanglement. For each partition A|B, quantified by $\eta_A = N_A/N$, subsystem B can be divided into additional $m-1$ parts of equal atom number (the inset shows an example). The regions where genuine m-partite entanglement is witnessed according to Eq. 5 are indicated by the blue shadings, where the corresponding m is given on the right. The upper (lower) panel shows the results for 60 ms (150 ms) of spin-mixing time. The lowest bound is given by the Heisenberg uncertainty limit for our observables in the full system. The error bars correspond to an estimation of the 1-SD interval.

partitioning the absorption signal into three parts of equal length. Analogous to the analysis above, we evaluated the inference variance $\Delta^2 N^-_{A|BC}/N_A$ for all permutations of ABC. Figure 3 shows that each part is steered by the remaining atomic cloud, confirming three-way steering (30). For partitions that are too small, spurious effects of the imaging technique become relevant. Because the position of each atom is mapped onto a spatially distributed absorption signal, classical correlations are dominant below a certain length scale. By analyzing a fully separable coherent spin state, we confirmed that for the partitions chosen in this work, classical correlations are negligible (27).

The observation of EPR steering allows for statements about multipartite entanglement. Specifically, the steering product can be used to construct a witness $\mathcal{W}_m$ for genuine m-partite

entanglement. For this, we partitioned the system into a subsystem A and the remainder B, which we divided into $m-1$ parts with equal atom numbers. Generalizing the derivation in (30), we find that genuine m-partite entanglement (31) is present if the inequality

$$\mathcal{W}_m = \frac{\eta_A}{1-\eta_A} \frac{(1-\sqrt{S_{A|B}})}{g(0)\,g(\pi/2)} < \frac{3-m}{m-1} \qquad (5)$$

is fulfilled, given that $g(0) \cdot g(\pi/2) < 0$ (27). Here, $\eta_A = N_A/N$ denotes the fraction of atoms in system A, and the inferences are $N^-_{A,\mathrm{inf}}(\phi) = g(\phi)N^-_B(\phi)$. If Eq. 5 holds, the quantum state of the system cannot be written as a mixture of states that are separable with respect to all possible bipartitions. This implies that each part, or conjunction of parts, is entangled with the rest of the system. In the limit $m \to \infty$, Eq. 5 cannot hold owing to the Heisenberg uncertainty limit of the full system. Experimentally, we partitioned the absorption data of the atomic cloud into two parts and varied the fraction η_A (Fig. 4, inset). In this way, we verified genuine 5-partite entanglement (Fig. 4).

Our results, combined with the well-developed toolbox for the manipulation of ultracold gases, offer new perspectives for applications, as well as for fundamental questions. Retrapping and storage of the produced states in tailored potentials enable quantum enhanced sensing of spatially varying external fields. With the possibility of local control, the deterministic generation of more general classes of entangled states, including cluster states that are useful for continuous-variable quantum computation, is in reach (32). Our general strategy for the detection of entanglement between spatially separated regions can be applied to fundamental questions concerning the role of entanglement in long-time dynamics and thermalization of quantum many-particle systems (33).

Complementarily to our work, the detection of spatial entanglement patterns is reported in (34), and the observation of entanglement of spatially separated modes is described in (35).

REFERENCES AND NOTES

1. M. D. Reid et al., Rev. Mod. Phys. 81, 1727–1751 (2009).
2. A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. 47, 777–780 (1935).
3. E. Schrödinger, M. Born, Math. Proc. Camb. Philos. Soc. 31, 555 (1935).
4. H. M. Wiseman, S. J. Jones, A. C. Doherty, Phys. Rev. Lett. 98, 140402 (2007).
5. V. Giovannetti, S. Lloyd, L. Maccone, Science 306, 1330–1336 (2004).
6. N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Rev. Mod. Phys. 74, 145–195 (2002).
7. D. Cavalcanti, P. Skrzypczyk, Rep. Prog. Phys. 80, 024001 (2017).
8. S. L. Braunstein, P. van Loock, Rev. Mod. Phys. 77, 513–577 (2005).
9. S. J. Freedman, J. F. Clauser, Phys. Rev. Lett. 28, 938–941 (1972).
10. A. Aspect, P. Grangier, G. Roger, Phys. Rev. Lett. 47, 460–463 (1981).
11. Z. Y. Ou, S. F. Pereira, H. J. Kimble, K. C. Peng, Phys. Rev. Lett. 68, 3663–3666 (1992).
12. B. Julsgaard, A. Kozhekin, E. S. Polzik, Nature 413, 400–403 (2001).
13. N. J. Cerf, G. Leuchs, E. S. Polzik, Quantum Information with Continuous Variables of Atoms and Light (Imperial College Press, 2007).
14. I. Bloch, Nature 453, 1016–1022 (2008).
15. H.-N. Dai et al., Nat. Phys. 12, 783–787 (2016).
16. J. Zeiher et al., Phys. Rev. X 5, 031015 (2015).
17. H. Labuhn et al., Nature 534, 667–670 (2016).
18. Q. Y. He et al., Phys. Rev. Lett. 106, 120405 (2011).
19. N. Bar-Gill, C. Gross, I. Mazets, M. Oberthaler, G. Kurizki, Phys. Rev. Lett. 106, 120404 (2011).
20. H. Kurkjian, K. Pawłowski, A. Sinatra, P. Treutlein, Phys. Rev. A 88, 043605 (2013).
21. J. Peise et al., Nat. Commun. 6, 8984 (2015).
22. L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, P. Treutlein, arXiv:1609.01609 [quant-ph] (6 September 2016).
23. P. Hyllus, L. Pezzé, A. Smerzi, G. Tóth, Phys. Rev. A 86, 012337 (2012).
24. N. Killoran, M. Cramer, M. B. Plenio, Phys. Rev. Lett. 112, 150501 (2014).
25. D. M. Stamper-Kurn, M. Ueda, Rev. Mod. Phys. 85, 1191–1244 (2013).
26. C. D. Hamley, C. S. Gerving, T. M. Hoang, E. M. Bookjans, M. S. Chapman, Nat. Phys. 8, 305–308 (2012).
27. Details are given in the supplementary materials.
28. M. D. Reid, Phys. Rev. A 40, 913–923 (1989).
29. M. D. Reid, Phys. Rev. A 88, 062108 (2013).
30. S. Armstrong et al., Nat. Phys. 11, 167–172 (2015).
31. R. Y. Teh, M. D. Reid, Phys. Rev. A 90, 062337 (2014).
32. N. C. Menicucci et al., Phys. Rev. Lett. 97, 110501 (2006).
33. M. Rigol, V. Dunjko, M. Olshanii, Nature 452, 854–858 (2008).
34. M. Fadel, T. Zibold, B. Décamps, P. Treutlein, Science 360, 409–413 (2018).
35. K. Lange et al., Science 360, 416–418 (2018).

QUANTUM ENTANGLEMENT

Entanglement between two spatially separated atomic modes

Karsten Lange,[1] Jan Peise,[1] Bernd Lücke,[1] Ilka Kruse,[1] Giuseppe Vitagliano,[2,3] Iagoba Apellaniz,[3] Matthias Kleinmann,[3,4] Géza Tóth,[3,5,6] Carsten Klempt[1]*

Modern quantum technologies in the fields of quantum computing, quantum simulation, and quantum metrology require the creation and control of large ensembles of entangled particles. In ultracold ensembles of neutral atoms, nonclassical states have been generated with mutual entanglement among thousands of particles. The entanglement generation relies on the fundamental particle-exchange symmetry in ensembles of identical particles, which lacks the standard notion of entanglement between clearly definable subsystems. Here, we present the generation of entanglement between two spatially separated clouds by splitting an ensemble of ultracold identical particles prepared in a twin Fock state. Because the clouds can be addressed individually, our experiments open a path to exploit the available entangled states of indistinguishable particles for quantum information applications.

The progress toward large ensembles of entangled particles is pursued along two different paths. In a bottom-up approach, the precise control and characterization of small systems of ions, atoms in optical lattices, and photons is pushed toward increasingly large system sizes, reaching entangled states of 14 ions (1), 4 atomic lattice sites (2), or 10 photons (3). Complementary, large numbers of up to 3000 mutually entangled ultracold atoms (4–6) can be generated, for which the state characterization is advanced top-down toward resolving correlations on the single-particle level. Because the atoms cannot be addressed individually, ultracold atomic ensembles are controlled by global ensemble parameters, such as the total spin. Ideally, the atoms are indistinguishable, either with respect to the observable, such as the spin in hot vapor cells (7), or in all quantum numbers in Bose-Einstein condensates (BECs) (8–13). Is it possible to make these particles distinguishable—and addressable—again, while keeping the high level of entanglement?

The generation of entanglement in these systems is deeply connected with the fundamental indistinguishability of the particles (14). For example, two indistinguishable bosons 1 and 2, which are prepared in two independent modes a and b, are described by an entangled triplet state $\frac{1}{\sqrt{2}}(|a\rangle_1|b\rangle_2 + |b\rangle_1|a\rangle_2)$, owing to bosonic symmetrization. Although this type of entanglement seems to be artificial, the state presents a resource for violating a Bell inequality (15). Equivalently, an ensemble containing the same number of distinguishable spin-up and spin-down atoms is not necessarily entangled, whereas a twin Fock state, the corresponding ensemble with indistinguishable bosons, exhibits full many-particle entanglement (16, 17). This form of entanglement is directly applicable for atom interferometry beyond the standard quantum limit (11). However, most quantum information tasks require an individual addressing of subsystems. Despite the experimental progress in entanglement creation in BECs—including the demonstration of Einstein-Podolsky-Rosen (EPR) correlations (18) and Bell correlations (19–21), as well as the demonstration of strongly correlated momentum states (22–24)—a proof of entanglement between spatially separated and individually addressable subsystems has not been realized so far. The possible applications of such a resource range from spatially resolved quantum metrology to tests for fundamental

sources of decoherence or Bell tests of quantum nonlocality.

Here, we report the creation of particle entanglement in an ensemble of up to 5000 indistinguishable atoms and prove entanglement between two spatially separated clouds.

Our experiments started with the preparation of a ^{87}Rb BEC in a crossed-beam optical dipole trap. The ensemble of 20,000 particles was transferred to the hyperfine level $F = 1$, $m_F = 0$. Spin-changing collisions created up to 5000 entangled atoms in the Zeeman levels $m_F = \pm 1$, which resided in a spatially excited mode of the dipole trap (25, 26) (Fig. 1). The output state consisted of a superposition of twin Fock states with varying total number of atoms $N = N_{+1} + N_{-1}$. Each twin Fock state is characterized by an equal number of atoms $N_{+1} = N_{-1}$ in the two Zeeman levels $m_F = \pm 1$ (11). Because the total number of particles N was measured during detection, the system was well described by a single twin Fock state with a definite particle number. Self-similar expansion (26) allowed for an imaging of the undistorted but magnified density profiles. An inhomogeneous magnetic field separated the atoms in order to record the atomic densities for each Zeeman level.

The spatially excited mode of the ensembles in $m_F = \pm 1$ provides a natural splitting into a left and right cloud along a line of zero density. Hence, we divided the initial twin Fock state into two spatially separated parts, $|a\rangle$ (left side) and $|b\rangle$ (right side). This process can be described as a beam splitter of the initially populated anti-symmetric input mode $\frac{1}{\sqrt{2}}(|a\rangle - |b\rangle)$. The splitting introduces additional quantum noise caused by a coupling with the (empty) symmetric input mode $\frac{1}{\sqrt{2}}(|a\rangle + |b\rangle)$. In principle, an ideal twin Fock state shows a maximal entanglement depth (16); all atoms that make up a twin Fock state are entangled with one another. Therefore, any splitting results in the appearance of quantum correlations between the clouds. Such a creation of entanglement is analogous to so-called vacuum-class entanglement (27) in optics, in which a single-mode squeezed laser beam is coupled to a vacuum state on a beam splitter, creating entanglement between the two output modes (28).

The resulting quantum correlations can be visualized on the multiparticle Bloch sphere (Fig. 2A). Here, the atoms in the levels $m_F = \pm 1$ are represented by spin-$\frac{1}{2}$ particles, whose spins $\mathbf{j}^{(k)}$ sum up to a total spin $\mathbf{J}$. On the Bloch sphere, the lines of latitude represent the number imbalance

Fig. 1. Generation of entanglement between two spatially separated atomic clouds. (**A**) A BEC of atoms in the Zeeman level $m_F = 0$ is prepared in a crossed-beam optical trap. Collisions generate entangled pairs of atoms in the levels $m_F = \pm 1$ (spin up/down) in the first spatially excited mode. The created multiparticle-entangled ensemble is naturally divided into two clouds (red and blue). (**B**) The atomic density profile obtained from an average over 3329 measurements is shown in the background. The entanglement between the two clouds (indicated schematically with green lines) can be detected by analyzing spin correlations.

[1]Institut für Quantenoptik, Leibniz Universität Hannover, Welfengarten 1, D-30167 Hannover, Germany. [2]Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria. [3]Department of Theoretical Physics, University of the Basque Country UPV/EHU, Post Office Box 644, E-48080 Bilbao, Spain. [4]Naturwissenschaftlich–Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, D-57068 Siegen, Germany. [5]Ikerbasque, Basque Foundation for Science, E-48013 Bilbao, Spain. [6]Wigner Research Centre for Physics, Hungarian Academy of Sciences, Post Office Box 49, H-1525 Budapest, Hungary.
*Corresponding author. Email: klempt@iqo.uni-hannover.de

between the two levels, and the lines of longitude represent the phase difference. An ensemble in a twin Fock state can be depicted as a ring on the Bloch sphere, characterized by a vanishing imbalance, $J_z = (N_{+1} - N_{-1})/2 = 0$, and an undetermined phase difference.

If we divide the cloud, the collective spins $\mathbf{J}^{(a)}$, $\mathbf{J}^{(b)}$ of the two parts have to sum up to the original collective spin of the full ensemble $\mathbf{J} = \mathbf{J}^{(a)} + \mathbf{J}^{(b)}$. Therefore, the z components of the collective spins are perfectly anticorrelated $J_z^{(a)} + J_z^{(b)} = 0$. Furthermore, because the spin length $|\mathbf{J}|$ is maximal, the collective spins of the two parts have to point in a similar direction in the x-y plane and thus have similar azimuthal angles $\phi^{(a)} \approx \phi^{(b)}$.

Hence, if the particle number difference of cloud b is measured to yield $J_z^{(b)}$, the conditioned state of cloud a satisfies $J_z^{(a)} = -J_z^{(b)}$. If the value $J_x^{(b)}$ is measured on cloud b, the state of cloud a has to fulfill $J_x^{(a)} \approx J_x^{(b)}$; if the value $J_y^{(b)}$ is measured on cloud b, the state of cloud a has to fulfill $J_y^{(a)} \approx J_y^{(b)}$. The different possible measurements on cloud b yield precise predictions for the measurement results of cloud a, which cannot be explained by a single quantum state that is independent of the chosen type of measurement. In this sense, the described system is analogous to the thought experiment by Einstein, Podolsky, and Rosen (29), in which entanglement is witnessed by the variances of the predictions (30, 31). Therefore, it should be possible to detect entanglement between the spatially separated parts of a twin Fock state.

To this end, we derived an entanglement criterion that optimally exploits the described spin correlations (32). The spin correlations are represented by prediction operators $\tilde{J}_z^+ = \tilde{J}_z^{(a)} + \tilde{J}_z^{(b)}$ and $\tilde{J}_m^- = \tilde{J}_m^{(a)} - \tilde{J}_m^{(b)}$ for $m = x, y$. Here, the x and y components are normalized so that the optimal value is 1, according to $\tilde{J}_m^{(n)} = \frac{J_m^{(n)}/j_n}{\mathcal{J}^{(n)}}$, with $\mathcal{J}^{(n)} = \left\langle \frac{(J_x^{(n)})^2 + (J_y^{(n)})^2}{j_n^2} \right\rangle^{\frac{1}{2}}$ and the spin length $j_n = N_n/2$ for $n = $ a, b. We arrive at a simple separability criterion

$$\left[(\Delta \tilde{J}_z^+)^2 + \frac{1}{2} \right] \times \left[\left\langle (\tilde{J}_x^-)^2 + (\tilde{J}_y^-)^2 \right\rangle \right] \geq f\left(\mathcal{J}^{(a)}, \mathcal{J}^{(b)} \right) \quad (1)$$

where $f(r,s) = \frac{(r^2 + s^2 - 1)^2}{rs}$. Any separable state, including mixtures of product states of the form $\sum_k p_k |\Psi_k^{(a)}\rangle\langle\Psi_k^{(a)}| \otimes |\Psi_k^{(b)}\rangle\langle\Psi_k^{(b)}|$ with a fluctuating number of particles, fulfills this inequality. A violation of this criterion indicates that the state is inseparable and therefore entangled. For perfectly symmetric states, as we would expect in the ideal case, the right-hand side (RHS) of Eq. 1 is equal to 1. Any deterioration from perfect symmetry is quantified by $\mathcal{J}^{(a)}$ and $\mathcal{J}^{(b)}$. Containing the characteristic product of the prediction uncertainties, our inequality for three-dimensional spins has similarities to the famous two-dimensional continuous-variable entanglement criteria (31, 33, 34), specifically to the EPR criterion (33). It presents a general entanglement criterion, which is particularly sensitive for a spatially separated twin Fock state.

An application of criterion in Eq. 1 requires an evaluation of the spin correlations between the two clouds a and b. The measurement results for $J_z^{(a)}$ and $J_z^{(b)}$ are readily obtained from the absorption images. The measurement of the orthogonal direction is performed with a sequence of resonant microwave pulses before the particle number detection (32). The pulses lead to an effective rotation of the spins by $\pi/2$. Because the microwave phase is independent of the atomic phases, the rotation yields a measurement of the spin component $J_\perp$ along an arbitrary angle in the x-y plane. Because our quantum state is symmetric under rotations around the z axis, both owing to the initial symmetry and the influence of magnetic field noise, the measured distributions of $J_\perp$ can be identified with both J_x and J_y.

The histograms of J_z and $J_\perp/j$ are shown in Fig. 2, B and C, for a mean total number of 3460 particles in both clouds together. The J_z data show the expected anticorrelation, whereas the $J_\perp$ measurements are strongly correlated. The strength of these correlations can be quantified by evaluating the prediction uncertainties—the width of the distributions in the diagonal directions in the histograms: $(\Delta J_z^+)^2$ and $\langle (\tilde{J}_\perp^-)^2 \rangle$.

The prediction variance $(\Delta J_z^+)^2$ is presented in Fig. 3A as a function of the total number of atoms. The shown fluctuations, obtained by subtracting independent detection noise (32), remain

Fig. 2. Spin correlations between the clouds a and b.

(A) A twin Fock state is represented by a narrow ring on the equator of the multiparticle Bloch sphere (top, orange). When the system is split into two parts, a (left, blue) and b (right, red), the states in the individual subsystems gain uncertainty. However, a measurement of $J_z^{(b)}$ or $\phi^{(b)}$ on cloud b allows for a prediction of the measurement outcome of cloud a, so that $J_z^{(a)} = -J_z^{(b)}$ or $\phi^{(a)} \approx \phi^{(b)}$. (B) Histogram of 506 measurements of $J_z^{(a)}$ and $J_z^{(b)}$ for a mean total number of 3460 atoms. The data show the anticipated anticorrelation between $J_z^{(a)}$ and $J_z^{(b)}$. (C) The strong correlations between the angles $\phi^{(a)}$ and $\phi^{(b)}$ are recorded by a measurement of their projection on an arbitrary axis $J_\perp$ in the x-y plane. The histogram of $J_\perp^{(a)}/j_a$ and $J_\perp^{(b)}/j_b$ is drawn from 487 measurements.

Fig. 3. Spin correlations as a function of the total number of atoms N.

(A) The prediction variance $(\Delta J_z^+)^2$ (green circles) surpasses the shot-noise limit (black solid line), indicating number squeezing of up to −11.0(5) dB. The number-dependent detection noise is modeled by a linear fit (gray dashed line). (B) The fluctuations $\langle (\tilde{J}_\perp^-)^2 \rangle = \langle (\tilde{J}_x^-)^2 \rangle = \langle (\tilde{J}_y^-)^2 \rangle$ (green circles), corresponding to the phase prediction variance in the experiment, show excess noise, which increases the standard deviation by a factor of 1.8 (gray dashed line) above the shot-noise limited case (black solid line). (C) $\mathcal{J}^{(a)}$ and $\mathcal{J}^{(b)}$ quantify the symmetry of the states. The value 1 for purely symmetric states is indicated with the black line, and the mean experimental value is indicated with the gray dashed line. In (A) to (C), we binned our experimental data into sets of 1000 atoms. N corresponds to the mean number of atoms in each set. The error bars represent 1 standard deviation of the statistical fluctuations and are obtained via a bootstrapping method (32).
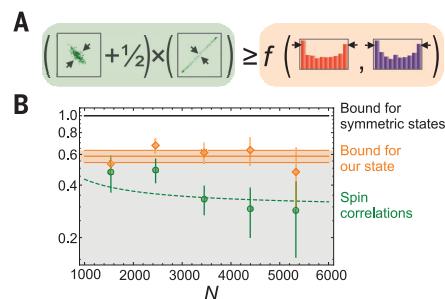
Fig. 4. Violation of the separability criterion as a function of the total number of atoms N.
(**A**) Pictorial representation of the criterion in Eq. 1. (**B**) The black line represents the RHS of Eq. 1 for the ideal case. The orange line represents the mean of the experimental results (orange diamonds) of the RHS of Eq. 1, where the spin length is reduced. The green circles show the experimental results for the LHS of Eq. 1 . The dashed green line indicates the prediction of the LHS corresponding to the gray lines in Fig. 3. The spin correlations clearly violate the criterion by 2.8 standard deviations at a mean total number of 3460 atoms. The error bars and shaded orange area indicate 1 standard deviation and are obtained via a bootstrapping method (32).

well below the shot-noise limit and are equivalent to a number squeezing of –11.0(5) dB. The orthogonal quantities (Fig. 3B) are slightly influenced by small position fluctuations of the clouds, increasing the standard deviation by a factor of 1.8 above shot noise. The quantities $\mathcal{J}^{(a)}$ and $\mathcal{J}^{(b)}$ are shown in Fig. 3C. We obtained a value of up to 0.94, close to the ideal value of 1, indicating a sufficiently clean preparation of an almost symmetric state.

From these results, we could test a violation of the separability criterion. In Fig. 4B, the orange diamonds correspond to an evaluation of the RHS of the criterion (Eq. 1), which would ideally be 1 (black line). For most values of N, the left-hand side (LHS), represented in Fig. 4B by the green circles, is well below the RHS, signaling entanglement in the system. At the best value of the total number of 3460 atoms, the experimental data violate the separability criterion by 2.8 standard deviations. Therefore, our measurements cannot result from classical correlations and prove the generation of entanglement

between spatially separated clouds from particle-entangled, indistinguishable atoms.

The presented creation of spatial entanglement opens a path to create highly entangled quantum states in spatially separated modes for a broad range of applications. This includes the field of quantum metrology, where our results can be applied in order to obtain an improved spatial resolution—for example, in the sensing of electromagnetic fields. Owing to the possibility of addressing the separated modes—for example, with Raman beams or spatially modulated light shifts—the created states are a resource for quantum information protocols. For example, it presents a resource for the synthesis of any pure symmetric state with only single-particle projective measurements (35, 36). Moreover, the spatially separated twin Fock state carries nonlocal EPR and Bell correlations. The presented experimental scheme can be extended to a multiparticle Bell test of quantum mechanics with the addition of local addressing of the two modes and a single-particle–resolving atom counting (37). Last, if the created twin Fock state is fully separated into single atoms—for example, in an optical lattice—all constituent particles become individually addressable.

Complementary to our work, in (38), spatially distributed multipartite entanglement has been observed, whereas spatial entanglement patterns were observed in (39).

REFERENCES AND NOTES

1. T. Monz et al., Phys. Rev. Lett. **106**, 130506 (2011).
2. R. Islam et al., Nature **528**, 77–83 (2015).
3. X.-L. Wang et al., Phys. Rev. Lett. **117**, 210502 (2016).
4. R. McConnell, H. Zhang, J. Hu, S. Ćuk, V. Vuletić, Nature **519**, 439–442 (2015).
5. F. Haas, J. Volz, R. Gehr, J. Reichel, J. Estève, Science **344**, 180–183 (2014).
6. O. Hosten, N. J. Engelsen, R. Krishnakumar, M. A. Kasevich, Nature **529**, 505–508 (2016).
7. B. Julsgaard, A. Kozhekin, E. S. Polzik, Nature **413**, 400–403 (2001).
8. J. Estève, C. Gross, A. Weller, S. Giovanazzi, M. K. Oberthaler, Nature **455**, 1216–1219 (2008).
9. C. Gross, T. Zibold, E. Nicklas, J. Estève, M. K. Oberthaler, Nature **464**, 1165–1169 (2010).
10. M. F. Riedel et al., Nature **464**, 1170–1173 (2010).
11. B. Lücke et al., Science **334**, 773–776 (2011).
12. C. D. Hamley, C. S. Gerving, T. M. Hoang, E. M. Bookjans, M. S. Chapman, Nat. Phys. **8**, 305–308 (2012).
13. T. Berrada et al., Nat. Commun. **4**, 2077 (2013).
14. N. Killoran, M. Cramer, M. B. Plenio, Phys. Rev. Lett. **112**, 150501 (2014).
15. B. Yurke, D. Stoler, Phys. Rev. A **46**, 2229–2234 (1992).
16. B. Lücke et al., Phys. Rev. Lett. **112**, 155304 (2014).
17. X.-Y. Luo et al., Science **355**, 620–623 (2017).
18. J. Peise et al., Nat. Commun. **6**, 8984 (2015).
19. J. Tura et al., Science **344**, 1256–1258 (2014).
20. J. Tura et al., Ann. Phys. **362**, 370–423 (2015).
21. R. Schmied et al., Science **352**, 441–444 (2016).
22. R. Bücker et al., Nat. Phys. **7**, 608–611 (2011).
23. J. Lopes et al., Nature **520**, 66–68 (2015).
24. P. Dussarrat et al., Phys. Rev. Lett. **119**, 173202 (2017).
25. C. Klempt et al., Phys. Rev. Lett. **103**, 195302 (2009).
26. M. Scherer et al., Phys. Rev. Lett. **105**, 135302 (2010).
27. J. DiGuglielmo, B. Hage, A. Franzen, J. Fiurášek, R. Schnabel, Phys. Rev. A **76**, 012323 (2007).
28. T. Eberle et al., Phys. Rev. A **83**, 052329 (2011).
29. A. Einstein, B. Podolsky, N. Rosen, Phys. Rev. **47**, 777–780 (1935).
30. L.-M. Duan, G. Giedke, J. I. Cirac, P. Zoller, Phys. Rev. Lett. **84**, 2722–2725 (2000).
31. R. Simon, Phys. Rev. Lett. **84**, 2726–2729 (2000).
32. Materials and methods are available as supplementary materials.
33. M. D. Reid, Phys. Rev. A **40**, 913–923 (1989).
34. L.-M. Duan, A. Sørensen, J. I. Cirac, P. Zoller, Phys. Rev. Lett. **85**, 3991–3994 (2000).
35. W. Wieczorek et al., Phys. Rev. Lett. **103**, 020504 (2009).
36. N. Kiesel et al., Phys. Rev. A **81**, 032316 (2010).
37. F. Laloë, W. J. Mullin, Eur. Phys. J. B **70**, 377–396 (2009).
38. P. Kunkel et al., Science **360**, 413–416 (2018).
39. M. Fadel, T. Zibold, B. Décamps, P. Treutlein, Science **360**, 409–413 (2018).

ORGANIC CHEMISTRY

Catalytic enantioselective Minisci-type addition to heteroarenes

Rupert S. J. Proctor, Holly J. Davis, Robert J. Phipps*

Basic heteroarenes are a ubiquitous feature of pharmaceuticals and bioactive molecules, and Minisci-type additions of radical nucleophiles are a leading method for their elaboration. Despite many Minisci-type protocols that result in the formation of stereocenters, exerting control over the absolute stereochemistry at these centers remains an unmet challenge. We report a process for addition of prochiral radicals, generated from amino acid derivatives, to pyridines and quinolines. Our method offers excellent control of both enantioselectivity and regioselectivity. An enantiopure chiral Brønsted acid catalyst serves both to activate the substrate and induce asymmetry, while an iridium photocatalyst mediates the required electron transfer processes. We anticipate that this method will expedite access to enantioenriched small-molecule building blocks bearing versatile basic heterocycles.

Heteroarenes with basic nitrogen centers, of which pyridines and quinolines are the most common classes, are ubiquitous in pharmaceuticals, agrochemicals, and small molecules of medicinal interest (*1*, *2*). The nitrogen's basicity precludes traditional electrophilic aromatic substitution pathways, necessitating alternative strategies for elaboration. One widely used protocol is the addition of free radicals to protonated heterocycles, pioneered by Minisci in the 1960s and 1970s and often referred to as "Minisci-type" chemistry (*3–6*). Recent renewed interest has led to numerous advances in radical generation, allowing for milder conditions and convenient radical precursors (*7–11*). Furthermore, photoredox catalysis (*12–14*) has provided exciting avenues for radical generation in Minisci-type additions with precursors that include alcohols (*15*, *16*), ethers (*17*), boronic acids (*18*), carboxylic acids (*19*), and redox-active esters (RAEs) (*20*, *21*). However, control over absolute stereochemistry in the addition of prochiral radicals has proven elusive, even though many of the aforementioned protocols result in the formation of a stereocenter at the benzylic position.

There is an increasing drive in the pharmaceutical industry for higher degrees of three-dimensionality in lead compounds, which necessitates asymmetric introduction of chiral centers (*22*). Realization of a catalytic enantioselective Minisci-type addition would constitute a powerful and efficient strategy for the enantioselective synthesis of basic heteroarenes bearing adjacent stereocenters. To put this prospect into context, a typical existing approach to enantioenriched α-pyridyl amines involves diastereoselective addition of organometallics to enantiopure pyridyl imines derived from Ellman's sulfonamide auxiliary (Fig. 1A, upper scheme) (*23*). Stereospecific coupling of pyrrolidine-derived

chiral boronic esters has also been demonstrated (*24*). Approaches using asymmetric catalysis are typically lengthy, involving asymmetric reduction of a pyridyl ketone, conversion to a leaving group, and azide displacement (lower scheme). An alternative catalytic route could involve installation of an enamide functionality onto the pyridine ring, followed by asymmetric hydrogenation (*25*). A recent strategy merged nickel catalysis and photoredox catalysis to access 3- and 4-substituted pyridines (*26*). In all cases, a functional group must first be installed on the pyridine scaffold in a regioselective manner, necessitating additional chemical steps.

Here, we disclose the direct addition of α-amino alkyl radicals to the 2-position of basic heteroarenes by virtue of a combination of asymmetric Brønsted acid catalysis and photoredox catalysis (Fig. 1B). This catalytic approach does not require prefunctionalization of the heterocycle and allows excellent control of both enantioselectivity and regioselectivity. The products generated possess structural features highly desirable in pharmaceutical compounds: a basic heteroarene, a protected primary amine, and a defined stereocenter, all in close proximity. In

addition, this work constitutes a rare case of noncovalent organocatalysis being applied to control enantioselectivity in a single-electron process (*27–33*).

In considering a strategy toward this goal, we surmised that Minisci chemistry should, in principle, be amenable to asymmetric Brønsted acid catalysis (*34*). The heteroarene's lowest unoccupied molecular orbital is considerably lowered upon protonation, and the conjugate anion of the chiral acid should remain associated with the pyridinium cation through electrostatic and hydrogen-bonding interactions, thereby providing substrate activation in a chiral environment. Shang, Fu, and co-workers recently reported (*20*, *21*) that catalytic amounts of achiral or racemic Brønsted and Lewis acids can be used to promote Minisci additions, in place of the stoichiometric acid that is typically used (*35*, *36*).

Upon careful consideration of the proposed mechanism of Minisci-type additions (*3*), we envisaged two plausible scenarios that may lead to product enantioenrichment using a chiral Brønsted acid. First, a bifunctional conjugate anion, such as a chiral phosphate, might engage in hydrogen-bonding interactions with a suitable radical nucleophile as well as its pyridinium counterion (Fig. 2A, *I*). In this manner, a well-defined ternary transition state for enantiodetermining radical addition may occur. After radical addition, the next step would be deprotonation of the aminium ion intermediate *II* before typically fast single-electron oxidation of the resulting neutral radical *III*. Inspiring studies by Minisci and co-workers have suggested that the addition of stabilized nucleophilic radicals to protonated heteroarenes can be reversible, and that in such cases the irreversible, product-determining step of the process is the subsequent deprotonation (*37*). This introduces the second scenario: that deprotonation of *II* could be product-determining and carried out by either the associated chiral phosphate or an external base in the chiral environment that the phosphate provides. In this case, reversible radical addition may result in a situation where both the relative stabilities of various diastereomers of *II* and the barriers to

Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.
*Corresponding author. Email: rjp71@cam.ac.uk
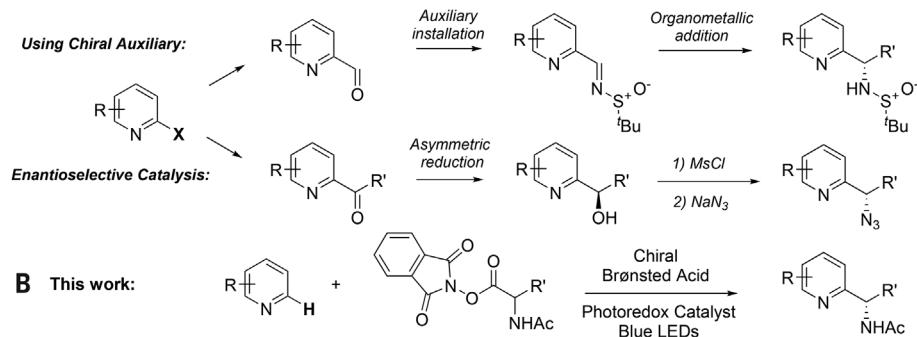
Fig. 1. Strategies to access enantioenriched α-heterocyclic amines. (**A**) Typical current approaches. (**B**) The method reported herein. Ms, methanesulfonyl; Ac, acetyl; *t*Bu, *tert*-butyl.

their deprotonation could play important roles in determining product enantioselectivity.

We elected to use RAEs derived from amino acids as precursors to *N*-acyl α-amino alkyl radicals (Fig. 2A, radical generation), which are prochiral, nucleophilic radicals that possess hydrogen bond donor functionality. It is well established that facile alkyl radical generation occurs from RAEs upon acceptance of a single electron (*38–40*), and there are a number of recent examples of photoredox catalysts being used as electron donors for this purpose, including in conjunction with RAEs derived from *N*-acyl amino acids (*20, 41, 42*) and to achieve Minisci-type additions (*20*). In line with these previous studies, we envisaged a reductive photoredox cycle, in which fragmentation of the RAE is induced via electron transfer from an Ir(II) species generated off-cycle (see figs. S1 to S4 for fluorescence-quenching studies). After regeneration of Ir(III) and subsequent excitation, the catalyst should function as the single-electron oxidant required to rearomatize the heteroarene (Fig. 2A, ***III*** to **3**).

To assess whether asymmetric induction could be achieved, we selected 4-methylquinoline as a test substrate. Irradiation from blue LEDs in the presence of 2 mole percent (mol %) of photocatalyst [Ir(dF(CF₃)ppy)₂(dtbpy)]PF₆ and 5 mol % of chiral Brønsted acid catalyst (*R*)-TRIP in ethyl acetate enabled efficient coupling of 4-methylquinoline (**2a**) and the RAE derived from *N*-acetyl-L-phenylalanine (**1a**) to give **3a** in 94% yield after 14 hours (Fig. 2B). The enantiomeric excess of the product was determined to be 94%; the absolute stereochemistry of the hydrochloride salt was determined by x-ray crystallography. Analysis of the RAE used in this reaction showed that it was racemic, indicating that the stereocenter had epimerized during either synthesis or purification. We evaluated a selection of *N*-protecting groups on the amino acid portion, and acetyl was found to be optimal in terms of both yield and enantioselectivity (table S1). Various solvents were compatible, with dioxane found to give highest enantioselectivities, and the photocatalyst loading could be reduced to 1 mol % (tables S2 and S3). Numerous *N*-acetyl amino acids are commercially available, as are both catalysts in this process. Control experiments without the photocatalyst, Brønsted acid catalyst, or visible light irradiation all gave no conversion to product (table S4).

Encouraged by this result, we evaluated a range of amino acid–derived RAEs (Fig. 3A) and found the scope to be broad. Nonaromatic amino acids reacted with comparably high enantioselectivity; for example, the RAE derived from *N*-Ac-alanine gave 94% enantiomeric excess (e.e.) (**3b**). Branching in the alkyl chain was well tolerated, as observed in the RAE derived from *N*-Ac-valine (**3c**, 97% e.e.). The catalyst overrode influence from an adjacent stereocenter, as demonstrated in the application of each enantiomer of TRIP to the RAE derived from *N*-Ac-isoleucine (**3d** and **3e**, both >20:1 diastereomeric ratio). Remote arenes, esters, and pro-
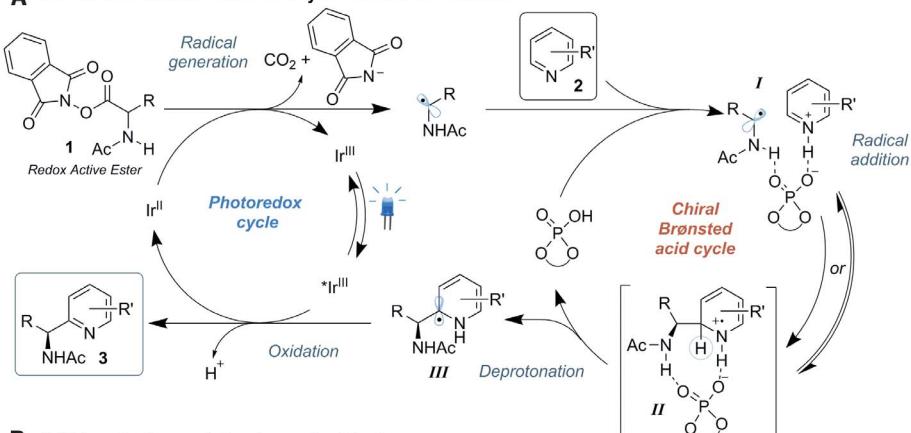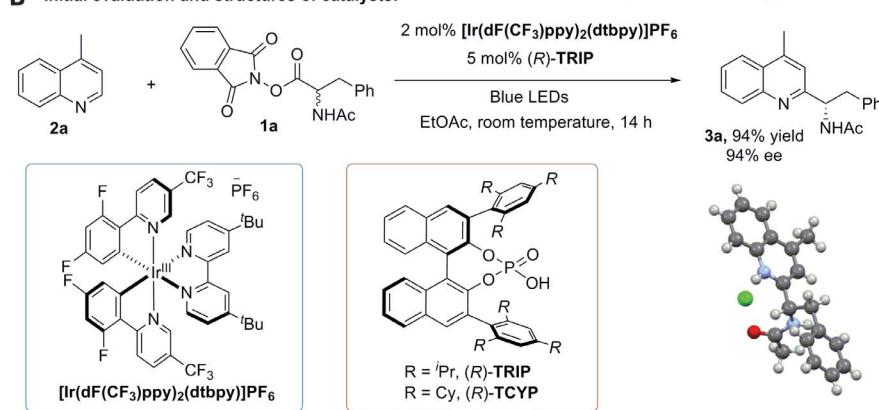
Fig. 2. Reaction development. (**A**) Hypothesis for enabling enantioselective Minisci-type addition. (**B**) Initial reaction evaluation on 4-methylquinoline, structures of the catalysts used in this study, and x-ray structure of **3a**. dF(CF₃)ppy = 2-(2,4-difluorophenyl)-5-(trifluoromethyl)-pyridinyl; dtbpy = 4,4'-bis(tert-butyl)bipyridine; Ph, phenyl; Et, ethyl; ᶦPr, isopropyl.

tected amines were well tolerated (**3f**, **3g**, and **3h**), as was the sulfide of methionine (**3i**). The carbon-iodine bond of iodotyrosine (**3j**) was untouched under the mild conditions, as was the electron-rich indole heterocycle in tryptophan (**3k**). *O*-Ac-tyrosine was also appended effectively (**3l**).

Although most of the RAEs used in Fig. 3 were derived from enantiopure amino acids, almost all exhibited negligible optical rotation, which we attribute to ready epimerization of the RAEs during synthesis or purification. The main exception to this was the RAE derived from valine, which appeared to be configurationally stable: The RAEs derived from L- and D-valine exhibited equal but opposite values of specific rotation, and chiral high-performance liquid chromatography (HPLC) analysis corroborated this observation. Control experiments with several substrates using (*R*)-TRIP showed that both RAEs gave the same enantiomer of product with the same enantiomeric excess, thus demonstrating that the stereochemistry of the RAE does not influence the stereochemical outcome (tables S5 and S6).

We next evaluated the scope of compatible quinolines (Fig. 3B). For unsubstituted quin-

oline, we anticipated that mixtures of regioisomers may arise as a result of competitive C2 and C4 addition. However, we were pleased to discover that under our conditions, very high (>20:1) regioselectivity for the C2 position was obtained, together with excellent enantioselectivity (**3m**, 96% e.e.). Methoxy (**3n**), chloride (**3o**), and fluoride (**3p**) substituents were tolerated on the benzo portion, and in all cases excellent regioselectivity was observed for addition at the C2 position. Regioselectivity in Minisci-type additions is known to be dependent on a range of variables, including solvent, acid, and the nature of the radical (*37, 43*). In our case, use of a polar solvent (*N,N*-dimethylacetamide) resulted in a switch to moderate C4 selectivity, in line with previous studies (*20, 37*). Interestingly, use of acetyl as the *N*-protecting group on the RAE and use of TRIP as catalyst were both found to be crucial to achieving the very high observed regioselectivity for C2 addition; variation of either resulted in only moderate C2 selectivity (tables S9 and S10). Substitution was well tolerated on the pyridine portion of the quinoline and could encompass 4-phenyl (**3q**), 3-methyl (**3r**), and a fused benzene ring (**3s**). An aryloxy group could be included in the 4-position, which
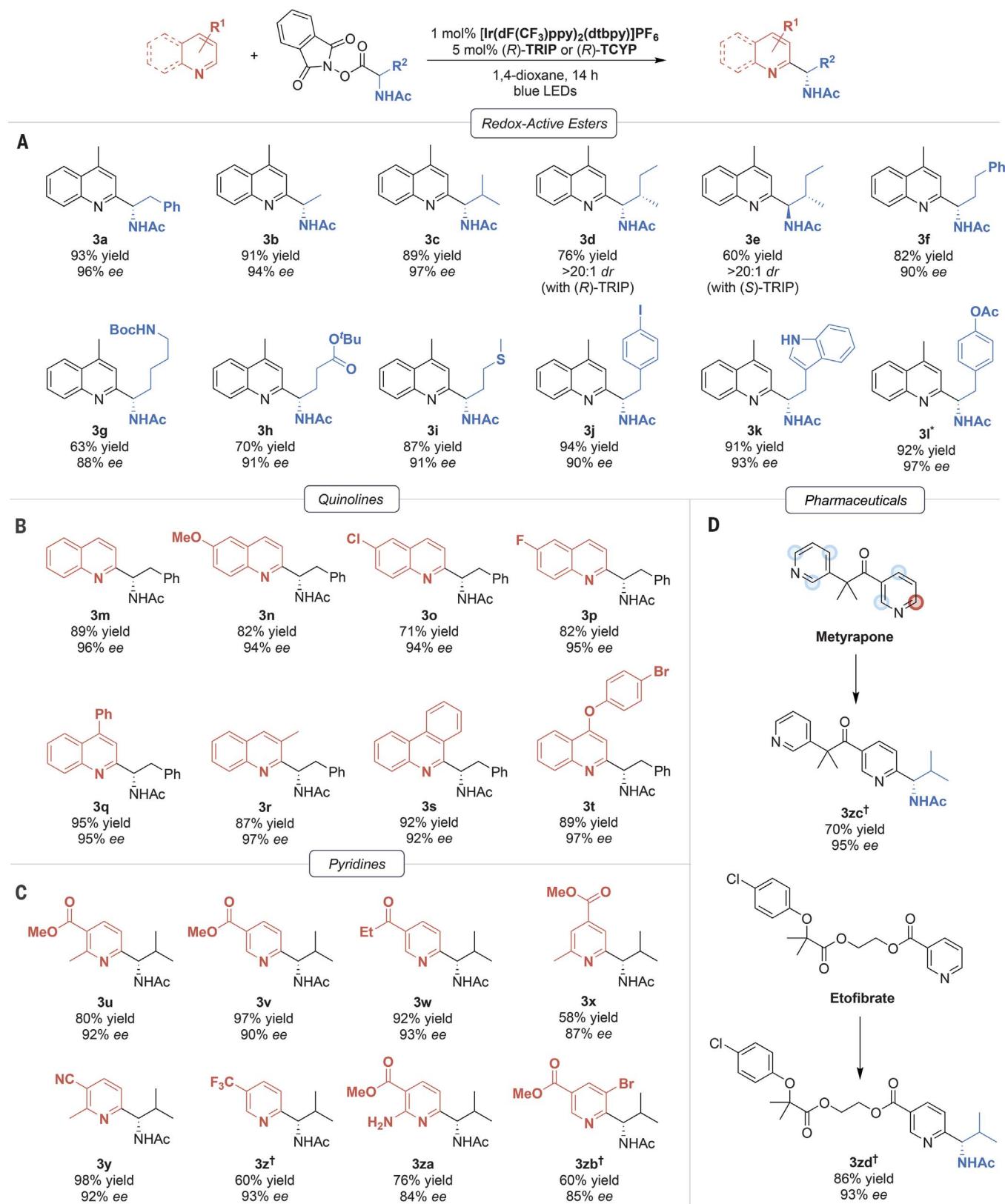
Fig. 3. Exploration of substrate scope. All yields are isolated yields; enantiomeric excesses were determined by HPLC. Absolute stereochemistries of products were assigned by analogy to **3a**. (**A**) Evaluation of scope of redox-active ester (RAE) component. (**B**) Scope of quinoline component.

(**C**) Scope of compatible pyridines. (**D**) Application to enantioselective late-stage functionalization of pharmaceuticals. Boc, *tert*-butyloxycarbonyl; Me, methyl. *O-Acetyl group was cleaved upon purification. †2 mol % photocatalyst and 10 mol % chiral Brønsted acid were used.

A Investigation of a RAE bearing no hydrogen bond donor

3ze
75% yield
4% *ee*
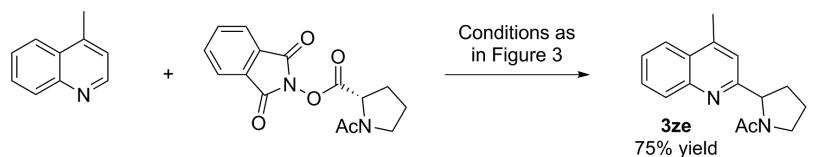
B Kinetic isotope effect experiment

3.6 : 1
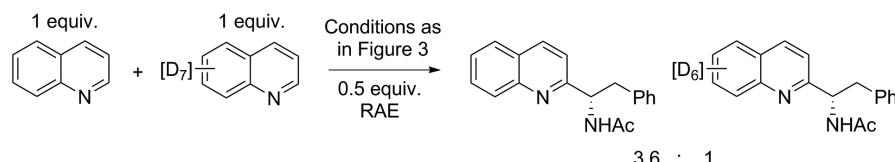
Fig. 4. Preliminary experiments to probe mechanism. (**A**) Outcome of reaction of a proline-derived RAE that bears no hydrogen bond donors. (**B**) Competition experiment to determine kinetic isotope effect.

is particularly relevant in agrochemicals because diaryloxy compounds encompass a privileged family of herbicides (**3t**) (*44*). In contrast to the excellent enantioselectivity obtained with quinoline, isoquinoline gave considerably lower enantioselectivity.

We then explored the scope of pyridines (Fig. 3C). These proved to be very amenable to our method; we observed that (i) an electron-withdrawing substituent is typically necessary to enable productive radical addition, and (ii) the highest enantioselectivities could be obtained by using the bulkier phosphoric acid TCYP (Fig. 2B and tables S7 and S8) (*45*). No addition at the C4 position was observed in any case. The substrate could feature ester (**3u**, **3v**, **3x**), ketone (**3w**), nitrile (**3y**), and trifluoromethyl (**3z**) groups at the 3- or 4-positions, together with other substituents around the ring, including methyls (**3u**, **3x**, **3y**), an unprotected amine (**3za**), and a bromide (**3zb**).

Minisci-type additions are one of the most valuable tools for late-stage functionalization of pharmaceuticals, and we found that our enantioselective addition works very effectively on two examples (Fig. 3D) (*46*). Metyrapone, an inhibitor of cortisol biosynthesis, was functionalized in a chemo-, regio-, and enantioselective manner with excellent enantiomeric excess (**3zc**). Etofibrate, a combination of clofibrate and niacin (**3zd**), also underwent highly selective functionalization at the pyridine ring.

To provide some mechanistic insight, we performed several preliminary experiments. First, the RAE derived from L-proline, which does not bear a hydrogen bond donor, delivered product **3ze** in good yield but negligible enantioselectivity (Fig. 4A). This result supports the crucial role of attractive noncovalent interactions between the phosphate and either the radical nucleophile or a reaction intermediate such as **II** (Fig. 2A). Second, we carried out a competition experiment between quinoline and quinoline-d_7, which demonstrated a primary kinetic isotope effect of 3.6. This outcome suggests that the second mechanistic scenario is most likely in operation: a potentially complex situation wherein radical addition may be reversible and the relative energies of diastereomeric intermediates **II** and the barriers to their deprotonation could both contribute to the excellent observed enantioselectivities (Fig. 4B) (*37*). Finally, we observed a nonlinear relationship between the enantiopurity of catalyst and product, which could suggest the involvement of two molecules of chiral phosphate in the deprotonation step (fig. S5).

Given the extraordinary pace of advancement of the field of photoredox catalysis and the range of molecules that are susceptible to Brønsted acid activation, we envisage that our successful merger of these strategies to address the challenge of enantioselective Minisci chemistry will have broad impact in both areas.

REFERENCES AND NOTES

1. A. F. Pozharskii, A. T. Soldatenkov, A. R. Katritzky, *Heterocycles in Life and Society: An Introduction to Heterocyclic Chemistry, Biochemistry and Applications* (Wiley, 2011).
2. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
3. F. Minisci, E. Vismara, F. Fontana, *Heterocycles* **28**, 489–519 (1989).
4. F. Minisci, F. Fontana, E. Vismara, *J. Heterocycl. Chem.* **27**, 79–96 (1990).
5. M. A. J. Duncton, *MedChemComm* **2**, 1135–1161 (2011).
6. J. Tauber, D. Imbri, T. Opatz, *Molecules* **19**, 16190–16222 (2014).
7. I. B. Seiple *et al.*, *J. Am. Chem. Soc.* **132**, 13194–13196 (2010).
8. Y. Ji *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14411–14415 (2011).
9. G. A. Molander, V. Colombel, V. A. Braz, *Org. Lett.* **13**, 1852–1855 (2011).
10. Y. Fujiwara *et al.*, *Nature* **492**, 95–99 (2012).
11. J. K. Matsui, D. N. Primer, G. A. Molander, *Chem. Sci.* **8**, 3512–3522 (2017).
12. C. K. Prier, D. A. Rankic, D. W. C. MacMillan, *Chem. Rev.* **113**, 5322–5363 (2013).
13. D. M. Schultz, T. P. Yoon, *Science* **343**, 1239176 (2014).
14. N. A. Romero, D. A. Nicewicz, *Chem. Rev.* **116**, 10075–10166 (2016).
15. J. Jin, D. W. C. MacMillan, *Nature* **525**, 87–90 (2015).
16. C. A. Huff *et al.*, *J. Org. Chem.* **81**, 6980–6987 (2016).
17. J. Jin, D. W. C. MacMillan, *Angew. Chem. Int. Ed.* **54**, 1565–1569 (2015).
18. G.-X. Li *et al.*, *Chem. Sci.* **7**, 6407–6412 (2016).
19. R. A. Garza-Sanchez, A. Tlahuext-Aca, G. Tavakoli, F. Glorius, *ACS Catal.* **7**, 4057–4061 (2017).
20. W.-M. Cheng, R. Shang, Y. Fu, *ACS Catal.* **7**, 907–911 (2017).
21. W.-M. Cheng, R. Shang, M.-C. Fu, Y. Fu, *Chem. Eur. J.* **23**, 2537–2541 (2017).
22. F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **52**, 6752–6756 (2009).
23. J. Maury, W. Zawodny, J. Clayden, *Org. Lett.* **19**, 472–475 (2017).
24. J. Llaveria, D. Leonori, V. K. Aggarwal, *J. Am. Chem. Soc.* **137**, 10958–10961 (2015).
25. P. D. O'Shea *et al.*, *J. Org. Chem.* **74**, 4547–4553 (2009).
26. Z. Zuo *et al.*, *J. Am. Chem. Soc.* **138**, 1832–1835 (2016).
27. M. Silvi, P. Melchiorre, *Nature* **554**, 41–49 (2018).
28. A. Bauer, F. Westkämper, S. Grimme, T. Bach, *Nature* **436**, 1139–1140 (2005).
29. L. J. Rono, H. G. Yayla, D. Y. Wang, M. F. Armstrong, R. R. Knowles, *J. Am. Chem. Soc.* **135**, 17735–17738 (2013).
30. G. Bergonzini, C. S. Schindler, C.-J. Wallentin, E. N. Jacobsen, C. R. J. Stephenson, *Chem. Sci.* **5**, 112–116 (2014).
31. D. Uraguchi, N. Kinoshita, T. Kizu, T. Ooi, *J. Am. Chem. Soc.* **137**, 13768–13771 (2015).
32. J.-S. Lin *et al.*, *J. Am. Chem. Soc.* **138**, 9357–9360 (2016).
33. L. Lin *et al.*, *Angew. Chem. Int. Ed.* **56**, 13842–13846 (2017).
34. M. Rueping, D. Parmar, E. Sugiono, *Asymmetric Brønsted Acid Catalysis* (Wiley, 2015).
35. C. A. Correia, L. Yang, C.-J. Li, *Org. Lett.* **13**, 4581–4583 (2011).
36. J. L. Biaco, S. L. Jones, T. J. Barker, *Heterocycles* **92**, 1687–1697 (2016).
37. F. Minisci *et al.*, *J. Org. Chem.* **52**, 730–736 (1987).
38. K. Okada, K. Okamoto, N. Morita, K. Okubo, M. Oda, *J. Am. Chem. Soc.* **113**, 9401–9402 (1991).
39. G. Pratsch, G. L. Lackner, L. E. Overman, *J. Org. Chem.* **80**, 6025–6036 (2015).
40. T. Qin *et al.*, *Science* **352**, 801–805 (2016).
41. J. Schwarz, B. Konig, *Green Chem.* **18**, 4743–4749 (2016).
42. Y. Jin, M. Jiang, H. Wang, H. Fu, *Sci. Rep.* **6**, 20068 (2016).
43. F. O'Hara, D. G. Blackmond, P. S. Baran, *J. Am. Chem. Soc.* **135**, 12122–12134 (2013).
44. E. Kennepohl, I. C. Munro, J. S. Bus, in *Hayes' Handbook of Pesticide Toxicology* (Academic Press, ed. 3, 2010), pp. 1829–1847.
45. V. Rauniyar, Z. J. Wang, H. E. Burks, F. D. Toste, *J. Am. Chem. Soc.* **133**, 8486–8489 (2011).
46. T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **45**, 546–576 (2016).

BIOPHYSICS

Quantitative mass imaging of single biological macromolecules

Gavin Young,[1]* Nikolas Hundt,[1]* Daniel Cole,[1] Adam Fineberg,[1] Joanna Andrecka,[1] Andrew Tyler,[1] Anna Olerinyova,[1] Ayla Ansari,[1] Erik G. Marklund,[2] Miranda P. Collier,[1] Shane A. Chandler,[1] Olga Tkachenko,[1] Joel Allen,[3]† Max Crispin,[3]† Neil Billington,[4] Yasuharu Takagi,[4] James R. Sellers,[4] Cédric Eichmann,[5] Philipp Selenko,[5] Lukas Frey,[6] Roland Riek,[6,7] Martin R. Galpin,[1] Weston B. Struwe,[1] Justin L. P. Benesch,[1]‡ Philipp Kukura[1]‡

The cellular processes underpinning life are orchestrated by proteins and their interactions. The associated structural and dynamic heterogeneity, despite being key to function, poses a fundamental challenge to existing analytical and structural methodologies. We used interferometric scattering microscopy to quantify the mass of single biomolecules in solution with 2% sequence mass accuracy, up to 19-kilodalton resolution, and 1-kilodalton precision. We resolved oligomeric distributions at high dynamic range, detected small-molecule binding, and mass-imaged proteins with associated lipids and sugars. These capabilities enabled us to characterize the molecular dynamics of processes as diverse as glycoprotein cross-linking, amyloidogenic protein aggregation, and actin polymerization. Interferometric scattering mass spectrometry allows spatiotemporally resolved measurement of a broad range of biomolecular interactions, one molecule at a time.

Biomolecular interactions and assemblies are central to a wide range of physiological and pathological processes spanning length scales from small complexes (1) to the mesoscale (2, 3). Despite considerable developments in techniques capable of providing high-resolution structural information (4), such techniques are typically static, involve averaging over many molecules in the sample, and are therefore often do not fully capture the diversity of structures and interactions. Solution-based ensemble methods enable dynamic studies but lack the resolution of separation required to distinguish different species (5–7). Single-molecule methods offer a means to circumvent heterogeneity in both structure and dynamics, and progress has been made in terms of characterizing interactions (8) and mechanisms (9, 10). So far, however, no single-molecule approach has been capable of quantifying and following the diversity

of interactions of biomolecules with the required spatiotemporal accuracy and resolution.

Given sufficient sensitivity, light scattering is an ideal means for detecting and characterizing molecules in low-scattering in vitro conditions because of its universal applicability. In an interferometric detection scheme (Fig. 1A), the scattering signal scales with the polarizability,

which is a function of the refractive index and proportional to the particle volume (11). Combining the approximation that single amino acids effectively behave like individual nano-objects with the observation that the specific volumes of amino acids and refractive indices of proteins vary by only ~1% (fig. S1 and table S1) suggests that the number of amino acids in a polypeptide, and thus its mass, are proportional to its scattering signal. This close relationship between mass and interferometric contrast, which has been predicted (12, 13) and observed (14, 15) to hold coarsely even at the single-molecule level, could thus in principle be used to achieve high mass accuracy.

Building on recent advances in the experimental approach (fig. S2) that improved imaging contrasts for interferometric scattering microscopy (15, 16), we were able to obtain high-quality images of single proteins as they diffused from solution to bind nonspecifically to the microscope coverslip/solution interface (Fig. 1, B and C, and movie S1). Reaching signal-to-noise ratios >10, even for small proteins such as bovine serum albumin (BSA), together with an optimized data analysis approach (16), allowed us to extract the scattering contrast for each molecular binding event with high precision (Fig. 1D and fig. S3). These data revealed clear signatures of different oligomeric states of BSA, with relative abundances for monomer to tetramer of 88.63, 9.94, 1.18, and 0.25% of the detected particles. For nonspecific binding to an unfunctionalized microscope coverslip, surface attachment was effectively irreversible (12,209 binding versus 372 unbinding events). As a result, we could determine

[1]Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK. [2]Department of Chemistry Biomedicinskt Centrum, Uppsala University, Box 576, 75123 Uppsala, Sweden. [3]Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. [4]Cell Biology and Physiology Center, National Heart, Lung and Blood Institute (NHLBI), Bethesda, MD 20892, USA. [5]In-Cell NMR Laboratory, Department of NMR-supported Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Robert-Rössle Straße 10, 13125 Berlin, Germany. [6]Laboratory of Physical Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland. [7]Department of Immunology and Microbiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.
*These authors contributed equally to this work.
†Present address: Centre for Biological Sciences and Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK.
‡Corresponding author. Email: justin.benesch@chem.ox.ac.uk (J.L.P.B.); philipp.kukura@chem.ox.ac.uk (P.K.)

Fig. 1. Concept of interferometric scattering mass spectrometry (iSCAMS). (A) Schematic of the experimental approach relying on immobilization of individual molecules near a refractive index interface. Oligomeric states are colored differently for clarity. (B) Differential interferometric scattering image of BSA. Scale bar, 0.5 μm. (C) Representative images of monomers (top left), dimers (bottom left), trimers (top right), and tetramers (bottom right) of BSA. Scale bars, 200 nm. (D) Scatter plot of single-molecule binding events and their scattering contrasts for 12 nM BSA from 14 movies (lower panel). Corresponding histogram (n = 12,209) with a zoomed-in view of the region for larger species (upper panel). The reduction in landing rate results from a drop in BSA concentration with time owing to the large surface-to-volume ratio of our sample cell (supplementary materials).

(bulk) binding rate constants, which generally exhibited only small variations with oligomeric state. These could be accommodated to obtain minor corrections to the recorded mass spectra and yield the solution distribution (fig. S4). Our results, including the detection and quantification of rare complexes such as BSA tetramers, demonstrate the ability of interferometric scattering mass spectrometry (iSCAMS) to characterize solution distributions of oligomeric species and molecular complexes at high dynamic range.

The regular spacing in the contrast histogram of BSA tentatively confirms the expected linear scaling between mass and interferometric contrast. Repeating these measurements for eight different proteins, spanning 53 to 803 kDa, validates the linear relationship (Fig. 2A and fig. S5A). The deviation between measured and sequence mass was <5 kDa, resulting in an average error of 1.9%, and this deviation showed no detectable correlation with refractivity in relation to the overall shape of the molecule (fig. S6A). Even for large structural differences, such as those between the extended and folded conformations of smooth-muscle myosin (530.6 kDa; Fig. 2A

and figs. S5B and S7), we did not find measurable differences in the apparent molecular mass beyond the increase expected for the addition of glutaraldehyde molecules used to cross-link myosin into the folded conformation (extended, 528.4 ± 16.2 kDa; folded, 579.4 ± 14.8 kDa; fig. S5B). The resolution, as defined by the full width at half-maximum of the measured contrast, reached 19 kDa for streptavidin. In all cases, the resolution was limited by photon shot noise and influenced by molecular mass, increasing from 19 kDa for streptavidin to 102 kDa for thyroglobulin (fig. S6, B and C). The <0.5% deviation from sequence mass for species of >100 kDa compares well with native mass spectrometry (17) and demonstrates the intrinsic utility of iSCAMS for the accurate mass measurement of biomolecules with oligomeric resolution.

Moving beyond species composed solely of amino acids, lipid nanodiscs are an ideal system for testing the broad applicability of iSCAMS owing to their flexibility in terms of polypeptide and lipid content (18). For nanodiscs composed of the MSP1D1 belt protein and DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine) lipids, we obtained a mass of 141.0 ± 1.6 kDa, in good

agreement with the range of masses spanning 124 to 158 kDa reported from other methods (Fig. 2B and fig. S5D). Replacing MSP1D1 with the smaller MSP1ΔH5 reduced the nanodisc diameter and the lipid content by ~20%, after accounting for the thickness of the protein belt (19). Given the masses of MSP1D1 and MSP1ΔH5 (47 and 42 kDa, respectively), we predicted a mass for the MSP1ΔH5 nanodisc of 113.6 kDa, in excellent agreement with our measurement (114.1 ± 1.9 kDa). Mass shifts associated with changes in lipid composition, such as those introduced by partially unsaturated lipids and cholesterol, matched those predicted from the assembly ratios (Fig. 2B and tables S2 to S6).

To see whether our approach also applies to solvent-exposed moieties that experience a different dielectric environment from those buried within a protein, we selected the HIV envelope glycoprotein complex (Env), which is a trimer of gp41-gp120 heterodimers. Env is extensively N-glycosylated, with the carbohydrates contributing almost half of its mass (20). For an Env trimer mimic expressed in the presence of kifunensine, a mannosidase inhibitor that leads predominantly to unprocessed $Man_9GlcNAc_2$ glycans (Man,
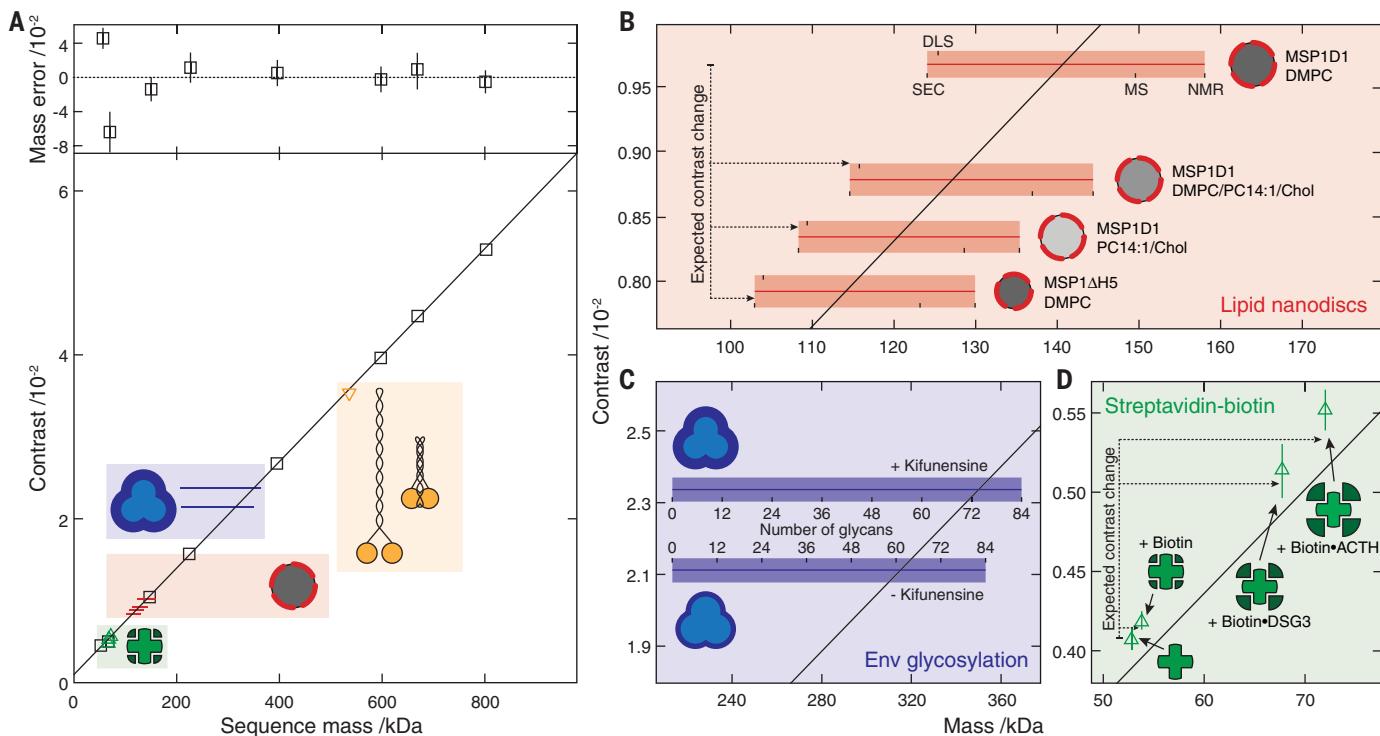
Fig. 2. Characterization of iSCAMS accuracy, precision, and dependence on molecular shape and identity. (**A**) Contrast versus molecular mass, including for proteins used for mass calibration (black), characterization of shape dependence (yellow), protein-ligand binding (green), lipid nanodisc composition (red), and glycosylation (blue). Mass error (upper panel) is given as a percentage of the sequence mass relative to the given linear fit. (**B**) Nanodisc mass measurement for different lipid compositions and protein belts. Masses obtained by alternative methodologies for MSP1D1/DMPC are marked and extrapolated to the other compositions. The horizontal bars indicate the expected mass range as a function of characterization technique, with the thin bars indicating the contrast measured and the thick bars representing the measurement uncertainty in terms of the standard error of the mean (SEM) for repeated experiments. For each sample, the upper text denotes the membrane scaffold protein (MSP) used, and the lower text indicates the lipids in the nanodisc. (**C**) Recorded differential contrast for Env expressed in the presence or absence of kifunensine and associated mass ranges expected for different glycosylation levels. (**D**) Mass-sensitive detection of ligand binding in the biotin-streptavidin system, according to the sequence mass of streptavidin and the masses of biotin and two biotinylated peptides relative to the calibration obtained from (A). Abbreviations are defined in table S8. In (A) and (D), error bars represent SEM.

mannose; GlcNAc, N-acetylglucosamine) (fig. S8), we recorded a mass of 350.0 ± 5.7 kDa. Making the crude approximation that glycans and amino acids have similar polarizabilities, this corresponds to a glycan occupancy of 74 ± 3 out of 84 possible sites (Fig. 2C and fig. S5E), consistent with recent observations of high occupancy for gp120 expressed with kifunensine (21). For Env expressed without kifunensine, we recorded a lower mass of 315.3 ± 10.5 kDa. The mass difference can be attributed only in part to the lower average mass of the processed glycans (fig. S8) and yields a total N-glycan occupancy of 61 ± 6. Although the exact values for occupancy are beholden to our calibration (Fig. 2A), the presence of unoccupied sites is consistent with their observation in proteomics data (22).

The high precision of 1.8 ± 0.5% with respect to the protein mass (Fig. 2A) indicates the potential for direct detection of small-molecule binding. To probe the current limits of iSCAMS in terms of precision, we therefore examined the biotin-streptavidin system (Fig. 2D and fig. S5C). We measured masses for streptavidin in the absence (55.7 ± 1.1 kDa) and presence (57.4 ± 0.9 kDa) of biotin, finding a difference of 1.7 ± 1.4 kDa, in good agreement with the expected 0.98 kDa for complete occupancy of the four binding sites. Upon addition of two different biotinylated peptides (3705.9 and 4767.4 Da), we found increases of 16.1 ± 2.8 and 22.0 ± 2.2 kDa (compared with the expected 14.8 and 19.1 kDa) (Fig. 2D and fig. S5C). These data show that iSCAMS can detect the association of kilodalton-sized ligands, demonstrating its suitability for sensitive ligand-binding studies in solution.

After having established the capabilities of iSCAMS, we sought to test it on more complex systems that are difficult to assess quantitatively with existing techniques as a consequence of heterogeneity and multistep assembly mechanisms (Fig. 3). In addition, we aimed to monitor nucleation and polymerization dynamics of mesoscopic structures down to the single-molecule level, which is challenging because of the simultaneous requirement for high dynamic range, high imaging speed, and direct correlation between the observed signals and the associated molecular events. The biotin-streptavidin system exhibits nearly covalent binding, raising the question of whether iSCAMS is capable not only of determining mass distributions, but also of quantifying weaker equilibria, as are often encountered for protein-protein interactions.

We therefore investigated the interaction of Env with the antiviral lectin BanLec, which neutralizes HIV by binding to surface N-glycans (23, 24) through an unknown mechanism. We were able to monitor the interactions and short-lived complexes before aggregation, with the addition of BanLec to Env resulting in a reduction of single Env units coupled to the appearance of dimers and higher-order assemblies (Fig. 3A). Describing the experimental oligomeric evolution with a simple model (Fig. 3B) enabled us to extract the underlying association constants [K_{BanLec} = 0.12 nM^{-1}; K_{Env} = 8 nM^{-1}; K'_{BanLec} = 0.4 nM^{-1} (as defined in Fig. 3B)], in good agreement with recent bulk studies (K_{BanLec} = 0.19 nM^{-1}), which also found signatures of a secondary binding event (2.85 nM^{-1}) (25). Our ability to follow and model the evolution of different oligomeric species allowed us to extract the interaction mechanism and the energetics underlying the lectin-glycoprotein interaction, despite the heterogeneity of this multicomponent system. As a result, we could show that binding of Env to BanLec that is already bound to Env (K_{Env}) is much stronger than to free BanLec (K_{BanLec}), a key characteristic of cooperative behavior. Moreover, the mass resolution of our approach enabled us to quantify the number of BanLecs bound per dimer (one to two), trimer (two to three), and tetramer (three to four) of Env, demonstrating bivalent binding. These results are directly relevant to the characterization and optimization of antiretrovirals, given that multivalency and aggregation have been proposed to be linked to neutralization potency (25). We anticipate similar quantitative insights to be achievable for other therapeutic target proteins and protein-protein interactions in general.

An advantage of our imaging-based approach is its ability to time-resolve mass changes in a position- and local concentration–sensitive manner. This enables us to examine surface-catalyzed nucleation events that may eventually lead to amyloid formation (26). Previous studies using fluorescence labeling found aggregates of ~0.6 μm in diameter within a minute of addition of the amyloidogenic protein α-synuclein at 10 μM to an appropriately charged bilayer (27). Upon adding α-synuclein to a planar, negatively charged DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine)/DOPS (1,2-dioleoyl-sn-glycero-3-phospho-L-serine) (3:1) membrane at physiological pH, we observed the appearance and growth of nanoscopic objects within seconds, even at low micromolar concentrations (Fig. 4A and movie S2). We were unable to determine the sizes of the initial nucleating species or individual assembly steps owing to the low molecular mass of α-synuclein (14 kDa), but we could monitor the nanoscale formation of associated structures in the range of hundreds of kilodaltons and determine the kinetics (Fig. 4B). Growth of these clusters was uniform across the field of view, with the initial rates following expectations for a first-order process (Fig. 4B and fig. S9A), pointing to a simple growth mechanism. We did not detect such structures on neutral, DOPC-only bilayers, and we found evidence for thioflavin T–positive aggregates after overnight incubation (fig. S9B), suggesting that our assay probes early stages of amyloid assembly.

At the extremes of its current sensitivity, iSCAMS enables mass imaging of mesoscopic self-assembly, molecule by molecule. In an actin polymerization assay, subtraction of the constant background revealed the growth of surface-immobilized filaments. In contrast to α-synuclein, where the growth of interest took place within a diffraction-limited spot, in this case we could quantify length changes of filaments larger than the diffraction limit upon the attachment and detachment of actin subunits (Fig. 4C, fig. S10C, and movie S3). We observed distinct, stepwise changes in the filament length (Fig. 4D; fig. S10, D to F; and movie S4), the most frequent forward and backward step sizes in the traces being 3.0 ± 0.8 and 2.7 ± 0.7 nm, respectively—very close to the expected length increase of 2.7 nm upon binding of a single actin subunit to a filament (Fig. 4E). Detection of larger step sizes represents
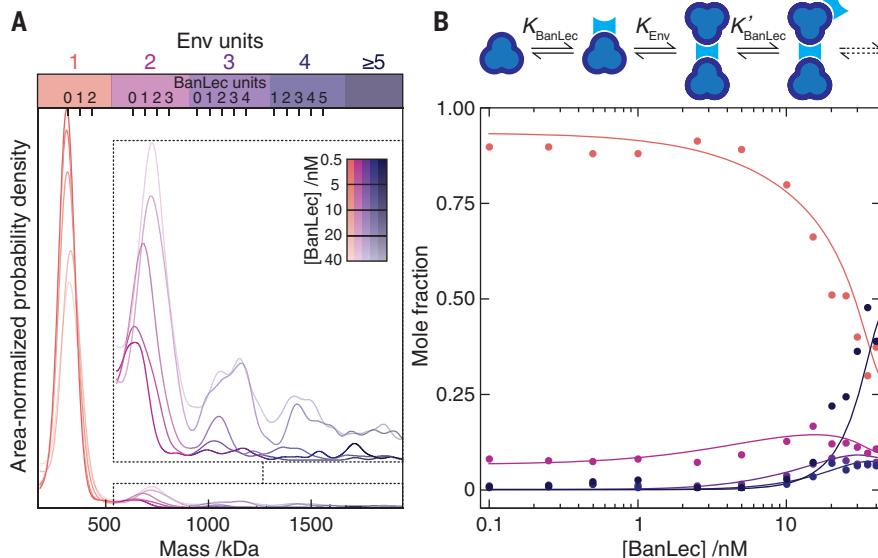
Fig. 3. Single-molecule mass analysis of heterogeneous protein assembly. (**A**) Mass distributions for Env in the presence of 0.5 to 40 nM BanLec monomer, alongside expected positions for multiples of bound BanLec tetramers. Inset, a zoomed-in view of the region for larger species. (**B**) Oligomeric fractions colored according to (A) versus BanLec concentration, including predictions (curves) using the model shown.
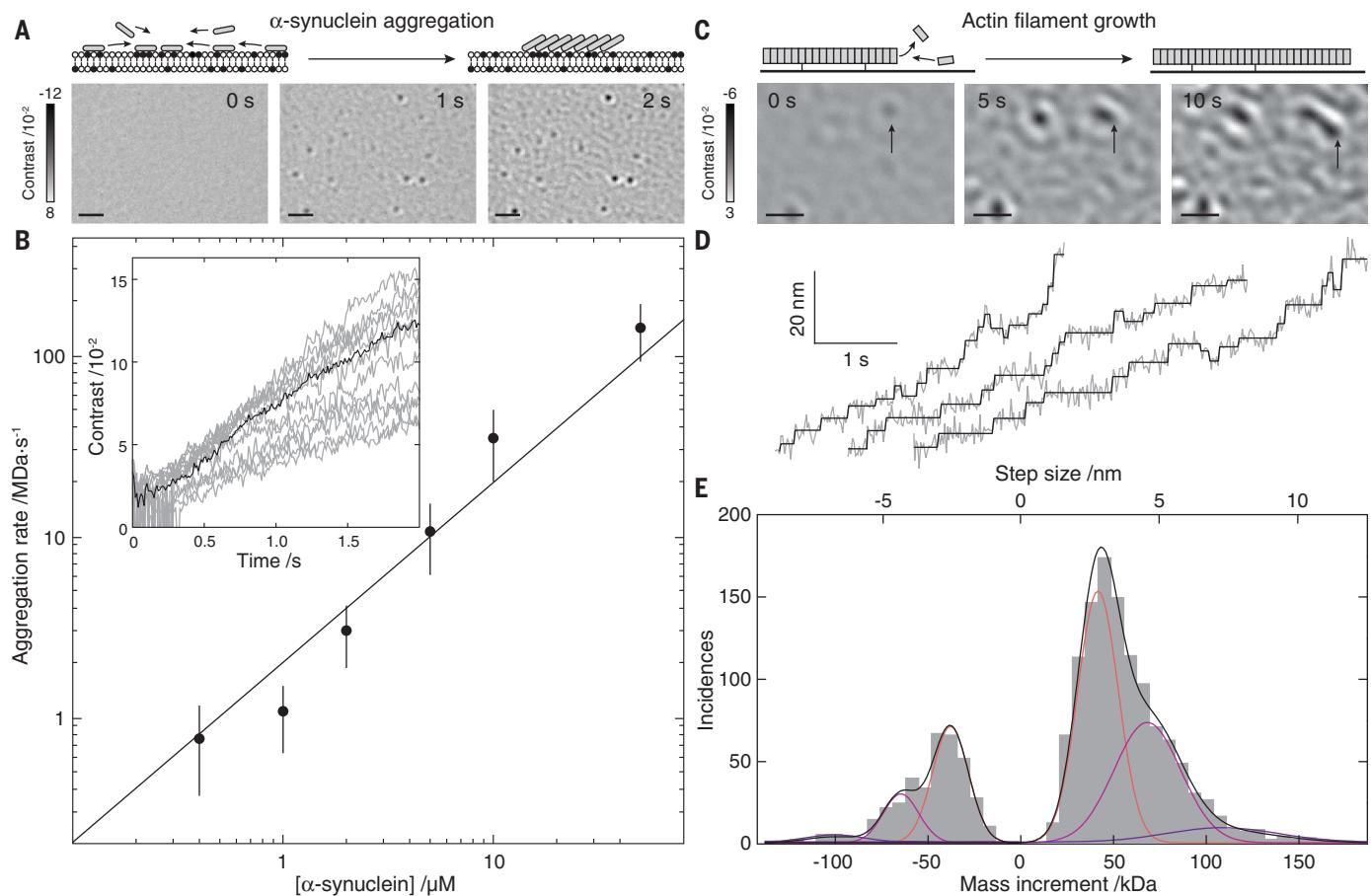
Fig. 4. Mass imaging of mesoscopic dynamics. (**A**) Schematic and iSCAMS images of α-synuclein (1 μM) aggregation on a negatively charged bilayer membrane. (**B**) Initial growth rate versus α-synuclein concentration, shown with the best fit assuming first-order kinetics. Error bars denote SEM for different particles. Inset, individual (gray) and average (black) growth trajectories for 21 particles from (A). (**C**) Schematic and iSCAMS images of actin polymerization. The arrow highlights a growing filament. (**D**) Representative traces of actin filament tip position (gray) and corresponding detected steps (black). (**E**) Step and mass histogram from 1523 steps and 33 filaments, including a fit to a Gaussian mixture model (black) and individual contributions (colored according to fig. S10G). Scale bars, 1 μm. In these experiments, background correction involved removal of the static background before acquisition, rather than continuous differential imaging as in Figs. 2 and 3 (supplementary materials).

the addition of multiple actin subunits within our detection time window. The contrast changes associated with the different step sizes corresponded to mass changes of one, two, or three actin monomers binding to (and unbinding from) the tip of the growing filaments during acquisition (fig. S10, G and H). Even though we cannot yet distinguish between models invoking monomer (*28*) or oligomer (*29*) addition to a growing filament at the current level of spatiotemporal resolution, these results demonstrate the capability of iSCAMS to quantitatively image mesoscopic dynamics and determine how they are influenced by associated proteins at the single-molecule level.

We anticipate that combining iSCAMS with established surface modifications (*30*) will dramatically expand its capabilities. Passivation decreases surface binding probabilities and thereby should provide access to much higher analyte concentrations (greater than micromolar), and surface activation will reduce measurement times at low concentrations (less than nanomolar).

Specific functionalization and immobilization of individual subunits or binding partners could allow for the determination of on and off rates, in addition to equilibrium constants, and enable targeted detection in the presence of other analytes (*14*). Although studies within complex three-dimensional environments such as the cell may prove to be beyond reach, the advances reported here will make iSCAMS a key approach for dynamic in vitro studies of biomolecular interactions, assembly, and structure at the single-molecule level.

REFERENCES AND NOTES

1. S. E. Ahnert, J. A. Marsh, H. Hernández, C. V. Robinson, S. A. Teichmann, *Science* **350**, aaa2245 (2015).
2. K. Rottner, J. Faix, S. Bogdan, S. Linder, E. Kerkhoff, *J. Cell Sci.* **130**, 3427–3435 (2017).
3. M. Hemmat, B. T. Castle, D. J. Odde, *Curr. Opin. Cell Biol.* **50**, 8–13 (2018).
4. H.-W. Wang, J.-W. Wang, *Protein Sci.* **26**, 32–39 (2017).
5. D. M. Kanno, M. Levitus, *J. Phys. Chem. B* **118**, 12404–12415 (2014).
6. M. Chakraborty *et al.*, *Biophys. J.* **103**, 949–958 (2012).
7. S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, T. P. J. Knowles, *J. Mol. Biol.* **421**, 160–171 (2012).
8. A. Jain *et al.*, *Nature* **473**, 484–488 (2011).
9. B. Schuler, W. A. Eaton, *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
10. W. J. Greenleaf, M. T. Woodside, S. M. Block, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 171–190 (2007).
11. C. F. Bohren, D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley Interscience, 1983).
12. P. Kukura *et al.*, *Nat. Methods* **6**, 923–927 (2009).
13. J. Ortega Arroyo *et al.*, *Nano Lett.* **14**, 2065–2070 (2014).
14. M. Piliarik, V. Sandoghdar, *Nat. Commun.* **5**, 4495 (2014).
15. M. Liebel, J. T. Hugall, N. F. van Hulst, *Nano Lett.* **17**, 1277–1281 (2017).
16. D. Cole, G. Young, A. Weigel, A. Sebesta, P. Kukura, *ACS Photonics* **4**, 211–216 (2017).
17. M. van de Waterbeemd *et al.*, *Nat. Methods* **14**, 283–286 (2017).
18. I. G. Denisov, S. G. Sligar, *Chem. Rev.* **117**, 4669–4713 (2017).
19. S. Bibow *et al.*, *Nat. Struct. Mol. Biol.* **24**, 187–193 (2017).
20. L. A. Lasky *et al.*, *Science* **233**, 209–212 (1986).
21. W. B. Struwe, A. Stuckmann, A.-J. Behrens, K. Pagel, M. Crispin, *ACS Chem. Biol.* **12**, 357–361 (2017).
22. L. Cao *et al.*, *Nat. Commun.* **8**, 14954 (2017).
23. M. D. Swanson, H. C. Winter, I. J. Goldstein, D. M. Markovitz, *J. Biol. Chem.* **285**, 8646–8655 (2010).
24. J. T. S. Hopper *et al.*, *Structure* **25**, 773–782.e5 (2017).

25. S. Lusvarghi et al., ACS Infect. Dis. **2**, 882–891 (2016).
26. C. Galvagnion et al., Nat. Chem. Biol. **11**, 229–234 (2015).
27. A. Iyer, N. Schilderink, M. M. A. E. Claessens, V. Subramaniam, Biophys. J. **111**, 2440–2449 (2016).
28. M. Kasai, S. Asakura, F. Oosawa, Biochim. Biophys. Acta **57**, 22–31 (1962).
29. H. P. Erickson, J. Mol. Biol. **206**, 465–474 (1989).
30. B. Hua et al., Nat. Methods **11**, 1233–1236 (2014).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6387/423/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S8
References (31–51)

Movies S1 to S4

26 November 2017; accepted 26 March 2018
10.1126/science.aar5839

OCEAN HYPOXIA

Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico

K. J. Van Meter,[1,2] P. Van Cappellen,[1,2,3] N. B. Basu[1,3,4]*

In August 2017, the Gulf of Mexico's hypoxic zone was declared to be the largest ever measured. It has been estimated that a 60% decrease in watershed nitrogen (N) loading may be necessary to adequately reduce eutrophication in the Gulf. However, to date there has been no rigorous assessment of the effect of N legacies on achieving water quality goals. In this study, we show that even if agricultural N use became 100% efficient, it would take decades to meet target N loads due to legacy N within the Mississippi River basin. Our results suggest that both long-term commitment and large-scale changes in agricultural management practices will be necessary to decrease Mississippi N loads and to meet current goals for reducing the size of the Gulf hypoxic zone.

The areal extent of the northern Gulf of Mexico's hypoxic zone has been measured annually since 1985 and has varied in size from a 1988 minimum of <100 km^2 to the 22,729 km^2—larger than the state of New Jersey—observed in 2017 (*1*, *2*). Problems of hypoxia have been noted along the Louisiana continental shelf in the Gulf of Mexico since the 1970s, and the sediment record indicates that oxygen stress has increased in severity since the 1950s (*1*), parallel to increases in commercial fertilizer use across the Mississippi River basin (MRB) (*3*, *4*).

In 2001, the Mississippi River/Gulf of Mexico Watershed Nutrient Task Force (WNTF) developed an action plan to reduce hypoxia in the Gulf of Mexico and to protect inland waters. The WNTF's stated goal for the Gulf was to reduce the average area of the hypoxic zone to <5000 km^2 by 2015—a reduction that would require an estimated 60% decrease in nitrogen (N) loading from the Mississippi River (*5*, *6*). Since 2001, millions of dollars have been spent on a range of conservation measures across the MRB, including wetland restoration, construction of riparian buffers, planting of cover crops, and improved nutrient management, targeting both surface and subsurface pathways of N loss (*7*). However, despite such interventions, the hypoxic zone in 2015 was three times the size called for by the WNTF, and the target year has since been pushed to 2035.

Such a failure to meet water quality goals is not unique to the MRB. In the Chesapeake Bay region, for example, policy measures to reduce N loading have been in place since the 1990s,

leading to improved nutrient management practices. Despite these efforts, annual N pollution to the Chesapeake Bay remains at more than 5000 kilotons above 2017 goals (*8*). Although there is further scope for implementation of conservation measures, one of the key causes of failures to reduce nutrient loading to surface waters is the existence of nutrient legacies within intensively managed watersheds (*9*). Such legacies, which we define as nutrients that remain within a watershed at least 1 year beyond their initial application at the land surface, exist within soil and groundwater and can serve as long-term sources to surface waters. As a result, current-year reductions in N use may have little effect on short-term changes in N loading at the catchment outlet. In other words, time lags in watershed response must be expected.

For the MRB, numerous modeling approaches have been used to identify the steps necessary to meet water quality goals for the Gulf. Most, however, do not adequately capture the legacy-related nonlinear relationships between N inputs and outputs typical of the problem of reducing N pollution to receiving water bodies (*10*). For example, many approaches rely on fixed, empirical relationships to estimate the landscape-to-stream delivery of nutrients, assuming that nutrients do not accumulate within the landscape (*11*–*16*). Although these relationships may be useful for characterizing spatially varying patterns in watershed dynamics, they are inherently unable to predict changes in loading over long time scales and under nonstationary management conditions (*17*, *18*). Indeed, even most mechanistic watershed models (*19*) do not satisfactorily consider the potentially long residence times for N within the subsurface and therefore do not allow us to account for the effects of N legacies on long-term nutrient loading.

In this study, we used the process-based ELEMeNT (Exploration of Long-tErM Nutrient Trajectories) model (*4*) to address the ways in which watershed legacy N may affect changes

in N loading under major changes in land use and management. ELEMeNT was specifically designed to account for non–steady-state conditions under which N may accumulate in both soils and groundwater, thus capturing the "memory effects" of past nutrient management (see the supplementary materials for more detail). We used the ELEMeNT model to address two primary questions: (i) Is it possible to achieve the desired reductions in loading to the Gulf of Mexico within the specified time horizons? (ii) To what extent are sources of legacy N within the MRB hindering our attempts to reduce N loading?

We have previously used ELEMeNT to hindcast a >200-year trajectory of MRB nitrate loading (1800–2015) (*4*), with measured 1955–2013 N loading data being used for model calibration and validation (see supplementary materials). In this work, we used sediment core chloropigment data from the Gulf of Mexico to further cross-validate the ability of ELEMeNT to predict changes in N loading across periods of changing land use and management (*20*) (Fig. 1). Our analysis suggests strong linearity between model-predicted N loads and inferred primary productivity along the Gulf's continental shelf during two key, dynamic periods of watershed history: (i) the period of European settlement and early watershed development (1820–1930), when pristine lands were being brought under cultivation at rates as high as 15,000 km^2/year, and (ii) from 1960 to 2000, when there was a greater than sixfold increase in commercial N fertilizer use (fig. S1).

Next, to determine whether it is possible to achieve desired reductions in N loading within the time horizons specified by the policy goals, we used ELEMeNT to model multiple scenarios of future watershed management for the MRB. In the first scenario, we assumed business-as-usual (BAU) practices, meaning that cropping patterns, nutrient use, and livestock management remain the same (see supplementary materials) (*12*). For the other scenarios, we imposed instantaneous decreases (25%, 75%, and 100%) in the MRB-wide agricultural N surplus. In these scenarios, we focused on simple reductions in the N surplus, which was calculated as the difference between N inputs to cropland (in the form of fertilizer, livestock manure, and atmospheric N deposition) and N outputs (in the form of livestock and crop production). Such reductions were used as a means of representing a variety of possible changes in land use and management, ranging from moderate reductions in fertilizer use to fundamental changes in the agricultural production system, as might occur with widespread conversion from row-crop agriculture to the planting of perennial biofuel crops such as switchgrass and *Miscanthus* (*21*). Also, when evaluating the results of our simulations, we assumed, based on analysis of historical data (fig. S3), that reductions in stream N from NO$_3$ (hereafter, NO$_3$-N) loads linearly correlate with reductions in total N loads (see supplementary materials).

Our results show that by 2050, the modeled scenarios could result in NO$_3$-N loading reductions

[1]Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [2]Ecohydrology Research Group, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [3]Water Institute, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [4]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
*Corresponding author. Email: nandita.basu@uwaterloo.ca

of anywhere from 11 to 55% of 1980–1996 values (Fig. 2). The BAU loading trajectory is particularly noteworthy because it suggests that, due to legacy N sources within the watershed, an 11% reduction in N loading could be achieved over the next four decades merely by maintaining the status quo. To come close to achieving the areal goal for the hypoxic zone, however, it would be necessary to bring the N surplus to zero, a change that would require 100% efficiency in agricultural N use. This 100% efficiency scenario is not only considered unrealistic but also inherently unsustainable, with researchers suggesting that, to avoid potential decreases in crop production and an increased risk of soil degradation, efficiency values should not exceed 90% (22). It is also important to note here not only the magnitude of the N surplus reduction necessary to achieve policy goals but also the time required to achieve the desired change. Under all of the management scenarios, from BAU to the 100% reduction, it takes more than 30 years for NO_3-N loading to reach new steady-state values (Fig. 2).

In addition, the time necessary to achieve a stated goal can vary depending on the magnitude of reductions in the N surplus. As an example, the WNTF set an interim goal to reduce N loading from the MRB by 20% by the year 2025. According to our model calculations, only an immediate reduction in the agricultural N surplus of 80% or more would allow us to reach this policy goal within the specified time frame (Fig. 3A). However, if policy goals were modified to allow more time to achieve the 20% improvement, more modest reductions in the N surplus could be implemented. The trade-offs between the speed at which improvements in water quality can be achieved and the extent to which watershed management must be modified are illustrated in Fig. 3B, which depicts the time-cost trade-off in a Pareto front framework (23). As indicated in the figure, the proposed 20% reduction in N loading could theoretically be achieved in 7 years with a 100% reduction in the N surplus (red arrows). Alternatively, it would take ~35 years to achieve the 20% reduction in N loading with an ~16% reduction in the N surplus (gray arrows).

With the ELEMeNT model it is possible to simulate the residence times of N within watershed soil profiles and along groundwater flowpaths and, hence, to calculate N age distributions at the mouth of the Mississippi River (see supplementary materials). Here, we define "age" as the time it takes for an N atom to travel from the landscape surface to the catchment outlet, including both time spent within the soil organic matter (biogeochemical lag) and time spent traveling along potentially slow groundwater transport pathways (hydrologic lag). The results suggest that, at present, more than 50% of NO_3-N exported from the watershed is >30 years of age, whereas only 25% is <5 years old (Fig. 4). As expected, the N age distributions also vary with time. For example, under the BAU scenario represented in Fig. 4, the proportion of NO_3-N greater than 30 years of age decreases from ~47 to 21% between 2010 and 2050. At the same time, the relative importance of younger N increases during this period, with the youngest N (0 to 5 years) increasing from 25 to 36% of loading.

The present findings suggest that past land use and agricultural management are primary drivers of current N loading from the Mississippi River to the northern Gulf of Mexico. From 1970 to 2010, cropped area within the MRB increased by ~20% (24, 25). During this period,

Fig. 1. N loading trajectories at the MRB outlet compared to sediment core pigment concentrations. The ELEMeNT model accurately predicts long-term trends in Mississippi nitrate loading, as demonstrated by the close relationship between the modeled annual loads (green line) and depth-varying chloropigment concentrations (purple and orange) within a sediment core obtained from the northern Gulf of Mexico (20). The panels represent two different periods of watershed land use: (A) European settlement and large-scale conversion of prairie lands to row crop (1820–1930) (zeaxanthin, coefficient of determination R^2 = 0.66, P < 0.001; β-carotene, R^2 = 0.57, P < 0.001) and (B) rapid 20th century increases in N fertilizer use (1960–2000) (zeaxanthin, R^2 = 0.0.86, P < 0.001; β-carotene, R^2 = 0.84, P < 0.001). Both chloropigments shown here (zeaxanthin and β-carotene) are known proxies of high phytoplankton biomass, with β-carotene being common to many types of algae and zeaxanthin being specifically associated with cyanobacteria. ktons, kilotons.


Fig. 2. Model-predicted NO_3-N export from the MRB under future scenarios. Scenarios include (**A**) business as usual (BAU) and then (**B**) 25%, (**C**) 75%, and (**D**) 100% decreases in the agricultural N surplus. The green dotted lines represent mean N loading for the period 1980–1996, and the red dotted lines represent target N loading to achieve water quality goals for the Gulf of Mexico. For these scenarios, reductions in N loading ranging from 11 to 55% will be achieved. Note that under all scenarios, it takes ~30 years to reach new steady-state loading levels after the 2017 shift in watershed management.
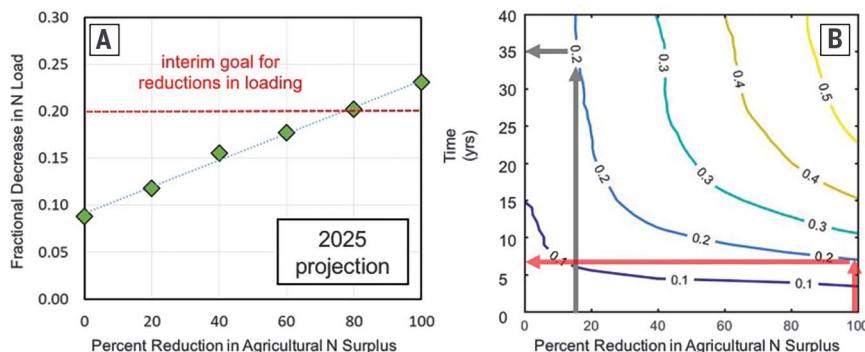
Fig. 3. Combined effects of watershed N surplus reductions and time on achieving N loading goals for the MRB. (**A**) Predicted reductions in N loading by 2025 as a function of the reduction in the agricultural N surplus imposed in 2017 (note: 2025 is the interim WNTF target year for achieving a 20% reduction in N loading). (**B**) Cost-time trade-offs in achieving reductions in MRB N loads. The contour lines represent fractional reductions in N loading as a function of both percent reductions in the agricultural N surplus and the time required to see changes in loading. The red and gray arrows, respectively, demonstrate that it may take between 7 and 35 years to achieve a 20% reduction in N loading, depending upon the extent to which N surplus values are decreased.
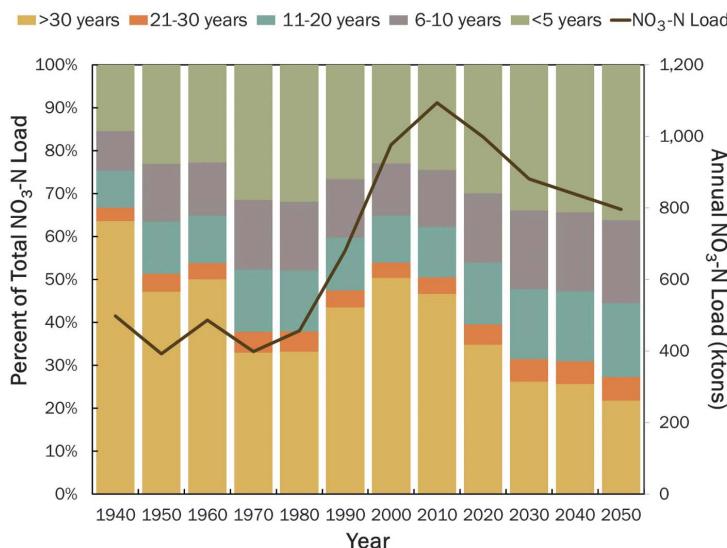
Fig. 4. Changing age distribution of MRB nitrate loadings to the Gulf of Mexico from 1940 to 2050 for the BAU scenario. Note that in 2010, close to 50% of the catchment NO_3-N load is made up of N that has resided within the watershed for more than 30 years. In contrast, by 2050, the magnitude of N in this age range has shrunk to only 21%. As demonstrated by the changing proportions of legacy N, decreases in N loading under the BAU scenario (dark brown line) are driven almost entirely by reductions in sources of legacy N over the simulation period.

fertilizer application rates peaked and livestock production intensified (3, 26), leading not only to short-term increases in riverine N loading but also to accumulations of excess N in soils and groundwater (27, 28). As steps are taken to improve nutrient management and implement new best management practices, we are beginning to see a decrease in N surplus values across the MRB and, in some areas, modest reductions in riverine N loading (5, 29, 30). However, the effects of any changes made today are inevitably modulated by legacy N already present within the watershed (29).

Our results first indicate that it will take decades to achieve the desired improvements in Gulf of Mexico water quality, even without the inevitable delays in implementation of selected conservation measures. Policy-makers must therefore consider this time lag to set realistic goals for the MRB. Additionally, our results show that to meet current goals for improvement—specifically the nearly 60% decrease in N loading deemed necessary to reduce the size of the hypoxic zone—would require action on multiple fronts. This action would include reductions in N inputs (e.g., more-targeted use of N fertilizer) as well as a

major shift in land-use practices—perhaps a widespread conversion to perennial biofuel crops. Furthermore, a much larger emphasis on restoration of riparian wetlands, which can enhance removal of both current-year and legacy N that has already left the fields, could speed reductions in loads and reduce time lags (31). Thus, although our analysis suggests that large reductions in N loading may be possible, these reductions will require not just minor changes in land management but a fundamental alteration of the agroecosystem.

REFERENCES AND NOTES

1. N. N. Rabalais, R. E. Turner, D. Scavia, *Bioscience* **52**, 129–142 (2002).
2. D. R. Obenour, D. Scavia, N. N. Rabalais, R. E. Turner, A. M. Michalak, *Environ. Sci. Technol.* **47**, 9808–9815 (2013).
3. M. B. David, L. E. Drinkwater, G. F. McIsaac, *J. Environ. Qual.* **39**, 1657–1667 (2010).
4. K. J. Van Meter, N. B. Basu, P. Van Cappellen, *Global Biogeochem. Cycles* **31**, 2–23 (2017).
5. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, "Action plan for reducing, mitigating, and controlling hypoxia in the northern Gulf of Mexico" [U.S. Environmental Protection Agency (EPA), 2001]; www.epa.gov/sites/production/files/2015-03/documents/2001_04_04_msbasin_actionplan2001.pdf.
6. D. Scavia et al., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8823–8828 (2017).
7. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, 2015 Report to Congress (U.S. EPA, 2015); www.epa.gov/sites/production/files/2015-10/documents/htf_report_to_congress_final_-_10.1.15.pdf.
8. ChesapeakeProgress, 2017 and 2025 Watershed Implementation Plans (WIPs); www.chesapeakeprogress.com/clean-water/watershed-implementation-plans.
9. S. E. Vero et al., *Hydrogeol. J.* **26**, 7–22 (2018).
10. J. J. Patterson, C. Smith, J. Bellamy, *J. Environ. Manage.* **128**, 441–452 (2013).
11. E. Mayorga et al., *Environ. Model. Softw.* **25**, 837–853 (2010).
12. M. L. McCrackin, E. J. Cooter, R. L. Dennis, J. A. Harrison, J. E. Compton, *Biogeochemistry* **133**, 263–277 (2017).
13. R. B. Alexander et al., *Environ. Sci. Technol.* **42**, 822–830 (2008).
14. B. Hong, D. P. Swaney, R. W. Howarth, *Environ. Sci. Technol.* **47**, 5199–5207 (2013).
15. E. Sinha, A. M. Michalak, *Environ. Sci. Technol.* **50**, 12874–12884 (2016).
16. E. Sinha, A. M. Michalak, V. Balaji, *Science* **357**, 405–408 (2017).
17. K. J. Van Meter, N. B. Basu, *Environ. Res. Lett.* **12**, 084017 (2017).
18. E. J. Gustafson, *Landsc. Ecol.* **28**, 1429–1437 (2013).
19. J. G. Arnold, R. Srinivasan, R. S. Muttiah, J. R. Williams, *J. Am. Water Resour. Assoc.* **34**, 73–89 (1998).
20. N. N. Rabalais, N. Atilla, C. Normandeau, R. E. Turner, *Mar. Pollut. Bull.* **49**, 537–547 (2004).
21. G. F. McIsaac, M. B. David, C. A. Mitchell, *J. Environ. Qual.* **39**, 1790–1799 (2010).
22. EU Nitrogen Expert Panel, "Nitrogen use efficiency (NUE) - An indicator for the utilization of nitrogen in agriculture and food systems" (Wageningen Univ., 2015); www.eunep.com/wp-content/uploads/2017/03/Report-NUE-Indicator-Nitrogen-Expert-Panel-18-12-2015.pdf.
23. K. J. Van Meter, N. B. Basu, *PLOS ONE* **10**, e0125971 (2015).
24. N. Ramankutty, J. A. Foley, *Global Biogeochem. Cycles* **13**, 997–1027 (1999).
25. D. G. Brown, K. M. Johnson, T. R. Loveland, D. M. Theobald, *Ecol. Appl.* **15**, 1851–1863 (2005).
26. L. Bouwman et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20882–20887 (2013).
27. K. J. Van Meter, N. B. Basu, J. J. Veenstra, C. L. Burras, *Environ. Res. Lett.* **11**, 035014 (2016).
28. L. J. Puckett, A. J. Tesoriero, N. M. Dubrovsky, *Environ. Sci. Technol.* **45**, 839–844 (2011).

29. J. C. Murphy, R. M. Hirsch, L. A. Sprague, "Nitrate in the Mississippi River and its tributaries, 1980-2010: An update" (Scientific Investigations Report 2013-5169, U.S. Geological Survey, 2013); https://pubs.er.usgs.gov/publication/sir20135169.
30. G. F. McIsaac, M. B. David, G. Z. Gertner, *J. Environ. Qual.* **45**, 1268–1275 (2016).
31. N. B. Basu *et al.*, *Geophys. Res. Lett.* **37**, L23404 (2010).

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

NEUROSCIENCE

Interregional synaptic maps among engram cells underlie memory formation

Jun-Hyeok Choi,* Su-Eon Sim,* Ji-il Kim,* Dong Il Choi,* Jihae Oh, Sanghyun Ye, Jaehyun Lee, TaeHyun Kim, Hyoung-Gon Ko, Chae-Seok Lim, Bong-Kiun Kaang†

Memory resides in engram cells distributed across the brain. However, the site-specific substrate within these engram cells remains theoretical, even though it is generally accepted that synaptic plasticity encodes memories. We developed the dual-eGRASP (green fluorescent protein reconstitution across synaptic partners) technique to examine synapses between engram cells to identify the specific neuronal site for memory storage. We found an increased number and size of spines on CA1 engram cells receiving input from CA3 engram cells. In contextual fear conditioning, this enhanced connectivity between engram cells encoded memory strength. CA3 engram to CA1 engram projections strongly occluded long-term potentiation. These results indicate that enhanced structural and functional connectivity between engram cells across two directly connected brain regions forms the synaptic correlate for memory formation.

Memory storage and retrieval require specific populations of neurons that show increased neuronal activity during memory formation. Several studies identified these engram cells throughout various brain regions and demonstrated that activated engram cells can induce artificial retrieval of stored memories (1–6). To explain how memory is encoded in the engram, Hebb proposed a hypothetical mechanism, often paraphrased as "fire together, wire together" (7). This hypothesis suggests that synaptic strengthening between coactivated neurons forms the neural substrate of memory. However, it has not been possible to delineate whether memory formation enhances synapses between engram cells in connected brain regions because we could not distinguish presynaptic regions originating from engram cells and nonengram cells.

To compare two different presynaptic populations that project to a single postsynaptic neuron, we modified the green fluorescent protein (GFP) reconstitution across synaptic partners (GRASP) technique (8, 9). GRASP uses two complementary mutant GFP fragments (10), which are expressed separately on presynaptic and postsynaptic membranes and reconstitute in the synaptic cleft to form functional GFP. This GFP signal indicates a formed synapse between the neuron expressing the presynaptic component and the neuron expressing the postsynaptic component. We developed an enhanced GRASP (eGRASP) technique, which exhibits increased GRASP signal intensity by introducing a weakly interacting domain that

facilitates GFP reconstitution and a single mutation commonly found on most advanced GFP variants (fig. S1) (11). We further evolved eGRASP to reconstitute cyan or yellow fluorescent protein (Fig. 1, A and B, and fig. S2) (12–14). Placing the color-determining domain in the presynaptic neuron (cyan/yellow pre-eGRASP) and the common domain to the postsynaptic neuron (post-eGRASP) enabled visualization of the two synaptic populations that originated from two different presynaptic neuron populations and projected to a single postsynaptic neuron. We named this technique dual-eGRASP (Fig. 1A). We demonstrated that two colors reveal the contact interface in human embryonic kidney (HEK) 293T cells expressing the common domain with cells expressing either of the color-determining domains (Fig. 1C). We successfully applied this technique to synapses on dentate gyrus (DG) granule cells originating from either the lateral entorhinal cortex (LEC) or the medial entorhinal cortex (MEC) that projected to the outer and middle molecular layers of the DG, respectively (Fig. 1D) (15). This technique can also separately label intermixed synapses that do not have a unique spatial distribution on CA1 pyramidal neurons that originate from either the contralateral CA3 or ipsilateral CA3 (Fig. 1E) (16). We confirmed that the eGRASP formation itself does not induce undesired strengthening of the synaptic transmission between the neurons expressing pre-eGRASP and post-eGRASP (fig. S3).

To apply dual-eGRASP on synaptic connections between engram cells from two different regions, we used a Fos promoter–driven reverse tetracycline–controlled transactivator (rtTA) delivered by adeno-associated virus (AAV) to express specific genes of interest in the engram cells

at particular time points (17–20). Doxycycline injection 2 hours before either seizure induction or contextual fear conditioning (CFC) successfully labeled the cells activated during these events (figs. S4 and S5). Using this Fos-rtTA system, we expressed post-eGRASP together with membrane-targeted mScarlet-I (21) unilaterally in CA1 engram cells and yellow pre-eGRASP in the contralateral CA3 engram cells to avoid possible coexpression of pre-eGRASP and post-eGRASP. This system labeled CA3 engram to CA1 engram (E-E) synapses with yellow eGRASP signals on red fluorescently labeled dendrites. To compare these synapses with other synapses [nonengram to engram (N-E), engram to nonengram (E-N), and nonengram to nonengram (N-N) synapses], we expressed post-eGRASP together with membrane-targeted iRFP670 (22) in a sparse neuronal population from the ipsilateral CA1, while expressing cyan pre-eGRASP in a random neuronal population from the contralateral CA3. We achieved strong expression in the random neuronal population using a high titer of double-floxed inverted open reading frame (DIO) AAV with a lower titer of Cre recombinase expressing AAV (Fig. 2A). We confirmed that yellow pre-eGRASP expression is doxycycline dependent, demonstrating that this system can label synapses originating from engram cells of a specific event (fig. S6). We successfully distinguished four types of synapses in the same brain slice after CFC. Based on the percentage of overlapping fluorescence, CA3 cells expressing cyan pre-eGRASP, yellow pre-eGRASP, CA1 cells expressing iRFP and mScarlet-I are estimated to be 78.38, 40.25, 11.61, and 20.93%, respectively (fig. S7). Cyan and yellow puncta on red (mScarlet-I) dendrites indicated N-E and E-E synapses, respectively, whereas cyan and yellow puncta on near-infrared (iRFP670) dendrites indicated N-N and E-N synapses (Fig. 2, B and C). We considered puncta expressing both cyan and yellow fluorescence as synapses originating from engram cells, because these synapses originate from CA3 cells expressing both cyan pre-eGRASP (randomly selected population) and yellow pre-eGRASP (engram cells). We found no significant differences between the density of N-N and N-E synapses (Fig. 2D and fig. S8, A and C); however, the density of E-E synapses was significantly higher than E-N synapses (Fig. 2D and fig. S8, B and D). This difference indicates that presynaptic terminals from CA3 engram cells predominantly synapsed on CA1 engram cells rather than CA1 nonengram cells. We also examined the size of spines in each synapse population. E-E spine head diameter and spine volume were significantly greater than N-E synaptic spines, whereas N-N and E-N did not show any significant differences (Fig. 2E).

Although the number of engram cells may remain constant across different memory strengths (23), we predicted that connectivity between pre- and post-engram cells could encode memory strength. We investigated whether memory strength correlates with connectivity between engram cells using the same combination of
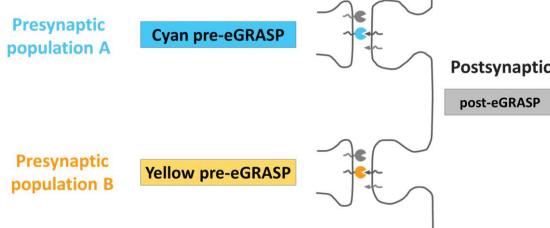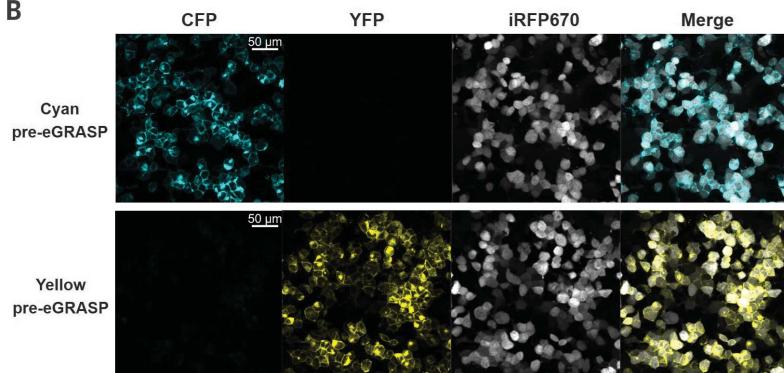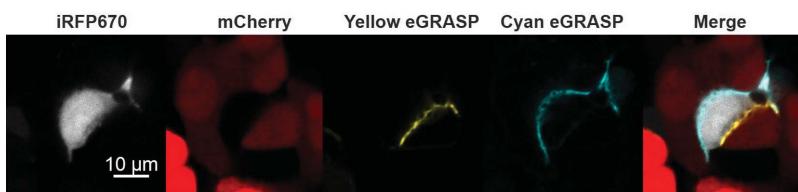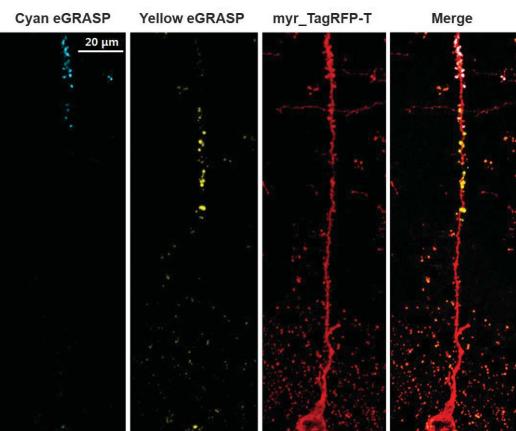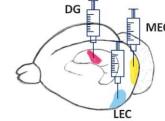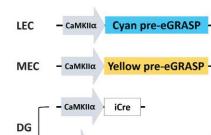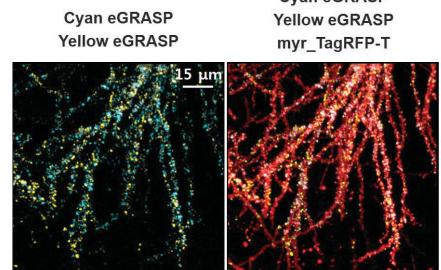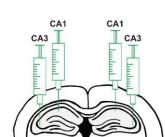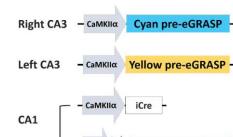
Fig. 1. Dual-eGRASP differentiates two population of synapses on a single neuron. (**A**) Schematic illustration of cyan and yellow eGRASP. Cyan pre-eGRASP and yellow pre-eGRASP are expressed in two different presynaptic population, and common post-eGRASP is expressed in a single postsynaptic cell. (**B**) Coexpression of either cyan or yellow pre-eGRASP with post-eGRASP and iRFP670 in HEK293T cells. (**C**) Three populations of HEK293T cells were separately transfected using nucleofection. One population expressed cyan pre-eGRASP and mCherry, another population expressed yellow pre-eGRASP and mCherry, and the third population expressed post-eGRASP and iRFP670. (**D**) Cyan pre-eGRASP and yellow pre-eGRASP were expressed in the LEC and MEC, respectively. Post-eGRASP was expressed together with myristoylated TagRFP-T (myr_TagRFP-T) in the DG. (**E**) Cyan pre-eGRASP and yellow pre-eGRASP were expressed in the right CA3 and left CA3, respectively. Post-eGRASP was expressed together with myr_TagRFP-T in CA1.

AAVs and injection sites (Fig. 3A) as described in Fig. 2. To induce different strengths of memory, we divided mice into three groups. Mice were exposed to either weak (one shock of 0.35 mA) or strong (three shocks of 0.75 mA) electric foot shocks during CFC, while mice in the context-only group were exposed to the context without any foot shocks (Fig. 3B). Increasing electric foot shock intensity during memory formation produced higher freezing levels (Fig. 3C). When we quantified the number of CA3 and CA1 engram cells, we found no significant differences among the three groups (fig. S9) (23). There were no significant differences between the density of N-N and N-E synapses in all groups. However, we found a significantly higher density of E-E syn-

apses in the strong shock group compared with the context only and weak shock group (Fig. 3, D and E). We further investigated whether the size of spines was positively correlated with memory strength. E-E spine head diameter and spine volume were significantly greater in the strong shock group than in the other groups, whereas N-N and E-N did not show any significant differences in all groups (Fig. 3F and fig. S10).

Because we found increased structural connectivity between CA3 and CA1 engram cells after memory formation, we investigated the synaptic strength of these synapses. We selectively stimulated two different inputs from CA3 neurons using two opsins, Chronos and ChrimsonR, that

can be independently activated using blue and yellow wavelength lasers, respectively (24). We expressed ChrimsonR in CA3 engram neurons using Fos-rtTA, while we expressed Chronos primarily in CA3 excitatory neurons under the calcium/calmodulin-dependent protein kinase type II alpha (CaMKIIα) promoter (Fig. 4A) (25). We labeled CA1 engram neurons with nucleus-targeted mEmerald (mEmerald-Nuc) using FosrtTA and then performed whole-cell recordings from either CA1 engram or nonengram neurons. We investigated the following four combinations of synaptic responses in a single hippocampal slice after CFC: total excitatory to nonengram (T-N), total excitatory to engram (T-E), engram to nonengram (E-N), and engram to engram (E-E)
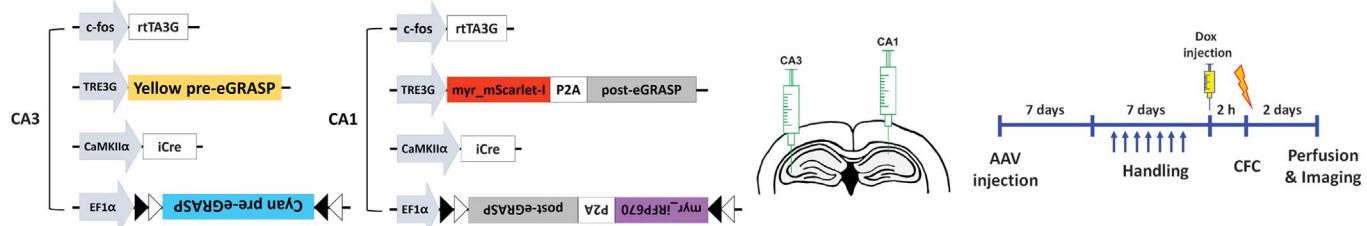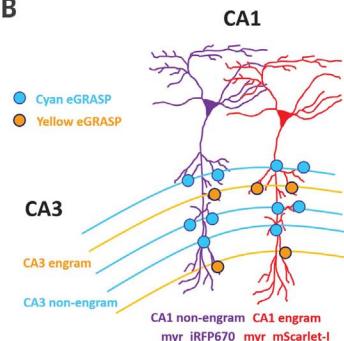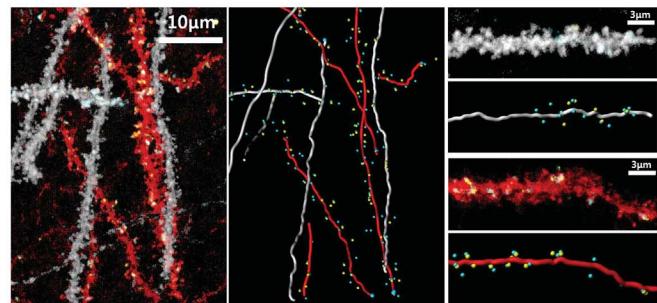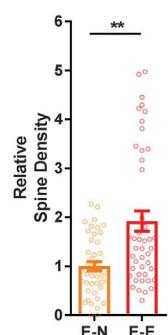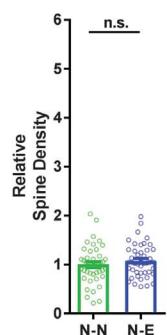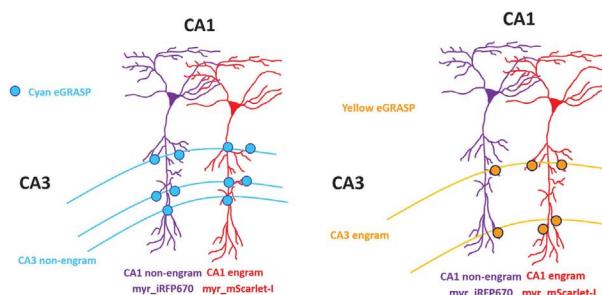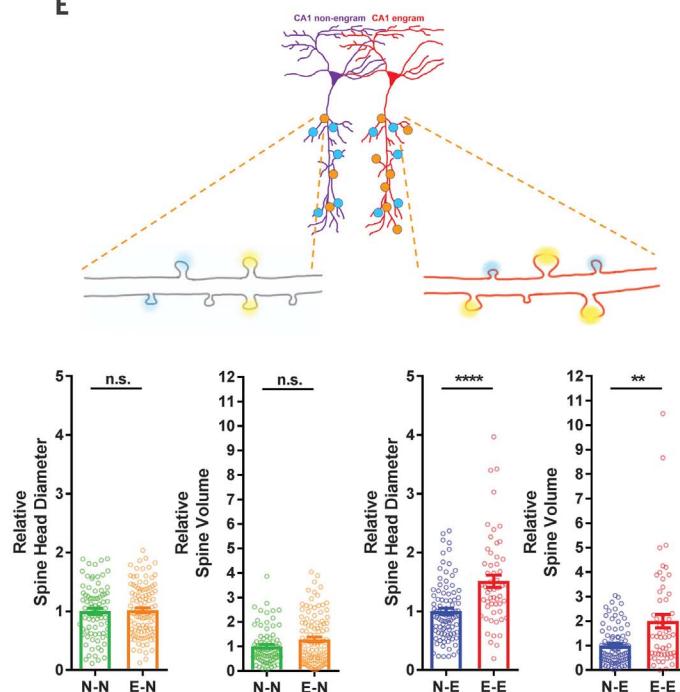
Fig. 2. CA3 engram to CA1 engram synapses exhibited higher synaptic density and larger spine size after memory formation.
(**A**) (Left) Schematic illustration of injected AAVs. (Middle) Illustration of virus injection sites. Injection in each site was performed with a complete cocktail of all the virus infected in each site. (Right) Schematic of the experimental protocol. (**B**) (Left) Schematic diagram of the four possible synapse populations among engram and nonengram cells. (Right) Classification of the four synaptic populations indicated by four colors. Green, N-N; orange, E-N; blue, N-E; red, E-E. The color for each group applies to Figs. 2 and 3. (**C**) Representative image with three-dimensional modeling for analysis. (**D**) Normalized cyan/yellow eGRASP per dendritic length. The densities of cyan-only (left) or yellow puncta (right) on red dendrites are normalized to the corresponding cyan-only or yellow puncta on near-infrared dendrites from the same images in order to exclude the effect of different number of CA3 cells expressing each presynaptic components. Each data point represents a dendrite. $n = 43$ for CA1 nonengram dendrites; $n = 45$ for CA1 engram dendrites; 9 images from 3 mice. Mann Whitney two-tailed test. n.s., not significant; **$P = 0.0017$. (**E**) Normalized spine head diameters and spine volumes of dendrites from CA1 nonengram cells (left) and engram cells (right) with schematic illustration. Sizes of the spines with yellow puncta were normalized to those of the spines with cyan-only puncta of the same dendrite. Each data point represents a spine. N-N, $n = 81$; E-N, $n = 107$; N-E, $n = 93$; E-E, $n = 55$. Mann Whitney two-tailed test. n.s., not significant; **$P = 0.0014$; ****$P < 0.0001$. Data are represented as mean ± SEM.
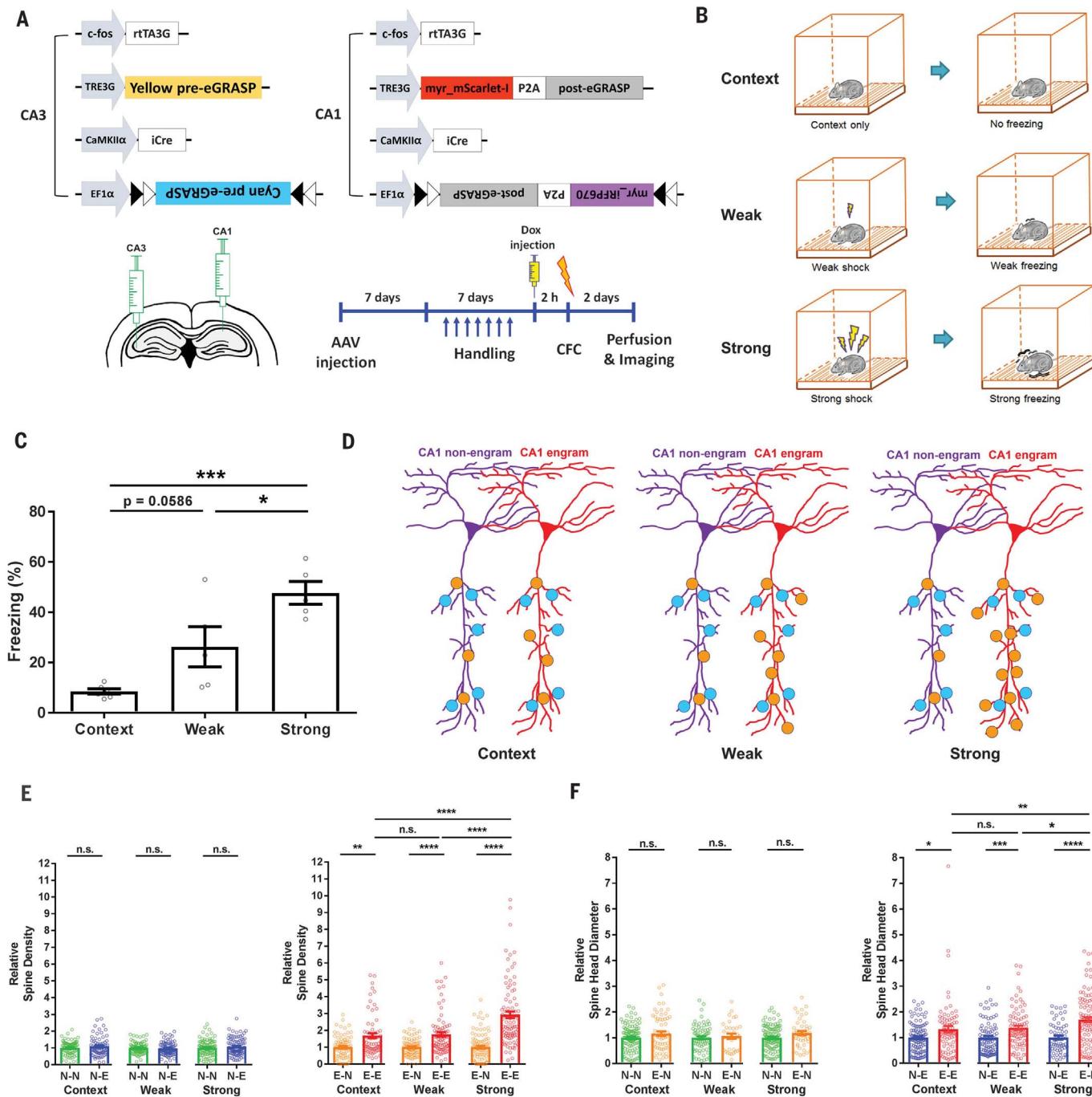
Fig. 3. Synaptic connectivity between pre- and post-engram cells is correlated to memory strength. (**A**) Schematic illustration of injected AAVs, illustration of virus injection sites, and experimental protocol. (**B**) Schematic illustration of the conditioning and retrieval process. (**C**) Freezing levels for each group: context, $n = 6$; weak, $n = 5$; strong, $n = 5$, Tukey's multiple comparison test after one-way analysis of variance (ANOVA); $F(2,13) = 15.85$; $*P < 0.05$; $***P < 0.001$. (**D**) Schematic illustrations of hypothesized results showing higher density of E-E synapses with increasing memory strength. (**E**) Synaptic density of each connections. $n = 74$, context N-N; $n = 67$, context N-E; $n = 79$, weak N-N; $n = 80$, weak N-E; $n = 92$, strong N-N; $n = 91$, strong N-E; $n = 74$, context E-N; $n = 67$, context E-E; $n = 79$, weak

E-N; $n = 80$, weak E-E; $n = 92$, strong E-N; $n = 91$, strong E-E. Fifteen images from six mice for context group. Sixteen images from five mice for weak group. Nineteen images from five mice for strong group. Mann-Whitney two-tailed test, n.s.: not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. (**F**) Spine head diameter of each connection. $n = 107$, context N-N; $n = 64$, context E-N; $n = 72$, weak N-N; $n = 34$, weak E-N; $n = 112$, strong N-N; $n = 46$, strong E-N; $n = 103$, context N-E; $n = 77$, context E-E; $n = 85$, weak N-E; $n = 84$, weak E-E; $n = 57$, strong N-E; $n = 110$, strong E-E, six mice for context group, five mice for weak shock group, five mice for strong shock group. Mann Whitney two-tailed test. n.s., not significant; $*P < 0.05$, $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$. Data are represented as mean ± SEM.
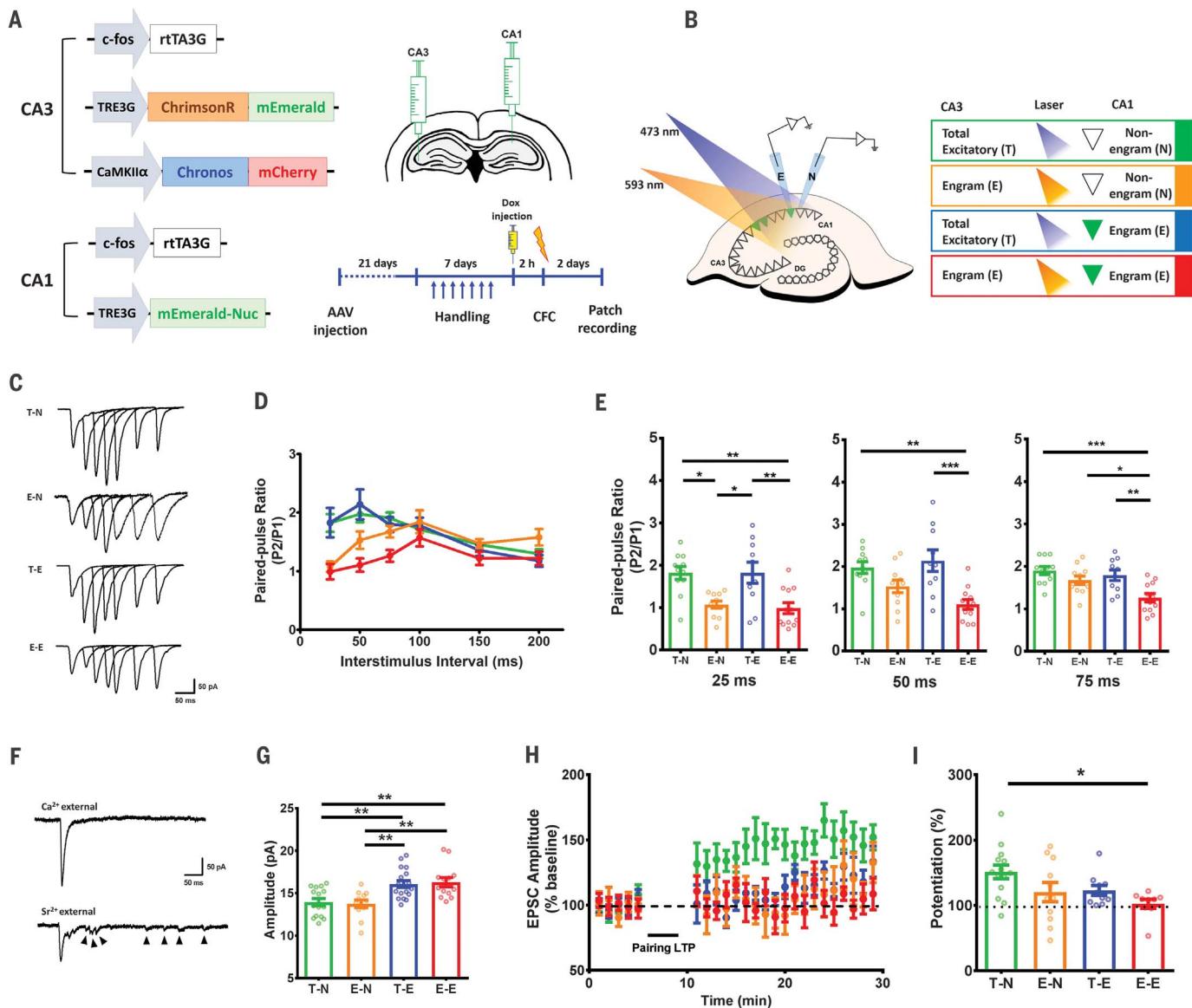
Fig. 4. Enhanced synaptic transmission between CA3 engram and CA1 engram cells through pre- and postsynaptic mechanisms.
(**A**) (Left) Schematic illustration of injected AAVs. (Right) Illustration of virus injection sites and experimental protocol. (**B**) (Left) Diagram of whole-cell recording experiments. (Right) Classification of the four synaptic populations indicated by four colors. Green, T-N; orange, E-N; blue, T-E; red, E-E. The color for each group applies to all the panels below. (**C**) Traces from PPR recordings. (**D**) Results from PPR recordings. T-N, $n = 11$; T-E, $n = 10$; E-N, $n = 11$; E-E, $n = 12$. (**E**) Average PPR at the indicated interstimulus intervals. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; Tukey's multiple comparison test after one-way ANOVA; (25 ms)

$F_{(3,40)} = 8.259$, $*P = 0.0276$; (50 ms) $F_{(3,40)} = 7.989$, $***P = 0.0003$; (75 ms) $F_{(3,40)} = 7.517$, $***P = 0.0004$. (**F**) Traces of Sr^{2+} light-evoked mEPSCs. Arrowheads indicate quantal release events. (**G**) Average amplitude of the Sr^{2+} light-evoked mEPSCs. T-N, $n = 15$; T-E, $n = 18$; E-N, $n = 12$; E-E, $n = 13$; $**P < 0.01$, Tukey's multiple comparison test after one-way ANOVA, $F_{(3,54)} = 8.540$, $***P < 0.0001$. (**H**) Pairing LTP with stimulus given after 5 min of baseline recording. T-N, $n = 14$; T-E, $n = 10$; E-N, $n = 11$; E-E, $n = 9$. (**I**) Average EPSC amplitude of the last 5 min of recording. $*P < 0.05$, Tukey's multiple comparison test after one-way ANOVA, $F_{(3,40)} = 3.683$, $*P = 0.0197$. Data are represented as mean ± SEM.

(Fig. 4B). First, we investigated presynaptic transmission using paired-pulse ratios (PPR) (Fig. 4, C and D). PPR from CA3 engram inputs were significantly decreased at 25-, 50-, and 75-ms interstimulus intervals, which suggests increased release probability from CA3 engram inputs to CA1. The decrease was most prominent in E-E synaptic responses (Fig. 4E). We then examined postsynaptic α-amino-3-hydroxy-5-methyl-

4-isoxazolepropionic acid (AMPA) receptor levels in individual synapses from the four combinations of synaptic responses by replacing Ca^{2+} with Sr^{2+} in the external recording solution (26, 27). Sr^{2+} desynchronized evoked release and induced prolonged asynchronous release, which enabled measurement of quantal synaptic response (Fig. 4F). We measured the amplitude of evoked miniature excitatory postsynaptic currents (mEPSCs) 60 to

400 ms after light stimulation. Synapses from CA1 engram cells exhibited significantly increased levels of postsynaptic AMPA receptors compared with CA1 nonengram cell levels (Fig. 4G). These results indicate that the synapses of CA1 engram cells were potentiated after memory formation but not the synapses of CA1 nonengram cells. Alterations in both presynaptic release probability and postsynaptic potentiation are important for long-term

potentiation (LTP) (*28*). To measure the existence of LTP during memory formation, we examined the extent of LTP occlusion by inducing pairing LTP separately in the four synaptic types (Fig. 4H) (*29*). After 5 min of baseline recording, we delivered pairing LTP stimuli. We found robustly potentiated T-N synaptic responses (~150%). T-E and E-N synaptic responses were potentiated to a lower extent than T-N synaptic responses (~120%), but these differences were not significant. Interestingly, we found that pairing LTP in E-E synaptic responses was completely blocked and potentiation was significantly lower than T-N synaptic responses (Fig. 4I).

Our finding that synaptic populations that fired together during memory formation showed the strongest connections demonstrates that classical Hebbian plasticity indeed occurs during the learning and memory process at CA3 to CA1 synapses (*7, 30*). It is possible that cells with higher connectivity are allocated together into a memory circuit, in contrast to enhanced connectivity after learning. However, the allocated cell number remains constant regardless of the memory strength, whereas the connectivity is significantly enhanced with a stronger memory. This finding indicates a significant contribution of post-learning enhancement over the predetermined connectivity. The relationship between memory strength and synaptic connectivity suggests that these specific connections between engram cells across two directly connected brain regions form the synaptic substrate for memory.

REFERENCES AND NOTES

1. C. A. Denny et al., *Neuron* **83**, 189–201 (2014).
2. J.-H. Han et al., *Science* **323**, 1492–1496 (2009).
3. X. Liu et al., *Nature* **484**, 381–385 (2012).
4. S. Ramirez et al., *Science* **341**, 387–391 (2013).
5. K. Z. Tanaka et al., *Neuron* **84**, 347–354 (2014).
6. Y. Zhou et al., *Nat. Neurosci.* **12**, 1438–1443 (2009).
7. D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory.* (Wiley, New York, NY, 1949).
8. J. Kim et al., *Nat. Methods* **9**, 96–102 (2011).
9. E. H. Feinberg et al., *Neuron* **57**, 353–363 (2008).
10. S. Cabantous, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **23**, 102–107 (2005).
11. M. T. Pisabarro, L. Serrano, *Biochemistry* **35**, 10634–10640 (1996).
12. J. Goedhart et al., *Nat. Commun.* **3**, 751 (2012).
13. M. L. Markwardt et al., *PLOS ONE* **6**, e17896 (2011).
14. R. M. Wachter, M.-A. Elsliger, K. Kallio, G. T. Hanson, S. J. Remington, *Structure* **6**, 1267–1277 (1998).
15. D. G. Amaral, H. E. Scharfman, P. Lavenex, *Prog. Brain Res.* **163**, 3–22 (2007).
16. G. T. Finnerty, J. G. Jefferys, *Neuroscience* **56**, 101–108 (1993).
17. L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230–1233 (2007).
18. G. van Haasteren, S. Li, S. Ryser, W. Schlegel, *Neuroendocrinology* **72**, 368–378 (2000).
19. X. Zhou, M. Vink, B. Klaver, B. Berkhout, A. T. Das, *Gene Ther.* **13**, 1382–1390 (2006).
20. R. Loew, N. Heinz, M. Hampf, H. Bujard, M. Gossen, *BMC Biotechnol.* **10**, 81 (2010).
21. D. S. Bindels et al., *Nat. Methods* **14**, 53–56 (2017).
22. D. M. Shcherbakova, V. V. Verkhusha, *Nat. Methods* **10**, 751–754 (2013).
23. D. J. Morrison et al., *Neurobiol. Learn. Mem.* **135**, 91–99 (2016).
24. N. C. Klapoetke et al., *Nat. Methods* **11**, 338–346 (2014).
25. J.-H. Choi et al., *Mol. Brain* **7**, 17 (2014).
26. F. A. Dodge Jr., R. Miledi, R. Rahamimoff, *J. Physiol.* **200**, 267–283 (1969).
27. M. A. Xu-Friedman, W. G. Regehr, *Biophys. J.* **76**, 2029–2042 (1999).
28. T. V. P. Bliss, G. L. Collingridge, *Mol. Brain* **6**, 5 (2013).
29. H.-X. Chen, N. Otmakhov, J. Lisman, *J. Neurophysiol.* **82**, 526–532 (1999).
30. N. Andersen, N. Krauth, S. Nabavi, *Curr. Opin. Neurobiol.* **45**, 188–192 (2017).

BIOTECHNOLOGY

CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity

Janice S. Chen,[1]* Enbo Ma,[1]* Lucas B. Harrington,[1]* Maria Da Costa,[2] Xinran Tian,[3] Joel M. Palefsky,[2] Jennifer A. Doudna[1,3,4,5,6]†

CRISPR-Cas12a (Cpf1) proteins are RNA-guided enzymes that bind and cut DNA as components of bacterial adaptive immune systems. Like CRISPR-Cas9, Cas12a has been harnessed for genome editing on the basis of its ability to generate targeted, double-stranded DNA breaks. Here we show that RNA-guided DNA binding unleashes indiscriminate single-stranded DNA (ssDNA) cleavage activity by Cas12a that completely degrades ssDNA molecules. We find that target-activated, nonspecific single-stranded deoxyribonuclease (ssDNase) cleavage is also a property of other type V CRISPR-Cas12 enzymes. By combining Cas12a ssDNase activation with isothermal amplification, we create a method termed DNA endonuclease-targeted CRISPR trans reporter (DETECTR), which achieves attomolar sensitivity for DNA detection. DETECTR enables rapid and specific detection of human papillomavirus in patient samples, thereby providing a simple platform for molecular diagnostics.

CRISPR-Cas adaptive immunity in bacteria and archaea uses RNA-guided nucleases to target and degrade foreign nucleic acids (1–3). The CRISPR-Cas9 family of proteins has been widely deployed for gene-editing applications (4, 5) because of the precision of double-stranded DNA (dsDNA) cleavage induced by two catalytic domains, RuvC and HNH, at sequences complementary to a guide RNA (6, 7). A second family of enzymes, CRISPR-Cas12a (Cpf1), uses a single RuvC catalytic domain for guide RNA–directed dsDNA cleavage (8–13) (Fig. 1A). Distinct from Cas9, Cas12a enzymes recognize a T nucleotide–rich protospacer-adjacent motif (PAM) (8), catalyze their own guide CRISPR RNA (crRNA) maturation (14), and generate a PAM-distal dsDNA break with staggered 5′ and 3′ ends (8), features that have attracted interest for gene-editing applications (15, 16). However, the substrate specificity and DNA cleavage mechanism of Cas12a remain to be fully elucidated.

While investigating substrate requirements for Cas12a activation, we tested *Lachnospiraceae* bacterium ND2006 Cas12a (LbCas12a) for guide RNA–directed single-stranded DNA (ssDNA) cleavage, a capability of diverse CRISPR-Cas9 orthologs (17, 18). Purified LbCas12a or *Streptococcus pyogenes* Cas9 (SpCas9) proteins (fig. S1) were assembled with guide RNA sequences targeting a circular, single-stranded M13 DNA phage. In contrast to SpCas9, we were surprised to find that LbCas12a induced rapid and complete degradation of M13 by a cleavage mechanism that could not be explained by sequence-specific DNA cutting (Fig. 1B). This ssDNA shredding activity, not observed with a catalytically inactive LbCas12a (D832A, with an Asp[832]→Ala substitution), raised the possibility that a target-bound LbCas12a could degrade any ssDNA sequence. Remarkably, LbCas12a also catalyzed M13 degradation in the presence of a different guide RNA and its complementary ssDNA "activator" that have no sequence homology to the M13 phage genome (Fig. 1C). These findings reveal that binding of the LbCas12a-crRNA complex to a guide-complementary ssDNA unleashes robust, nonspecific ssDNA trans-cleavage activity.

We next investigated the requirements for LbCas12a-catalyzed trans-cleavage activity. Using a fluorophore quencher (FQ)–labeled reporter assay (19), we assembled LbCas12a with its crRNA and either a complementary ssDNA, dsDNA, or single-stranded RNA (ssRNA) and introduced an unrelated ssDNA- or ssRNA-FQ reporter in trans (fig. S2). Both the crRNA-complementary ssDNA or dsDNA (the activator) triggered LbCas12a to cleave the ssDNA-FQ reporter substrate (fig. S2A). However, ssRNA was neither capable of activating trans cleavage nor susceptible to degradation by LbCas12a (fig. S2B), confirming that LbCas12a harbors DNA-activated general deoxyribonuclease (DNase) activity.

To determine how LbCas12a-catalyzed ssDNA cleavage relates to site-specific dsDNA cutting, we tested the target strand (TS) and nontarget strand (NTS) requirements for LbCas12a activation. Although TS cutting occurred irrespective of the NTS length (fig. S3A), NTS cleavage occurred only when the TS contained at least 15 nucleotides of complementarity with the crRNA (fig. S3B). This shows that TS recognition is a prerequisite for NTS cutting. To test whether LbCas12a remains active for nonspecific ssDNA cleavage after
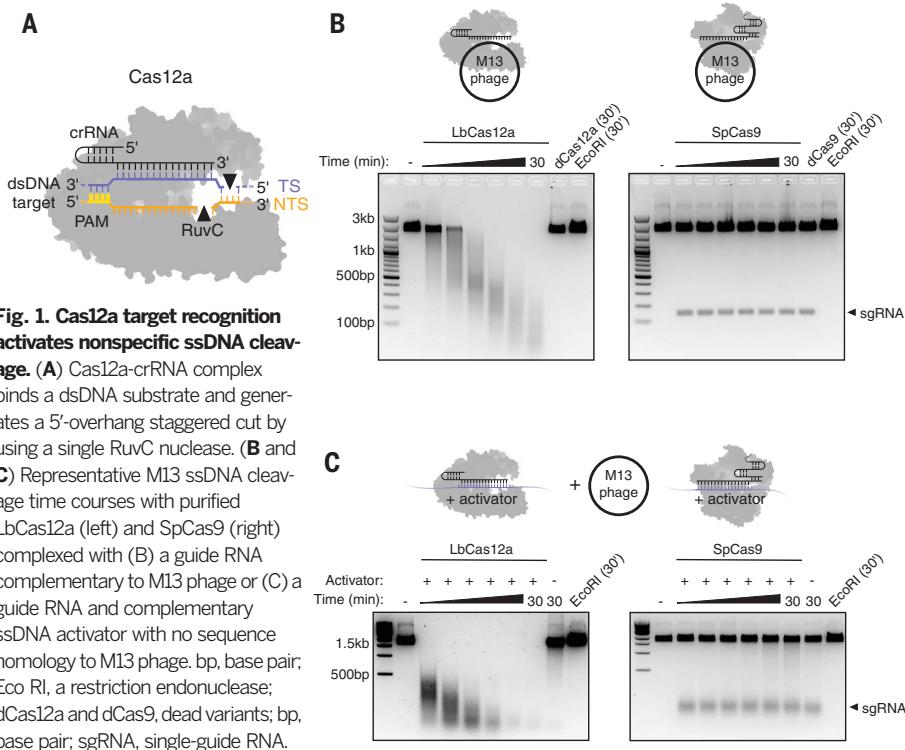
[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA. [3]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA. [4]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA 94704, USA. [5]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA 94720, USA. [6]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: doudna@berkeley.edu

Fig. 1. Cas12a target recognition activates nonspecific ssDNA cleavage. (A) Cas12a-crRNA complex binds a dsDNA substrate and generates a 5′-overhang staggered cut by using a single RuvC nuclease. (B and C) Representative M13 ssDNA cleavage time courses with purified LbCas12a (left) and SpCas9 (right) complexed with (B) a guide RNA complementary to M13 phage or (C) a guide RNA and complementary ssDNA activator with no sequence homology to M13 phage. bp, base pair; Eco RI, a restriction endonuclease; dCas12a and dCas9, dead variants; bp, base pair; sgRNA, single-guide RNA.

dsDNA target cleavage, we first cut a dsDNA plasmid with LbCas12a-crRNA and then added an unrelated dsDNA or ssDNA to the reaction (Fig. 2A). Whereas the nonspecific dsDNA substrate remained intact, the ssDNA was rapidly degraded in a RuvC domain–dependent manner (Fig. 2A and figs. S4 and S5). Using truncated activators that are too short to be cleaved, we determined that only target-DNA binding is required to activate trans-ssDNA cleavage (fig. S6). Together, these results show that RNA-guided DNA binding activates LbCas12a for both site-specific dsDNA cutting and nonspecific ssDNA trans cleavage.

The rapid degradation of a trans substrate suggested that the kinetics of LbCas12a-catalyzed site-specific dsDNA (cis) cleavage and nonspecific ssDNA (trans) cleavage are fundamentally different. Stoichiometric titration experiments showed that cis cleavage is single turnover (20) (Fig. 2B), whereas trans cleavage is multiple turnover (Fig. 2C). Although the LbCas12a-crRNA complex remains bound to the dsDNA target after cis cleavage, the complex releases its PAM-distal cleavage products from the RuvC active site (20), enabling ssDNA substrate access and turnover. We found that LbCas12a-crRNA bound to a ssDNA activator molecule catalyzed trans-ssDNA cleavage at a rate of ~250 turnovers per second and a catalytic efficiency (k_{cat}/K_m) of 5.1×10^8 s^{-1} M^{-1}. When bound to a dsDNA activator, LbCas12a-crRNA catalyzed ~1250 turnovers per second with a catalytic efficiency approaching the rate of diffusion (21), a k_{cat}/K_m of 1.7×10^9 s^{-1} M^{-1} (Fig. 2D and fig. S7). These differences suggest that the NTS of the dsDNA activator helps stabilize the Cas12a complex in an optimal conformation for trans-ssDNA cutting.

We next tested the specificity of trans-cleavage activation using either a ssDNA or dsDNA activator. We found that the PAM sequence required for dsDNA binding by CRISPR-Cas12a (20) is critical for catalytic activation by a crRNA-complementary dsDNA (9) but not for a crRNA-complementary ssDNA (Fig. 3A). Mismatches between the crRNA and activator slowed the trans-cleavage rate, with PAM mutations or mismatches in the PAM-adjacent "seed region" inhibiting trans-ssDNA cleavage only for a dsDNA activator (Fig. 3B and fig. S8). Consistent with the mismatch tolerance pattern observed in Cas12a off-target studies (22, 23), these data corroborate PAM-mediated target-dsDNA binding and the role of base pairing between the crRNA and target strand to activate trans-ssDNA cutting.

We wondered if this trans-ssDNA–cutting activity might be a property shared by type V
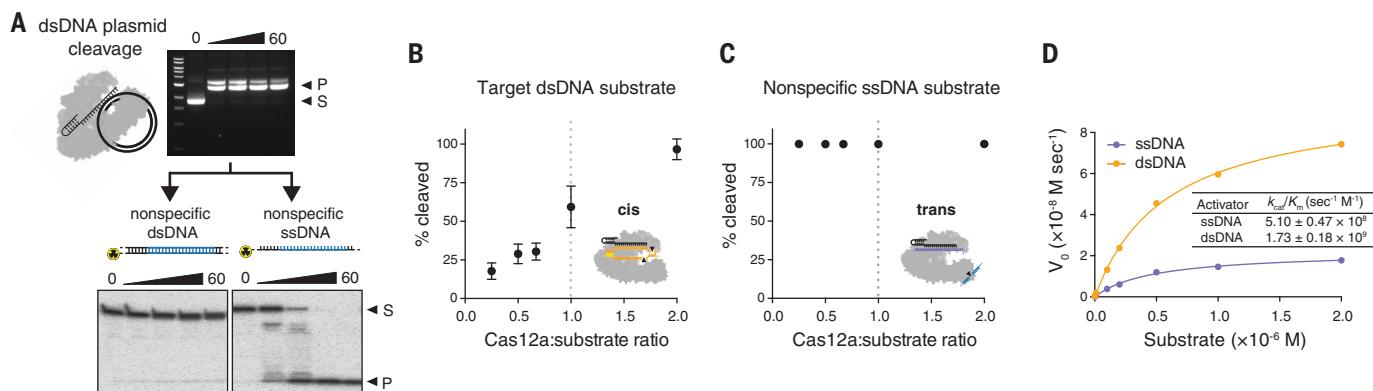
Fig. 2. Kinetics of Cas12a ssDNA trans cleavage. (**A**) Sequence-specific plasmid DNA cleavage reactions by LbCas12a-crRNA complex (top) were introduced to a separate radiolabeled dsDNA or ssDNA substrate of unrelated sequence (bottom); time course represented in minutes. Substrate (S) and nucleotide products (P) were resolved by agarose gel electrophoresis (top) or denaturing polyacrylamide gel electrophoresis (PAGE) (bottom). (**B**) Target dsDNA or (**C**) nonspecific ssDNA incubated with molar ratios of LbCas12a-crRNA, as indicated. Each point represents the mean quantified percent cleavage after 30 min at 37°C, at which time the reaction was at completion. Error bars represent mean ± SD, where $n = 3$ replicates. Yellow coloring indicates PAM. Arrowheads indicate cleavage sites. (**D**) Representative Michaelis-Menten plot for LbCas12a-catalyzed ssDNA trans cleavage using a dsDNA or ssDNA activator. Measured k_{cat}/K_m values reported as mean ± SD, where $n = 3$ Michaelis-Menten fits. V_0, rate of catalysis. Color scheme in (B) and (C) is the same as for (D). In (C), blue indicates nonspecific DNA.
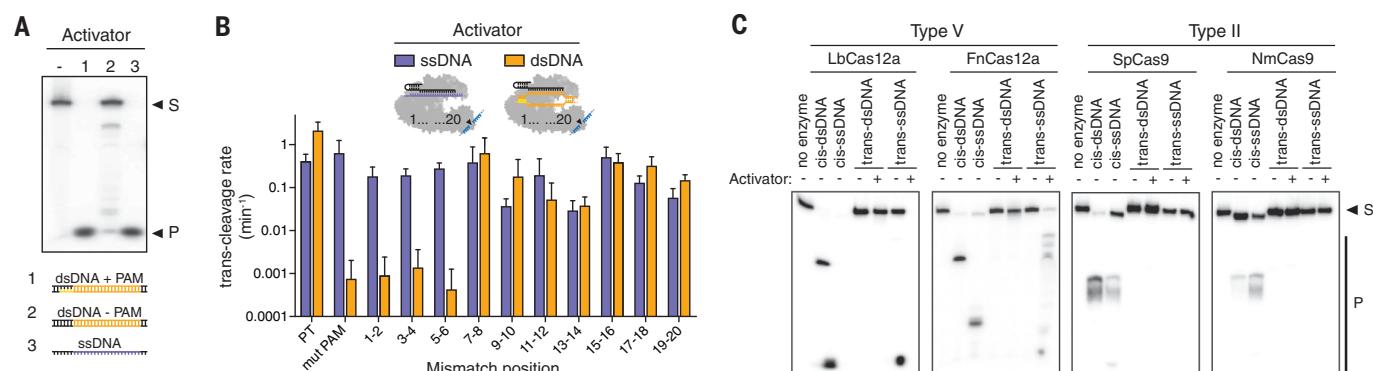
Fig. 3. Specificity and conservation of trans-cleavage activation.
(**A**) LbCas12a-crRNA in the absence or presence of indicated activator, incubated with a radiolabeled nonspecific ssDNA substrate (S) for 30 min at 37°C; products (P) resolved by denaturing PAGE. (**B**) Observed trans-cleavage rates for LbCas12a using a ssDNA or dsDNA activator with indicated mismatches; rates represent the average of three different targets measured in triplicate, and error bars represent mean + SD, where $n = 9$ (three replicates for three independent targets). PT, perfect target; mut PAM, mutated PAM. (**C**) Radiolabeled cis (complementary) or trans (noncomplementary) substrates were incubated with Cas12a-crRNA or Cas9-sgRNA in the presence or absence of a ssDNA activator for 30 min at 37°C; a cis-dsDNA substrate was used in the "no enzyme" lanes. Substrate (S) and nucleotide products (P) were resolved by denaturing PAGE. NmCas9 is *Neisseria meningitidis* Cas9.

Fig. 4. Rapid identification of HPV16 and HPV18 in human samples by DETECTR. (**A**) Diagram of HPV16 and HPV18 sequences within the hypervariable loop V of the L1-encoding gene targeted by Cas12a; yellow highlighted bases indicate 5′ PAM sequence. (**B**) Heatmap represents normalized mean fluorescence values of HPV16 and HPV18 detected in human cell lines by DETECTR; normalized scale shown in (D). (**C**) Schematic outlining DNA extraction from human anal samples to HPV identification by DETECTR. RPA, recombinase polymerase amplification; F, fluorophore; Q, quencher. (**D**) Identification of HPV16 and HPV18 in 25 patient samples by PCR (left) and DETECTR (right). DETECTR heatmap represents normalized mean fluorescence values.
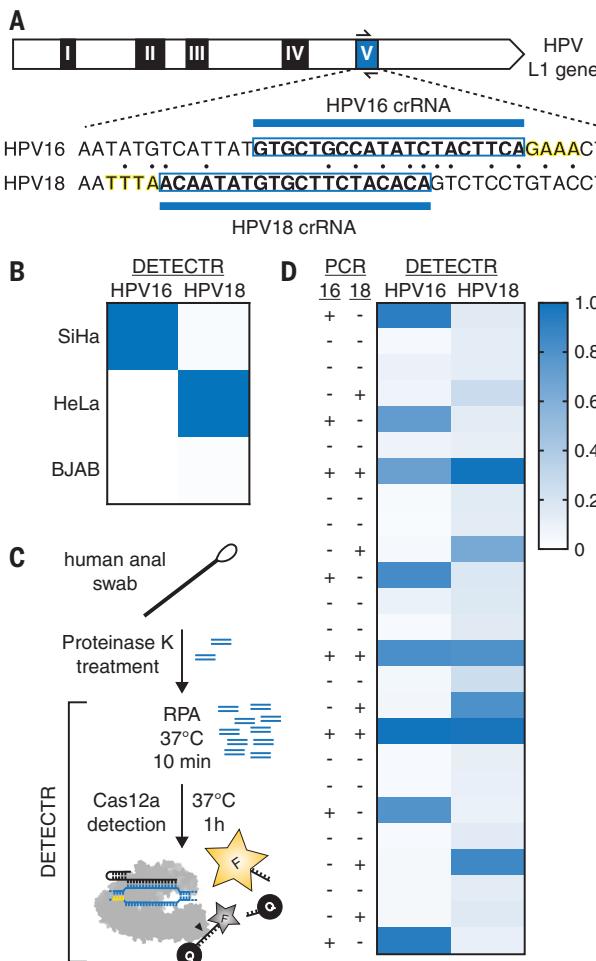
CRISPR effector proteins, considering that all Cas12 proteins contain a single RuvC nuclease domain (*3*, *24*). Consistent with this possibility, purified Cas12a orthologs from *Acidaminococcus* sp. (AsCas12a) and *Francisella novicida* (FnCas12a), as well as a Cas12b protein from *Alicyclobacillus acidoterrestris* (AaCas12b), all catalyzed nonspecific ssDNA cleavage when assembled with a crRNA and complementary ssDNA activator (Fig. 3C and fig. S9). By contrast, none of the type II CRISPR-Cas9 proteins tested showed evidence for trans cleavage (Fig. 3C and fig. S9). These results reveal the unexpected functional convergence of Cas12 enzymes with the type III CRISPR-Csm/Cmr and type VI CRISPR-Cas13 effectors, which also exhibit target-activated, nonspecific ssDNase or single-stranded ribonuclease (ssRNase) activity (*25*, *26*).

We next explored whether LbCas12a could be repurposed as a DNA-detection platform. In particular, accurate and rapid detection of human papillomavirus (HPV) is critical for identifying those at risk of HPV-related cancers, with types 16 (HPV16) and 18 (HPV18) accounting for most precancerous lesions (*27*). To test if LbCas12a could distinguish between these two dsDNA viruses, we selected a target sequence next to a TTTA PAM that varied by six base pairs between the two HPV genotypes (fig. S10). HPV16- or

HPV18-containing plasmids were incubated with LbCas12a-crRNA that targeted either the HPV16 or HPV18 fragment and a ssDNA-FQ reporter, which produced a signal only in the presence of the cognate target (fig. S10). To enhance sensitivity, we coupled isothermal amplification by recombinase polymerase amplification with LbCas12a to develop a one-pot detection method termed DNA endonuclease-targeted CRISPR trans reporter (DETECTR) (fig. S11A). When programmed to its cognate plasmid, DETECTR identified targets with attomolar sensitivity (fig. S11B).

To assess whether we could detect HPV in complex mixtures, DNA extracted from cultured human cells infected with HPV16 (SiHa cells) or HPV18 (HeLa cells) or without HPV (BJAB cells) was added to LbCas12a-crRNA targeting the hypervariable loop V of the L1-encoding gene within HPV16 or HPV18 (Fig. 4A). Whereas LbCas12a-crRNA alone lacked the sensitivity to detect HPV, DETECTR unambiguously identified HPV16 and HPV18 only in SiHa and HeLa cells, respectively (Fig. 4B and fig. S12, A and B). To investigate the utility of DETECTR for analyzing patient samples, we tested crude DNA extractions from 25 human anal swabs previously analyzed by a polymerase chain reaction (PCR)–based method for HPV infection (fig. S13) (*28*). Within 1 hour, DETECTR accurately identified HPV16

(25/25 agreement) and HPV18 (23/25 agreement) in patient samples containing a heterogeneous mixture of HPV types, with good correlation between the PCR-based intensity and DETECTR signal (Fig. 4, C and D, and figs. S12, C and D, and S13). These results demonstrate a new platform for CRISPR-based diagnostics, similar to that developed for RNA detection using CRISPR-Cas13a (*19*, *29*), and suggest that DETECTR could, in principle, detect any DNA sequence with high sensitivity and specificity (fig. S14).

Together, these findings support a mechanism of target interference that begins with the Cas12a–guide RNA complex binding to a complementary DNA sequence in a PAM-dependent (dsDNA) or PAM-independent (ssDNA) manner (fig. S15). Within a host bacterium, such enzyme activation could provide simultaneous protection from both dsDNA and ssDNA phages and could also target ssDNA sequences that arise temporarily during phage replication or transcription (*30*). In a genome-editing context, target-activated ssDNA cutting by Cas12a has the potential to cleave transiently exposed ssDNA at replication forks (*31*), R-loops (*32*), and transcription bubbles (*33*) or ssDNA templates used for homology-directed repair (*34*). Finally, unleashing the ssDNase activity of Cas12 proteins offers a new strategy to improve the speed, sensitivity, and specificity of molecular diagnostic applications.

REFERENCES AND NOTES

1. R. Barrangou *et al.*, *Science* **315**, 1709–1712 (2007).
2. L. A. Marraffini, E. J. Sontheimer, *Science* **322**, 1843–1845 (2008).
3. E. V. Koonin, K. S. Makarova, F. Zhang, *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
4. J. A. Doudna, E. Charpentier, *Science* **346**, 1258096 (2014).
5. R. Barrangou, P. Horvath, *Nat. Microbiol.* **2**, 17092 (2017).
6. M. Jinek *et al.*, *Science* **337**, 816–821 (2012).
7. J. S. Chen, J. A. Doudna, *Nat. Rev. Chem.* **1**, 0078 (2017).
8. B. Zetsche *et al.*, *Cell* **163**, 759–771 (2015).
9. D. C. Swarts, J. van der Oost, M. Jinek, *Mol. Cell* **66**, 221–233.e4 (2017).
10. D. Dong *et al.*, *Nature* **532**, 522–526 (2016).
11. P. Gao, H. Yang, K. R. Rajashankar, Z. Huang, D. J. Patel, *Cell Res.* **26**, 901–913 (2016).
12. S. Stella, P. Alcón, G. Montoya, *Nature* **546**, 559–563 (2017).
13. T. Yamano *et al.*, *Cell* **165**, 949–962 (2016).
14. I. Fonfara, H. Richter, M. Bratovič, A. Le Rhun, E. Charpentier, *Nature* **532**, 517–521 (2016).
15. B. Zetsche *et al.*, *Nat. Biotechnol.* **35**, 31–34 (2017).
16. X. Tang *et al.*, *Nat. Plants* **3**, 17018 (2017).
17. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. J. Sontheimer, *Mol. Cell* **60**, 242–255 (2015).
18. E. Ma, L. B. Harrington, M. R. O'Connell, K. Zhou, J. A. Doudna, *Mol. Cell* **60**, 398–407 (2015).
19. A. East-Seletsky *et al.*, *Nature* **538**, 270–273 (2016).
20. D. Singh, J. Mallon, A. Poddar, Y. Wang, R. Tipanna, O. Yang, S. Bailey, T. Ha, Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1. BioRxiv 205575 [Preprint]. 18 October 2017.
21. R. A. Alberty, G. G. Hammes, *J. Phys. Chem.* **62**, 154–159 (1958).
22. D. Kim *et al.*, *Nat. Biotechnol.* **34**, 863–868 (2016).
23. B. P. Kleinstiver *et al.*, *Nat. Biotechnol.* **34**, 869–874 (2016).
24. S. Shmakov *et al.*, *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
25. J. R. Elmore *et al.*, *Genes Dev.* **30**, 447–459 (2016).
26. O. O. Abudayyeh *et al.*, *Science* **353**, aaf5573 (2016).
27. C. B. Woodman, S. I. Collins, L. S. Young, *Nat. Rev. Cancer* **7**, 11–22 (2007).

28. J. M. Palefsky, E. A. Holly, M. L. Ralston, N. Jay, *J. Infect. Dis.*
    177, 361–367 (1998).
29. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
30. P. Samai *et al.*, *Cell* **161**, 1164–1174 (2015).
31. J. M. Sogo, M. Lopes, M. Foiani, *Science* **297**, 599–602 (2002).
32. V. Bhatia, E. Herrera-Moyano, A. Aguilera, B. Gómez-González,
    Genes (Basel) **8**, E171 (2017).
33. Y. Zhang *et al.*, *Science* **338**, 1076–1080 (2012).
34. C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, J. E. Corn,
    Nat. Biotechnol. **34**, 339–344 (2016).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6387/436/suppl/DC1
Materials and Methods
Figs. S1 to S15
Table S1
References (*35–37*)

BIOTECHNOLOGY

Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6

Jonathan S. Gootenberg,[1,2,3,4,5]* Omar O. Abudayyeh,[1,2,3,4,6]* Max J. Kellner,[1] Julia Joung,[1,2,3,4] James J. Collins,[1,4,6,7,8] Feng Zhang[1,2,3,4]†

Rapid detection of nucleic acids is integral for clinical diagnostics and biotechnological applications. We recently developed a platform termed SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) that combines isothermal preamplification with Cas13 to detect single molecules of RNA or DNA. Through characterization of CRISPR enzymology and application development, we report here four advances integrated into SHERLOCK version 2 (SHERLOCKv2) (i) four-channel single-reaction multiplexing with orthogonal CRISPR enzymes; (ii) quantitative measurement of input as low as 2 attomolar; (iii) 3.5-fold increase in signal sensitivity by combining Cas13 with Csm6, an auxiliary CRISPR-associated enzyme; and (iv) lateral-flow readout. SHERLOCKv2 can detect Dengue or Zika virus single-stranded RNA as well as mutations in patient liquid biopsy samples via lateral flow, highlighting its potential as a multiplexable, portable, rapid, and quantitative detection platform of nucleic acids.

Versatile, rapid, and portable sensing of nucleic acids is vital for applications in human health. The RNA-targeting CRISPR-associated enzyme Cas13 (*1*, *2*) has recently been adapted for such purpose. This detection platform, termed SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) (*3*), can discriminate between inputs that differ by a single nucleotide at very low concentrations and can be lyophilized for portable deployment. However, this technology has several limitations, including the lack of quantitation and reliance on fluorescence detection equipment for readout. Here, we extend the SHERLOCK technology to address these limitations and further develop the utility of this platform.

Many applications require detection of more than one target molecule in a single reaction, and we therefore sought to create a multiplexed platform that relies on specific cleavage preferences of Cas enzymes (*2*–*5*). To identify possible candidate enzymes compatible with multiplexing, we biochemically characterized three members of the CRISPR-Cas13a family and 14 members of the CRISPR-Cas13b family (*6*, *7*) (figs. S1 and S2 and table S1). We profiled cleavage preferences on homopolymer reporters and found that most orthologs preferred either uridine, a combination of bases, or adenine (fig. S3 and

tables S2 to S5) and that cleavage could be improved with buffer and CRISPR RNA (crRNA) design optimization (figs. S4 to S7 and supplementary methods). Among the adenine-cleaving enzymes, Cas13b from *Prevotella* sp. MA2016 (PsmCas13b) was more sensitive than Cas13a from *Lachnospiraceae* bacterium NK4A179 (LbaCas13a) (fig. S8). We refined the cleavage sequence preferences by evaluating collateral activity across dinucleotide motifs (Fig. 1A), finding a large diversity of dinucleotide cleavage motif preferences (figs. S9 and S10 and supplementary methods). From these dinucleotide cleavage screens, we found that the activities of LwaCas13a, Cas13b from *Capnocytophaga canimorsus* Cc5 (CcaCas13b), LbaCas13a, and PsmCas13b could all be independently measured with the four dinucleotide reporters AU, UC, AC, and GA, respectively (Fig. 1B and fig. S11). Additionally, using a random in vitro RNA library motif cleavage screen, we identified numerous RNA oligomers of 6 bases that allowed for further orthogonality between Cas13 enzymes (figs. S12 to S15 and supplementary methods).

Using these specific cleavage preferences, we could detect synthetic Zika virus (ZIKV) ssRNA in the HEX channel and synthetic Dengue virus (DENV) ssRNA in the FAM channel in the same reaction (fig. S16). To expand the in-sample multiplexing capabilities of SHERLOCK, we engineered a detection system based on Cas12a (Cpf1), which also exhibits collateral activity (*8*) (Fig. 1C). Although Cas12a from *Acidaminococcus* sp. BV3L6 (AsCas12a) collateral activity did not produce a detectable signal at input concentrations <10 nM, preamplification with recombinase polymerase amplification (RPA) enabled single-molecule detection at 2 aM (Fig. 1D and fig. S17) (unless otherwise noted, all SHERLOCK reactions that involve preamplification are performed in two steps, with the RPA reaction mixture being

directly added to the Cas13 assay without any purification step). For triplex detection, we designed a LwaCas13a uridine reporter in the Cy5 channel, a PsmCas13b adenine reporter in the FAM channel, and an AsCas12a ssDNA reporter in the HEX channel (fig. S18A). We were able to detect three targets (a synthetic ssDNA target, ZIKV ssRNA, and DENV ssRNA) in a single reaction (fig. S18B). We further extended detection to four targets by leveraging orthogonal dinucleotide motifs, with reporters for LwaCas13a, PsmCas13b, CcaCas13b, and AsCas12a in FAM, TEX, Cy5, and HEX channels, respectively (Fig. 1E), and could distinguish all combinations of targets (Fig. 1F). When combined with RPA, we detected two DNA targets (the *Pseudomonas aeruginosa* acyltransferase gene and the *Staphylococcus aureus* thermonuclease gene) (Fig. 1G) at concentrations as low as the attomolar range (Fig. 1H). Similarly, multiplexed SHERLOCK with PsmCas13b and LwaCas13a achieved attomolar multiplexed detection of ZIKV and DENV RNA dilutions as well as allele-specific genotyping of human saliva samples (fig. S19). These advances on in-sample multiplexing via orthogonal base preferences allow for many targets to be detected at scale and for cheaper cost.

We next focused on tuning the output of the SHERLOCK signal to make it more quantitative, sensitive, and robust to broaden the utility of the technology. SHERLOCK relies on an exponential preamplification, which saturates quickly and hinders accurate quantitation, but we observed that more dilute primer concentrations increased both raw signal and quantitative accuracy, indicating that at lower primer concentrations, the reaction does not saturate (Fig. 2, A and B, and fig. S20, A to E). We tested a range of primer concentrations and found that 240 nM exhibited the greatest correlation between signal and input (fig. S20F), and quantitation was sustainable across a large range of sample concentrations as low as the attomolar range (Fig. 2C and fig. S20G). Many applications of nucleic acid detection, such as HIV detection (*9*, *10*), require single-molecule-per-milliliter sensitivity, and we therefore tested whether the detection limit could be pushed beyond 2 aM, allowing for more dilute sample inputs into SHERLOCK. By scaling up the preamplification RPA step, we found that LwaCas13a could produce a detection signal for 200, 80, and 8 zM input samples and allow for single-molecule volume inputs of 250 and 540 μl (fig. S21, A and B), and PsmCas13b could detect 200 zM input samples in 250-μl reactions (fig. S21C).

To amplify the detection signal, we leveraged the CRISPR type III effector nuclease Csm6 (*11*–*17*), which is activated by cyclic adenylate molecules or linear adenine homopolymers terminated with a 2′,3′-cyclic phosphate (*18*, *19*). LwaCas13a and PsmCas13b collateral activity generates cleavage products with hydroxylated 5′ ends and 2′,3′-cyclic phosphate ends (fig. S22), suggesting that Cas13 collateral activity could generate Csm6 activating species, which would allow for amplified signal detection in the SHERLOCK

[1]Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA 02142, USA. [2]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA. [3]Department of Brain and Cognitive Science, MIT, Cambridge, MA 02139, USA. [4]Department of Biological Engineering, MIT, Cambridge, MA 02139, USA. [5]Department of Systems Biology, Harvard University, Boston, MA 02115, USA. [6]Department of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA. [7]Institute for Medical Engineering and Science, MIT, Cambridge, MA 02139, USA. [8]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: zhang@broadinstitute.org
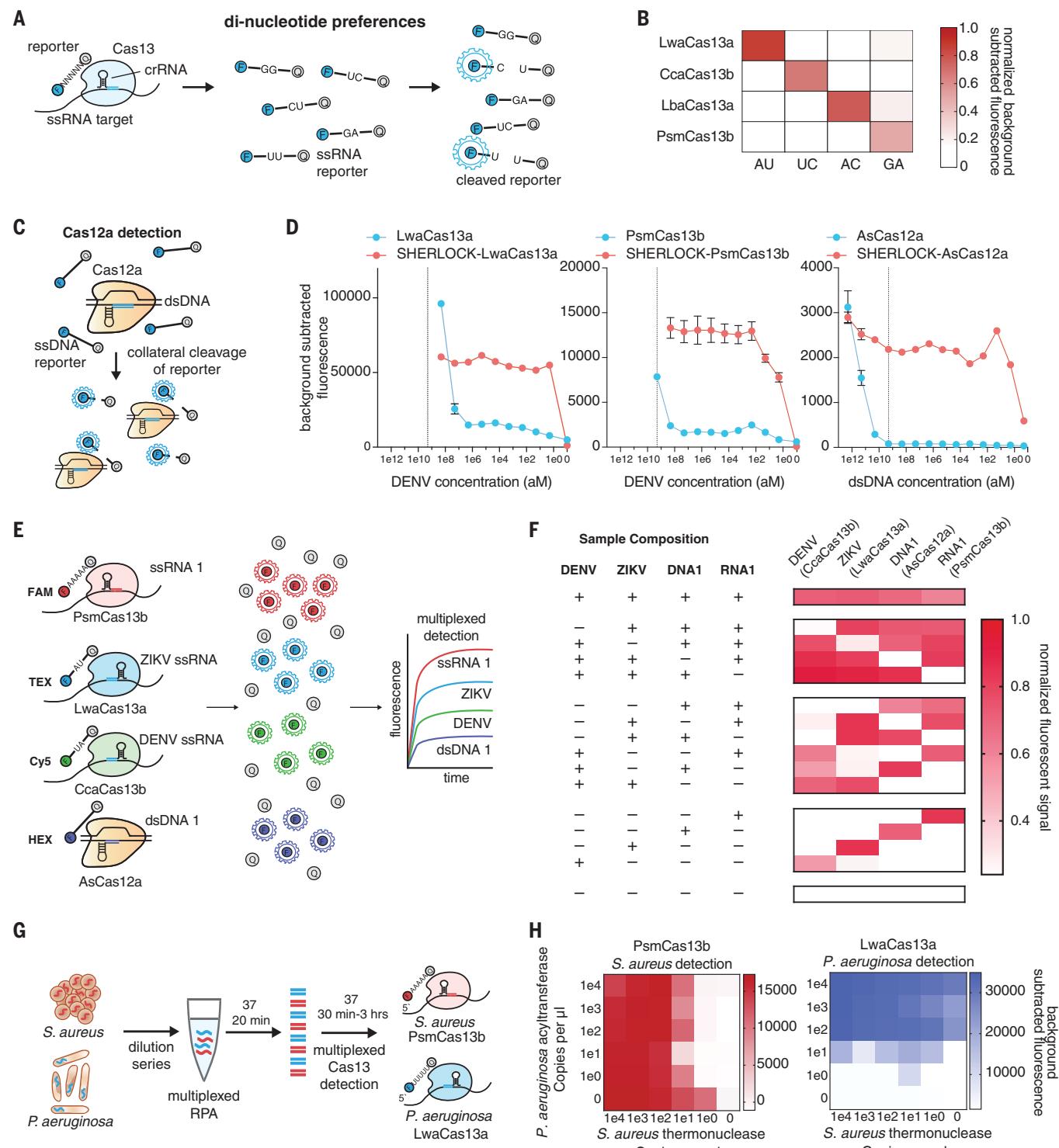
Fig. 1. Multiplexed SHERLOCK detection with orthogonal collateral activity of class 2 enzymes. (**A**) Schematic of assay for determining dinucleotide preferences of Cas13a/b enzymes. (**B**) Collateral activity of LwaCas13a, CcaCas13b, LbaCas13a, and PsmCas13b on orthogonal dinucleotide reporters. (**C**) Schematic of collateral activity of Cas12a activated by double-stranded DNA (dsDNA). (**D**) Comparison of collateral activity and preamplification enhanced collateral activity (SHERLOCK) of LwaCas13a, PsmCas13b, and AsCas12a. The dotted line denotes 2e9 (aM), the limit of AsCas12a sensitivity without preamplification.

Values represent mean ± SEM. (**E**) Schematic of in-sample four-channel multiplexing with orthogonal Cas13 and Cas12a enzymes. (**F**) In-sample multiplexed detection of ZIKV ssRNA, ssRNA 1, DENV ssRNA, and dsDNA 1 with LwaCas13a, PsmCas13b, CcaCas13b, and AsCas12a, respectively. (**G**) Schematic of in-sample multiplexed detection of *S. aureus* thermonuclease and *P. aeruginosa* acyltransferase synthetic targets with LwaCas13a and PsmCas13b. (**H**) In-sample multiplexed RPA and collateral detection at decreasing concentrations of *S. aureus* thermonuclease and *P. aeruginosa* acyltransferase synthetic targets with LwaCas13a and PsmCas13b.

assay. By testing RNA adenylate molecules of different lengths and 3′-end modifications (figs. S23 and S24A and table S6), we found that Csm6 from *Enterococcus italicus* (EiCsm6) and Csm6 from *Lactobacillus salivarius* (LsCsm6) were efficiently activated by hexadenylates containing 2′,3′-cyclic phosphate ends (fig. S24, B and C). Moreover, EiCsm6, LsCsm6, and Csm6 from *Thermus thermophilus* (TtCsm6) demonstrated a strong cleavage preference for A- and C-rich reporters based on reporter screening, enabling independent measurements of LwaCas13a and Csm6 cleavage activity in separate channels (Fig. 2D and figs. S24, B to D, S25, and S26, A to E).

To couple the activity of Cas13 with Csm6 activation, we designed protected RNA activators that contained a polyadenylate [poly(A)] stretch followed by a protecting poly(U) stretch that could be cleaved by a uracil-preferring Cas13 enzyme, with the rationale that LwaCas13a could degrade all the uridines down to the homopolymeric A stretch because it had robust activity on UU and AU two-base motifs (fig. S9). We found that, upon addition of target and LwaCas13a-crRNA complex, EiCsm6 and LsCsm6 were activated by the $(A)_6$-$(U)_5$ activator, con-

sistent with the finding that the A_6 activator is optimal for Csm6 activation and confirmed by mass spectrometry (Fig. 2E and figs. S26F, S27, and S28). We combined the reporters for both Csm6 and Cas13 in the same reaction within the same fluorescence channel and found that increasing the activator concentration increased the synergistic activation of Csm6 by Cas13 for DENV ssRNA detection (Fig. 2F), and that increasing the Csm6-specific poly(A) reporter also increased the Csm6 signal, leading to a larger increase in signal upon activator addition (fig. S29, A and B). After optimization (fig. S30), we found that Csm6-enhanced LwaCas13a increased the overall signal and kinetics of synthetic acyltransferase gene detection by SHERLOCK (Fig. 2G).

Another goal of SHERLOCKv2 was engineering a visual readout of activity requiring no additional instrumentation. We first tested a colorimetric ribonuclease (RNase) reporter based on gold nanoparticle cluster disaggregation (20, 21), but this readout required a level of RNase activity beyond what Cas13 collateral activity could achieve (fig. S31). We then designed a lateral-flow readout that was based on the destruction of a FAM-biotin reporter, allowing for

detection on commercial lateral flow strips. Abundant reporter accumulates anti-FAM antibody-gold nanoparticle conjugates at the first line on the strip, preventing binding of the antibody-gold conjugates to protein A on the second line; cleavage of reporter would reduce accumulation at the first line and result in signal on the second line (Fig. 3A). We tested this design for instrument-free detection of ZIKV or DENV ssRNA and found that detection was possible in <90 min with sensitivities as low as 2 aM (Fig. 3, B and C, and fig. S32). Moreover, we found that we could do rapid (<10 min) genomic DNA extraction from human saliva and input this directly into SHERLOCK without purification for rapid genotyping in <23 min by fluorescence and 2 hours by lateral flow (fig. S33). This exemplifies a closed-tube assay format in which the entire SHERLOCK reaction is performed in a one-pot assay without any sample purification.

We also applied the system to create a rapid and portable paper test for detecting mutations in liquid biopsies of non–small cell lung cancer (NSCLC) patients. We designed SHERLOCK assays to detect either the epidermal growth factor receptor (EGFR) Leu→Arg (L858R) mutation or
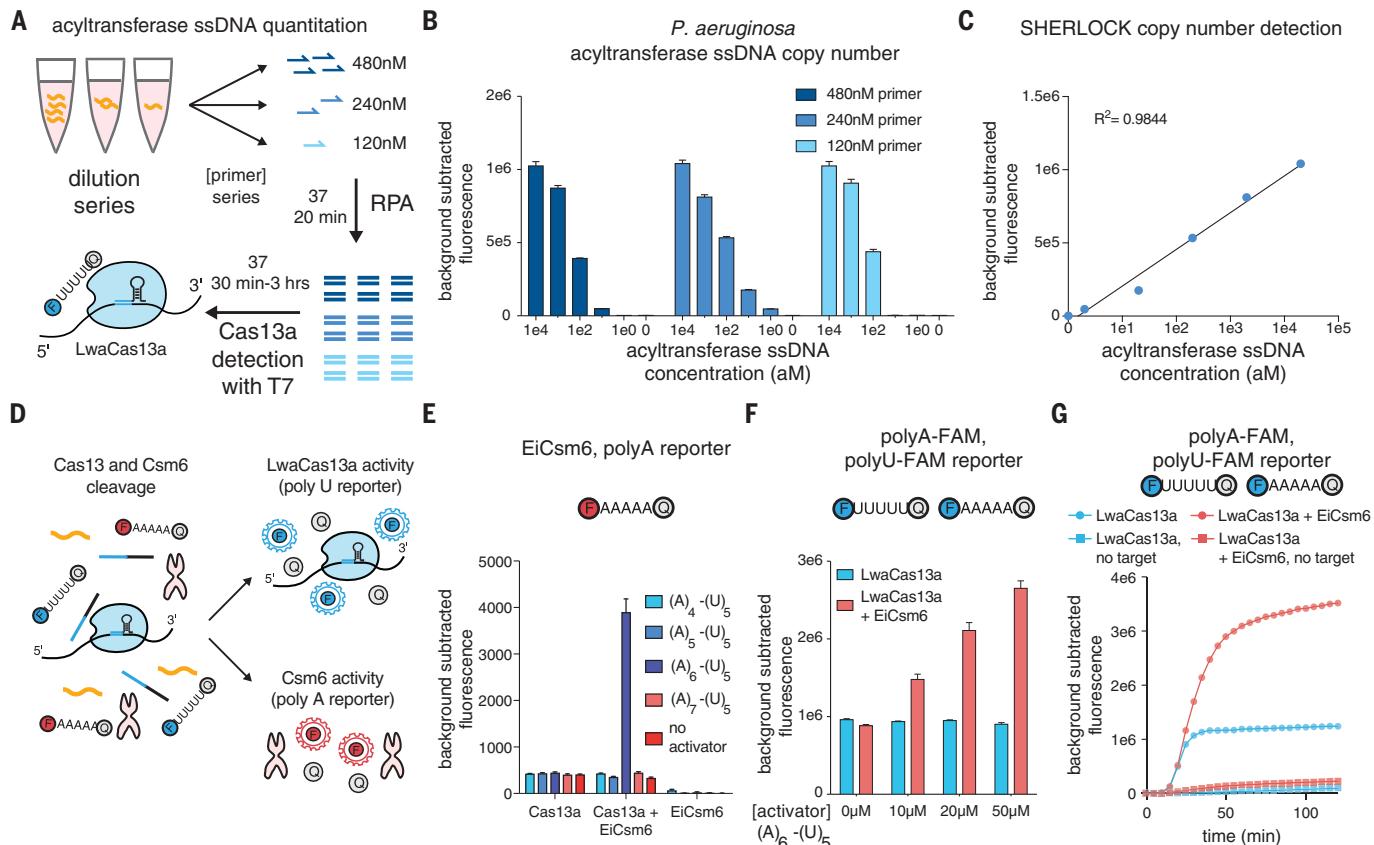
Fig. 2. Single-molecule quantitation and enhanced signal with SHERLOCK and Csm6. (**A**) Schematic of DNA reaction scheme for quantitation of *P. aeruginosa* synthetic DNA. (**B**) Quantitation of *P. aeruginosa* synthetic DNA at various RPA primer concentrations. Values represent mean ± SEM. (**C**) Correlation of *P. aeruginosa* synthetic DNA concentration with detected fluorescence. Values represent mean ± SEM. (**D**) Schematic of independent readout of LwaCas13a and Csm6 cleavage activity with orthogonal reporters. (**E**) Activation of EiCsm6 by LwaCas13a cleavage of adenine-uridine activators with adenine tracts of different lengths. LwaCas13a is targeting synthetic DENV ssRNA. Values represent mean ± SEM. (**F**) Combined LwaCas13a and EiCsm6 signal for increasing concentrations of $(A)_6$-$(U)_5$ activator detecting 20 nM of DENV ssRNA. Values represent mean ± SEM. (**G**) Kinetics of EiCsm6-enhanced LwaCas13a SHERLOCK detection of *P. aeruginosa* acyltransferase synthetic target.

the exon 19 deletion (five amino acids) and isolated cell-free DNA (cfDNA) from patients with or without these mutations (Fig. 3D), as verified by targeted sequencing (table S7). SHERLOCK successfully detected these mutations, both with fluorescence-based readout (Fig. 3, E and H)

and lateral flow–based readout (Fig. 3, F, G, I, and J, and fig. S34, A to D). Fluorescence-based SHERLOCK could also detect a different common EGFR mutation, Thr→Met (T790M), in synthetic and patient cfDNA liquid biopsy samples (fig. S34, E and F).

To improve the robustness of the detection and reduce the likelihood of a false-positive readout, we combined Csm6 with Cas13 detection on lateral flow (Fig. 3K). We tested lateral-flow reporters of various sequences and lengths in the presence of Csm6 and activator and found
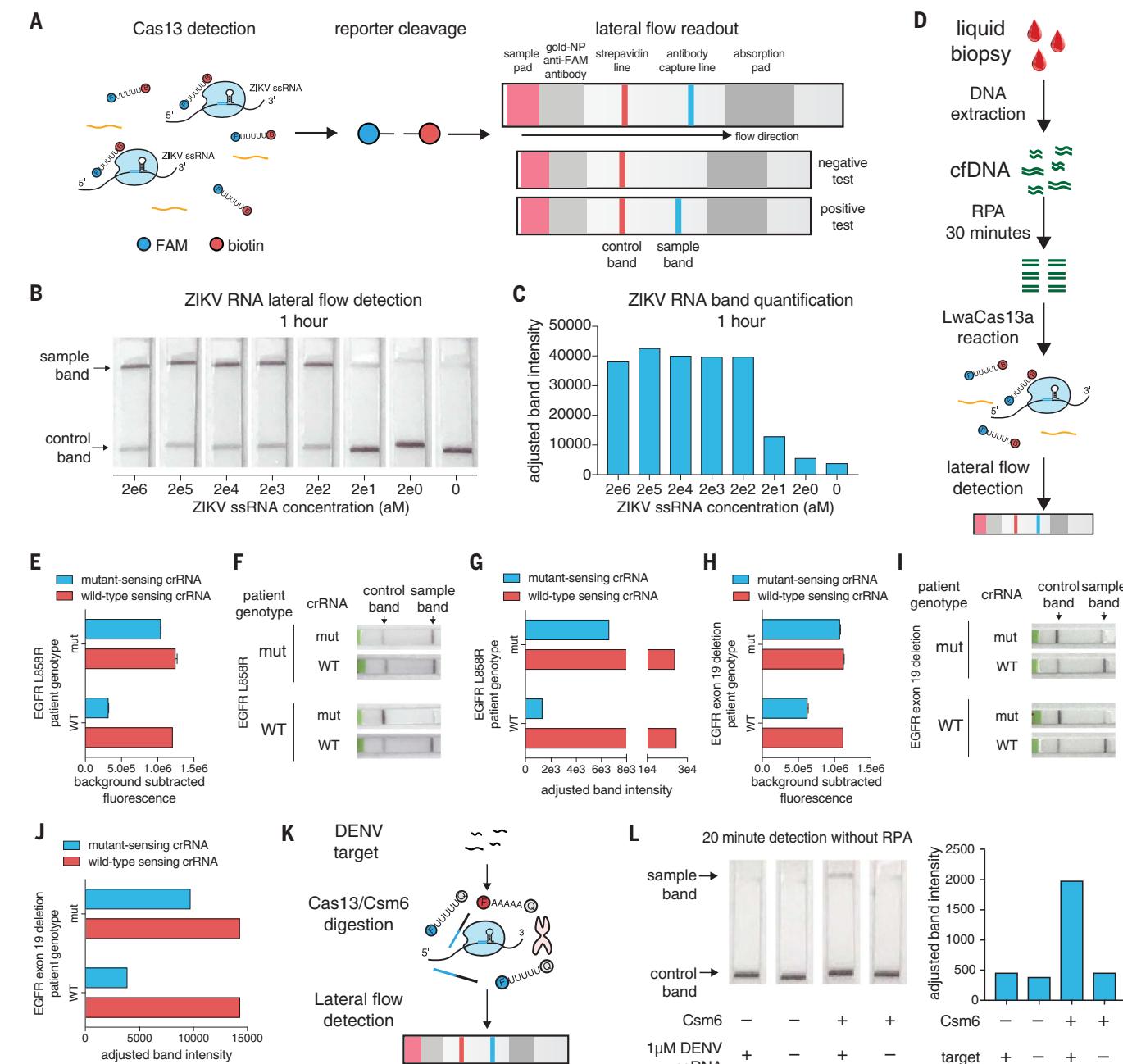

Fig. 3. Adapting SHERLOCK for lateral flow detection. (**A**) Schematic of lateral-flow detection with SHERLOCK. (**B**) Detection of synthetic ZIKV ssRNA using lateral-flow SHERLOCK with 1 hour of LwaCas13a reaction. (**C**) Quantitation of band intensity from detection in (B). (**D**) Schematic of lateral flow detection of therapeutically relevant EGFR mutations from patient liquid biopsy samples. (**E**) Detection of EGFR L858R mutation in patient-derived cfDNA samples with either L858R or wild-type (WT) alleles. Values represent mean ± SEM. (**F**) Lateral-flow detection of EGFR L858R mutation in patient-derived cfDNA samples with either L858R or WT alleles.

(**G**) Quantitation of band intensity from detection in (E). (**H**) Detection of EGFR exon 19 deletion mutation in patient-derived cfDNA samples with either exon 19 deletion or WT alleles. Values represent mean ± SEM. (**I**) Lateral-flow detection of EGFR exon 19 deletion mutation in patient-derived cfDNA samples with either exon 19 deletion or WT alleles. (**J**) Quantitation of band intensity from detection in (H). (**K**) Schematic of lateral-flow readout of EiCsm6-enhanced LwaCas13a detection of DENV ssRNA. (**L**) EiCsm6-enhanced lateral-flow detection of synthetic DENV RNA in combination with LwaCas13a without preamplification by RPA. Band intensity quantitation is shown to the right.

that a long A-C reporter demonstrated strong cleavage signal (fig. S35, A and B). We used this reporter in combination with the Cas13 lateral-flow reporter for rapid detection of DENV ssRNA, relying solely on Csm6 for amplification (i.e., in the absence of RPA) (Fig. 3L). We subsequently combined RPA, Cas13/Csm6, and lateral-flow readout to detect an acyltransferase target and found that the increase in signal conferred by Csm6 allowed for more rapid detection by lateral flow (fig. S35, C and D) with reduced background.

Finally, we applied SHERLOCKv2 in a simulated approach that involves Cas13 serving as both a companion diagnostic and the therapy itself, as Cas13 has been developed for a variety of applications in mammalian cells, including RNA knockdown, imaging, and editing (22, 23) (Fig. 4A and table S8). We recently harnessed Cas13b from *Prevotella* sp. P5-125 (PspCas13b) to correct mutations underlying genetic diseases by using a system called RNA Editing for Programmable A-to-I Replacement (REPAIR) (23). To direct and monitor the outcome of a treatment, we tested if SHERLOCK could be used both for genotyping to inform the REPAIR treatment and as a readout of the edited RNA to track the efficiency of the therapy. We used a mutation in *APC* (APC:c.1262G>A) implicated in familial adenomatous polyposis 1 (Fig. 4, B and C) (24) and transfected synthetic healthy and mutant cDNAs of the fragment surrounding the mutation into human embryonic kidney (HEK) 293FT cells. We harvested DNA from these cells and successfully genotyped the correct samples by using single-sample multiplexed SHERLOCK with LwaCas13a and PsmCas13b (Fig. 4D). Concurrently, we designed and cloned guide RNAs for the REPAIR system and transfected cells that had the diseased genotype with the guide RNA and dPspCas13b-ADAR2$_{dd}$(E488Q) REPAIR system. We confirmed editing by next-generation sequencing analysis, finding that 43% editing was achieved with the REPAIR system (Fig. 4E), and we could detect this editing with SHERLOCK (Fig. 4F and fig. S36).

The additional refinements presented here for Cas13-based detection allow for quantitative, visual, more sensitive, and multiplexed readouts, enabling additional applications for nucleic acid detection, especially in settings where portable and instrument-free analysis is necessary (table S9). SHERLOCKv2 can be used for multiplexed genotyping to inform pharmacogenomic therapeutic development and application, detecting genetically modified organisms in the field, or determining the presence of co-occurring pathogens. Moreover, the rapid, isothermal readout of SHERLOCKv2, enabled by lateral flow and Csm6, provides an opportunity for detection in settings where power or portable readers are unavailable, even for rare species like circulating DNA. In the future, it might be possible to make solution-based colorimetric readouts and multiplex lateral-flow assays containing multiple test lines for different targets. Improved CRISPR-based diagnostic (CRISPR-dx) nucleic acid tests make it easier

to detect the presence of nucleic acids in a range of applications across biotechnology and health and are now field-ready for rapid and portable deployment.
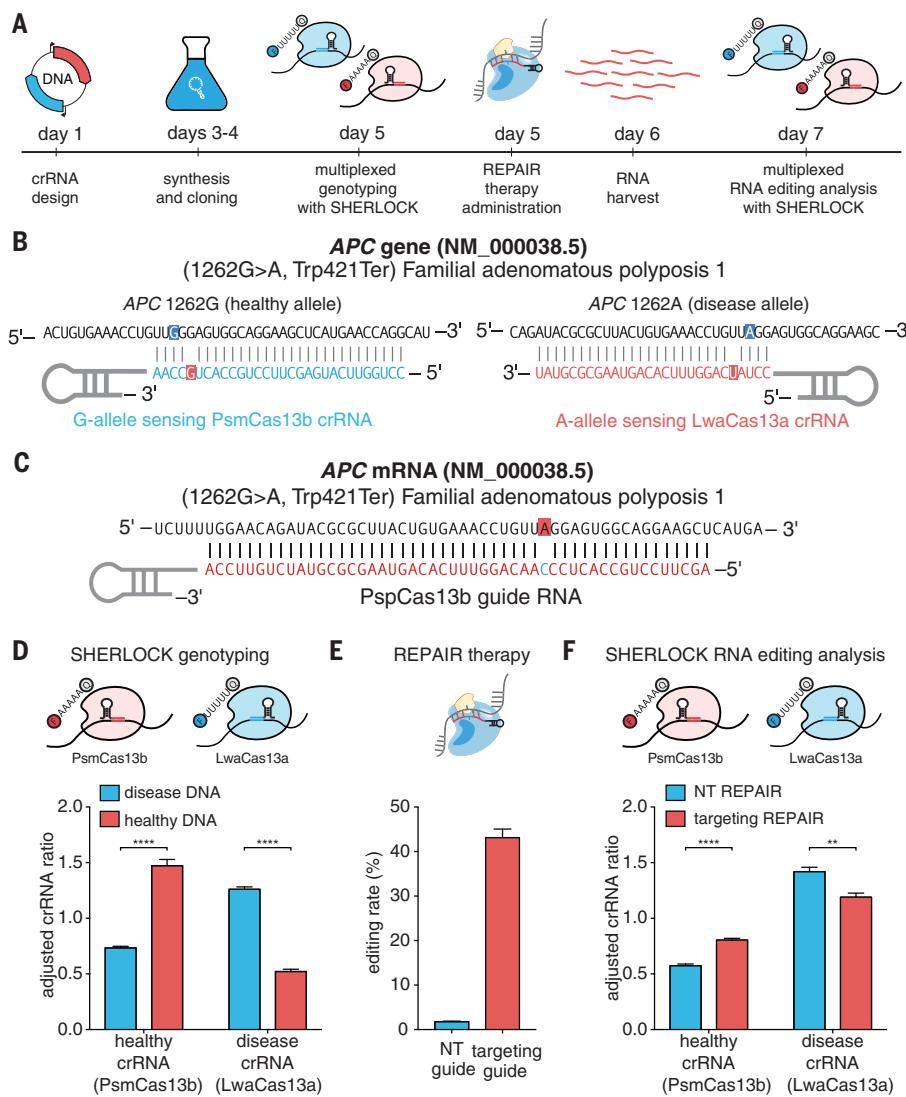
Fig. 4. Combined therapeutics and diagnostics with Cas13 enzymes. (**A**) Schematic of time line for detection of disease alleles, correction with REPAIR, and assessment of REPAIR correction. (**B**) Sequences of targets and crRNA designs used for detection of *APC* alleles. (**C**) Sequences of target and REPAIR guide design used for correction of *APC* alleles. (**D**) In-sample multiplexed detection of *APC* alleles from healthy- and disease-simulating samples with LwaCas13a and PsmCas13b. Adjusted crRNA ratio allows for comparisons between different crRNAs that will have different overall signal levels (see supplementary methods for more details). Values represent mean ± SEM. (**E**) Quantitation of REPAIR editing efficiency at the targeted *APC* mutation. Values represent mean ± SEM. (**F**) In-sample multiplexed detection of *APC* alleles from REPAIR targeting and nontargeting samples with LwaCas13a and PsmCas13b. Values represent mean ± SEM.

REFERENCES AND NOTES
1. S. Shmakov *et al.*, *Mol. Cell* **60**, 385–397 (2015).
2. O. O. Abudayyeh *et al.*, *Science* **353**, aaf5573 (2016).
3. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
4. A. East-Seletsky *et al.*, *Nature* **538**, 270–273 (2016).
5. A. East-Seletsky, M. R. O'Connell, D. Burstein, G. J. Knott, J. A. Doudna, *Mol. Cell* **66**, 373–383.e3 (2017).
6. S. Shmakov *et al.*, *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
7. A. A. Smargon *et al.*, *Mol. Cell* **65**, 618–630.e7 (2017).
8. J. S. Chen, E. Ma, L. B. Harrington, X. Tian, J. A. Doudna, *bioRxiv* [preprint]. 29 November 2017. www.biorxiv.org/content/early/2017/11/29/226993.
9. W. H. Organization, in *Guidelines for Using HIV Testing Technologies in Surveillance: Selection, Evaluation and Implementation: 2009 Update* (Geneva, 2009).
10. J. M. Barletta, D. C. Edelman, N. T. Constantine, *Am. J. Clin. Pathol.* **122**, 20–27 (2004).
11. L. Deng, R. A. Garrett, S. A. Shah, X. Peng, Q. She, *Mol. Microbiol.* **87**, 1088–1099 (2013).
12. G. W. Goldberg, W. Jiang, D. Bikard, L. A. Marraffini, *Nature* **514**, 633–637 (2014).
13. W. Jiang, P. Samai, L. A. Marraffini, *Cell* **164**, 710–721 (2016).
14. O. Niewoehner, M. Jinek, *RNA* **22**, 318–329 (2016).
15. P. Samai *et al.*, *Cell* **161**, 1164–1174 (2015).
16. R. H. Staals *et al.*, *Mol. Cell* **56**, 518–530 (2014).
17. G. Tamulaitis *et al.*, *Mol. Cell* **56**, 506–517 (2014).

18. M. Kazlauskiene, G. Kostiuk, Č. Venclovas, G. Tamulaitis, V. Siksnys, *Science* **357**, 605–609 (2017).
19. O. Niewoehner *et al.*, *Nature* **548**, 543–548 (2017).
20. W. Zhao, M. M. Ali, S. D. Aguirre, M. A. Brook, Y. Li, *Anal. Chem.* **80**, 8431–8437 (2008).
21. W. Zhao, J. C. Lam, W. Chiuman, M. A. Brook, Y. Li, *Small* **4**, 810–816 (2008).
22. O. O. Abudayyeh *et al.*, *Nature* **550**, 280–284 (2017).
23. D. B. T. Cox *et al.*, *Science* **358**, 1019–1027 (2017).
24. S. Cottrell, W. F. Bodmer, D. Bicknell, L. Kaklamanis, *Lancet* **340**, 626–630 (1992).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6387/439/suppl/DC1
Materials and Methods
Figs. S1 to S36
Tables S1 to S8
References (*25–27*)

BIOTECHNOLOGY

Field-deployable viral diagnostics using CRISPR-Cas13

Cameron Myhrvold,[1,2]*† Catherine A. Freije,[1,2,3]*† Jonathan S. Gootenberg,[1,4,5,6,7]‡
Omar O. Abudayyeh,[1,5,6,7,8]‡ Hayden C. Metsky,[1,9] Ann F. Durbin,[3,10] Max J. Kellner,[1]
Amanda L. Tan,[11] Lauren M. Paul,[11] Leda A. Parham,[12] Kimberly F. Garcia,[12]
Kayla G. Barnes,[1,2,13] Bridget Chak,[1,2] Adriano Mondini,[14] Mauricio L. Nogueira,[15]
Sharon Isern,[11] Scott F. Michael,[11] Ivette Lorenzana,[12] Nathan L. Yozwiak,[1,2]
Bronwyn L. MacInnis,[1,13] Irene Bosch,[10,16] Lee Gehrke,[3,10,17]
Feng Zhang,[1,5,6,7] Pardis C. Sabeti[1,2,3,13,18]†

Mitigating global infectious disease requires diagnostic tools that are sensitive, specific, and rapidly field deployable. In this study, we demonstrate that the Cas13-based SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) platform can detect Zika virus (ZIKV) and dengue virus (DENV) in patient samples at concentrations as low as 1 copy per microliter. We developed HUDSON (heating unextracted diagnostic samples to obliterate nucleases), a protocol that pairs with SHERLOCK for viral detection directly from bodily fluids, enabling instrument-free DENV detection directly from patient samples in <2 hours. We further demonstrate that SHERLOCK can distinguish the four DENV serotypes, as well as region-specific strains of ZIKV from the 2015–2016 pandemic. Finally, we report the rapid (<1 week) design and testing of instrument-free assays to detect clinically relevant viral single-nucleotide polymorphisms.

Recent viral outbreaks have highlighted the challenges of diagnosing viral infections, particularly in areas far from clinical laboratories. Viral diagnosis was especially difficult during the 2015–2016 Zika virus (ZIKV) pandemic; low viral titers and transient infection (1, 2), combined with limitations of existing diagnostic technologies, contributed to ZIKV circulating for months before the first cases of infection were confirmed clinically (3–5). An additional challenge for viral diagnostics is differentiating between related viruses that cause infections with similar symptoms, like ZIKV and dengue virus (DENV) (1). Existing nucleic acid detection methods are very sensitive and rapidly adaptable, but most require extensive sample manipulation and expensive machinery (1, 6–8). In contrast, antigen-based rapid diagnostic tests require minimal equipment but have lower sensitivity and specificity, and assay development can take months (9–11). An ideal diagnostic would combine the sensitivity, specificity, and flexibility of nucleic acid diagnostics with the speed and ease of use of antigen-based tests. Such a diagnostic could be rapidly developed and deployed in the face of emerging viral outbreaks and would be suitable for disease surveillance or routine clinical use in any context.

The Cas13-based nucleic acid detection platform SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) has the potential to address the key challenges associated with viral diagnostics. SHERLOCK combines isothermal amplification via recombinase polymerase amplification (RPA) (12) with highly specific Cas13-based detection (Fig. 1A) (13). Cas13, an RNA-guided ribonuclease, provides specificity through CRISPR RNA (crRNA)–target pairing and additional sensitivity due to signal amplification by Cas13's collateral cleavage activity (14, 15).

For SHERLOCK to excel at viral detection in any context, it should be paired with methods enabling a visual readout. In this study, we tested the performance of SHERLOCK for ZIKV and DENV detection in patient samples and developed HUDSON (heating unextracted diagnostic samples to obliterate nucleases), a method to enable rapid, sensitive detection of ZIKV and DENV directly from bodily fluids with a colorimetric readout, demonstrated as part of SHERLOCKv2 (16). Additionally, we designed SHERLOCK assays to distinguish multiple viral species and strains and identify clinically relevant mutations.

Detection of ZIKV and DENV in patient samples provides a stringent test of the sensitivity of SHERLOCK and its tolerance of viral diversity. Our ZIKV SHERLOCK assay had single-copy [1 copy (cp)/μl] sensitivity when tested on seed stock cDNA (fig. S1). We evaluated its performance on 40 cDNAs derived from samples collected during the 2015–2016 ZIKV pandemic, 37 from samples obtained from patients with suspected ZIKV infections and 3 from mosquito pools (Fig. 1B, fig. S2, and table S1). For 16 samples from these patients, we benchmarked SHERLOCK by comparing its sensitivity and specificity to those of other nucleic acid amplification tests, including the commercially available Altona Realstar ZIKV reverse transcription polymerase chain reaction (RT-PCR) assay (Fig. 1C, figs. S3 to S5, table S2, and supplementary text). Of the 10 samples that tested positive by the Altona assay, all 10 were detected by SHERLOCK (100% sensitivity); the other 6 samples were negative by both assays (100% specificity, 100% concordance). Our ZIKV assay had no false positives when tested on healthy urine and water (Fig. 1B). We then validated the ability of SHERLOCK to detect DENV, a related but more diverse flavivirus that causes symptoms similar to those of ZIKV infection. All 24 RT-PCR–positive DENV RNA samples were confirmed to be positive for DENV after 1 hour of detection (Fig. 1D, figs. S6 and S7, and table S3). SHERLOCK sensitively and specifically detects viral nucleic acids extracted from ZIKV and DENV patient samples.

Although SHERLOCK excels at detecting extracted nucleic acids, a field-deployable, rapid diagnostic test should not require an extraction step to detect viral nucleic acid in bodily fluids. Many viruses, including ZIKV and DENV, are shed in urine or saliva, and sampling is not invasive (2, 7). To detect viral nucleic acid directly from bodily fluids via SHERLOCK, we developed HUDSON, a method to lyse viral particles and inactivate the high levels of ribonucleases found in bodily fluids with the use of heat and chemical reduction (Fig. 2A and fig. S8) (17). HUDSON-treated urine or saliva could be directly added to RPA reaction mixtures with no dilution or purification step [blood products were diluted 1:3 in phosphate-buffered saline (PBS) to avoid solidification during HUDSON] without inhibiting subsequent amplification or detection. HUDSON and SHERLOCK enabled sensitive detection of free ZIKV nucleic acid spiked into

Fig. 1. ZIKV and DENV detection from patient samples and clinical isolates. (**A**) Schematic of SHERLOCK. Nucleic acid is extracted from clinical samples, and the target is amplified by RPA with either RNA or DNA as the input (RT-RPA or RPA, respectively). RPA products are detected in a reaction mixture containing T7 RNA polymerase, Cas13, a target-specific crRNA, and an RNA reporter that fluoresces when cleaved. We tested SHERLOCK on (**B**) cDNAs derived from 37 patient samples collected during the 2015–2016 ZIKV pandemic and (**C**) cDNAs from 16 patient samples for which results were compared head-to-head to those from the Altona RealStar ZIKV RT-PCR assay. +, ZIKV seed stock cDNA (3×10^2 cp/µl); ×, no input; a.u., arbitrary units. (**D**) SHERLOCK testing of RNA extracted from 24 DENV-positive patient samples and clinical isolates. Dashed blue line: threshold for detecting the presence or absence of ZIKV or DENV (see methods in the supplementary materials).
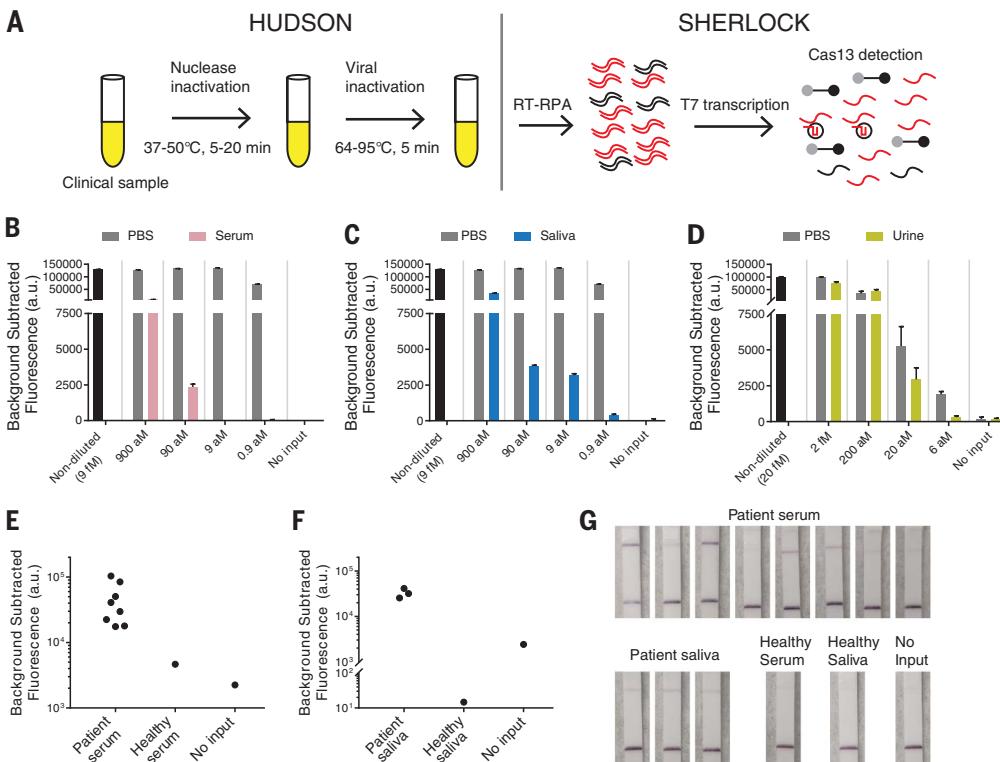
Fig. 2. Direct detection of ZIKV and DENV in bodily fluids with HUDSON and SHERLOCK. (**A**) Schematic of direct viral detection by HUDSON and SHERLOCK. (**B** and **C**) Detection of ZIKV RNA in particles diluted in healthy human serum (B) or healthy human saliva (C). The same PBS control was used for (B) and (C) as experiments were performed together. (**D**) Detection of ZIKV RNA in particles diluted in healthy human urine. Error bars indicate 1 SD for three technical replicates. (**E** and **F**) Detection of DENV RNA directly from patient serum (E) and saliva (F) samples. (**G**) Lateral-flow detection of DENV from the samples represented in (E) and (F). All samples were treated with tris(2-carboxyethyl)phosphine hydrochloride (TCEP)–EDTA before being heated.

urine, whole blood, plasma, serum, or saliva (figs. S9 and S12). To mimic clinical infection, where viral nucleic acid is encapsulated in infectious particles, we spiked infectious ZIKV particles into bodily fluids. HUDSON combined with SHERLOCK (figs. S13 and S14) permitted sensitive detection of ZIKV RNA from infectious particles at 90 aM (45 cp/µl) in whole blood (fig. S15) or serum (Fig. 2B), 0.9 aM (~1 cp/µl) in saliva (Fig. 2C), and 20 aM (10 cp/µl) in urine. The total turnaround time was <2 hours with fluorescent and colorimetric readouts (Fig. 2D and fig. S16). The sensitivity of HUDSON and SHERLOCK is comparable to ZIKV RNA concentrations observed in patient samples, which range from 1 to 1,000 cp/µl (1, 2). HUDSON, paired with the pan-DENV SHERLOCK assay, detected DENV in whole blood, serum, and saliva (figs. S17 and S18). DENV was detected directly from eight of eight patient serum samples (Fig. 2E) and three of three patient saliva samples (Fig. 2F) tested, with

Fig. 3. Multivirus panels can be used to differentiate viral species and serotypes. (**A** to **C**) Panel of four related flaviviruses (A) used to detect individual viral targets (B) or paired viral targets (C) with species-specific crRNAs after 3 hours. Z, ZIKV; D, DENV; W, WNV; Y, YFV. (**D** and **E**) Identification of DENV serotypes 1 through 4 with the use of serotype-specific crRNAs (D), tested with the use of synthetic targets after 3 hours (E). (**F**) Identification of DENV serotypes in RNA extracted from patient samples. Each row represents a sample, each column represents a crRNA, and target-specific fluorescence values are normalized by row. Purple: DENV serotype identified. Synthetic targets were used at 10^4 cp/μl. Error bars indicate 1 SD for three technical replicates. We expect off-target crRNAs to have close to zero target-specific fluorescence (see methods in the supplementary materials). Primer, crRNA, and target sequences are in tables S5 to S7.
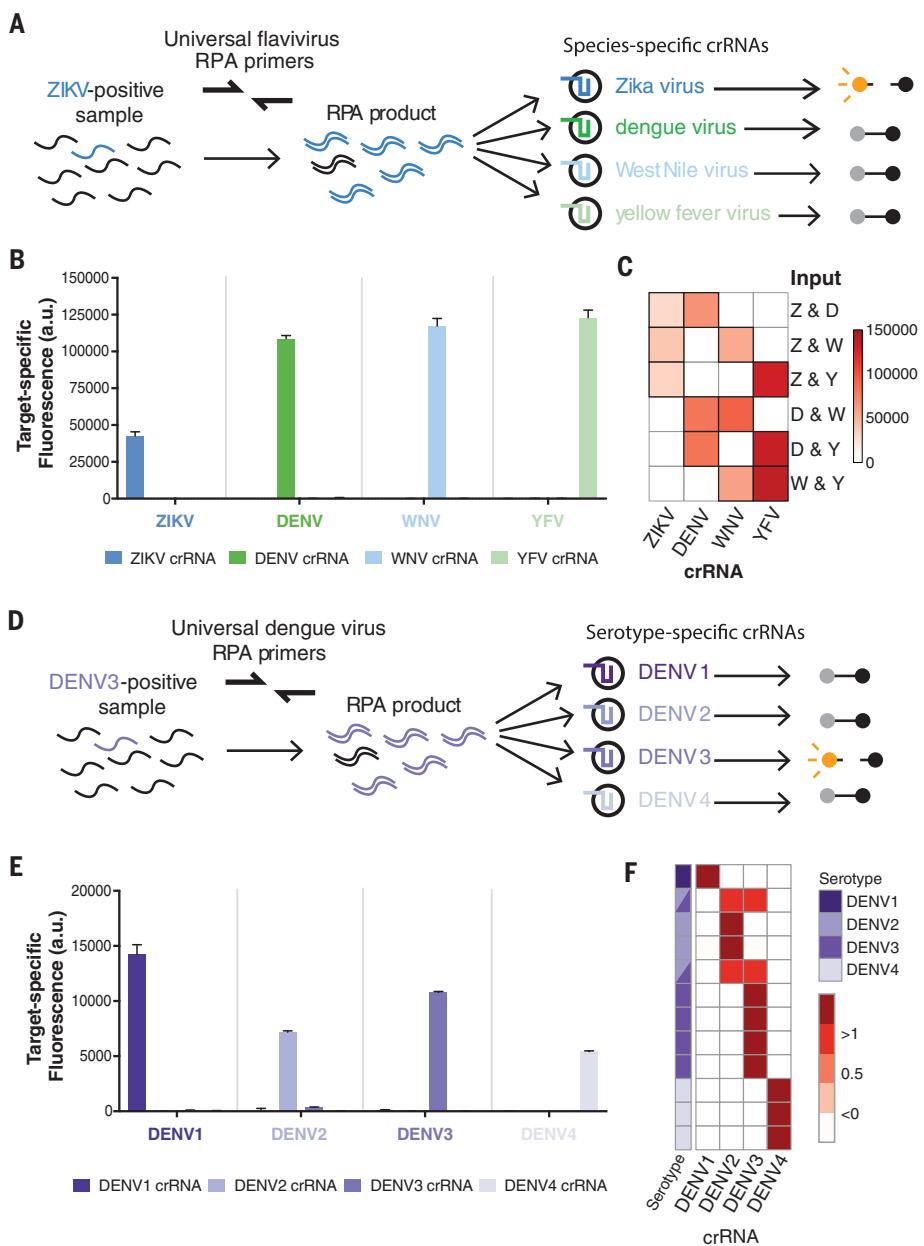
a total turnaround time of <1 hour for saliva despite lower viral titers than those in serum (7). We directly detected DENV with a colorimetric readout using lateral-flow strips (Fig. 2G), showcasing a HUDSON-to-SHERLOCK pipeline that can detect ZIKV or DENV directly from bodily fluids with minimal equipment.

Because many genetically and antigenically similar flaviviruses cocirculate and cause similar symptoms, we developed diagnostic panels to distinguish related viral species and serotypes. We identified conserved regions within ZIKV, DENV, West Nile virus (WNV), and yellow fever virus (YFV) genomes and designed a flavivirus panel with universal flavivirus RPA primers that can amplify any of the four viruses and species-specific crRNAs (Fig. 3A). This panel detected synthetic ZIKV, DENV, WNV, and YFV DNA targets with <0.22% off-target fluorescence (Fig. 3B,

figs. S19 and S20, and methods in the supplementary materials) and identified the presence of all pairwise combinations of these four viruses, demonstrating the ability to detect mock coinfections (Fig. 3C and figs. S21 and S22). We also designed a DENV panel with DENV-specific RPA primers and serotype-specific crRNAs (Fig. 3D) that could distinguish between DENV serotypes 1 through 4 with <3.2% off-target fluorescence (Fig. 3E and figs. S23 and S24). This low level of off-target fluorescence allows for 100% specificity in differentiating among serotypes, providing an alternative to current serotype identification approaches (8). The DENV panel confirmed the serotypes of 12 RT-PCR–serotyped patient samples or clinical isolates (Fig. 3F and fig. S25) and identified two clinical isolates with mixed infection, a commonly observed phenomenon (18). SHERLOCK can therefore be extended

to differentiate between related viruses or serotypes with a single amplification reaction.

SHERLOCK is poised for field-deployable variant identification, which would allow real-time tracking of microbial threats. Genotyping of single-nucleotide polymorphisms (SNPs) typically involves PCR and either fluorescence- or mass spectrometry–based detection, requiring extensive sample processing and expensive equipment and limiting field deployability (19). SHERLOCK can identify SNPs by placing a synthetic mismatch in the crRNA near the SNP and testing each target with ancestral target-specific and derived target-specific crRNAs (Fig. 4A) (13). We designed diagnostics for three region-specific SNPs from the 2015–2016 ZIKV pandemic (Fig. 4B) and identified these SNPs in synthetic targets, a viral seed stock, and cDNA samples from Honduras, the Dominican Republic, and the United States (Fig. 4, C to E). These results
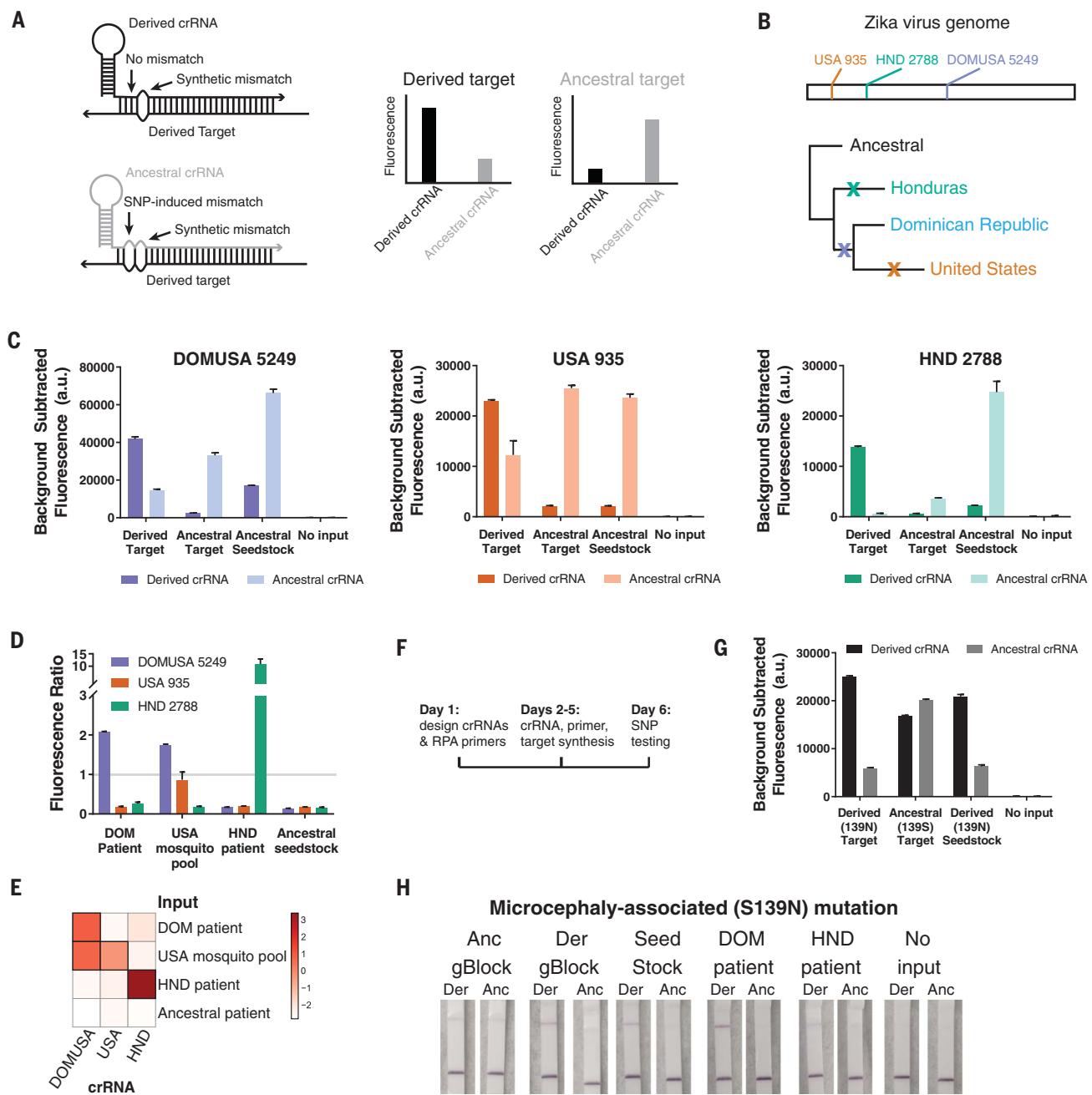
Fig. 4. Identification of adaptive and functional ZIKV mutations.
(**A**) SHERLOCK assays for SNP identification. (**B** and **C**) Three region-specific SNPs from the 2015–2016 ZIKV pandemic, including genomic locations of the SNPs and a simplified phylogenetic tree, tested with the use of synthetic targets (10^4 cp/μl) and viral seed stock cDNA (3×10^2 cp/μl). (**D** and **E**) Identification of region-specific SNPs in ZIKV cDNA samples from the Dominican Republic (DOM), the United States, and Honduras (HND). The fluorescence ratio (derived crRNA fluorescence divided by ancestral crRNA fluorescence) for each SNP in each sample is shown in a bar plot ($\log_2$-transformed data in a heatmap). (**F**) Timeline for developing a SHERLOCK assay for a new SNP (more detail in fig. S26). (**G** and **H**) Identification of a microcephaly-associated ZIKV mutation (PrM S139N) by fluorescence and colorimetric detection. In all panels, error bars indicate 1 SD for three technical replicates. Anc, ancestral; Der, derived; gBlock, gene fragment.

demonstrate that SHERLOCK can identify SNPs in samples from the ZIKV pandemic and highlight the single-nucleotide specificity of SHERLOCK.

Rapid identification of emerging drug resistance and other clinically relevant mutations for viruses such as ZIKV and HIV would have great utility. A ZIKV point mutation in the PrM protein region [Ser[139]→Asn (S139N)] recently reported to be associated with fetal microcephaly (20) was used as a test case for the rapid development of assays for variant identification. Within a week of the report's publication (Fig. 4F and fig. S26), we developed multiple SHERLOCK assays for the S139N mutation (Fig. 4G and fig. S27) and

could identify the mutation in patient samples from the 2015–2016 ZIKV pandemic with a visual readout (Fig. 4H). To further illustrate the ease of developing SHERLOCK diagnostics for many clinically relevant mutations, we designed and tested assays for the six most commonly observed drug resistance mutations in HIV reverse transcriptase

(*21*) in 1 week (fig. S28). These examples underscore the potential for SHERLOCK to be used for monitoring clinically relevant variants in near real time.

Combining HUDSON and SHERLOCK, we have created a field-deployable viral diagnostic platform with high performance and minimal equipment or sample processing requirements. This platform is as sensitive and specific as amplification-based nucleic acid diagnostics (*12*, *22–26*), with speed and equipment requirements similar to those of rapid antigen tests (*9–11*). Furthermore, this approach can be easily adapted to detect virtually any virus present in bodily fluids and scaled to enable multiplexed detection (*16*), and the reagents can be lyophilized for cold-chain independence (*13*). Cas13-based detection is a promising next-generation diagnostic strategy with the potential to be implemented almost anywhere in the world to enable effective, rapid diagnosis of viral infections.

REFERENCES AND NOTES

1. O. Faye *et al.*, *J. Clin. Virol.* **43**, 96–101 (2008).
2. G. Paz-Bailey *et al.*, *N. Engl. J. Med.* NEJMoa1613108 (2017).
3. H. C. Metsky *et al.*, *Nature* **546**, 411–415 (2017).
4. N. R. Faria *et al.*, *Nature* **546**, 406–410 (2017).
5. N. D. Grubaugh *et al.*, *Nature* **546**, 401–405 (2017).
6. O. Faye *et al.*, *Virol. J.* **10**, 311 (2013).
7. A.-C. Andries *et al.*, *PLOS Negl. Trop. Dis.* **9**, e0004100 (2015).
8. J. J. Waggoner *et al.*, *J. Clin. Microbiol.* **51**, 3418–3420 (2013).
9. I. Bosch *et al.*, *Sci. Transl. Med.* **9**, eaan1589 (2017).
10. A. Balmaseda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8384–8389 (2017).
11. L. Priyamvada *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7852–7857 (2016).
12. O. Piepenburg, C. H. Williams, D. L. Stemple, N. A. Armes, *PLOS Biol.* **4**, e204 (2006).
13. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
14. O. O. Abudayyeh *et al.*, *Science* **353**, aaf5573 (2016).
15. A. East-Seletsky *et al.*, *Nature* **538**, 270–273 (2016).
16. J. S. Gootenberg *et al.*, *Science* **360**, 439–444 (2018).
17. J. L. Weickmann, D. G. Glitz, *J. Biol. Chem.* **257**, 8705–8710 (1982).
18. R. Requena-Castro, M. Á. Reyes-López, R. E. Rodríguez-Reyna, P. Palma-Nicolás, V. Bocanegra-García, *Mem. Inst. Oswaldo Cruz* **112**, 520–522 (2017).
19. S. Kim, A. Misra, *Annu. Rev. Biomed. Eng.* **9**, 289–320 (2007).
20. L. Yuan *et al.*, *Science* **358**, 933–936 (2017).
21. S.-Y. Rhee *et al.*, *Nucleic Acids Res.* **31**, 298–303 (2003).
22. Y. Du *et al.*, *Angew. Chem. Int. Ed.* **56**, 992–996 (2017).
23. K. E. Eboigbodin, M. Brummer, T. Ojalehto, M. Hoser, *Diagn. Microbiol. Infect. Dis.* **86**, 369–371 (2016).
24. K. Pardee *et al.*, *Cell* **165**, 1255–1266 (2016).
25. N. Chotiwan *et al.*, *Sci. Transl. Med.* **9**, eaag0538 (2017).
26. J. Van Ness, L. K. Van Ness, D. J. Galas, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4504–4509 (2003).

IMMUNOLOGY

Dimethyl fumarate targets GAPDH and aerobic glycolysis to modulate immunity

Michael D. Kornberg,[1] Pavan Bhargava,[1] Paul M. Kim,[2] Vasanta Putluri,[3] Adele M. Snowman,[4] Nagireddy Putluri,[3,5] Peter A. Calabresi,[1,4] Solomon H. Snyder[2,4,6]*

Activated immune cells undergo a metabolic switch to aerobic glycolysis akin to the Warburg effect, thereby presenting a potential therapeutic target in autoimmune disease. Dimethyl fumarate (DMF), a derivative of the Krebs cycle intermediate fumarate, is an immunomodulatory drug used to treat multiple sclerosis and psoriasis. Although its therapeutic mechanism remains uncertain, DMF covalently modifies cysteine residues in a process termed succination. We found that DMF succinates and inactivates the catalytic cysteine of the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in mice and humans, both in vitro and in vivo. It thereby down-regulates aerobic glycolysis in activated myeloid and lymphoid cells, which mediates its anti-inflammatory effects. Our results provide mechanistic insight into immune modulation by DMF and represent a proof of concept that aerobic glycolysis is a therapeutic target in autoimmunity.

Pro-inflammatory stimuli induce a metabolic switch in both myeloid and lymphoid cells, leading to a Warburg-like up-regulation of aerobic glycolysis that regulates the balance between inflammatory and regulatory immune phenotypes (1, 2). Classically activated macrophages and effector lymphocytes such as T helper 1 (T_H1) and T_H17 cells require glycolysis for their survival, differentiation, and effector functions (3–9), whereas oxidative metabolism favors the differentiation of alternatively activated (M2) macrophages and regulatory T (T_{reg}) cells (10, 11).

Dimethyl fumarate (DMF) is a serendipitously discovered immunomodulatory drug used to treat psoriasis and multiple sclerosis (MS) (12). Although its mechanisms of action remain incompletely understood, it is known to covalently modify cysteine residues in a process termed succination (not to be confused with lysine succinylation) (fig. S1) (13, 14). DMF succinates kelch-like ECH-associated protein 1 (KEAP1), which activates nuclear factor (erythroid-derived 2)–related factor 2 (Nrf2) to produce antioxidant and anti-inflammatory effects that nonetheless fail to fully account for the drug's actions (15). Endogenous fumarate also succinates proteins, with a primary target being the active-site cysteine of the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (16). Here, we show that DMF and its clinically relevant metabolite monomethyl fumarate (MMF) target GAPDH and inactivate its enzyme activity, both in vitro and after oral treatment in mice and humans. In turn, GAPDH inhibition down-regulates aerobic glycolysis in myeloid and lymphoid cells, preventing immune activation and shifting the balance between inflammatory and regulatory cell types.

To determine whether GAPDH was succinated by DMF and MMF, we performed liquid chromatography–tandem mass spectrometry (LC–MS/MS). Treatment of recombinant human GAPDH with MMF led to monomethyl succination (2-monomethyl succinyl-cysteine) at its active-site cysteine (Cys^{152} in human) as well as at Cys^{156} and Cys^{247}, whereas DMF produced a combination of dimethyl succination (2-dimethyl succinyl-cysteine) and monomethyl succination at the same cysteines (table S1 and fig. S2). Neither of these modifications was observed at any cysteine in vehicle-treated GAPDH. In mice, oral treatment with DMF led to both monomethyl and dimethyl succination exclusively of the active-site cysteine (Cys^{150} in mouse) of GAPDH purified from the spleen and brain, with no such modifications observed on other cysteines or in vehicle-treated mice (Fig. 1A and table S1). Although only the monomethyl form of the drug is detected in serum (17), our finding that dimethyl succination occurred after oral administration was consistent with prior work (18, 19). Monomethyl succination of GAPDH Cys^{152} was also identified in peripheral blood mononuclear cells (PBMCs) from MS patients treated with DMF (Fig. 1B and table S1) but not on other cysteines or in PBMCs from MS patients not treated with DMF. GAPDH succination by endogenous fumarate occurred physiologically, as we identified modification by fumarate (2-succinyl-cysteine) at the active-site cysteine in vehicle-treated mice and healthy human controls (fig. S3).

Covalent modification of its catalytic cysteine should irreversibly inactivate GAPDH. We found that both DMF and MMF decreased the catalytic activity of recombinant GAPDH in a dose- and time-dependent manner (Fig. 1C). This inhibition was irreversible, as desalting failed to restore activity (fig. S4A). Additionally, this inhibition was mediated by active-site binding of DMF or MMF, as the drug effect was blocked by pre-incubation with 10-fold excess of GAPDH substrates (fig. S4B). GAPDH inhibition was biphasic, with an initial fast phase and a secondary slow phase. We used the Kitz-Wilson method (20) to calculate the kinetics of inhibition for both phases (fig. S4C). GAPDH activity was similarly inhibited in cultured mouse peritoneal macrophages (mPMs) treated overnight with 25 μM DMF (Fig. 1D). In mice, oral treatment with DMF decreased GAPDH activity measured from both the spleen and small intestine (Fig. 1E). This effect was particularly profound in the small intestine, which may be relevant given the role of the gut immune system in autoimmune disorders such as MS (21).

We next asked whether GAPDH inhibition by DMF affected aerobic glycolysis in activated immune cells. Using lactate production as a proxy measure, co-treatment with DMF significantly impaired glycolysis in mPMs stimulated in vitro for 24 hours with lipopolysaccharide (LPS; 1 μg/ml) (Fig. 2A). MMF had a similar effect but with lower potency. The influence on glycolysis was not due to cytotoxicity (fig. S5). Measurements of the extracellular acidification rate revealed a similar inhibition of aerobic glycolysis by DMF and MMF in LPS-stimulated mPMs (Fig. 2B). In activated mouse and human $CD4^+$ lymphocytes, DMF and MMF decreased basal glycolysis, with an even greater effect on maximal glycolytic capacity (Fig. 2C and fig. S6). In LPS-stimulated mPMs, treatment with DMF produced a blockade in glycolytic flux at GAPDH (Fig. 2D and fig. S7), providing evidence that GAPDH inactivation mediates the down-regulation of glycolysis by DMF.

DMF had no effect on glycolysis in unstimulated mPMs (Fig. 2E), raising the possibility that DMF acts not only on GAPDH but also on the signaling pathway required for glycolytic up-regulation. However, DMF had no effect on the activity of mechanistic target of rapamycin (mTOR, as measured via p70-S6 kinase phosphorylation) or levels of hypoxia-inducible factor 1α (HIF-1α) (Fig. 2F and fig. S8). This selective inhibition of glycolysis in LPS-stimulated mPMs is consistent with recent evidence that GAPDH only becomes a rate-limiting enzyme when glycolysis is up-regulated in the setting of Warburg physiology (22–24), as it is in cancer and activated immune cells. This likely explains why DMF is not generally toxic.

We also examined whether DMF affected oxidative phosphorylation (OXPHOS). DMF increased OXPHOS in mPMs under both resting and LPS-stimulated conditions (fig. S9A). The inhibition of glycolysis did not depend on the up-regulation of OXPHOS (fig. S9, B and C).

[1]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. [2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. [3]Advanced Technology Core, Baylor College of Medicine, Houston, TX 77030, USA. [4]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [5]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA. [6]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
*Corresponding author. Email: ssnyder@jhmi.edu

Fig. 1. DMF and MMF succinate and inactivate GAPDH in vitro and after oral treatment. (**A**) Representative LC-MS/MS spectra demonstrating covalent modification of the catalytic cysteine (Cys[150] in mouse) by either MMF (2-monomethyl succinyl-cysteine, +130 Da) or DMF (2-dimethyl succinyl-cysteine, +144 Da) in GAPDH immuno-precipitated from splenic lysates of mice treated orally with DMF (100 mg/kg) daily for 5 days. Because samples were reduced and treated with iodoacetate, nonsuccinated cysteines were modified with carbamido-methyl (+57 Da). Pooled samples were analyzed from two vehicle-treated and two DMF-treated animals. (**B**) Representative LC-MS/MS spectrum demonstrating monomethyl succination of the catalytic cysteine (Cys[152] in human) of GAPDH immunoprecipitated from PBMC lysates of MS patients treated with DMF for 3 months. Pooled samples were analyzed from three DMF-treated and two non–DMF-treated patients. (**C**) Dose- and time-dependent inactivation of GAPDH enzyme activity in vitro. Recombinant GAPDH was treated with the indicated drug concentrations or vehicle alone. Aliquots were removed at the specified time points, followed by enzyme activity assay. Data were pooled from four experiments performed in duplicate and represent mean ± SEM for each time point. See fig. S4C for associated Kitz-Wilson plots and kinetic parameters. (**D**) Peritoneal macrophages were treated overnight with DMF. Cell lysates were used for GAPDH enzyme activity assay. Data represent mean ± SEM of three experiments performed in duplicate. (**E**) Mice were treated with DMF (100 mg/kg) daily for 5 days by oral gavage. On day 5, mice were killed and lysates from spleen and small intestine were used for GAPDH enzyme activity assay. Data represent mean ± SEM of five mice per group, with assays run in triplicate. *$P < 0.05$, **$P < 0.01$ (two-tailed Student t test).
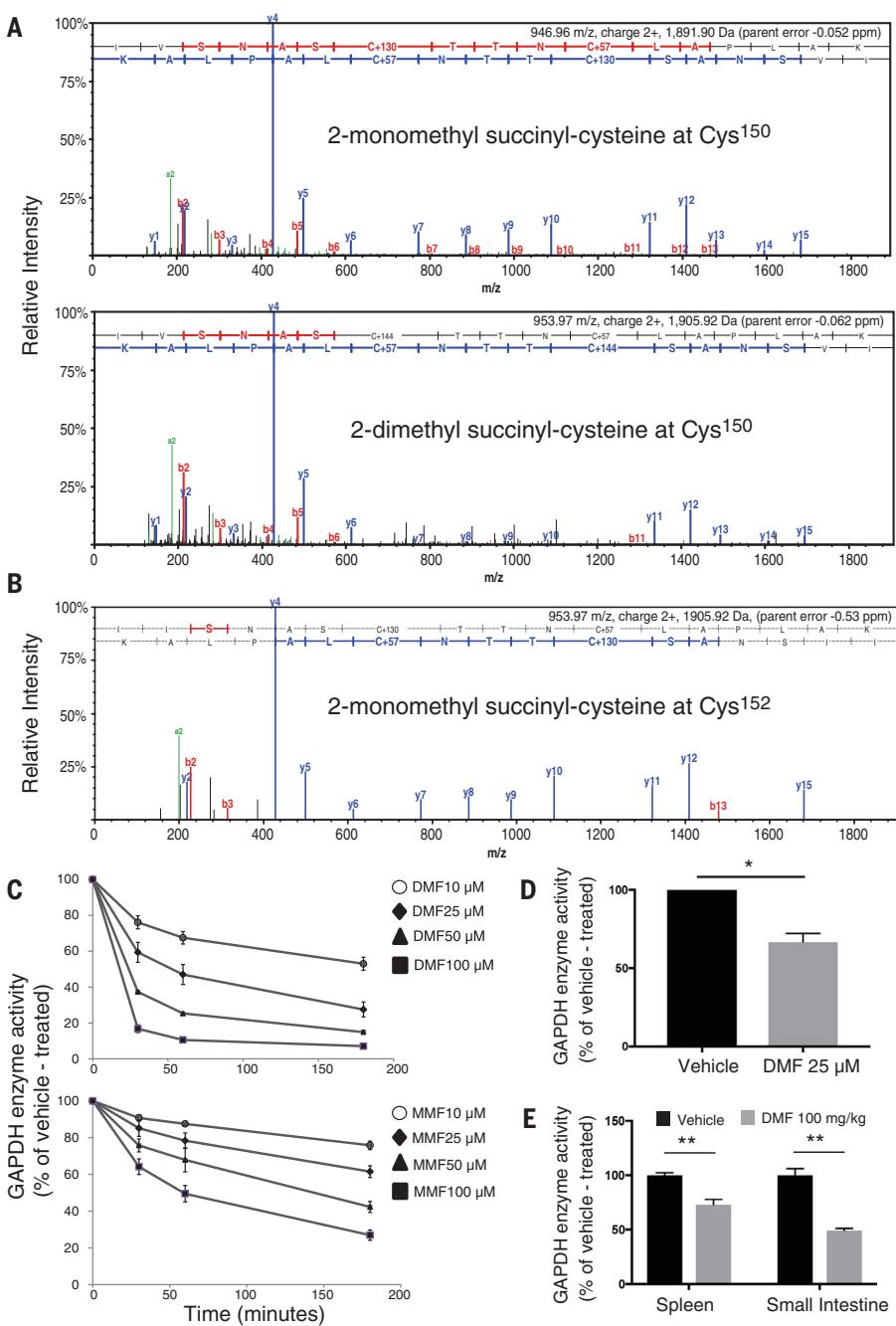
We next asked whether inhibition of GAPDH and aerobic glycolysis mediated the immunologic actions of DMF. We first addressed this question in macrophages. We reproduced previous findings that both DMF and glycolytic blockade prevent classical macrophage activation (fig. S10, A to D) (*3*, *25*) and then determined that the inhibition of cytokine production by DMF was unrelated to its effects on OXPHOS (fig. S10E). We next measured the production of interleukin (IL)–1β under low (0.5 mM) or high (10 mM) glucose concentrations and found that DMF was much less effective in the presence of high glucose (Fig. 3A); this result suggests that its anti-inflammatory effect can be overcome by driving

glycolysis higher with saturating concentrations of glucose. DMF augmented the IL-4–induced expression of arginase-1 (Arg-1), a marker of M2 alternative activation (fig. S10F), but the relative importance of DMF's effects on aerobic glycolysis versus OXPHOS in promoting alternative activation was not examined.

Heptelidic acid (also known as koningic acid) is a GAPDH inhibitor that binds the active site and covalently modifies the catalytic cysteine (*23*, *26*). The treatment of mPMs with heptelidic acid replicated the effects of DMF on IL-1β secretion (Fig. 3B), inducible nitric oxide synthase (iNOS) expression (Fig. 3C and fig. S11A), and nuclear translocation of nuclear factor κB (NF-κB) (Fig.

3D and fig. S11B). Conversely, the overexpression of wild-type GAPDH, but not catalytically inactive GAPDH mutated at Cys[150], mitigated the effect of DMF on IL-1β production (Fig. 3E and fig. S12). Thus, the immunologic actions of DMF were replicated by GAPDH inhibition and reversed by increasing GAPDH expression.

We next examined the effects of DMF and MMF on lymphocyte differentiation and function. We activated mouse naïve CD4+ T cells under T_H1, T_H17, or T_{reg} cell–polarizing conditions for 4 days with or without DMF or MMF, with treatment at the start of polarization. Consistent with the known effects of glycolytic blockade (*7*, *8*), DMF had a disproportionate impact on the survival of
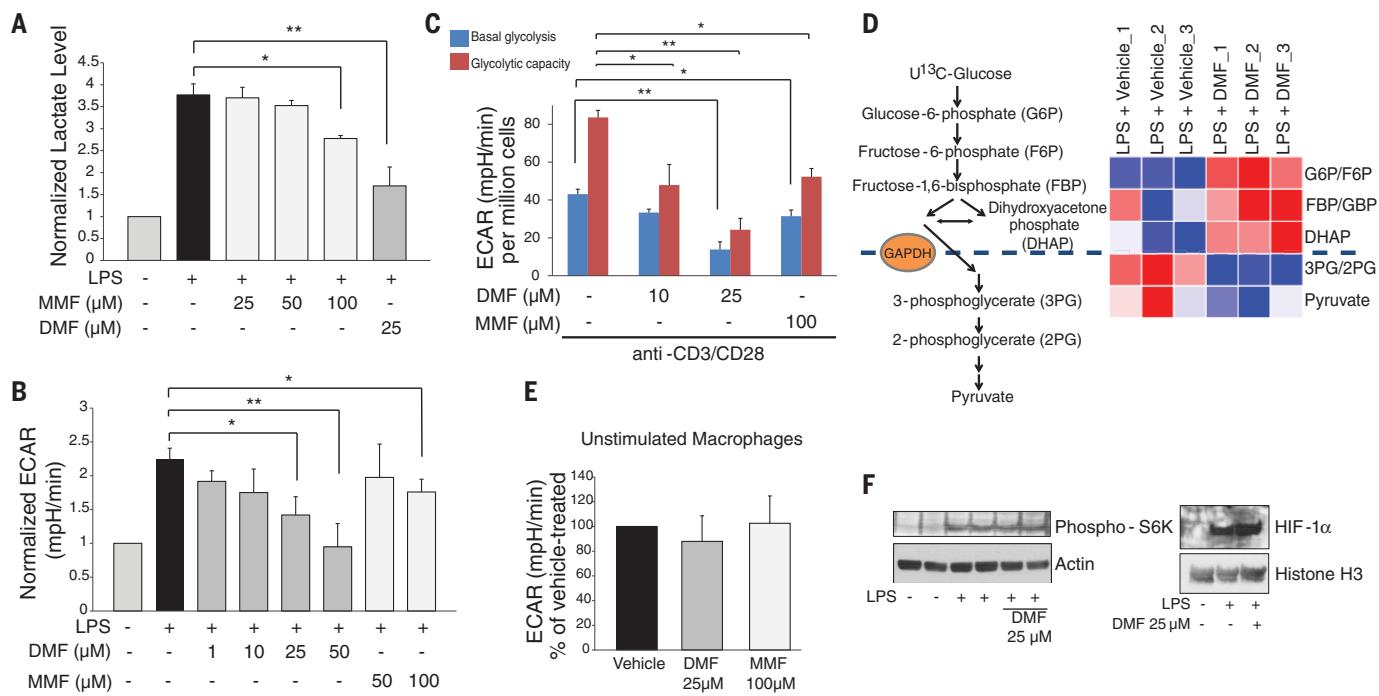
Fig. 2. GAPDH inactivation by DMF and MMF inhibits glycolysis in activated, but not resting, macrophages and lymphocytes. (**A**) mPMs were treated with LPS with or without DMF or MMF for 24 hours, followed by measurement of lactate (a proxy measure of glycolysis) in culture media by colorimetric assay. Data represent mean ± SEM of three experiments performed in duplicate. (**B**) mPMs were treated as in (A), and glycolysis was measured as extracellular acidification rate (ECAR) using a Seahorse extracellular flux analyzer. Data represent mean ± SEM of five experiments performed in quadruplicate. (**C**) Glycolysis was measured via Seahorse extracellular flux analyzer in mouse naïve CD4[+] lymphocytes activated overnight with antibodies to CD3 and CD28 with or without DMF or MMF. Data represent mean ± SEM of four

experiments performed in triplicate. (**D**) mPMs were stimulated with LPS for 24 hours with or without 25 μM DMF, in triplicate. Cells were then labeled with U[13]C-glucose, and ^{13}C labeling of glycolytic intermediates was measured from lysates via LC-MS. Heat map shows blockade of glycolytic flux at the level of GAPDH. (**E**) DMF or MMF had no effect on glycolysis in unstimulated mPMs, measured as ECAR. Data represent mean ± SEM of four experiments performed in quadruplicate. (**F**) Representative immunoblots showing no effect of DMF on phospho-S6K (a marker of mTOR activity) (N = 2 experiments performed in duplicate) or HIF-1α levels (N = 3 experiments) in LPS-stimulated mPMs. Data are quantified in fig. S8. *P < 0.05, **P < 0.01 [one-way analysis of variance (ANOVA) with Dunnett multiple-comparisons test].
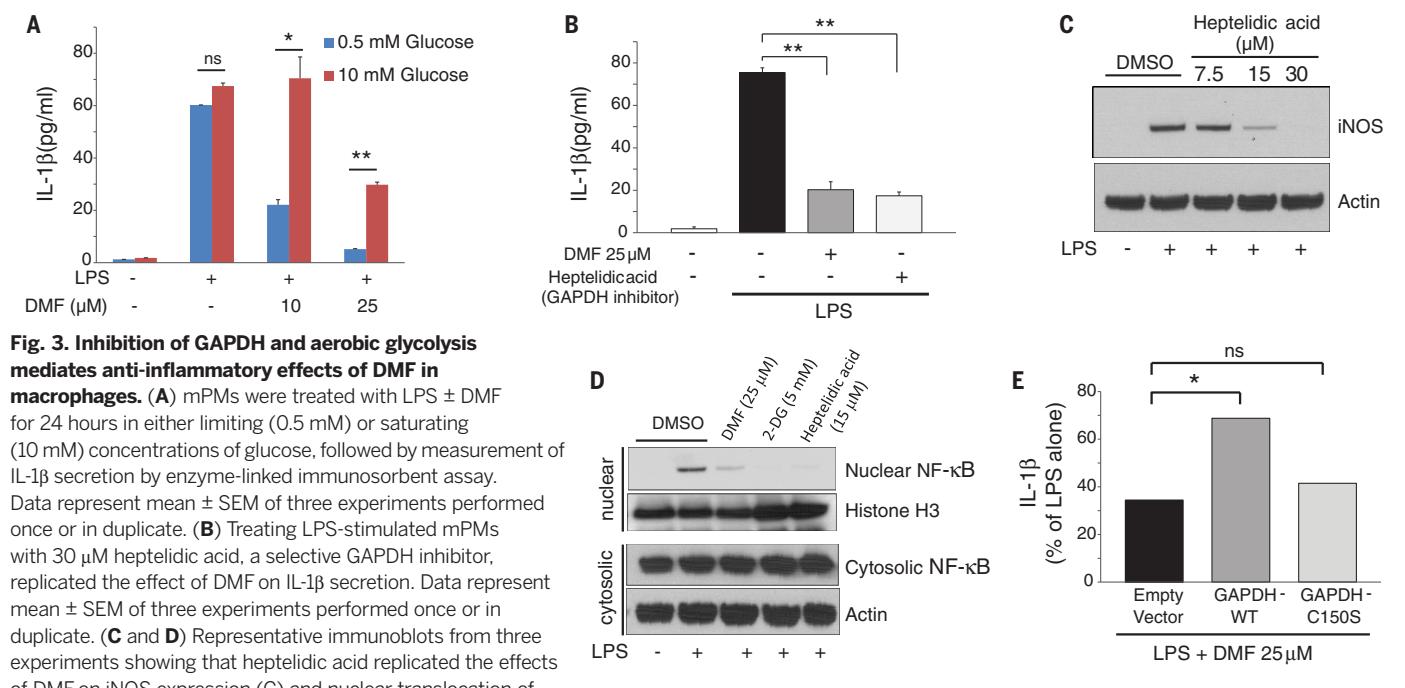
Fig. 3. Inhibition of GAPDH and aerobic glycolysis mediates anti-inflammatory effects of DMF in macrophages. (**A**) mPMs were treated with LPS ± DMF for 24 hours in either limiting (0.5 mM) or saturating (10 mM) concentrations of glucose, followed by measurement of IL-1β secretion by enzyme-linked immunosorbent assay. Data represent mean ± SEM of three experiments performed once or in duplicate. (**B**) Treating LPS-stimulated mPMs with 30 μM heptelidic acid, a selective GAPDH inhibitor, replicated the effect of DMF on IL-1β secretion. Data represent mean ± SEM of three experiments performed once or in duplicate. (**C** and **D**) Representative immunoblots from three experiments showing that heptelidic acid replicated the effects of DMF on iNOS expression (C) and nuclear translocation of

NF-κB (D) in LPS-stimulated mPMs. Data are quantified in fig. S11. DMSO, dimethyl sulfoxide. (**E**) Overexpression of wild-type GAPDH (GAPDH-WT), but not a Cys[150] → Ser mutant (GAPDH-C150S), mitigated the effect of DMF on IL-1β secretion in mPMs. Data represent the mean of two experiments performed in duplicate. *P < 0.05, **P < 0.01 [two-tailed Student t test in (A), one-way ANOVA with Dunnett multiple-comparisons test in (B) and (E)]; ns, nonsignificant.

T$_H$1 and T$_H$17 versus T$_{reg}$ cells (Fig. 4A and fig. S13A). Similarly, DMF and MMF inhibited both differentiation and cytokine production under T$_H$1- and T$_H$17-polarizing conditions, an effect replicated by heptelidic acid (Fig. 4, B to D, and fig.

S13B). As reported with HIF-1α deficiency and 2-deoxyglucose treatment (11), DMF promoted T$_{reg}$ cell differentiation under T$_{reg}$-polarizing conditions (Fig. 4E and fig. S13B), and both DMF and MMF reciprocally inhibited T$_H$17 and promoted

T$_{reg}$ cell development under T$_H$17-polarizing conditions (Fig. 4F). When added after 3 days of polarization, DMF, MMF, and low-dose heptelidic acid had no effect on the viability (fig. S14A) or differentiation (fig. S14B) of T$_H$1 or T$_H$17 cells.
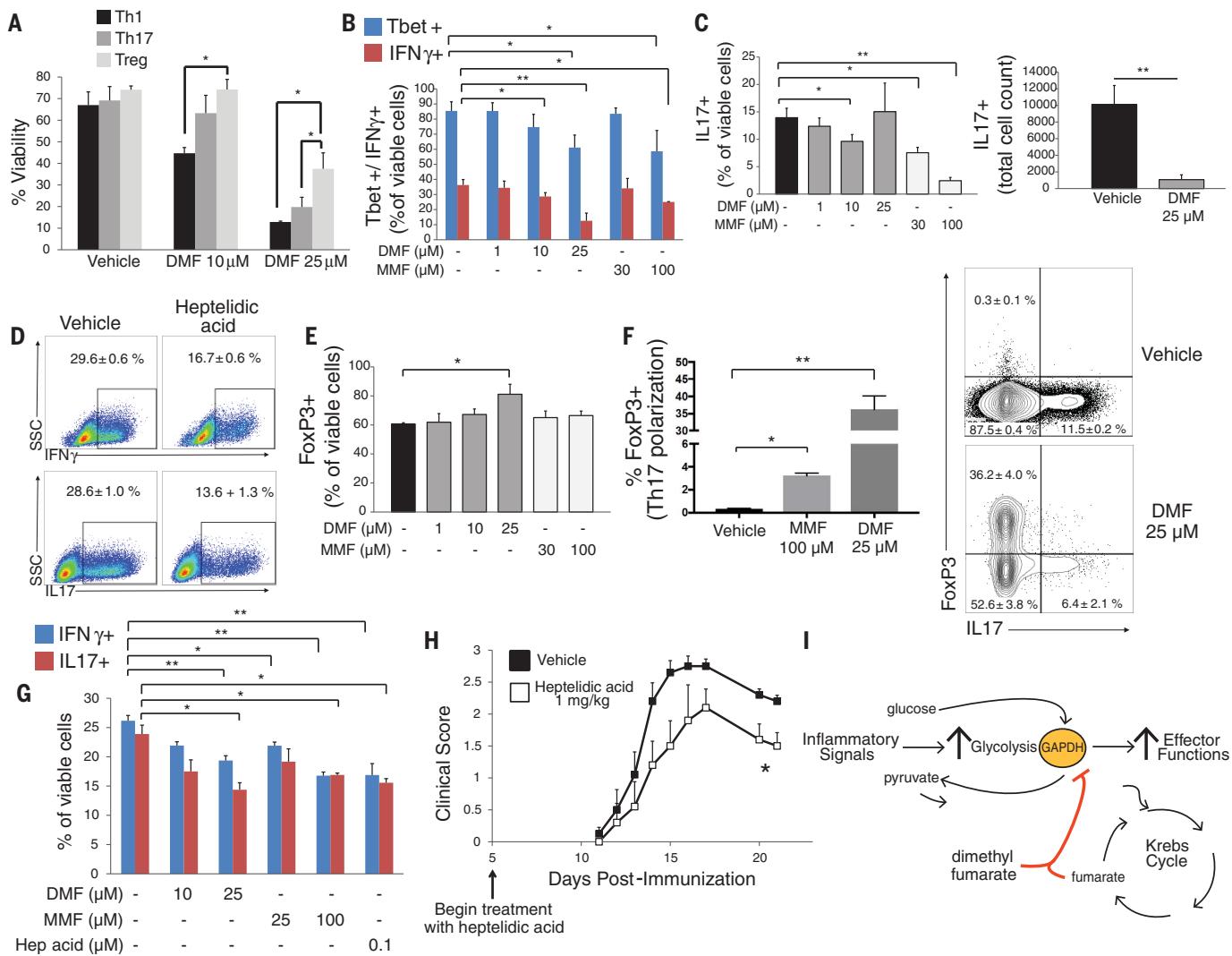
Fig. 4. Impacts of DMF and MMF on survival, differentiation, and effector function of metabolically distinct lymphocyte subsets.
(A to F) Mouse naïve CD4$^+$ lymphocytes were activated for 4 days with antibodies to CD3 and CD28 under T$_H$1 cell–, T$_H$17 cell–, or T$_{reg}$ cell–polarizing conditions. Cells were treated with the indicated doses of DMF, MMF, or heptelidic acid on day 0 and assayed by flow cytometry on day 4. (A) DMF disproportionately decreased survival under T$_H$1 and T$_H$17 cell– versus T$_{reg}$ cell–polarizing conditions, as assessed by viability staining. Data represent mean ± SEM of three experiments performed in duplicate or triplicate. DMF and MMF decreased the proportion of Tbet$^+$ and IFN-γ$^+$ cells under T$_H$1 cell–polarizing conditions (B) and of IL-17$^+$ cells under T$_H$17 cell–polarizing conditions [(C), left]. DMF (25 μM) produced variable results under T$_H$17 cell–polarizing conditions, likely due to high toxicity at that dose, but nonetheless caused a significant decrease in total IL-17$^+$ cell count [(C), right]. Data represent mean ± SEM of three experiments performed in duplicate or triplicate. (D) Representative flow cytometric plots demonstrating that low-dose (0.5 μM) heptelidic acid replicated the effect of DMF or MMF on IFN-γ and IL-17 expression under T$_H$1 cell– and T$_H$17 cell–polarizing conditions, respectively. Toxicity limited

the testing of higher doses. Values represent mean ± SEM of a triplicate experiment. (E) In contrast to effects on T$_H$1 and T$_H$17 cells, DMF increased the proportion of FoxP3$^+$ cells under T$_{reg}$ cell–polarizing conditions. Data represent mean ± SEM of three experiments performed in duplicate or triplicate. (F) DMF or MMF produced a reciprocal increase in FoxP3$^+$ cells under T$_H$17 cell–polarizing conditions. Bar graph (left) and representative flow cytometric plot (right) show results from a triplicate experiment. Data represent mean ± SEM. (G) Mouse naïve CD4$^+$ lymphocytes were activated under T$_H$1 cell– or T$_H$17 cell–polarizing conditions for 3 days and then treated overnight with the indicated drug. Expression of IFN-γ and IL-17 was then assessed by flow cytometry. Data represent mean ± SEM of a triplicate experiment. (H) Daily intraperitoneal treatment with heptelidic acid attenuated the course of EAE. Data were pooled from five mice per group and represent mean ± SEM for each time point. (I) Proposed model of immune modulation by DMF, which may exploit a physiologic negative feedback function of endogenous fumarate. *$P < 0.05$, **$P < 0.01$ [one-way ANOVA with Tukey multiple-comparisons test in (A); one-way ANOVA with Dunnett multiple-comparisons in (B), (C), (E), (F), and (G); Mann-Whitney U test in (H)].

However, all three drugs inhibited the expression of interferon-γ (IFN-γ) and IL-17 (Fig. 4G), suggesting an effect on cytokine production independent of survival and differentiation.

Because GAPDH binding to mRNA underlies posttranscriptional regulation of cytokine production (9), we tested the effect of GAPDH succination on RNA binding (fig. S15). Pretreatment with DMF or heptelidic acid decreased GAPDH-RNA binding, as reported with other modifications of the active-site cysteine (27). This effect was small, however, as a similar decrease was produced by nicotinamide adenine dinucleotide (NAD^+) at one-tenth the concentration. Thus, the alteration of such binding does not appear to underlie the immunologic actions of DMF.

Finally, to ascertain whether GAPDH inhibition produced anti-inflammatory actions in vivo, we examined the effect of heptelidic acid in experimental autoimmune encephalomyelitis (EAE), a mouse model of MS, and found that it attenuated the disease in these mice (Fig. 4H).

By demonstrating that a known immunomodulatory drug acts by inhibiting aerobic glycolysis, our findings provide a proof of concept that metabolism is a viable therapeutic target in autoimmunity. They may also explain important observations of DMF therapy in patients. DMF differentially affects distinct lymphocyte subsets, producing lymphopenia that selectively depletes highly glycolytic effector T cells while sparing oxidative naïve T cells and T_{reg} cells (28, 29). Our findings suggest that the inhibition of aerobic glycolysis underlies these selective effects. It is also notable that DMF is simply a derivative of fumarate, which is a metabolic intermediate of the Krebs cycle, lying downstream of glycolysis in cellular energy production. We hypothesize that fumarate-induced inactivation of GAPDH represents an endogenous negative feedback loop. DMF—a more cell-permeable and electro-philic derivative of fumarate—may simply exploit this physiologic pathway to produce its immunologic actions (Fig. 4I). However, additional targets of succination (in addition to KEAP1) are likely relevant to both the therapeutic and toxic effects of the drug.

REFERENCES AND NOTES

1. E. L. Pearce, E. J. Pearce, *Immunity* **38**, 633–643 (2013).
2. B. Kelly, L. A. O'Neill, *Cell Res.* **25**, 771–784 (2015).
3. G. M. Tannahill et al., *Nature* **496**, 238–242 (2013).
4. C. M. Cham, T. F. Gajewski, *J. Immunol.* **174**, 4670–4677 (2005).
5. C. M. Cham, G. Driessens, J. P. O'Keefe, T. F. Gajewski, *Eur. J. Immunol.* **38**, 2438–2450 (2008).
6. R. Wang et al., *Immunity* **35**, 871–882 (2011).
7. A. N. Macintyre et al., *Cell Metab.* **20**, 61–72 (2014).
8. V. A. Gerriets et al., *J. Clin. Invest.* **125**, 194–207 (2015).
9. C. H. Chang et al., *Cell* **153**, 1239–1251 (2013).
10. D. Vats et al., *Cell Metab.* **4**, 13–24 (2006).
11. L. Z. Shi et al., *J. Exp. Med.* **208**, 1367–1376 (2011).
12. R. A. Linker, A. Haghikia, *Ther. Adv. Chronic Dis.* **7**, 198–207 (2016).
13. R. A. Linker et al., *Brain* **134**, 678–692 (2011).
14. M. M. Blewett et al., *Sci. Signal.* **9**, rs10 (2016).
15. U. Schulze-Topphoff et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4777–4782 (2016).
16. M. Blatnik, N. Frizzell, S. R. Thorpe, J. W. Baynes, *Diabetes* **57**, 41–49 (2008).
17. N. H. Litjens et al., *Br. J. Clin. Pharmacol.* **58**, 429–432 (2004).
18. M. Rostami-Yazdi, B. Clement, T. J. Schmidt, D. Schinor, U. Mrowietz, *J. Invest. Dermatol.* **129**, 231–234 (2009).
19. H. Peng et al., *J. Neuroimmunol.* **299**, 35–44 (2016).
20. Z. D. Parsons, K. S. Gates, *Methods Enzymol.* **528**, 129–154 (2013).
21. N. Kamada, S. U. Seo, G. Y. Chen, G. Núñez, *Nat. Rev. Immunol.* **13**, 321–335 (2013).
22. A. A. Shestov et al., *eLife* **3**, 03342 (2014).
23. M. V. Liberti et al., *Cell Metab.* **26**, 648–659.e8 (2017).
24. J. Yun et al., *Science* **350**, 1391–1396 (2015).
25. M. A. Michell-Robinson et al., *Ann. Clin. Transl. Neurol.* **3**, 27–41 (2015).
26. K. Sakai, K. Hasumi, A. Endo, *Biochim. Biophys. Acta* **1077**, 192–196 (1991).
27. E. Nagy, W. F. Rigby, *J. Biol. Chem.* **270**, 2755–2763 (1995).
28. C. M. Spencer, E. C. Crabtree-Hartman, K. Lehmann-Horn, B. A. Cree, S. S. Zamvil, *Neurol. Neuroimmunol. Neuroinflamm.* **2**, e76 (2015).
29. C. C. Gross et al., *Neurol. Neuroimmunol. Neuroinflamm.* **3**, e183 (2015).

PINS

RESTORING HOPE

MODEL G102R

MODEL G102

MODEL G101A

Since the first Deep Brain Stimulation initiative of Tsinghua Univeristy in 2000, PINS Medical has gradually established a multinational corporation with headquarter based in Beijing and international business center in Singapore. As an innovative high-tech enterprise with focus on neuromodulation, a variety of clinical products have been developed to date, which include stimulators for deep brain, vagus nerve, spinal cord and sacral nerve stimulation therapies. PINS Medical devotes itself to providing cutting-edge treatments for patients who suffer from neurological disorders such as Parkinson's Disease, Epilepsy, Chronic Pain and OAB, etc.

As part of the "National Engineering Laboratory for Neuromodulation", PINS Medical works in close cooperation with Tsinghua University and the numerous affiliated clinical centers, becoming a center of attraction for a wide range of professional talents in areas of clinical research, innovative R&D and business management. Since 2008, PINS Medical has developed rapidly in becoming a leading brand in neuromodulation within the Chinese market, due to the success of its creative research platform that efficiency links basic research, R&D of novel products, clinical testing and market entry.

With an outstanding reputation as a high-tech healthcare corporation, PINS Medical has a primary mission for providing innovative, high-quality products and services for patients to improve quality of life. PINS, which stands for Programmable Implanted Neuromodulation Stimulator, is also an abbreviation of "Patient Is No.1 alwayS". This clearly presents the goal of PINS Medical for "restoring hope", not simply as an innovation company but also across society to citizens.

Looking into the future with the continuous rise in incidence of neuropsychiatric disorders and increased social burden across the globe, PINS Medical along with local governments, research centers, companies and top academic scientists, are now developing and promoting innovative therapies worldwide.

www.pinsmedical.com   info@pinsmedical.cn

Control vision with software.

Not only THINGS but even the BRAIN and CELLS connect to the Internet. For example, it is no longer a dream for blind people to regain their vision through software control. The NEW.VISION project is now on track to make it a reality. Check out our website for details.

SOFTWARE-BASED VISION RESTORATION PROJECT

NEW.VISION

HTTPS://WWW.NEWVISION-PRJ.COM/EN

JIG-SAW

Entry-Level Heating Block Starter Kit

Asynt's DrySyn Classic Starter Kit gives you a great opportunity to experience the outstanding performance of this popular laboratory heating block, which is compatible with the most commonly used round-bottomed flask sizes (50 mL, 100 mL, 250 mL, 500 mL, and 1000 mL). Also included are Asynt heat-resistant handles that enable fast, easy, and safe lifting of your heated reaction flask. Used in combination with a standard hotplate stirrer, DrySyn Classic units are widely proven to outperform the heat-conducting properties of oil baths. They pose a far-lower fire risk and eliminate the messy, time-consuming task of removing residual oil contamination from the outside of your glassware. Their solid anodized aluminum construction delivers outstanding thermal and magnetic transfer efficiency and the durability suitable for use in any laboratory environment. Designed by chemists for chemists, DrySyn Classic units allow unhindered visibility of your reaction.

Asynt
For info: +44-(0)-1638-781709
www.asynt.com/product/drysyn-classic-starter-kit

384-Well Filtration Plates

A range of 384-well filtration plates from Porvair Sciences has a unique drip geometry that provides unmatched sample recovery. By cleverly reshaping the drip director and giving it a compound curve structure, the new filtration plates virtually eliminate the traditional "hanging filtrate" effect, whereby surface tension causes excess filtrate to cling to the outside of the structure. In addition to improving total sample recovery, the new filtration-plate design helps minimize work-surface and sample contamination and improve productivity when used with automated liquid handlers. Available with four different membranes in a useful 140-µL volume, these microplates are made from high-grade polypropylene to minimize extractables. They are available with glass-fiber membranes in three different pore sizes (0.7 µm, 1.2 µm, and 5.0 µm) to enable optimum filtration for a wide range of high biomolecule-binding applications.

Porvair Sciences
For info: +44-(0)-1978-666222
www.porvair-sciences.com

Flow Cytometer

The BD LSRFortessa X-20 cell analyzer delivers high-performance, multicolor analysis with the most compact footprint in its class at 30 in. x 29 in. (76.2 cm × 73.7 cm) and a height of just 30 in. (76.2 cm). It can be configured with up to five lasers, enabling detection of up to 20 parameters simultaneously. The most common laser choices are blue, red, violet, yellow-green, and UV. Additional laser choices are available and may be configured at the time of purchase or at a later date to meet evolving application requirements. The unit is supported by a full range of BD Biosciences reagents, including the latest BD Horizon Brilliant Violet polymer conjugates. Based on the BD LSRFortessa analyzer, the BD LSRFortessa X-20 integrates the newest laser and light technologies in a greatly reduced footprint.

BD Biosciences
For info: 877-232-8995
www.bdbiosciences.com/us/go/lsrfortessa/bdlsrfortessax-20

Ozone/UV Sterilizer

The CoolCLAVE Plus sterilizer supplies 20x to 25x more sterilization power than its popular predecessor (CoolCLAVE), by using the power of combined UV and ozone cycles plus a new heating function. It is highly effective in sterilizing—or keeping sterile—contaminated pipettes, pipette tips, gloves, plates, and small instruments as well as personal items like keys, glasses, and wallets. To disinfect items, simply place them inside the CoolCLAVE Plus and press the appropriate cycle button. The simultaneous ozone gas and UV light provide a powerful dual-sanitizing effect that can eliminate more than 98% of common pathogenic organisms (bacteria, fungus, virus) in just 30 min. The antimicrobial properties of ozone also eliminate odors produced by some organisms. Best of all, the CoolCLAVE Plus is safe and efficient, uses no harsh chemicals, does not damage any surfaces, and leaves no cleaning-chemical residues.

AMS Biotechnology
For info: +44-(0)-1235-828-200
www.amsbio.com/coolclave.aspx

Plasmid DNA Purification Kits

NucleoBond Xtra Maxi and Maxi Plus are anion-exchange plasmid purification kits based on the patented NucleoBond technology. The typical yields of >1,000 µg of ultrapure plasmid DNA can be obtained in about half of the time needed with other Maxi kits that use anion-exchange chromatography. The entire plasmid purification procedure takes 80 min with NucleoBond Xtra Maxi and 34 min with NucleoBond Xtra Maxi Plus. Specially designed column filters are inserted in the NucleoBond Xtra columns, allowing a convenient, time-saving parallel clarification of bacterial lysate and sample loading onto the column—minimizing the risk of clogging. The enlarged columns lead to lower silica resin beds and enable faster flow of lysate and buffers through the columns. Optimized buffer compositions result in improved alkaline lysis and increased column flow rates. The NucleoBond Finalizer technology replaces the tedious, time-consuming centrifugation step that follows every isopropanol precipitation of diluted DNA samples.

Takara
For info: 800-662-2566
www.takarabio.com

Mycoplasma PCR Detection Kits

e-Myco, e-Myco plus, and e-Myco VALiD Mycoplasma PCR Detection Kits greatly simplify testing and detection of mycoplasma contamination in cell cultures. With PCR testing, reliable results are obtained within a few hours, since the presence of contaminant mycoplasmas can be easily and sensitively detected by verifying the amplified DNA fragments after gel electrophoresis. These kits can identify the five most common cell culture–infecting species of mycoplasma, all the way down to 10 colony-forming units (CFU)/mL (e-Myco VALiD) and 20 CFU/mL (e-Myco and e-Myco plus). They also contain an internal control (primer set and template DNA) to identify false negatives caused by faulty PCR conditions, as well as 8-methoxypsoralen (8-MOP) to prevent cross-contamination by PCR products from earlier experiments. Each tube contains all the components for PCR (lyophilized) except for templates: DNA polymerase, deoxynucleotides, PCR buffer, primer sets, 8-MOP, and internal control. Just add your templates and perform PCR.

Bulldog Bio
For info: 603-570-4248
www.bulldog-bio.com/eMyco_plus.html

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Research
OFFICIAL JOURNAL OF CAST

An Open Access Journal

Research is a comprehensive, interdisciplinary, and selective **Open Access journal from the China Association of Science and Technology (CAST)**, published quarterly and distributed by the **American Association for the Advancement of Science (AAAS)**. *Research* provides an international platform for academic exchange, research collaboration, and technological advancements. The journal will publish fundamental research in the life and physical sciences as well as important findings or issues in engineering and applied science.

The Science Partner Journals (SPJ) program was established by AAAS, the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only open access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS expands its efforts to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility and publishing expertise that AAAS is uniquely positioned **to offer as the world's largest general science membership society**.

CHANGING LIVES.

ELEVATING CAREERS.

Apply for the *Science* & SciLifeLab Prize for Young Scientists —
an annual prize awarded to early-career scientists. The prize is
presented in four categories: Cell and Molecular Biology, Genomics
and Proteomics, Ecology and Environment, and Translational Medicine.

The winners will have their essays published by *Science,* win up
to USD 30,000 and be invited to a week in Sweden to attend
the award ceremony. Get ready for a life-changing moment in
your scientific career.

SCIENCEPRIZE.SCILIFELAB.SE

Science &
SciLifeLab
PRIZE
—
For Young Scientists

Knut och Alice
Wallenbergs
Stiftelse

Science
AAAS

SciLifeLab

Hangzhou Normal University

Serving China through world-class research and education

Sponsored by

杭州师范大学110周年校庆
The 110th Anniversary Of Hangzhou Normal University

Produced by the
Science/AAAS Custom
Publishing Office

Science

AAAS

Hangzhou Normal University: Serving China through world-class research and education

Hangzhou Normal University: Serving China through world-class research and education

Driving research and education for the next century

Nestled next to the natural beauty of West Lake, a UNESCO World Heritage site, the city of Hangzhou is the most populous and popular destination in Zhejiang Province in Eastern China. About an hour southwest of Shanghai by train, Hangzhou has enjoyed remarkable economic and cultural growth in the past decade, and has been frequently listed as one of China's "happiest cities" by the English-language *China Daily* newspaper.

The city provides the backdrop for one of China's oldest universities, Hangzhou Normal University (HZNU). Established in 1908 as the Zhejiang Official Secondary Normal School, it has grown to become a respected educational institution encompassing two faculties (Education and Medical Science), 18 schools, and a teaching hospital, with over 17,000 undergraduate and 2,700 graduate students.

Despite its size and accomplishments over the past 110 years, the university is still growing, developing, and improving. With a focus on both education and research, it is constantly striving to inspire the highest-quality research from its faculty and to graduate students who are eminently qualified to join the workforce as productive and useful members of society. One example of this is Hangzhou alumnus and businessman Ma Yun (Jack Ma), cofounder and executive chairman of the Alibaba Group, one of the largest and wealthiest companies in the world.

Within this supplement, we have striven to provide a representative sampling of the exceptional research currently being conducted at HZNU. Scanning through the Table of Contents, it's clear that the range of work carried out at the university is both deep and broad, from understanding ecological variation in plants to finding new applications for nanotechnology, and from enhancing cybersecurity to developing cancer therapies. A common theme running through this booklet is that the authors are attempting to push scientific boundaries while remaining focused on asking questions and finding solutions that will benefit their society and drive progress in China.

Although it has a long history, HZNU is clearly looking forward to the future. Dedicated to training the next generation of leaders, its passion for providing a world-class education will no doubt serve it and its students well, helping to raise its profile as it moves toward greater international recognition.

Sean Sanders, Ph.D.
Jackie Oberst, Ph.D.
Science/AAAS Custom Publishing Office

Hangzhou Normal University: Over a century of extraordinary achievements

Hangzhou Normal University is located in Hangzhou, a famous historical and cultural center nestled in a picturesque natural landscape. The university's predecessor can be traced back to the Zhejiang Official Secondary Normal School, opened in 1908 as one of the six founding advanced normal schools in China. In 1978, Hangzhou Teachers College was established on the approval of the State Council of the People's Republic of China, and in 2007, the college was renamed Hangzhou Normal University. The university currently has three campuses: Cangqian, Xiasha, and Yuhuangshan, located near the beautiful West Lake, the Qiantang River, and the Xixi Wetland Park, respectively, together covering a total area of 1,980,000 square meters. The university is made up of 18 schools, two public teaching units, an affiliated hospital (Hangzhou No. 2 People's Hospital), and a state-owned, civilian-run independent college—Qianjiang College.

Boasting an exemplary history of over 100 years, Hangzhou Normal University has always attached equal importance to research and teaching, and has cultivated many generations of exceptional students. The university embodies its motto of "diligence, discretion, honesty and tolerance, erudition and enhancement" by empowering its graduates with an innovative and entrepreneurial spirit that will serve them well for the future, and by developing a comprehensive and complete education and training system across a broad range of disciplines. The university offers 71 undergraduate majors (including five national-level characteristic majors) that fall into 10 disciplines. In recent years, the university has ranked among the top in Zhejiang Province for the quality of its undergraduate education and has won six national awards for its teaching achievements. It attaches great importance to developing the entrepreneurial and

business acumen of students and is a testing ground for this type of applied education, in line with recent teaching reforms in China. It is a national standard university for the local implementation of an innovation and entrepreneurship training program, and counts some outstanding business people among its graduates, including Ma Yun (Jack Ma), executive chairman of the Alibaba Group. The university currently has over 21,000 full-time students (excluding independent colleges), over 17,000 undergraduates, more than 2,700 doctoral and Master's degree candidates, and over 1,400 international students. The university staff includes 2,200 teaching and administrative employees, of whom over 50% are full-time teachers who hold academic titles and doctoral degrees. The staff includes four Chinese Academy of Sciences and Chinese Academy of Engineering academicians, six candidates from the National Recruitment Program of Global Experts (also known as the Thousand Talents Plan), three Changjiang Scholars of the Ministry of Education, five recipients of the National Natural Science Foundation of China Young Scientist Fund, seven national candidates from the New Century Talents Project, and 11 candidates from the New Century Outstanding Talents Supporting Scheme of the National Ministry of Education.

The university takes great pride in its curriculum and continually strives to improve it, along with its reputation for scientific research. At present, it has a single doctoral training program to serve the country's special needs, two doctoral degree programs run jointly with partner universities, 19 first-level Master's programs, and six categories of professional Master's degree. It has been designated as a general institute of higher education whose new graduates can enter postgraduate courses without the requirement to take additional entrance examinations.

To date the university has had 14 disciplines included in Zhejiang Province's "Double First-Class" Plan (modeled on China's "Double First-Class" Initiative–China's largest educational development scheme, implemented in 2017 to create world-class universities and disciplines by the end of 2050), namely, eight disciplines categorized as type A (world-class discipline with a top ranking in China) and six disciplines as type B (highly ranked in China, having significant domestic impact).

Since 2007, the university's important achievements have included obtaining research grants supported by the Major Program of the National Social Science Fund of China, the Key Program of the National Natural Science Fund of China, and the National Science Fund for Distinguished Young Scholars. Furthermore, the university has created the Innovative Research Team–which received approval from the Ministry of Education–and also established the Engineering Research Center, funded by the same ministry. Additionally, it has won Outstanding Achievement awards in the Humanities and Social Sciences, also bestowed by the ministry.

At present, the university has launched one national level sci-tech business incubator; a total of 14 provincial laboratories, engineering research centers, research bases, and sci-tech innovation service platforms; and nine ministerial and provincial innovation teams. In 2017, the university was ranked 99th among domestic colleges in the Essential Science Indicators rankings and 69th among the top 100 universities in mainland China in the *Nature* Index. It was

HISTORY OF HANGZHOU NORMAL UNIVERSITY

1908: Zhejiang Official Secondary Normal School

1912: Zhejiang Provincial Secondary Normal School

1913: Zhejiang Provincial No.1 Normal School

1923: Merger of the middle school and the normal school

1931: Zhejiang Provincial Hangzhou Normal School

1938: Zhejiang Provincial Interim United Middle School, Teacher Training Department

1939: During the Second Sino-Japanese War, named Zhejiang Provincial United Normal School

1946: Zhejiang Provincial Hangzhou Normal School

1956: Hangzhou Normal School of Zhejiang Province

1978: Hangzhou Teachers College

2000: Hangzhou Normal School, Hangzhou Education College merged into Hangzhou Teachers College

2000: Hangzhou Law School merged into Hangzhou Teachers College

2000: Hangzhou Industrial Arts School merged into Hangzhou Teachers College

2001: Hangzhou Junior Medical College merged into Hangzhou Teachers College

2007: Hangzhou Normal University

listed in the *US News* Best Global Universities for the first time in 2017, ranking 1,100th in the world, 99th in China, and fourth in Zhejiang Province. The university is also ranked as one of the top in Zhejiang Province in terms of the number of approved grants from the National Natural Science Foundation and National Social Science Foundation. Additionally, the number of high-quality research papers has increased each year for three consecutive years, based on the NPI (*Nature* Publishing index) ranking of the top 100 universities in mainland China.

The university has established academic exchanges and cooperative relationships with over 30 institutions of higher learning in the United States, United Kingdom, Korea, Japan, Australia, and Taiwan, and jointly runs the Confucius Institute with Middle Tennessee State University in the United States. It is certified to enroll international students and is one of the first universities to pilot a program through which international students can receive their certification in China.

Since the initiation of the 13th Five-Year Plan (2016–2020), Hangzhou Normal University has been united in its goal to actively cater to the demands of local economic and social development while also advancing the development and improvement of the university, constantly striving to create a unique, distinctive, internationally recognized, and comprehensive educational institution.

Chilwee Group Headquarters

Chilwee Group New Energy Square

Chilwee Group:
A green energy leader

Chilwee Group was established in 1998 and listed on the main board of the Hong Kong stock exchange in 2010. Chilwee has over 100 subsidiaries worldwide and focuses on R&D and production of new energy engines, energy storage solutions, and high-power batteries, and is a leading global supplier of advanced green energy solutions. From its inception, Chilwee has pursued its mission of advocating for green energy to improve human life, aiming to become a leading company in the global new energy industry. Currently, Chilwee ranks 162nd in China's top 500

enterprises, 67th in its top 500 manufacturing companies, 283rd in *Fortune China* magazine's Top 500 Enterprises list, and sixth among China's top 100 light-industry companies. It currently ranks first in the battery industry in China.

Green energy is vitally important to Chilwee. Chilwee's first chairman, Mingming Zhou, formulated the company's mission statement, which is to "advocate for green energy and to perfect human life." The company always has this mission in mind, never forgetting its industrial, social, and environmental obligation to produce clean, green energy while keeping carbon emissions low and actively recycling waste. Additionally, Chilwee helped to found the National Environmental Protection Lead-

PHOTOS: QINGQING DING

Acid Battery Production and Recycling Pollution Prevention and Control Engineering Center, the China Battery Industry Green Recycling Development Alliance, and the first Battery Pollution Prevention and Relief Special Fund, making positive contributions to scientific progress and to the ecofriendly development of the industry, while demonstrating the highest level of corporate responsibility.

Innovation is the driving force behind Chilwee's success, creating a powerful synergy with the company's expertise in science and technology and its dedicated, passionate workers. Some of Chilwee's innovation platforms include the National Enterprise Technology Center, the National Accreditation Laboratory, the National Environmental Engineering Center, and the Provincial Key Enterprise Research Institute. The company also has 17 R&D centers spread around the world, including in the United States, Germany, Russia, and France, and has undertaken more than 80 major national and provincial

science and technology projects. Chilwee's dedication to innovation has won it the China Gold Patent Award, the China Baosteel Environmental Award, and the China Industry Recognition Award, among others—the company also holds over 1,400 patents.

The exceptional talent of Chilwee's 3,000-plus technicians is the source of inspiration for many of the company's products. In its research work, Chilwee cooperates closely with six well-known Chinese academicians and more than 60 top international experts within the new energy industry. Moreover, Chilwee employs five experts from the National Thousand Talents Program, one from the National Ten Thousand Talents Program, five from the Zhejiang Province Thousand Talents Program, and seven from the Zhejiang Province 151 Talents Program.

Innovation is the driving force behind Chilwee's success, creating a powerful synergy with the company's expertise in science and technology.

Technology exchange between scientists from the Recruitment Program of Global Experts

Technicians debugging automated production equipment

The company culture at Chilwee is unique, with its practice of harmony and cooperation, built on a foundation of generosity in the sharing of resources for mutual benefit, common development, and long-term coexistence.

Continual growth and strategic transformation is seen as critical to Chilwee's success.

This culture has allowed Chilwee to actively innovate and realize the goal of establishing a new global energy field of research based on the ideals laid out by Chairman Zhou. Through these cultural strengths, Chilwee can attract exceptional and enthusiastic workers, fuel their passion for innovation and business, and cultivate entrepreneurs who will promote the rapid growth of the enterprise. Chilwee has won recognition for its culture by being distinguished with the title of "National Civilized Unit" and "National Enterprise Culture Demonstration Base."

Continual growth and strategic transformation is seen as critical to Chilwee's success. Currently, Chilwee has a traditional business segment represented by lead-acid and lithium-ion batteries as well as emerging businesses that include sodium batteries, Plante electric vehicles, energy storage (through their Chilwee Power subsidiary), and green energy technology. In its global R&D centers and subsidiaries, Chilwee conducts projects in cooperation with well-known institutes and enterprises, such as General Electric and Stanford University in the United States, and the volabo company in Germany.

In the future, Chilwee will continue its development along two paths. The first is to advance from working on battery power to developing complete power systems, and eventually to the manufacture of a fully electric vehicle. The second is to grow from battery storage solutions to advanced storage stations and systems. Through its partnerships, Chilwee aims to create an entirely new company based on new technologies, materials, and products, evolving into a novel energy manufacturer, operator, and service provider that contributes positively to global new energy development and promotes the advancement of human life.

Mingming Zhou
Chairman, Chilwee Group

PHOTOS: QINGQING DING

Chapter 1
Chemistry and Nanotechnology
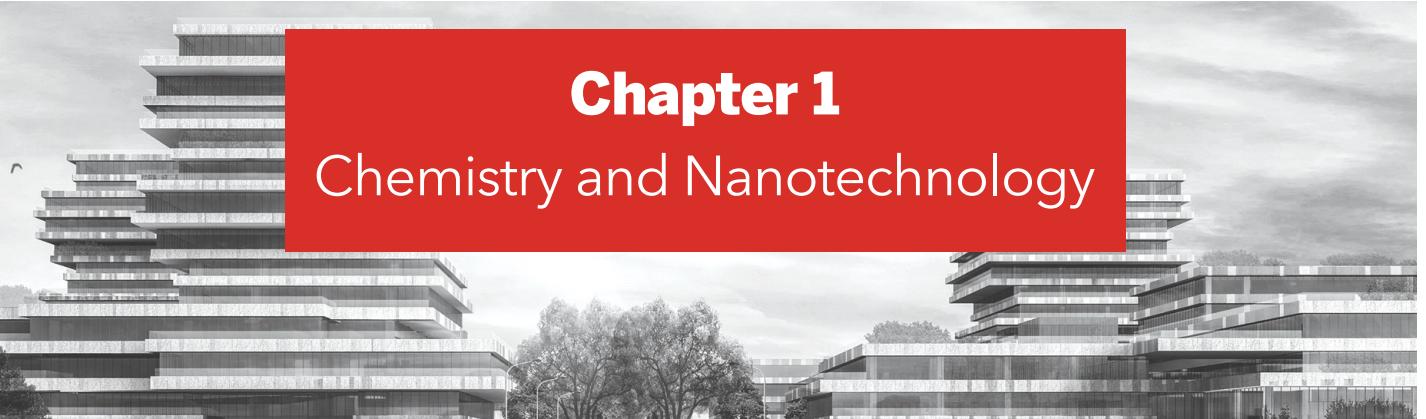
Novel multistereogenic ligand-controlled catalytic asymmetric reactions

Zheng Xu, Xing-Feng Bai, Long-Sheng Zheng, Tao Song, Guang Gao, Ji-Yuan Lv, Fei Ye, Zhan-Jiang Zheng, Li Li, Ke-Fang Yang, and Li-Wen Xu*

Recent successes in organic chemistry have led to chiral ligands playing a key role in metal-based asymmetric catalysis (*1*). Traditionally, chiral ligands can be classified into three main types based on their molecular skeleton and stereogenic centers: axial chirality (such as chiral spiro ligands and chiral biaryl ligands), planar chirality (such as metallocene-based chiral ligands), and ligands bearing only central chirality on the sp^3-carbon or sp^3-heteroatom center (*2*). In particular, biaryl-type atropisomeric ligands with axial chirality have become increasingly more important due to the excellent stability and high stereoselectivity imparted by binaphthyl-based atropisomers as ligands in numerous organic reactions (*3*). To date, a number of these ligands have been widely applied in transition metal catalysis and can be regarded as "privileged ligands" (*4*), which allow facile, efficient synthesis of optically active and structurally diverse products in high enantiopurities. However, the design, preparation, and application of novel chiral phosphorus (P)-ligands that meet the demands of highly enantioselective transformations remain formidable challenges in asymmetric catalysis. Enzymatic catalysis has motivated us to design and search for novel, synthetically practical approaches to construct structurally diverse and multifunctional backbones for the modulation of multifunctional and multicenter (MFMC) catalyst systems controlled by ligands bearing multiple stereogenic centers (i.e., multistereogenic ligands).

One of the primary challenges inherent in certain transition metal complex–based asymmetric catalysis systems is the rational design and facile construction of an effective heteroatom-containing ligand to enhance reaction efficiencies and stereoselectivities (*5*). Indeed, ligand design remains a key challenge in many catalytic asymmetric reactions (*6*). Therefore, the development of a simple, synthetically practical strategy to yield effective multistereogenic ligands is essential. In this context, the rational construction of multiple stereogenic centers that can take advantage of the electronic and steric properties of ligands could be achieved by two general strategies: chirality matching and chirality transfer. More specifically, the chirality transfer of one stereogenic center of a chiral molecule to construct new stereogenic centers is a simple, effective approach. In other words, the introduction of additional new stereogenic centers to modular ligands bearing multiple stereogenic centers can lead to the facile construction of an MFMC catalyst system that exhibits good enantioselectivity.

In the past decade, we found that the new concept of MFMC- or multistereogenic ligand–controlled catalytic asymmetric reactions has resulted in several elegant examples that display high levels of enantioselectivity and enhanced efficiencies (*1*). For example, in 2013, we designed and prepared a novel multistereogenic P-ligand called Fei-Phos, which is a chiral *trans*-1, 2-diaminocyclohexane-derived multifunctional and multidentate *N,P*-ligand bearing multiple stereogenic centers (*7*). Fei-Phos was found to exhibit excellent enantioselectivities [≤99% enantiomeric excess (*ee*)] and good yields in palladium-catalyzed allylic substitutions of benzyl alcohols and silanols with allylic acetates. This scaffold exhibits several beneficial features, including its facile accessibility, chirality matching of four sp^3-carbon stereogenic centers, and incorporation of a multifunctional diamine group onto the extremely air-stable diphosphine ligand.

In recent years, we have performed a series of investigations into the design and synthesis of novel

Key Laboratory of Organosilicon Chemistry and Material Technology of Ministry of Education, Hangzhou Normal University, Hangzhou, China
*Corresponding author: liwenxu@hznu.edu.cn

multistereogenic ligands bearing axial and sp³-carbon central chirality for the MFMC-controlled catalytic synthesis of fine chemicals, which provided several potentially useful approaches to the synthesis of chiral molecules (*8–18*). In this mini-review, we focus on our efforts to construct novel atropisomeric and multistereogenic ligands (Figure 1) bearing multiple sp³-carbon central and axial chirality for application in enantioselective carbon–carbon and carbon–heteroatom bond-forming reactions. More specifically, we focus on Ar-BINMOLs (1,1'-binaphthalene-2-α-arylmethanol-2'-ols) and their derived atropisomers bearing additional sp³-carbon central chirality (such as Ar-BINMOL-derived salan ligands) (*8–11*), Tao-Phos (*12, 13*), HZNU-Phos (*14, 15*), and Xing-Phos (*16–20*). Interestingly, our results demonstrate that the integration of multistereogenic ligands into asymmetric catalysis systems is an extremely powerful strategy.

Ar-BINMOLs and their derivatives

1,1'-binaphthyl-2,2'-diol (BINOL) and its derivatives have received significant attention in the past decades and are recognized as extremely important molecules in organic chemistry (*21, 22*). However, the development of a practical synthetic route to optically pure BINOL derivatives bearing sp³-carbon central and axial chirality is not an easy task. Thus, to develop a novel P-ligand bearing axial chirality and sp³-carbon central chirality through the modification of BINOL, we successfully explored the neighboring lithium-assisted [1,2]-Wittig rearrangement (NLAWR) strategy for the synthesis of optically pure Ar-BINMOLs via axial-to-central chirality transfer (*8*). Ar-BINMOLs and their derivatives, including the Ar-NNPs ((1-(2-arylnaphthalen-1-yl)naphthalen-2-yl)diphenylphosphines) (*3*), played an important role in several reactions, including the 1,2-addition of organometallic reagents to aldehydes or ketones (*1, 9*), providing the desired products in excellent enantioselectivities. Other groups have also utilized our

Ar-BINMOL ligands to promote several catalytic asymmetric transformations (*23*).

Considering the importance of the salen and salan coordination ligands in asymmetric catalysis, we selected the chiral Ar-BINMOLs as a starting material to prepare structurally diverse Ar-BINMOL–derived salen and salan ligands (Figure 2). We found that the corresponding salan–Cu complexes could be used to discriminate between the different reactivities of non- or halide-substituted aromatic aldehydes compared to other aldehydes (e.g., aliphatic aldehydes) in the Henry reaction (*10*). In addition, the same salan–Cu complex was found to be an efficient catalyst in the enantioselective fluorination of β-keto amides (*11*). It should also be noted that both the axial chirality and the benzyl moiety were crucial for achieving a high enantioselectivity (>99% *ee*).

As the Ar-BINMOLs exhibit a unique, attractive structure containing multiple stereogenic centers, we synthesized a series of chiral Ar-BINMOL–derived P-ligands (Ar-BINMOL-Phos) using our NLAWR strategy (*12*). Indeed, the synthesis of chiral Ar-BINMOL–based P-ligands with sp³-carbon central and axial chirality was achieved in good yields and stereoselectivities [>99% diastereomeric excess (*de*)]. As expected, both the phosphine moiety and the secondary benzyl alcohol on the sp³-carbon stereogenic center were key to improving the catalytic activity of the Ar-BINMOL-
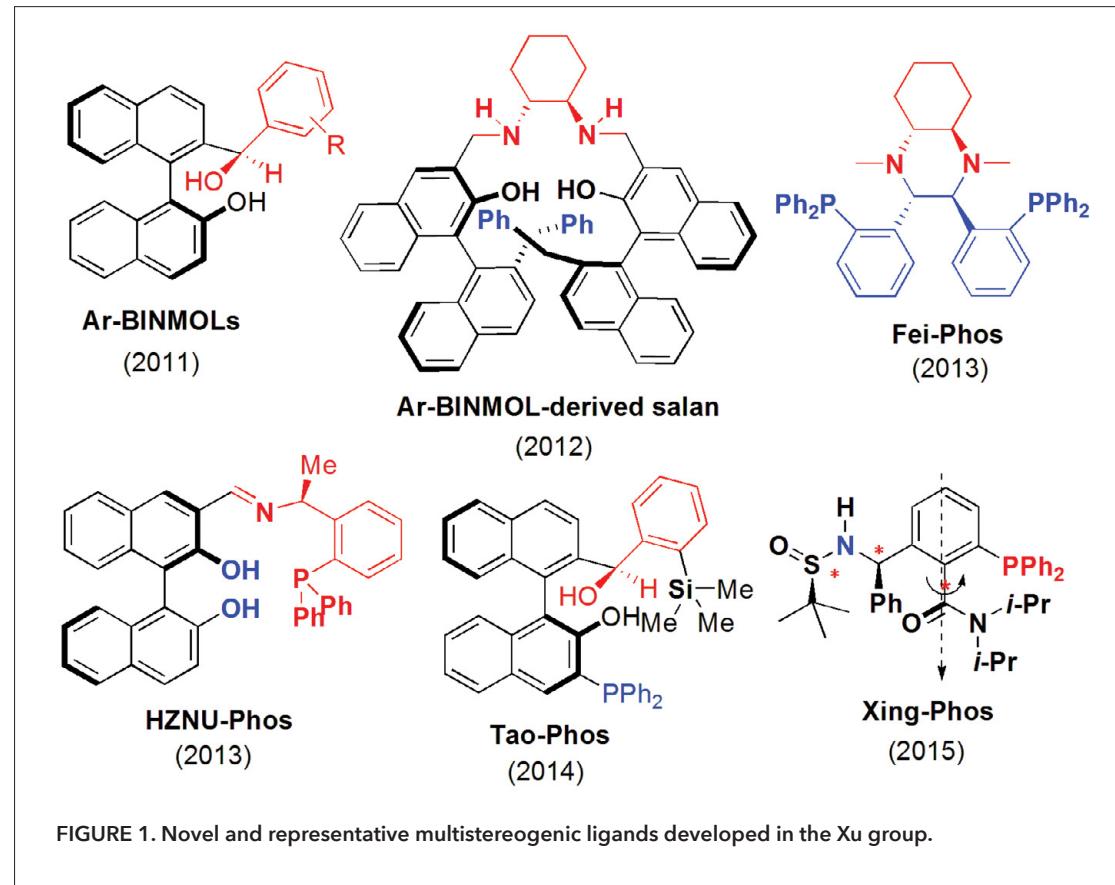
FIGURE 1. Novel and representative multistereogenic ligands developed in the Xu group.

Ar-BINMOLs (2011)

Ar-BINMOL-derived salan (2012)

Fei-Phos (2013)

HZNU-Phos (2013)

Tao-Phos (2014)

Xing-Phos (2015)
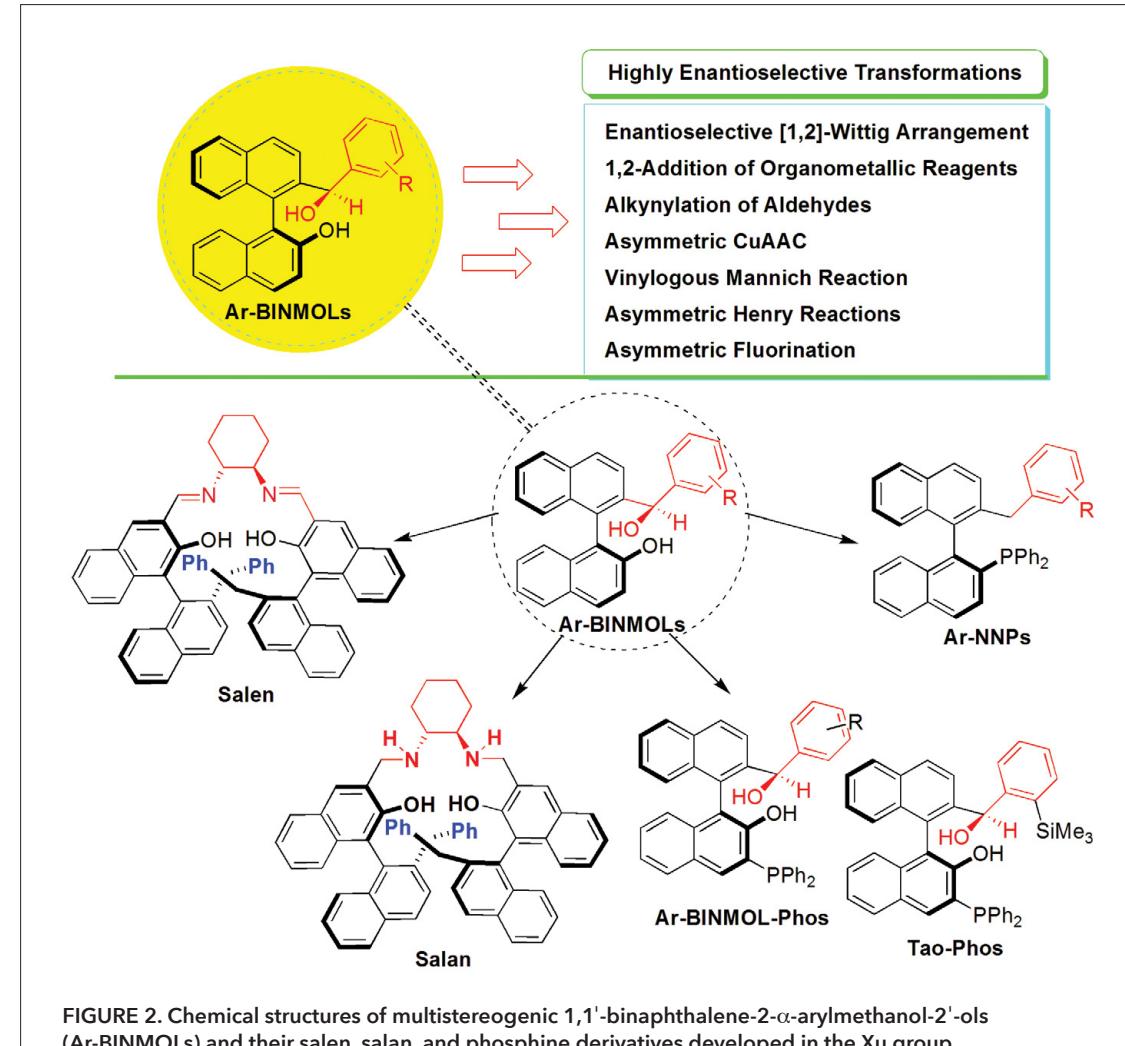
FIGURE 2. Chemical structures of multistereogenic 1,1'-binaphthalene-2-α-arylmethanol-2'-ols (Ar-BINMOLs) and their salen, salan, and phosphine derivatives developed in the Xu group.

Highly Enantioselective Transformations

Enantioselective [1,2]-Wittig Arrangement
1,2-Addition of Organometallic Reagents
Alkynylation of Aldehydes
Asymmetric CuAAC
Vinylogous Mannich Reaction
Asymmetric Henry Reactions
Asymmetric Fluorination

Ar-BINMOLs — Salen — Salan — Ar-NNPs — Ar-BINMOL-Phos — Tao-Phos

Phos–controlled alkynylation of aldehydes with alkynes. Notably, the combination of Ar-BINMOL-Phos with ZnMe₂ predominantly afforded the *S*-configured products (<94% yield, <70% *ee* of propargyl alcohols), while the addition of catalytic amounts of CaH₂ and *n*-BuLi gave the *R*-configured products (>99% *ee*) in high yields (*12*). In addition, the novel Ar-BINMOL–derived phosphine ligand (simplified as Tao-Phos) bearing a bulky silicon-based group was found to be highly efficient in the asymmetric copper-catalyzed Huisgen [3+2] cycloaddition of succinimide-based bisalkynes and azides (*13*). Among the chiral ligands evaluated in our studies, only Tao-Phos generated the corresponding succinimide-based triazole products in moderate to good yields (<80%) and satisfactory stereoselectivities (85:15 to >99:1 enantiomeric ratio). Therefore, the catalytic activities of the Ar-BINMOLs and their Ar-BINMOL-Phos derivatives in several organic transformations suggested the potential of ligands bearing multiple stereogenic centers and their

corresponding MFMC catalyst systems. Furthermore, it is possible that novel ligands based on the modification of the Ar-BINMOL scaffold can be easily tuned in both an electronic and a steric context due to its axial and sp³ central chirality.

HZNU-Phos

Successful application of the Ar-BINMOLs and their salan and phosphine derivatives bearing multiple stereogenic centers prompted us to develop a structurally novel, highly efficient phosphine ligand through the modulation of C₂-axial and sp³-carbon central chirality on a single P-ligand. We hypothesized that the BINOL-derived *N,P*-ligand bearing both axial and sp³ central chirality would be an ideal *N,O,P*-ligand for the formation of an MFMC catalyst system using a range of transition metals. In 2013, we reported the successful preparation of BINOL-derived *N,P*-ligand (HZNU-Phos) bearing a chiral amino moiety in addition to a phosphine group containing multiple stereogenic centers (Figure 3). More specifically, (*R,S*)-HZNU-Phos could be obtained easily from an (*S*)-1-phenylethanamine-derived phosphine and an (*R*)-BINOL-derived aldehyde. The novel BINOL-linked phosphine-containing chiral Schiff base could then be employed to form multinuclear and multifunctional transition metal catalysts. Indeed, HZNU-Phos is one of the best performing ligands in the copper-catalyzed 1,4-conjugate addition of the organozinc reagent Et₂Zn to enones (>99% *ee*). The resulting MFMC-type Cu–Zn complex formed in situ with HZNU-Phos was found to be an excellent bimetallic catalyst (*14*). In addition, this catalyst system exhibited an excellent catalytic performance in the 1,4-conjugate addition reaction,

FIGURE 3. Application of the HZNU-Phos ligand in catalytic asymmetric conjugate addition and allylic alkylation reactions.

1,4-Conjugate addition

Asymmetric allylic alkylation

HZNU-Phos
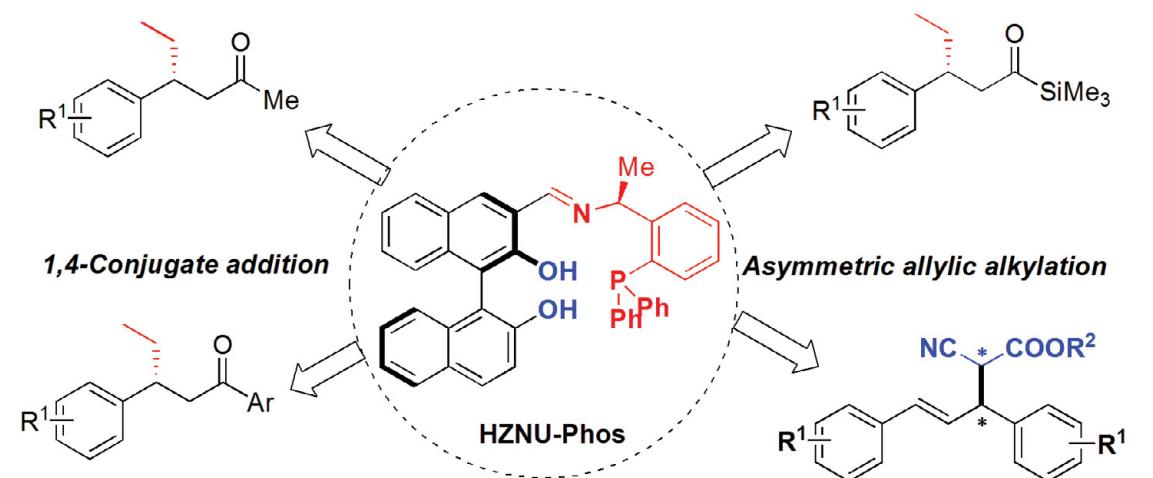
FIGURE 4. Application of the Xing-Phos ligand in the catalytic asymmetric synthesis of various N-heterocycles bearing at least two stereogenic centers.

Xing-Phos
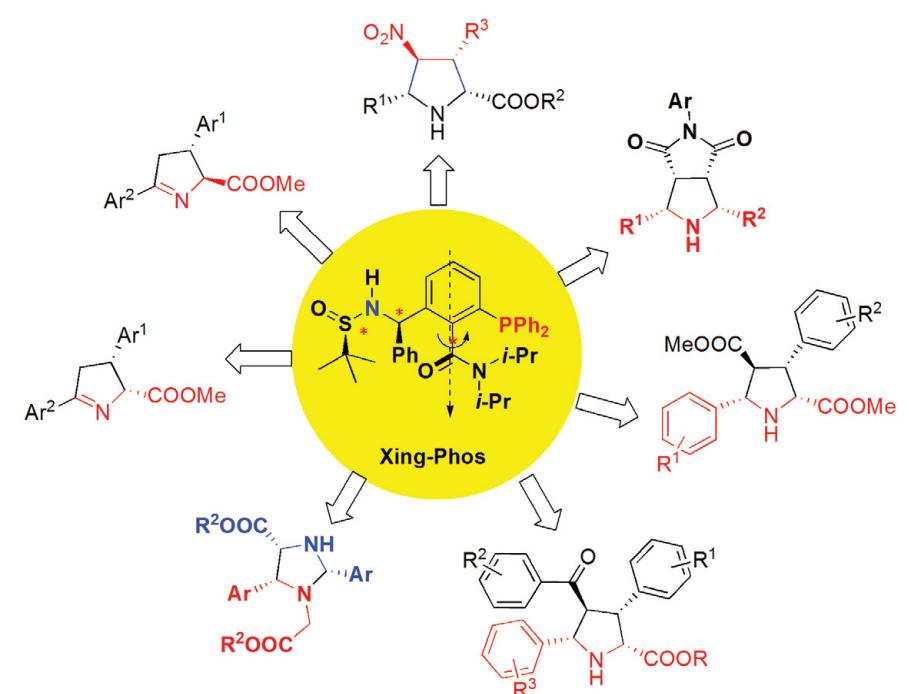
with a high enantioselectivity achieved in the presence of 0.05 mol% catalyst (turnover number = 17,600). Furthermore, the BINOL-derived HZNU-Phos ligand bearing multiple stereogenic centers also promoted a dramatic acceleration in the enantioselective allylic substitution of simple alkyl 2-cyanoacetates, generating the corresponding substituted alkyl 2-cyanoacetate products at excellent yields and with good stereoselectivities (15).These results confirmed that the HZNU-Phos ligand with matched chirality is particularly attractive as it allows the formation of a stereoselectively beneficial and substrate-sensitive

metal-centered MFMC catalyst system with multiple stereogenic centers.

Xing-Phos

The Ar-BINMOL and HZNU-Phos ligands described above are biaryl atropisomers. In contrast, the construction and application of nonbiaryl atropisomers is challenging because of their poor thermal stabilities, which results in racemization via axial rotation at high temperatures. As described in previous reports, although many chiral P-ligands have been employed in various catalytic asymmetric reactions, in 2015, we reported a novel type

of aromatic amide–derived atropisomer bearing multiple stereogenic centers with axial and S- and C-based central chirality (Xing-Phos ligands) (16) for use in asymmetric silver-catalyzed [3+2] cycloaddition reactions (Figure 4). This constitutes the first example of the catalytic application of aromatic amide–derived nonbiaryl atropisomeric P,O-ligands in highly enantioselective cycloaddition reactions. In addition, the catalytic asymmetric dipolar [3+2] cycloaddition of aldiminoesters with nitroalkenes is a particularly efficient strategy for the facile synthesis of nitro-substituted pyrrolidines in high stereoselectivities [>99:1 dr (diastereomeric ratio), 95–99% ee]. Notably, the effect of different functional groups or chirality (e.g., the S(=O)-NH group, the carbon stereogenic center, and the amide stereochemistry) on the catalytic activity and enantioselectivity of Xing-Phos ligands revealed the potential of multistereogenic ligands in this reaction. We subsequently found that the novel silver (Ag)/Xing-Phos catalyst system is particularly effective in the catalytic asymmetric dipolar [3+2] cycloaddition of various alkenes and glycine aldimino esters, including maleimides (17), chalcones, and the less-reactive methyl cinnamates (18), to yield the corresponding pyrrolidines bearing four carbon stereogenic centers (<98% ee) and in high diastereoselectivities (>98:2 dr). Remarkably, this constitutes the first example of such a high enantioselectivity in the silver-catalyzed dipolar [3+2] cycloaddition of methyl cinnamates and glycine aldimino esters (18).

Recently, we found that the Ag/Xing-Phos-catalyzed asymmetric dipolar [3+2] cycloaddition was a Michael addition–initiated intramolecular cyclization reaction, especially in the case of the dipolar [3+2] cycloaddition of enones with aldimino esters. Fortunately, the detection of trace amounts of Michael adduct side products in this dipolar [3+2] cycloaddition reaction inspired us to design the stereoselective Ag/Xing-Phos-catalyzed creation of multisubstituted and stereodivergent Δ(1)-pyrrolines. As expected, this novel Michael addition reaction allowed the stereoselective synthesis of chiral pyrrolines from simple enones and glycine. Furthermore, a series of cis-isomers of Δ(1)-pyrroline derivatives was prepared in excellent chemo- and stereoselectivities (<98% ee) (19), and interestingly, different workup procedures allowed the one-pot synthesis of trans isomers of the chiral Δ(1)-pyrroline derivatives (<98% ee).

Moreover, we recently described an enantioselective homo-1,3-dipolar [3+2] cycloaddition of glycine aldimino esters. Among the phosphine ligands evaluated in this work, Xing-Phos is the only effective P-ligand in the silver-catalyzed asymmetric homo-1,3-dipolar [3+2] cycloaddition reaction. Indeed, in this new reaction, a variety of chiral imidazolidines displayed high enantioselectivities (<99% ee) and diastereoselectivities (<96:4 dr) (20). Notably, the silver/Xing-Phos system was found to be a highly effective catalyst for the polymerization of glycine aldimino ester monomers, with the corresponding polymers bearing imidazolidine moieties on their main chains being easily obtained in high yields

(91%–96%) and with a narrow molecular weight distribution (20). For example, the silicon-tethered poly(imidazolidine) was easily prepared with a particularly narrow molecular weight (M_w) distribution, where M_w/M_n = 1.08 and M_n (number average molar mass) >33,000.

We could therefore conclude that the successful use of multistereogenic Xing-Phos in asymmetric catalysis indicates the potential of nonbiaryl and atropisomeric phosphines to behave as chiral ligands. We therefore believe that further research into Xing-Phos and the other multistereogenic ligands described here would be beneficial to the development of multistereogenic, ligand-oriented or MFMC-controlled asymmetric catalysis systems in addition to highly efficient organic transformations with excellent enantioselectivities and high yields, which will in turn allow new and synthetically useful approaches to be established for the asymmetric synthesis of optically pure molecules. The design and synthesis of a series of novel multistereogenic ligands will soon be evaluated to determine the efficiency and enantioselectivity of catalytic asymmetric reactions using the MFMC system paradigm.

References
1. Z. Xu, L.-W. Xu, *Chem. Rec.* **15**, 925–948 (2015).
2. L. W. Xu *et al.*, *Chin. J. Org. Chem.* **23**, 919–932 (2003).
3. L. S. Zheng *et al.*, *Tetrahedron* **69**, 8777–8784 (2013).
4. Q. L. Zhou, Ed., *Privileged Chiral Ligands and Catalysts* (Wiley-VCH, Singapore, 2011).
5. E. N. Jacobsen, A. Pfaltz, H. Yamamoto, Eds., *Comprehensive Asymmetric Catalysis*; Vol. I–III, Suppl. I–II (Springer, New York, 1999).
6. S. L. Shi *et al.*, *J. Am. Chem. Soc.* **132**, 19, 6638–6639 (2010).
7. F. Ye *et al.*, *Chem. Eur. J.* **19**, 15452–15457 (2013).
8. G. Gao *et al.*, *Chem. Eur. J.* **17**, 2698–2703 (2011).
9. L. S. Zheng *et al.*, *Eur. J. Org. Chem.* **2013**, 748–755 (2013).
10. F. Li *et al.*, *Chem. Asian J.* **7**, 2008–2013 (2012).
11. L. S. Zheng *et al.*, *Adv. Synth. Catal.* **356**, 3769–3776 (2014).
12. T. Song *et al.*, *Adv. Synth. Catal.* **356**, 1708–1718 (2014).
13. T. Song *et al.*, *Chem. Eur. J.* **21**, 554–558 (2015).
14. F. Ye *et al.*, *Chem. Asian J.* **8**, 2242–2253 (2013).
15. W. H. Deng *et al.*, *ChemCatChem* **7**, 75–79 (2015).
16. X. F. Bai *et al.*, *Angew. Chem. Int. Ed.* **54**, 5255–5259 (2015).
17. X. F. Bai *et al.*, *Tetrahedron* **72**, 2690–2699 (2016).
18. X. F. Bai *et al.*, *ACS Catal.* **5**, 6016–6020 (2015).
19. X. F. Bai *et al.*, *Chem. Eur. J.* **22**, 10399–10404 (2016).
20. B. Yu *et al.*, *Adv. Synth. Catal.* **359**, 3577–3584 (2017).
21. J. M. Brunel, *Chem. Rev.* **105**, 857–897 (2015).
22. J. M. Brunel, *Chem. Rev.* **107**, PR1–PR45 (2007).
23. J. F. Collados *et al.*, *ACS Catal.* **6**, 1952–1970 (2016).

Self-assembly directed and regulated by metal–ligand coordination

Shijun Li*, Zibin Zhang, Ben-Yue Zhu, Shu-Ping Wang, Bin Zhu, Yang Ye, and Jing Wu

Self-assembly is ubiquitous in nature. Indeed, relatively simple molecules can assemble into extremely complicated biomolecules, which is inspirational for chemists seeking to construct artificial chemical systems as complex as those found in nature. Supramolecular assembly, via reversible covalent bonding or through noncovalent interactions, is an important strategy for mimicking natural systems. Noncovalent interactions include metal coordination, hydrogen bonding, van der Waals forces, π–π stacking, and hydrophobic and electrostatic interactions (1). Metal coordination has distinct advantages, including strong bond energies, and optimal directionalities, multivalencies, and multiplicities (2).

Coordination-driven self-assembly—self-assembly via metal coordination interactions—is a powerful method for fabricating complex supramolecular entities. Many discrete supramolecular coordination complexes (SCCs) with well-defined sizes and shapes have been constructed through the combinatorial use of transition metals and rigid or flexible organic ligands (3, 4). The rapid appearance of new methods and new architectures has encouraged the use of SCCs in host-guest chemistry, chemical sensors, catalysis, new materials, and biological systems (2–5). During the past few years, our group has focused mainly on cooperative self-assembly by employing the orthogonality of metal–ligand coordination and other recognition motifs for the construction of supramolecules with impressive architectures and specific functions.

### Construction of mechanically interlocked molecules

Mechanically interlocked molecules (MIMs), such as rotaxanes and catenanes,
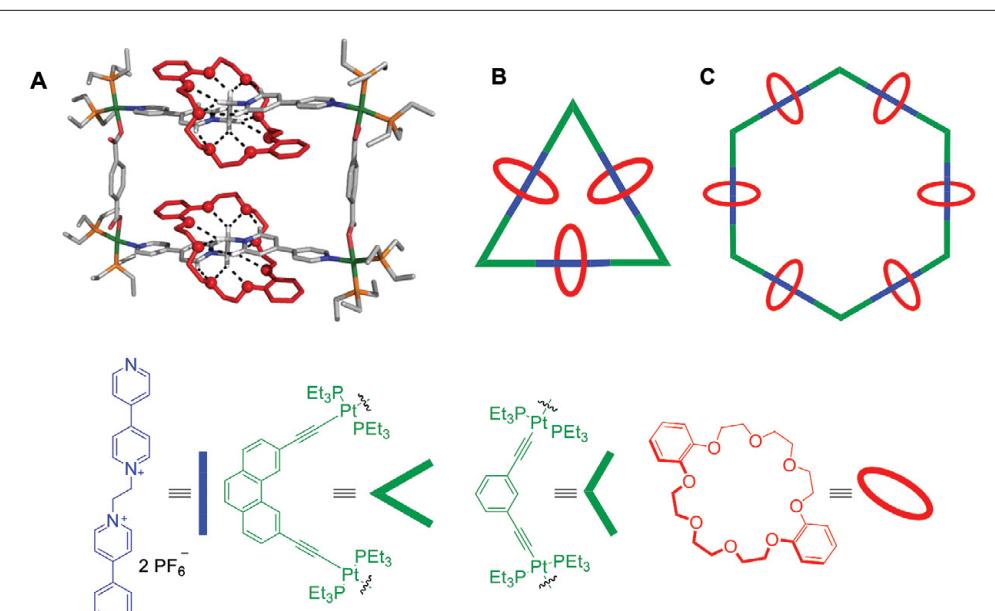
College of Material, Chemistry and Chemical Engineering, Hangzhou Normal University, Hangzhou, China
*Corresponding author: l_shijun@hznu.edu.cn

have drawn much attention because they are topologically important and have great potential in diverse applications, including molecular machines, molecular devices, smart materials, and biomaterials (6, 7). Although various template-directed methods have been exploited for the syntheses of MIMs, the construction of complicated higher-order MIMs remains challenging. To address this, we unified the coordination-driven self-assembly and host–guest chemistry themes using DB24C8/1,2-bis(pyridinium)ethane (DB24C8 = dibenzo[24]crown-8), a well-understood recognition motif, for the fabrication of a series of intricate MIMs (Figure 1). Several doubly threaded [3]catenanes (Figure 1A) were prepared using 10 total building blocks from four unique species by a dynamic decomplexation–threading–recomplexation protocol (8). These [3]catenanes were fabricated by first assembling a pseudorotaxane, followed by the addition of a 90° platinum(II) acceptor and a dicarboxylate ligand. The dynamic catenations of the [3]catenanes were further shown by the successive addition of KPF_6 and 18-crown-6; DB24C8 slipped reversibly in and out of the lock through a rectangular metallacycle in a controllable manner.

In another report, two complicated catenanes were efficiently fabricated using a similar hierarchical-assembly strategy and the DB24C8/1,2-bis(pyridinium) ethane recognition motif (9). Triangular and hexagonal metallacycles were initially assembled by reacting a pyridyl-derived 1,2-bis(pyridinium)ethane salt with 60° and 120° platinum(II) acceptors, respectively. These large metallacycles were then threaded by DB24C8 via host–guest interactions without disrupting the initial

FIGURE 1. Structures of catenanes formed by coordination-driven self-assembly. (A) Crystal structure of a self-assembled [3]catenane. **(B)** Representative structure of the [4]molecular necklace. **(C)** Representative structure of the [7]molecular necklace.
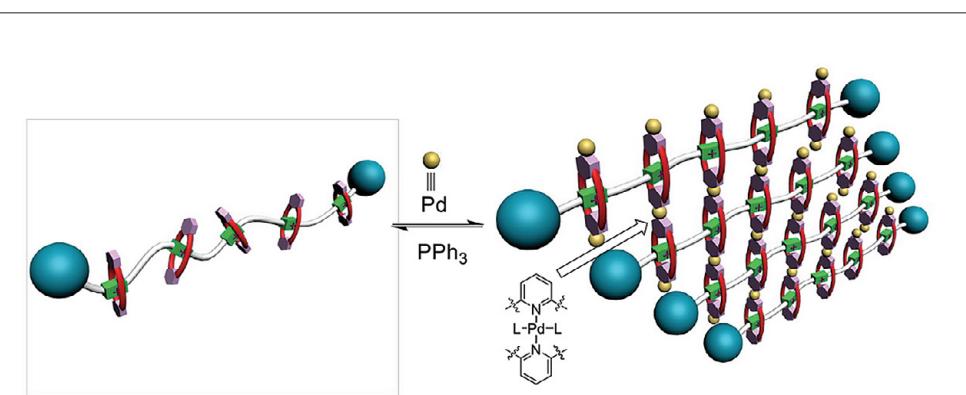
FIGURE 2. Transition between a linear and crosslinked polyrotaxane controlled by reversible palladium (Pd) coordination and decoordination.
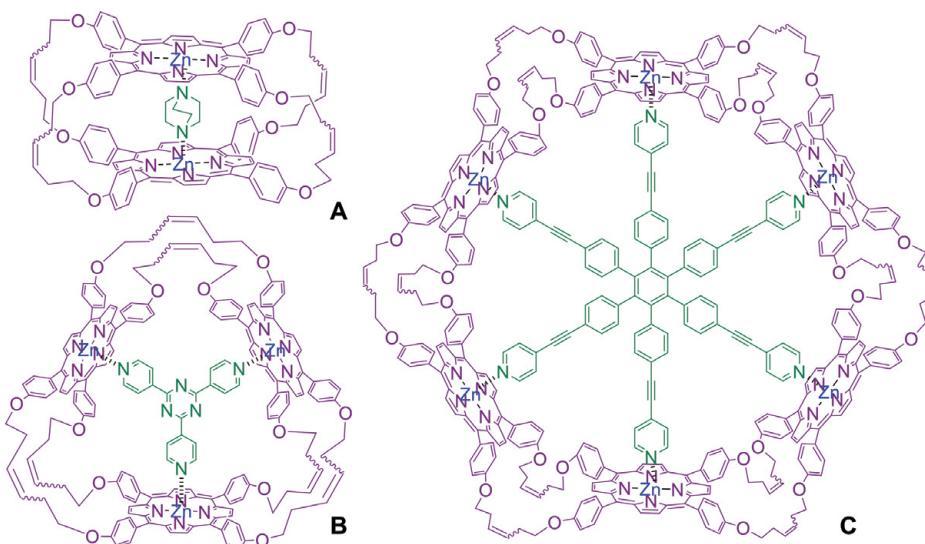
FIGURE 3. Structures of a template-coordinated porphyrin nanocage and porphyrin nanorings. (A) Nanocage obtained using 1,4-diazabicyclo[2.2.2]octane template. **(B)** Triporphyrin nanoring derived from tripyridyl template. **(C)** Hexporphyrin nanoring resulting from hexapyridyl template.

metal–ligand coordination, furnishing [4]- (Figure 1B) and [7]- molecular (Figure 1C) necklaces. In the case of the [7] molecular necklace, up to 18 individual components from three unique species were quantitatively assembled to deliver the final topologically complex supramolecule.

This hierarchical-assembly strategy was further extended to the fabrication of MIMs using the cryptand/paraquat recognition motif (10). With a paraquat-derived carboxylate ligand as the building block, supramolecular rectangles and triangles were first constructed through coordination-driven self-assembly. Two [3]catenanes and a [4]molecular necklace were then formed by the addition of a bis(m-phenylene)-32-crown-10-based cryptand. Because this cryptand exhibits significantly higher binding constants for complexations with paraquat derivatives than simple crown

ethers, the cryptand/paraquat recognition motif provided a more powerful impetus for MIM self-assembly. The addition of a slight excess of host was sufficient to complete the threading process. These hierarchical self-assembly methods greatly facilitate the construction of MIMs with high topological complexities from simple building blocks, thereby avoiding onerous traditional synthetic procedures.

Construction and regulation of supramolecular polymers

Supramolecular polymers, whose monomeric units are held together through reversible covalent bonds or noncovalent interactions, have distinct advantages over traditional polymers due to their dynamic properties (11). Metal–ligand coordination has been shown to be highly effective for directing supramolecular-polymer construction. Poly[2]pseudorotaxanes and a crosslinked supramolecular polymer were previously constructed using the orthogonality of Pt(II)-based metal–ligand coordination and cryptand-based molecular recognition (12–14). Furthermore, metal–ligand coordination facilitates the regulation of supramolecular polymers. A new main-chain–type polyrotaxane, a class of mechanically interlocked supramolecular polymer, was fabricated through one-pot olefin metathesis polymerization using the DP24C8/dialkylammonium recognition motif (15). The pyridyl groups of DP24C8 were then exploited to achieve a reversible transition between the linear and crosslinked polyrotaxane through palladium coordination and decoordination (Figure 2). In these studies, metal–ligand coordination provided a convenient tool for controlling topological structures and realizing stimuli-responsive smart materials.

Template-directed synthesis of a porphyrin nanocage and porphyrin nanorings

The coordination bonds between transitional metals and nitrogen ligands have also been used as templates for directing the syntheses of porphyrin nanocages and

nanorings (*16, 17*). A defined amount of a tetraalkene-derived zinc-porphyrin monomer was first self-assembled through nitrogen coordination involving multiple pyridyl ligands. The subsequent olefin metathesis reaction was triggered by the addition of Grubbs' catalyst, with exchanges of the terminal alkenes on the porphyrin monomers producing the cyclic porphyrin oligomer. A nanocage (Figure 3A) was obtained when 1,4-diazabicyclo[2.2.2]octane was used as the template, while triporphyrin (Figure 3B) and hexaporphyrin (Figure 3C) nanorings were furnished using a tripyridyl or hexapyridyl template, respectively (*17*). The empty triporphyrin and hexaporphyrin nanorings were applied as hosts for the complexations of fullerenes, with high binding constants.

FIGURE 4. Structures of chiral metallacycles and metallacages. (A and B) Rhomboids. (C) Squares. (D and E) Tetragonal prisms. (F) A hexagonal prism.

Chiral self-assembly from a BINOL-derived ligand

The modularity and efficiency of coordination-driven self-assembly has made it a powerful strategy for the fabrication of chiral supramolecular entities (*18*). We demonstrated the efficient construction of chiral metallacycles and chiral metallacages by using enantiomerically pure (*S*)- or (*R*)-BINOL-derived dicarboxylates as donors that possess adaptive coordination angles. Two rhomboids (Figures 4A and 4B), two squares (Figure 4C), two tetragonal prisms (Figures 4D and 4E), and a hexagonal prism (Figure 4F) were produced when the chiral BINOL-derived dicarboxylates were assembled with 180°, 120°, and 90° Pt(II)-based acceptors, in which the BINOL-derived donors adapted their directionalities to match the requirements of each architecture (Figure 4).

Conclusions

Our studies not only provide convenient pathways to new supramolecular architectures through metal–ligand coordination, but also offer an enhanced understanding of selective self-assembly. Recently, we explored the applications of these self-assemblies in host–guest chemistry, chiral recognition, drug delivery, and supramolecular catalysis (*19*).

References
1. V. Marcos *et al.*, *Science* **352**, 1555–1559 (2016).
2. T. R. Cook, Y.-R. Zheng, P. J. Stang, *Chem. Rev.* **113**, 734–777 (2013).
3. R. Chakrabarty, P. S. Mukherjee, P. J. Stang, *Chem. Rev.* **111**, 6810–6918 (2011).
4. T. R. Cook, P. J. Stang, *Chem. Rev.* **115**, 7001–7045 (2015).
5. T. R. Cook, V. Vajpayee, M. H. Lee, P. J. Stang, K.-W. Chi, *Acc. Chem. Res.* **46**, 2464–2474 (2013).
6. M. Xue, Y. Yang, X. Chi, X. Yan, F. Huang, *Chem. Rev.* **115**, 7398–7501 (2015).
7. G. Gil-Ramírez, D. A. Leigh, A. J. Stephens, *Angew. Chem. Int. Ed.* **54**, 6110–6150 (2015).
8. S. Li *et al.*, *J. Am. Chem. Soc.* **135**, 2084–2087 (2013).
9. S. Li *et al.*, *J. Am. Chem. Soc.* **136**, 5908–5911 (2014).
10. Y. Ye *et al.*, *Org. Lett.* **17**, 2804–2807 (2015).
11. P. Wei, X. Yan, F. Huang, *Chem. Soc. Rev.* **44**, 815–832 (2015).
12. P. Wei, J. Li, X. Yan, Q. Zhou, *Org. Lett.* **16**, 126–129 (2014).
13. H. Xing, P. Wei, X. Yan, *Org. Lett.* **16**, 2850–2853 (2014).
14. P. Wei, S. Li, Y. Zhang, Y. Yu, X. Yan, *Polym. Chem.* **5**, 3972–3976 (2014).
15. S. Li *et al.*, *Polym. Chem.* **5**, 3994–4001 (2014).
16. S.-P. Wang, Y.-F. Shen, B.-Y. Zhu, J. Wu, S. Li, *Chem. Commun.* **52**, 10205–10216 (2016).
17. B. Zhu *et al.*, *J. Am. Chem. Soc.* **136**, 15126–15129 (2014).
18. Y. Ye *et al.*, *J. Am. Chem. Soc.* **137**, 11896–11899 (2015).
19. X.-C. Zhang *et al.*, *Chem. Sci.* **7**, 4594–4599 (2016).

New pathways to high-performance immiscible polymer blends by reactive compatibilization: Comb-like molecular, nanomicellar, and nanoparticulate strategies

Yongjin Li*, Wenyong Dong, Hengti Wang, Meifeng He, Fanglu Ren, Junqing Zhang, Depei Chen, Zhiang Fu, Xuewen Zhao, and Xin Yang

Polymer blending is an economic, versatile route to develop new materials with desirable properties by combining selected characteristics of the constituent polymers (*1*). However, the inherent immiscibility between polymer pairs causes phase separation at the micrometer scale, which results in poor physical properties (*1, 2*). Reactive compatibilization is one of the most accessible strategies to improve the compatibility of polymer blends. This technique is based on the in situ formation of block/graft copolymers at the polymer–polymer interface, which lowers the interface tension, suppresses particle coalescence, and ultimately improves interfacial adhesion (*1, 3*). Although some researchers have shown that the stability of the copolymers

College of Material, Chemistry and Chemical Engineering, Hangzhou Normal University, Hangzhou, China
*Corresponding author: yongjin-li@hznu.edu.cn
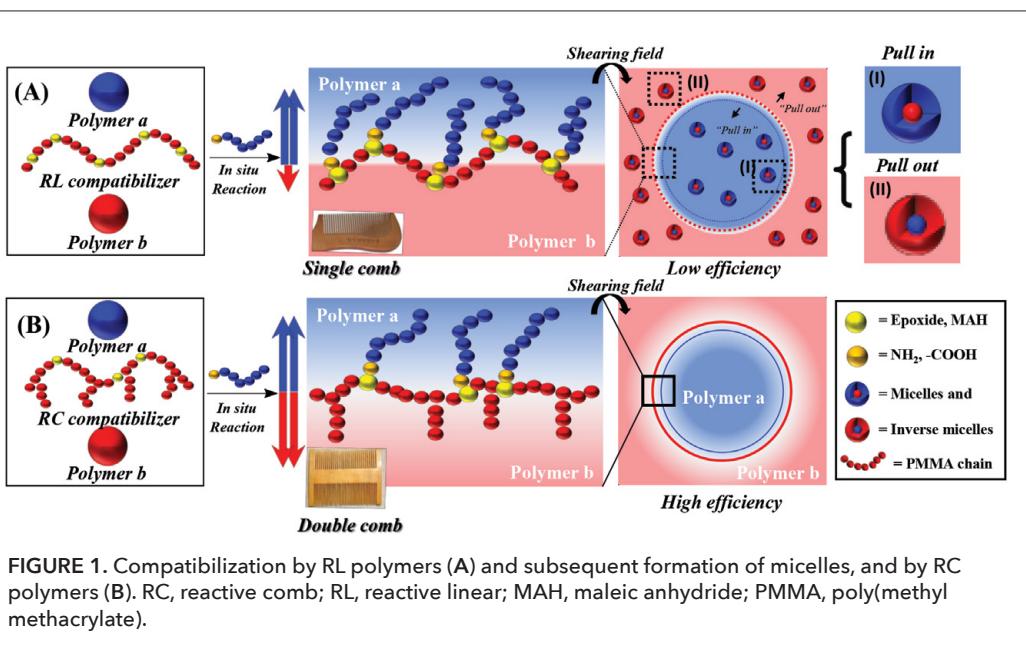
at the interface is significantly influenced by the molecular structures, only a limited number of theoretical and experimental works have been devoted to this topic.

To expand the stability window of graft polymers at the interface, we had previously synthesized a series of comb-like ["reactive comb" (RC)] polymers as compatibilizers (*1–3*). We anticipate that, compared to the reactive linear (RL) polymers, the side chains (the teeth of the "comb") can compensate for the weakened backbone in one phase due to graft reaction, thereby improving the interfacial stability. However, under external shear, traditional compatibilizers are inevitably dragged from the interface to form micelles in one phase, which would eliminate their compatibilization effect. Nevertheless, it is important to verify if this is always the case, in other words, that micelles are invariably poor compatibilizers for immiscible polymer blends.

In a previous study, we reported a new compatibilization mechanism using Janus nanomicelles (JNMs) formed in situ at the polymer–polymer interface in typical immiscible blends (*4–6*). In that study, we established a new compatibilization mechanism called interfacial micelle compatibilization (IMC) to achieve interfacial emulsification using JNMs, which is key to the compatibilization of polymer alloys. Meanwhile, the positional stability of the compatibilizers at the interface remains a challenge due to the intrinsic flexibility of the compatibilizer molecules under aggressive processing conditions. More recently, we overcame this challenge by thermodynamically stabilizing the nanoparticles (NPs) at the interface through melt reactive blending (*7–10*). This interfacial nanoparticle compatibilization (INC) mechanism paves new pathways for the use of various NPs as both an effective compatibilizer and a functional filler for immiscible polymer blends.

Reactive comb-like molecular compatibilization

We synthesized both RL and RC polymers by copolymerization with methyl methacrylate (MMA) with monomers containing reactive groups [such as glycidyl methacrylate (GMA) or maleic anhydride (MAH)], and an MMA macromer (*1*). The RL and RC polymers were tested as compatibilizers in several typical immiscible systems (*1–3*). During melt blending, the epoxide or MAH groups distributed randomly along the poly(methyl methacrylate) (PMMA) backbone can readily react with terminal carboxyl or amino groups of one polymer component.


FIGURE 1. Compatibilization by RL polymers (A) and subsequent formation of micelles, and by RC polymers (B). RC, reactive comb; RL, reactive linear; MAH, maleic anhydride; PMMA, poly(methyl methacrylate).

FIGURE 2. (A–B) Compatibilization by reactive PS-type copolymer and (C–G) the IMC mechanism using interfacial JNMs. IMC, interfacial micelle compatibilization; JNMs, Janus nanomicelles; PLLA, poly-L-lactic acid; PMMA, poly(methyl methacrylate); PVDF, poly(vinylidene fluoride); PS, polystyrene.

FIGURE 3. (A–B) Compatibilization by reactive nanoparticles and (C–G) the INC mechanism using interfacial Janus NPs. INC, interfacial nanoparticle compatibilization; NP, nanoparticle; LM-NP, PLLA/PMMA-grafted nanoparticle; PLLA, poly-L-lactic acid; PMMA, poly(methyl methacrylate); PVDF, poly(vinylidene fluoride).

Meanwhile, the PMMA backbone and side chains of RC are designed to be miscible with the other polymer.

For the RL polymers (Figure 1A), as the grafting reaction proceeds, the number of grafted chains obviously increases and the backbone length between two adjacent grafting sites decreases correspondingly. These factors significantly influence the interactions of the backbone with Polymer b and that of the grafted side chains with Polymer a in the graft polymers. Thus, the graft polymers formed in situ by the RL compatibilizer (single comb copolymer) do not have balanced molecular structures, and are prone to be "pulled in" or "pulled out" from the interface to one of the phases under shear condition to form micelles (I) or inverse micelles (II), respectively. In contrast, as shown in Figure 1B, RC polymers show better positional stability at the interface compared with their RL counterparts. After the in situ reaction, the molecular structure in the RC reaction is symmetrical, with two kinds of side chains (double comb copolymer). The pregrafted chains (e.g., PMMA) penetrate Polymer b [e.g., poly(vinylidene fluoride) (PVDF)], while the in situ-formed side chains penetrate Polymer a. This could compensate for the weakened interaction between the PMMA backbone and the PVDF phase due to the grafting reaction. Indeed, the RC compatibilizer showed a better emulsifying effect than the RL one, as it can significantly increase the stability of the in situ-formed polymers at the interface (1). Moreover, high compatibilization efficiency can be achieved by adjusting the molecular parameters of the RC polymer, namely the length of the backbone (M), the content of the reactive groups (E), and the distance between the side chains (D) and their length (L) (2). The RC compatibilizer can be applied to many typical immiscible

blends such as poly-L-lactic acid (PLLA)/PVDF (1, 2) and PLLA/acrylonitrile-butadiene-styrene (ABS) systems (3).

Reactive nanomicellar compatibilization

As mentioned above, the in situ-formed graft copolymers usually self-assemble into micelles in one phase of the blend. Significant efforts have been made to minimize the formation of these micelles and enhance the compatibilization efficiency. We discovered a new compatibilization mechanism using the JNMs formed in situ at the polymer–polymer interface in immiscible PLLA/PVDF blends (4).

As shown in Figure 2 (A–B), we first synthesized a polystyrene-type (PS-type) copolymer by the "grafting through" strategy. During blending, the epoxide groups in the PS-type copolymer react with the carboxyl ends of PLLA to form a double-grafted structure with both PLLA and PMMA side chains (Figure 2C). The long PMMA side chains are miscible with PVDF because of their carbonyl and CF_2 groups. Thus, the in situ-formed copolymer can self-assemble into JNMs owing to the spreading coefficient (Figure 2D). These JNMs have a spherical structure with one side of the surface constituted of PMMA chains and the other of PLLA chains, and these chains on the nanomicelles on each side are readily entangled among themselves. Therefore, the JNMs act as effective compatibilizers at the PLLA/PVDF interface, as observed from transmission electron microscopy (TEM) images (Figure 2E). Scanning transmission electron microscopy (STEM) images and elemental mapping images (EMI) of the PVDF/PLLA/PS-type copolymer blends are displayed in Figure 2F and Figure 2G, respectively. Our previous work also showed that the
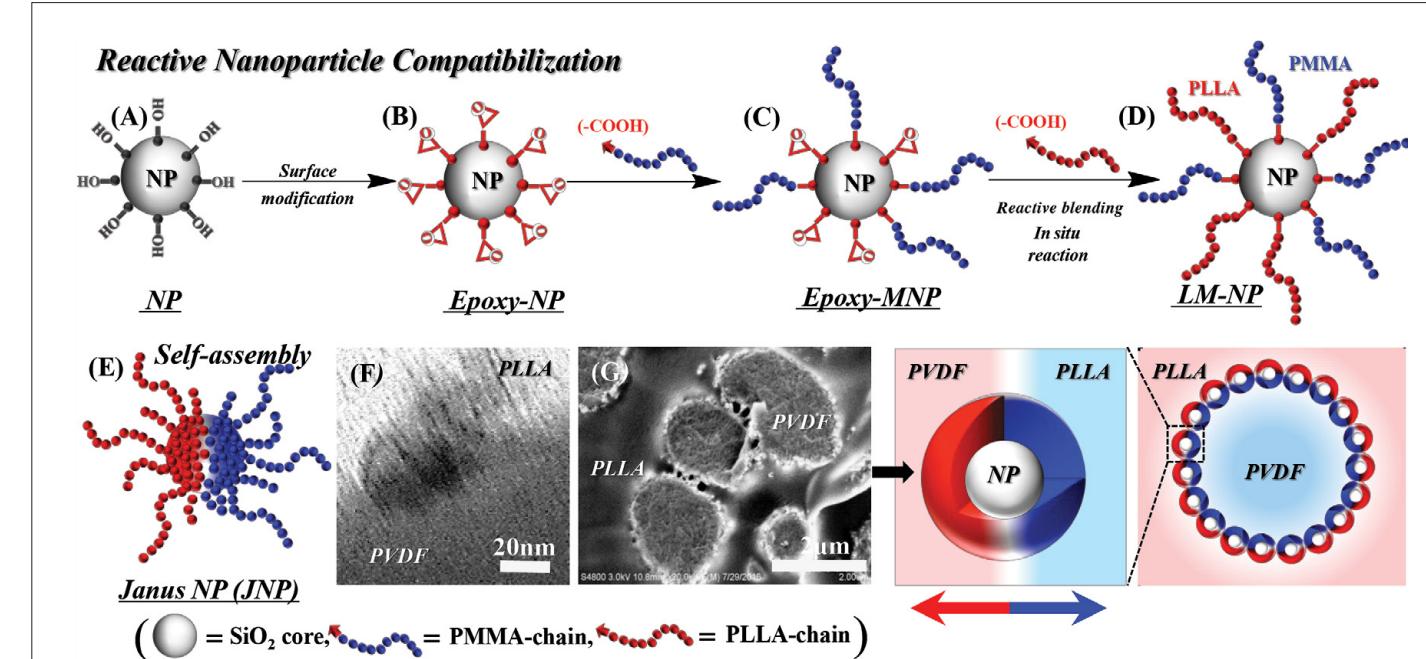
formation and locations of the JNMs depend not only on the molecular architecture of the PS-type copolymer (namely the length of PMMA side chains, the density of reactive epoxide groups, and the molecular weight of PS backbone), but also the melt processing sequence and molecular weight of the components (5, 6). Therefore, instead of a hindrance, certain micelles can be used as efficient compatibilizer for immiscible polymer blends.

Reactive nanoparticle compatibilization

Here, we describe the INC mechanism to emulsify immiscible binary polymer blends (7–10).

As shown in Figure 3 (A–C), new kinds of reactive nanoparticles were synthesized through the "attaching onto" approach. Because they contain both reactive epoxide groups and long PMMA chains on a solid NP core, such nanoparticles (termed epoxy-MNPs) can further react with the terminal carboxyl groups of PLLA during melt compounding (Figure 3C). Through this reaction, the two polymers become randomly attached to the exterior surface of the NP by chemical bonds, successively forming a heterogeneous polymer corona (Figure 3D). Owing to the strong immiscibility between the PLLA and PMMA grafts, a Janus PMMA/PLLA corona is spontaneously formed on the exterior surface of the NP (Figure 3E), as determined by the wetting and spreading coefficients. TEM observations (Figure 3F) confirmed that the Janus NP (JNP) formed featured a solid NP core and equally weighted PMMA/PLLA hemispheres. On one side of the JNP, the in situ-formed PLLA tails extend and interact with the PLLA polymer phase,

and the long PMMA chains on the other side interact with the PVDF phase to maximize the favorable contacts. Meanwhile, the solid core serves as a stiff nanofiller to improve interfacial stability. Thus, the interfacial JNP could effectively reinforce interfacial adhesion and facilitate compatibilization, effects that can be confirmed from the scanning electron microscopy images (Figure 3G). This INC mechanism creates new possibilities for the compatibilization of immiscible polymer blends. Various types of NPs, such as polyhedral oligomeric silsesquioxane (7, 8) and SiO_2 (9, 10), could be used in this way as both an effective compatibilizer and a functional filler for multiphase polymer materials.

References
1. W. Dong *et al.*, *Ind. Eng. Chem. Res*. **54**, 2081–2089 (2015).
2. W. Dong *et al.*, *ACS Sustain. Chem. Eng*. **4**, 4480–4489 (2016).
3. W. Dong *et al.*, *ACS Sustain. Chem. Eng*. **3**, 2542–2550 (2015).
4. H. Wang, W. Dong, Y. J. Li, *ACS Macro Lett*. **4**, 1398–1403 (2015).
5. H. Wang, Z. Fu, W. Dong, Y. J. Li, J. Y. Li, *J. Phys. Chem. B* **120**, 9240–9252 (2016).
6. D. Chen, H. Wang, Y. J. Li, *ACS Appl. Mater. Interfaces* **9**, 33091–33099 (2017).
7. Z. Fu, H. Wang, X. Zhao, S. Horiuchi, Y. J. Li, *Polymer* **132**, 353–361 (2017).
8. Z. Fu, H. Wang, W. Dong, Y. J. Li, *Acta Polym. Sin*. **2**, 334–341 (2017).
9. H. Wang, Z. Fu, X. Zhao, Y. J. Li, J. Y. Li, *ACS Appl. Mater. Interfaces* **9**, 14358–14370 (2017).
10. H. Wang *et al.*, *Macromolecules* **50**, 9494–9506 (2017).

4-quinolone derivatives: Synthesis and antitumor activity

Chao Shen, Jun Xu, Chengcai Xia, Yong Yang, Hongyun Shen, Beibei Ying, Xiaolei Zhu, and Pengfei Zhang*

4-quinolones have received significant attention in the field of medicine due to their antibacterial properties, with numerous studies indicating that the substituents on the 4-quinolone framework are responsible for altering the biological and pharmacological activities of this compound. Indeed, many novel biological properties have been reported, including antimalarial, antiviral, antitumor, and antibacterial activities. The synthesis of quinolone derivatives and the evaluation of their biological activities are therefore important. To date, quinolones have been generated by a range of synthetic methods, including classical routes to substituted quinolones such as the Gould–Jacobs reaction, the Conrad–Limpach reaction, and Camps cyclization. However, these approaches require the use of strong bases, and the cyclization process needs high temperatures. Functionalization of the 4-quinolone skeleton, by contrast, is a simple, efficient strategy to obtain such molecules.

Recently, our group has focused on developing novel strategies for the modification of 4-quinolone derivatives to improve reaction efficiencies and reduce side product formation. To date, several methods have been established for the modification of both the benzenoid ring and the pyridinone ring of the quinolone moiety.

College of Material, Chemistry and Chemical Engineering, Hangzhou Normal University, Hangzhou, China
*Corresponding author: chxyzpf@hotmail.com
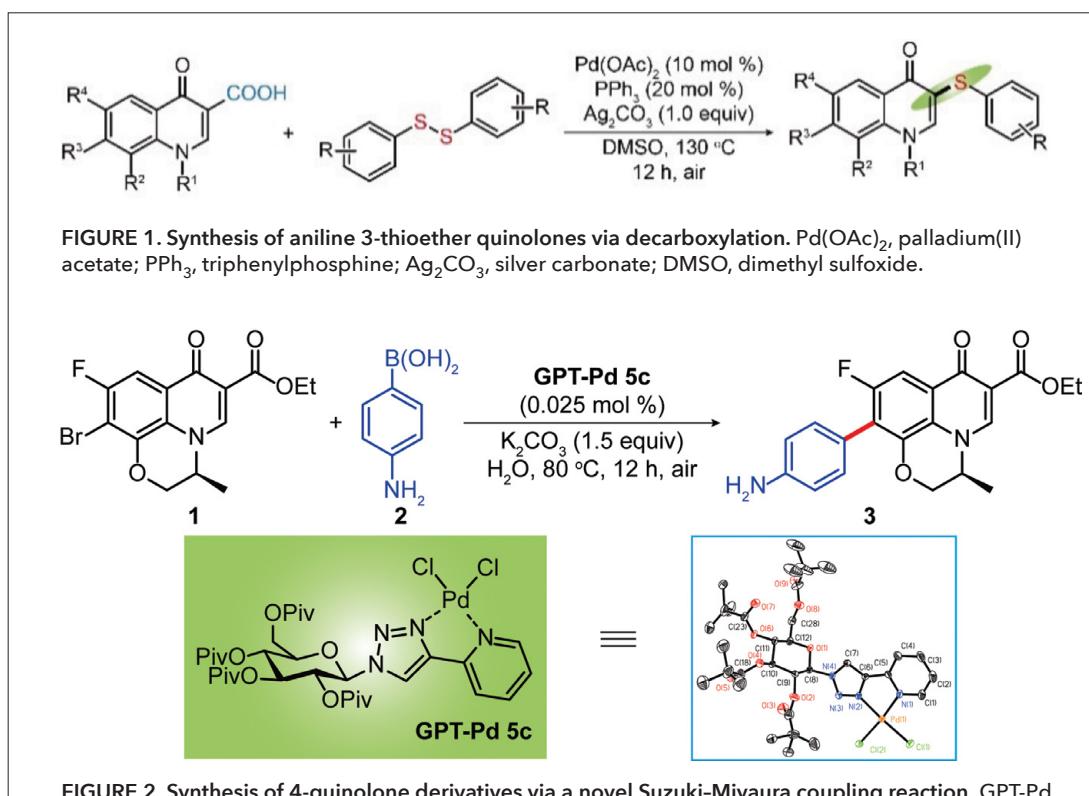
Synthesis of 4-quinolone derivatives by decarboxylation

Construction of novel compounds using easily accessible carboxylic acids via transition metal–catalyzed C–C and C–heteroatom bond-forming reactions has been considered the most sustainable, straightforward strategy for the preparation of 4-quinolone derivatives. Such reactions tend to involve an initial cleavage of the carboxyl C–C bond prior to the construction of new chemical bonds at the cleavage site. Based on our previous experience, we developed a novel protocol for preparing 3-arylsulfide–substituted quinolones via this route. A series of 3-arylsulfide–substituted quinolones was acquired in moderate to good yields through this new strategy (Figure 1) (*1*). We expect that this approach represents a promising alternative to existing protocols for the synthesis of 3-arylsulfide–substituted quinolone derivatives, and could be applied in the synthesis of key drug intermediates.

Synthesis of 4-quinolone derivatives via a novel Suzuki–Miyaura coupling reaction

As one of the most useful and widely studied strategies for the formation of C–C bonds, Suzuki–Miyaura coupling plays a significant role in organic synthesis, including in the total synthesis of numerous natural products, organic building blocks, and biologically active compounds. Although the application of transition metal–based homogenous

FIGURE 1. Synthesis of aniline 3-thioether quinolones via decarboxylation. Pd(OAc)₂, palladium(II) acetate; PPh₃, triphenylphosphine; Ag₂CO₃, silver carbonate; DMSO, dimethyl sulfoxide.

FIGURE 2. Synthesis of 4-quinolone derivatives via a novel Suzuki–Miyaura coupling reaction. GPT-Pd, glycosyl pyridyl-triazole@palladium nanoparticles; K₂CO₃, potassium carbonate.
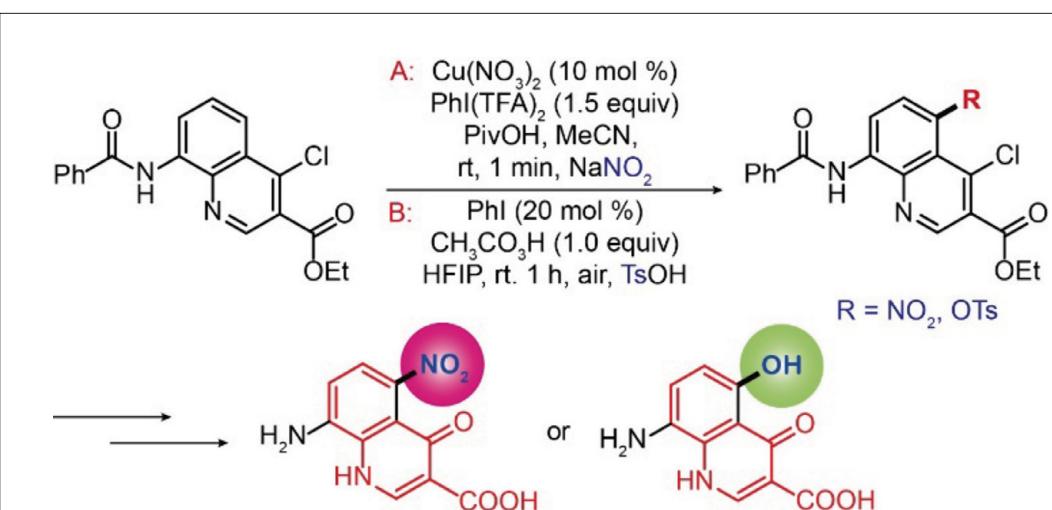
FIGURE 3. Synthesis of 4-quinolone derivatives via C–H functionalization. Cu(NO₃)₂, cupric nitrate; PhI(TFA)₂, [bis(trifluoroacetoxy)iodo]benzene; PivOH, pivalic acid; MeCN, acetonitrile; rt, room temperature; NaNO₂, sodium nitrite; PhI, iodobenzene; CH₃CO₃H, peracetic acid; HFIP, hexafluoroisopropanol; TsOH, p-toluenesulfonic acid.
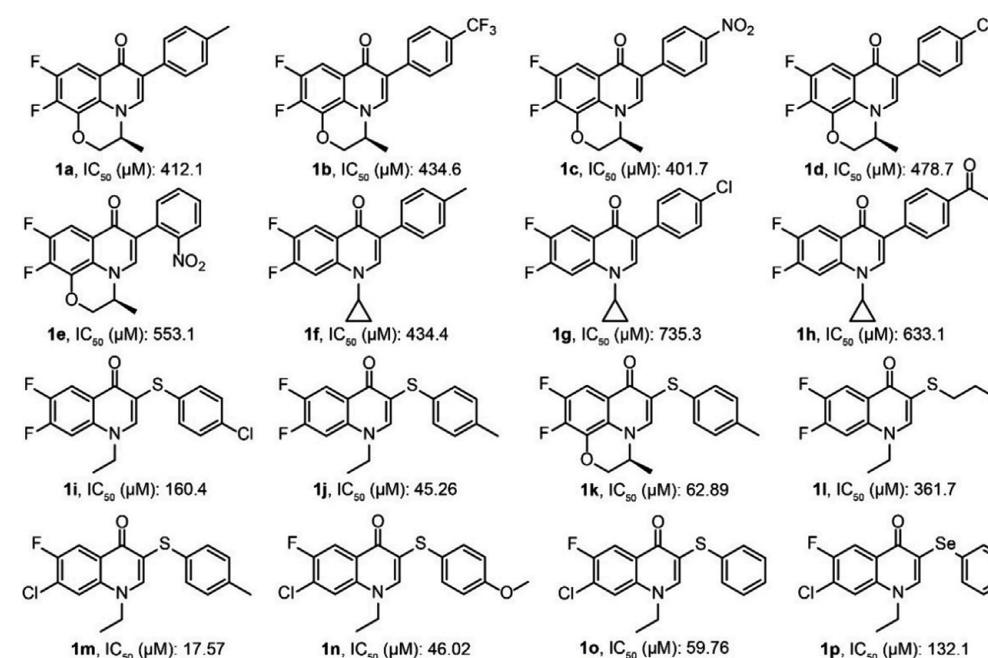
FIGURE 4. Cytotoxicities of selected compounds against Huh-7 cells in vitro. IC₅₀, inhibitory concentration 50%.

catalysts in these coupling reactions has received considerable attention, the issue of separating the catalysts from the reaction medium limits its development for applications in the chemical industry. Furthermore, the presence of metal residues originating from the transition-metal catalyst affects final product quality. Therefore, the use of metal catalysts supported on heterogeneous carriers has recently gained

widespread attention for application in C–C cross-coupling reactions. Such heterogeneous systems have many advantages compared to their homogeneous alternatives, including improved steric control of the reaction intermediates, catalyst reusability, and ease of product separation from the reaction mixture. We therefore prepared a series of novel heterogeneous catalysts from cheap, nontoxic materials to address the above issues with homogeneous systems, and the catalytic activities of our novel catalysts were tested in a range of coupling reactions (*2, 3*). With the optimal conditions in hand, we applied our newly established method to the efficient synthesis of novel fluoroquinolone derivatives, which were obtained in ≤63% yields after recrystallization. In addition, analysis by inductively coupled plasma atomic emission spectroscopy confirmed that palladium contamination was nonexistent in the reaction solvent (residual Pd content <20 ppm) (Figure 2) (*4*).

Synthesis of 4-quinolone derivatives by C–H functionalization

C–H functionalization presents a simpler and greener route to 4-quinolone derivatives than traditional coupling reactions. Since direct C–H bond cleavage is difficult for thermodynamic reasons, novel transition-metal complexes that can overcome the thermodynamic and kinetic barriers to achieve C–H bond cleavage with high precision are required to expand this field of chemistry. Indeed, the recent development of transition-metal catalysts in C–H functionalization constitutes a significant advance in organic synthesis. As such,

our ongoing studies focus on novel methods for the synthesis of 4-quinolone derivatives via C–H functionalization (Figure 3). Indeed, we recently established a number of C–H functionalization protocols for the formation of various chemical bonds, such as C–S bonds (5–8), C–O bonds (9, 10), C–N bonds (11), C–C bonds (12–14), and C–halogen bonds (15). The results of these studies provide insight into the C–H functionalization of organic molecules, while providing a new route for rational design and synthesis of 4-quinolone derivatives.

Anticancer activity of 4-quinolone derivatives

Many quinolone derivatives are known to exhibit anticancer properties. To investigate such properties in greater detail, Huh-7 or HepG2 cells were treated with various quinolone derivatives (Figures 4 and 5) (1). Interestingly, the quinolone concentration had no significant effect on cell growth. In addition, the MMT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) cell proliferation assay was conducted for various quinolone derivatives, and results indicated significant antiproliferative activity for compound 1m against Huh-7 cells, which exhibited an IC_{50} (inhibitory concentration 50%) value of 17.57 μM (Figure 4). Furthermore, these quinolone derivatives also displayed good inhibitory activity against HepG2 cells, with compound 1m again displaying a superior inhibitory activity, with an IC_{50} value of 3.05 μM (Figure 5).

Conclusions

Ongoing advances in the development of novel routes to 4-quinolone derivatives are particularly important due to the unique biological and pharmaceutical activities of these compounds. In recent years, our group has developed a number of novel synthetic strategies for this purpose, including decarboxylation, a novel Suzuki–Miyaura coupling, and C–H functionalization protocols. These approaches are more convenient and efficient than traditional methods, and thereby provide a richer intellectual basis for future R&D. Furthermore, we have also evaluated the antitumor activities of a selection of 4-quinolone derivatives, with some interesting results. However, further studies into the preparation and application of quinolone derivatives are ongoing.


FIGURE 5. Cytotoxicities of selected compounds against HepG2 cells in vitro. IC_{50}, inhibitory concentration 50%.

References

1. C. Xia et al., Chem. Asian J. **11**, 360–366 (2016).
2. C. Shen, J. Xu, W. Yu, P. Zhang, Green Chem. **16**, 3007–3012 (2014).
3. C. Shen, H. Shen, M. Yang, C. Xia, P. Zhang, Green Chem. **17**, 225–230 (2015).
4. H. Shen, C. Shen, C. Chen, A. Wang, P. Zhang, Catal. Sci. Technol. **5**, 2065–2071 (2015).
5. C. Shen et al., Chem. Soc. Rev. **44**, 291–314 (2015).
6. C. Shen et al., Chem. Sci. **3**, 2388–2393 (2012).
7. J. Xu et al., Chem. Asian J. **11**, 882–892 (2016).
8. Y. Yang et al., ChemCatChem **8**, 304–307 (2016).
9. B. Ying, J. Xu, X. Zhu, C. Shen, P. Zhang, ChemCatChem **8**, 2604–2608 (2016).
10. C. Shen et al., RSC Adv. **7**, 49436–49439 (2017).
11. X. Zhu et al., RSC Adv. **6**, 89979–89983 (2016).
12. C. Shen, J. Xu, B. Ying, P. Zhang, ChemCatChem **8**, 3560–3564 (2016).
13. J. Xu et al., Org. Lett. **19**, 5661–5664 (2017).
14. J. Xu et al., Org. Chem. Front. **4**, 1116–1120 (2017).
15. J. Xu et al., Org. Biomol. Chem. **14**, 3016–3021 (2016).

Transition-metal–catalyzed stereocontrolled vinylation/allylation at C_{sp2} and organocatalytic asymmetric reactions

Jian Zhang, Xiaofei Zeng, Limin Yang, and Guofu Zhong*

Introduction

The 1,3-butadiene moiety represents an exceptionally useful functional group, offering numerous applications in synthetic and polymer chemistry (1). One facile preparation with high atomic and step economy is the direct cross-coupling of alkenes via double C_{vinyl}–H bond activation (2). However, such cross-coupling reactions still suffer from great limitations. For example, very few directing groups are in practical use, and a stoichiometric amount of oxidant is generally required. Thus, a series of highly desirable, novel, and efficient protocols for the stereoselective construction of butadienes have been developed. In addition, a catalytic system combining a transition-metal catalyst with an organocatalyst for the asymmetric dearomatization of readily available naphthols via allylation has been discovered, and some important organocatalytic enantioselective reactions have been reported.

Oxidant-free protocols for the synthesis of linear (Z,E)-butadienes

Recently, we developed a ruthenium-catalyzed vinylic C–H bond alkenylation of acrylamides (Figure

College of Materials, Chemistry and Chemical Engineering, Hangzhou Normal University, Hangzhou 310036, China
*Corresponding author: zgf@hznu.edu.cn

1A, 1) using allyl acetate. This external oxidant-free protocol efficiently produced (Z,E)-butadiene skeletons (Figure 1A, 2) with good-to-excellent stereo- and regioselectivities using aminocarbonyl directing groups (3). The same conditions were also successfully applied to the synthesis of styrenes from aromatic amides and allyl acetate (4). Alkenylation of acrylamides (Figure 1B, 3) was also achieved using CONH(OMe) as an oxidizing directing group. This protocol provided an oxidant-free and mild route for preparing valuable (E,Z)-butadienes (Figure 1B, 4) in excellent selectivities (5).

Synthesis of linear (Z,E)-butadienes using labile directing groups

Oxidative cross-coupling reactions between alkenes were also developed using easily modified directing groups, such as Weinreb amides (Figure 2A, 5) and enol carbamates (Figure 2B, 6). Both reagents led to efficient synthesis of (Z,E)-configured butadienes (Figure 2A, 7 and Figure 2B, 8) in excellent stereoselectivities (6, 7).

Synthesis of branched (Z,Z)-butadienes using alkynes

Ruthenium-catalyzed cross-coupling reactions between olefins (Figure 3A, 9 and Figure 3B, 6) and alkynes via aminocarbonyl- or carbamate-directed olefinic C–H bond alkenylation have also been reported. These oxidant-free and atom-economic approaches efficiently provided branched (Z,Z)-butadienes (Figure 3A, 10 and Figure 3B, 11) in excellent stereoselectivities (7, 8).

Synthesis of γ-alkylidene butenolides from simple vinyl carboxylic acids and alkenes

γ-alkylidene butenolides are common organic molecules of biological significance. We previously developed the first atom-economic rhodium-catalyzed preparation of a wide spectrum of γ-alkylidene butenolides from simple vinyl carboxylic acids (Figure 4, 12). This protocol proceeded via a tandem cross-coupling/Michael cyclization/dehydrogenation process to give products 13 and 14 in 0%–60% yields (9).

FIGURE 1. Synthesis of butadienes without an external oxidant. (A) Cross-coupling between acrylamides (or benzamides) and allyl acetate. **(B)** C–H alkenylation of acrylamides using an oxidizing directing group. Ratios of 2/2' employed are given in square brackets. Z,E/Z,Z ratios of isomers 2 and 4 are given in parentheses. [Ru(p-cymene)Cl$_2$]$_2$, dichloro(p-cymene)ruthenium(II) dimer; DCE, 1,2-dichloroethane; DMF, N,N-dimethylformamide.

FIGURE 2. Synthesis of 1,3-butadiene skeletons using easily modified directing groups. (A) Cross-coupling reactions of alkenes using Weinreb amide as the directing group. (B) Cross-coupling reactions between enol carbamates and alkenes. Z,E/Z,Z ratios of products 7 and 8 are given in parentheses. [RhCp*Cl₂]₂, pentamethylcyclopentadienylrhodium(III) chloride dimer; [Ru(p-cymene)Cl₂]₂, dichloro(p-cymene) ruthenium(II) dimer; DCE, 1,2-dichloroethane; DMF, N,N-dimethylformamide.

FIGURE 3. Synthesis of branched Z,Z-butadienes using alkynes. (A) Cross-coupling reactions between acrylamides and alkynes. (B) Cross-coupling reactions between enol carbamates and alkynes. Z,Z/Z,E ratios of isomers 10 and 11 are given in parentheses. [Ru(p-cymene)Cl₂]₂, dichloro(p-cymene)ruthenium(II) dimer; THF, tetrahydrofuran; DME, 1,2-dimethoxyethane.

FIGURE 4. Preparation of γ-alkylidene butenolides using vinyl carboxylic acids. Z/E ratios of product 13 are given in parentheses. [RhCp*Cl₂]₂, pentamethylcyclopentadienylrhodium(III) chloride dimer.

FIGURE 5. Synthesis of chiral β-naphthalenones via the asymmetric allylic dearomatization of β-naphthols by cooperative iridium and phosphoric acid catalysis. [Ir(cod)Cl]₂, chloro (1,5-cyclooctadiene) iridium(I) dimer.

FIGURE 6. Asymmetric aza-Rauhut–Currier reaction catalyzed by an amino phosphine organocatalyst. cat., catalyst; ee, enantiomeric excess; dr, diastereomeric ratio; aza-RC, aza-Rauhut–Currier reaction.

Asymmetric dearomatization of ß-naphthols by cooperative iridium and phosphoric acid catalysis

Enantioselective dearomatization of β-naphthols (Figure 5, 16) was developed using easily accessible secondary allylic alcohols (Figure 5, 15) via an allylation reaction catalyzed by a dual catalytic system comprising a chiral iridium complex and phosphoric acid. β-naphthalenones bearing a quaternary carbon center (Figure 5, 17) were efficiently afforded in high enantio- and chemoselectivities. The experimental findings suggested that the chiral iridium complex and phosphoric acid components work cooperatively to activate the substrates, with the chiral iridium complex controlling the absolute configuration in the enantiodetermining step (10).

Facile access to highly functionalized chiral tetrahydropyridines via an amino-phosphine–catalyzed asymmetric [4+2] annulation

An asymmetric [4+2] annulation reaction initiated by the aza-Rauhut–Currier reaction was developed using an amino phosphine catalyst derived from a natural amino acid. Based on the reaction between vinyl ketones and N-sulfonyl-1-aza-1,3-dienes (Figure 6, 18), this method afforded a novel, facile preparation of highly functionalized tetrahydropyridine derivatives (Figure 6, 19) in high stereoselectivities (11, 12).

Organocatalytic asymmetric access to chiral spiro skeletons

We previously developed many organocatalytic approaches for the synthesis of chiral spiro skeletons (13–15). For example, the 3,3′-pyrrolidonyl spirooxindole skeleton, a common motif of bioactive molecules, was efficiently constructed by an efficient quinine-derivative–catalyzed [3+2] cycloaddition reaction, using α-isothiocyanato imides as dienophiles. Notably, three contiguous stereocenters, including one all-carbon spiro quaternary center, were elegantly established under mild conditions. Additionally, a highly enantioselective

preparation of unprotected spiroindoline pyrans was achieved using a cooperative catalytic system based on N-heterocyclic carbenes and Brønsted acid (≤95% yield, ≤93% enantiomeric excess). Furthermore, a [3+3] cycloaddition reaction of imines and oxindole-derived enals was developed for the rapid construction of spirocyclic oxindoles, where a chiral N-heterocyclic carbene acted as the catalyst. In general, high yields and excellent enantioselectivities were obtained with a short reaction time (i.e., 1 h). Finally, the enantioselective construction of spirocyclic oxindole derivatives with multiple stereocenters was achieved via a Michael/aldol/hemiacetalization cascade reaction (between 20 and 21 in Figure 7) catalyzed by a prolinol-derived organocatalyst, to give product 22 in good yields and with good enantioselectivities (13–15).

N-heterocyclic-carbene-catalyzed activation of N-hydroxyphthalimide esters to yield chiral dihydropyridinones bearing an all-carbon quaternary stereogenic center

We also disclosed the [3+3] cyclization of N-hydroxyphthalimide acrylates (Figure 8, 23) and N-Ts (Ts = p-toluenesulfonyl) ketimines (Figure 8, 21). Valuable chiral dihydropyridinones (Figure 8, 24) were obtained in good yields with excellent enantioselectivities, demonstrating the efficiency of N-hydroxyphthalimide acrylates in N-heterocyclic carbene catalysis (>99% enantiomeric excess) (17).

In conclusion, we have developed a series of useful methodologies utilizing transition-metal catalysis for the stereospecific synthesis of 1,3-butadienes. In addition,
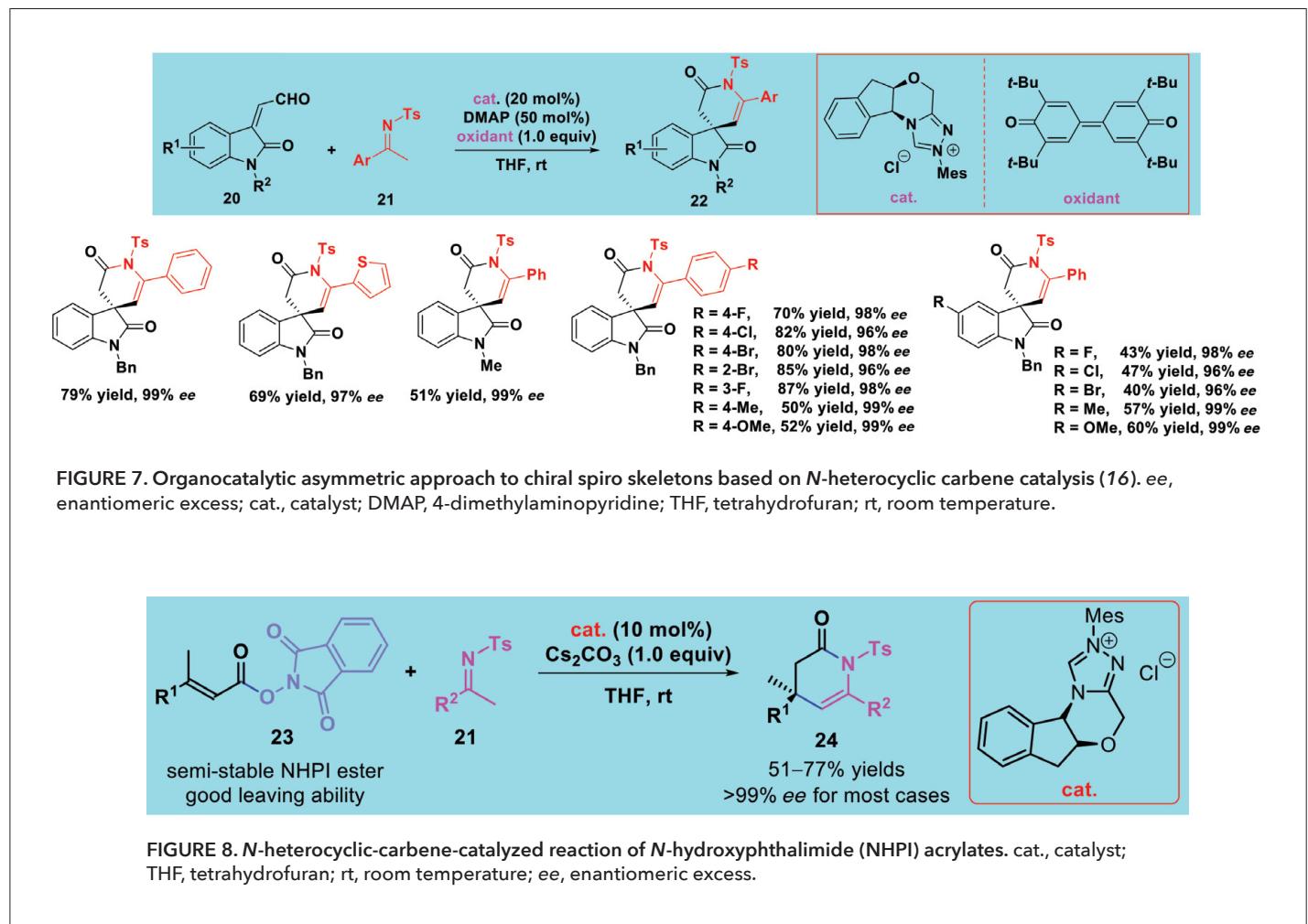
FIGURE 7. Organocatalytic asymmetric approach to chiral spiro skeletons based on *N*-heterocyclic carbene catalysis (*16*). ee, enantiomeric excess; cat., catalyst; DMAP, 4-dimethylaminopyridine; THF, tetrahydrofuran; rt, room temperature.
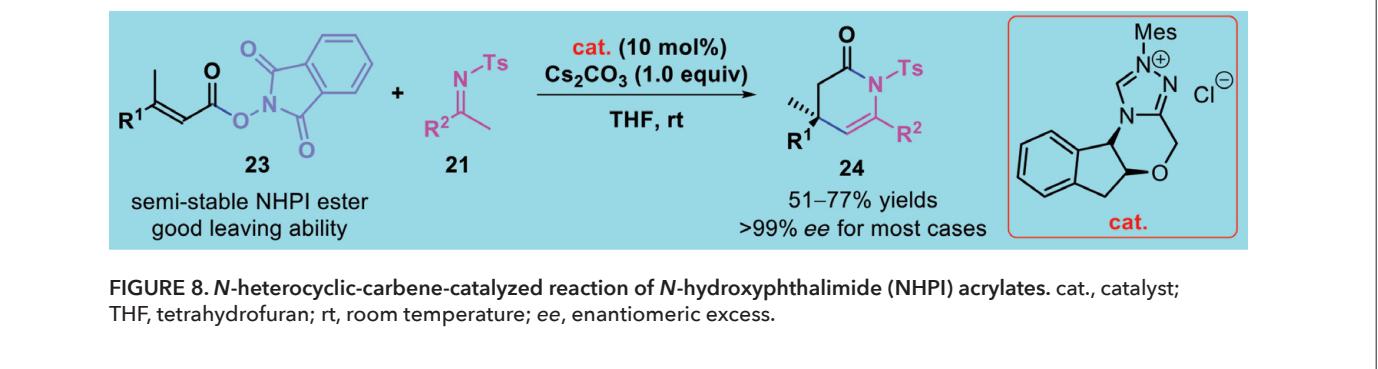
FIGURE 8. *N*-heterocyclic-carbene-catalyzed reaction of *N*-hydroxyphthalimide (NHPI) acrylates. cat., catalyst; THF, tetrahydrofuran; rt, room temperature; ee, enantiomeric excess.

an asymmetric dearomative allylation of β-naphthols has been realized by a cooperative chiral iridium catalysis and acid catalysis, and a series of organocatalytic enantioselective reactions have been discovered.

References

1. E.-i. Negishi *et al*., *Acc. Chem. Res*. **41**, 1474–1485 (2008).
2. X. Shang, Z.-Q. Liu, *Chem. Soc. Rev*. **42**, 3253–3260 (2013).
3. F. Li, C. Yu, J. Zhang, G. Zhong, *Org. Lett*. **18**, 4582–4585 (2016).
4. F. Li *et al*., *Adv. Synth. Catal*. **358**, 3932–3937 (2016).
5. C. Yu, F. Li, J. Zhang, G. Zhong, *Chem. Commun*. **53**, 533–536 (2017).
6. F. Li, C. Yu, J. Zhang, G. Zhong, *Org. Biomol. Chem*. **15**, 1236–1244 (2017).
7. T. Li *et al*. *Chem. Commun*. **53**, 12926–12929 (2017).
8. K. Meng *et al*., *Org. Lett*. **19**, 2498–2501 (2017).
9. C. Yu, J. Zhang, G. Zhong, *Chem. Commun*. **53**, 9902–9905 (2017).
10. D. Shen, Q. Chen, P. Yan, X.-F. Zeng, G. Zhong, *Angew. Chem. Int. Ed*. **56**, 3242–3246 (2017).
11. Z. Shi, Q. Tong, W. W. Y. Leong, G. Zhong, *Chem. Eur. J*. **18**, 9802–9806 (2012).
12. Z. Shi, P. Yu, T.-P. Loh, G. Zhong, *Angew. Chem. Int. Ed*. **51**, 7825–7829 (2012).
13. B. Tan *et al*., *Chem. Eur. J*. **18**, 63–67 (2012).
14. Y. Lin, L. Yang, Y. Deng, G. Zhong, *Chem. Commun*. **51**, 8330–8333 (2015).
15. L. Zhu *et al*., *Org. Lett*. **18**, 2387–2390 (2016).
16. D. Xie *et al*., *Org. Lett*. **17**, 2318–2321 (2015).
17. Z. Zhang *et al*., *Org. Lett*. **17**, 5052–5055 (2015).
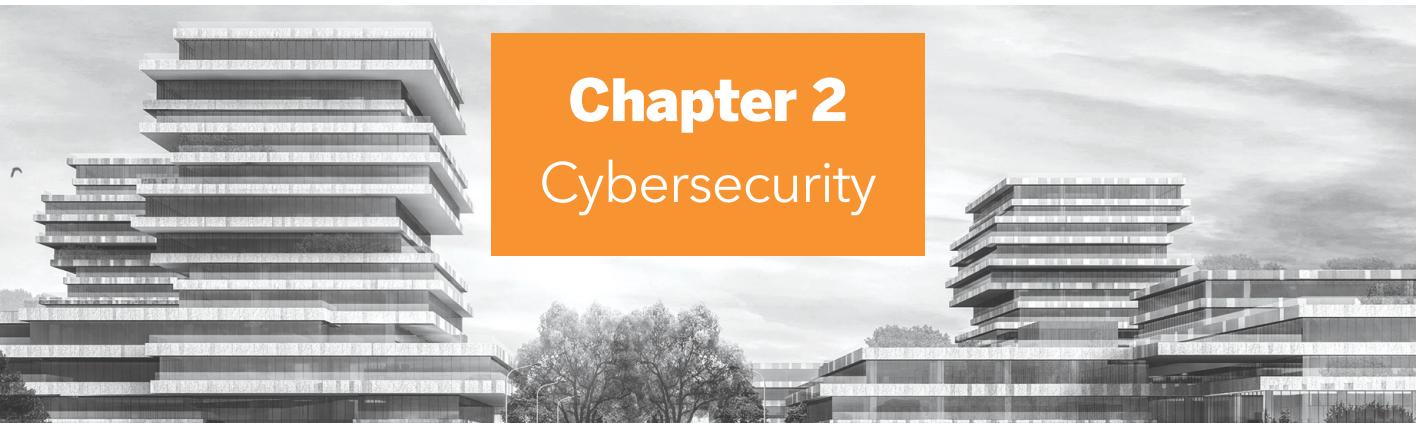
Chapter 2
Cybersecurity

Provably secure anonymous authenticated key agreement protocols with enhanced security models

Qi Xie[1*], Xiao Tan[1], Bin Hu[1],
Guilin Wang[2], Duncan S. Wong[3],
Kefei Chen[1], and XiuYuan Yu[1]

FIGURE 1. Proposed anonymous two-factor AKE scheme (*1*). AKE, authenticated key exchange. Card: smart card with stored information. Server: user registration information storage repository.

The rapid advancement of network and communication technologies has enabled mobile phone users to perform electronic transactions online by subscribing to a variety of remote services. For almost all transactions with remote servers, two key security objectives are required before the key exchange phase: mutual authentication and user privacy. In practice, it is generally necessary to adopt a specifically designed authenticated key exchange (AKE) protocol to realize the two objectives. In mutual authentication, both the remote server and user should be able to authenticate each other's respective identities to avoid the risk of potential nefarious interference, such as man-in-the-middle attacks. User privacy refers to the anonymity and untraceability of the user's identity in multiple executions of the AKE protocol. Anonymity is intended to conceal the user's location and online activities. Untraceability is also called unlinkability because there is a negligible probability that an adversary could link two sessions to the same user by cryptanalysis.

To date, several AKE schemes with provable security have been developed to address privacy concerns across multiple application scenarios. The authors of this article have made primary contributions to this field, including: (1) an enhanced AKE security model that supports user anonymity and provides many useful features in the conventional two-party setting; (2) a novel framework for anonymous AKE in the three-party setting (3PAKE) that can be directly applied to telecare medical information systems (TMIS); and (3) a one-round authentication protocol model optimized for cloud computing.

Anonymous two-factor AKE protocols

We previously proposed an anonymous two-factor AKE protocol (*1*) with many desirable properties including two-factor security, a universal requirement for AKE protocols. To enable a higher level of security, secret user keys are physically divided into two parts: a short phrase or password, and a portable

[1]Key Laboratory of Cryptography and Network Security, Hangzhou Normal University, Hangzhou, China
[2]Shield Laboratory, Central Research Institute, Huawei International Pte., Ltd., Singapore
[3]Hong Kong Applied Science and Technology Research Institute, Hong Kong, China
*Corresponding author: qixie68@126.com

computing device (such as a smart card) that stores personal information to identify the user. Each secret key part is deemed a single factor. This two-factor system makes AKE protocols resistant to two primary types of attacks. The first is the loss or theft of a smart card, in which a hacker can gain unauthorized access to private data on a corrupted smart card (*2*, *3*). The second is an offline dictionary attack, in which an attempt is made to guess short passwords that have a nonuniform distribution (i.e., the passwords occur with a higher frequency than chance, such as those using known words).

We extended the AKE security model (*4*) to support user anonymity and untraceability, while preserving security against both types of attacks described above. The scheme is based on elliptic-curve cryptography (ECC), a very lightweight protocol that can be easily deployed on mobile computing applications. The key concepts of the model are its encryption of the user's identity to enable user anonymity and untraceability; encryption of smart card data to help prevent a smart-card attack; and employment of cryptographic nonces (single-use random or pseudorandom numbers) in each message sent to prevent offline dictionary attacks.

This scheme is depicted in Figure 1. It preserves anonymity in the abovementioned extended model in order to avoid a powerful adversary, by enabling the criminal to gain access to specially designed oracles (cryptographic tools that appear to provide specific system information to an attacker, and can be used to probe the security of a system), as outlined by these five oracle examples:

1. Execute(): This oracle simulates scenarios in which the adversary mounts a passive attack by eavesdropping on several protocol executions.
2. Send(): The oracle simulates scenarios in which the adversary mounts an active attack by intercepting messages (either from a user or from a server) in a running session.
3. Reveal(): This oracle enables the adversary to learn the session key of a user in any running session.
4. Corrupt(): The adversary is allowed to corrupt one out of the two factors, that is, either the user's long-term password or the information stored on the user's smart card, but not both.
5. TestID(): This oracle provides an opportunity for the adversary to distinguish a user's real identity from a random value in the identity space.

It can be shown that the above five oracles address the key vulnerabilities of an offline dictionary attack, an impersonation attack, a known key attack, two-factor security, and user anonymity, respectively. Further details, scheme specifications, and formal proofs and performance evaluations refer to (*1*).

FIGURE 2. Proposed anonymous 3PAKE scheme (*5*). 3PAKE, three-party authenticated key exchange; *A,* user *A*, the patient; *B,* user *B,* the doctor; *TS,* trusted medical server.
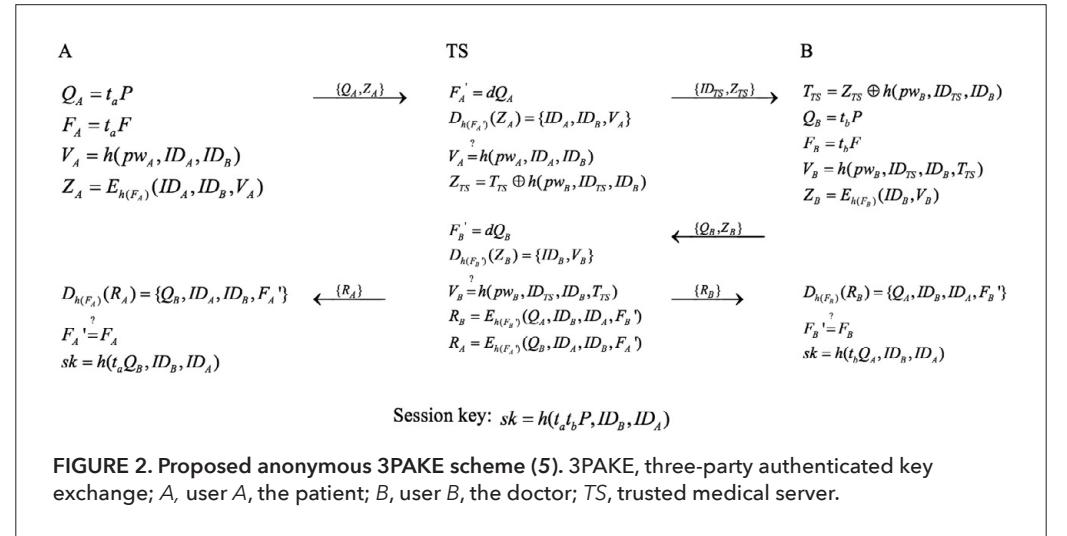
Three-party AKE

Another important part of our work involves the first anonymous three-party AKE (3PAKE) model and a protocol based on ECC for TMIS (*5*). In a traditional medical setting, a patient may have to drive several hours to a hospital or clinic and wait in a long queue to consult with a doctor in person. TMIS can make the process more effective, reliable, and efficient, saving time for the patient and enabling more convenient access to doctors. Moreover, when using TMIS, a patient's health files can be easily shared among different hospitals (*6*).

With TMIS, patients and doctors register with a trusted medical server and then authenticate using passwords. Patients would be able to consult with a doctor by communication via a secure video channel. The 3PAKE protocol was developed to fulfill mutual authentication between doctor and patient with the help of the medical server. It ensures that a potential hacker does not obtain access to the real identities of either the patient or the doctor. Additionally, 3PAKE can establish a session key shared by those parties running a 3PAKE protocol. To the best of our knowledge, no other 3PAKE scheme fully preserves patient privacy, since the user's identity is usually communicated in plain text during the authentication process. Furthermore, a scheme that utilizes ECC has the advantages of short secret keys and fast computational speed, factors that are greatly needed.

Figure 2 illustrates our anonymous 3PAKE scheme. The ProVerif (*7*) pi-calculus-based (*8*) formal verification tool was used to demonstrate that mutual authentication, session key security, and user anonymity can be achieved.

One-round authentication

The final part of our work is the dual achievement of a provably secure convertible proxy signcryption (CPS) scheme and a one-round authentication protocol for cloud computing scenarios based on this CPS scheme. In those scenarios, the user's authentication by the visited cloud (VC; the cloud service that a roaming user logs in to) can be conducted without interference from the home service cloud (HSC) if the user has already registered with the HSC (*9*).

FIGURE 3. Authentication protocol model for cloud computing (*9*). HSC, home service cloud; VC, visited cloud.
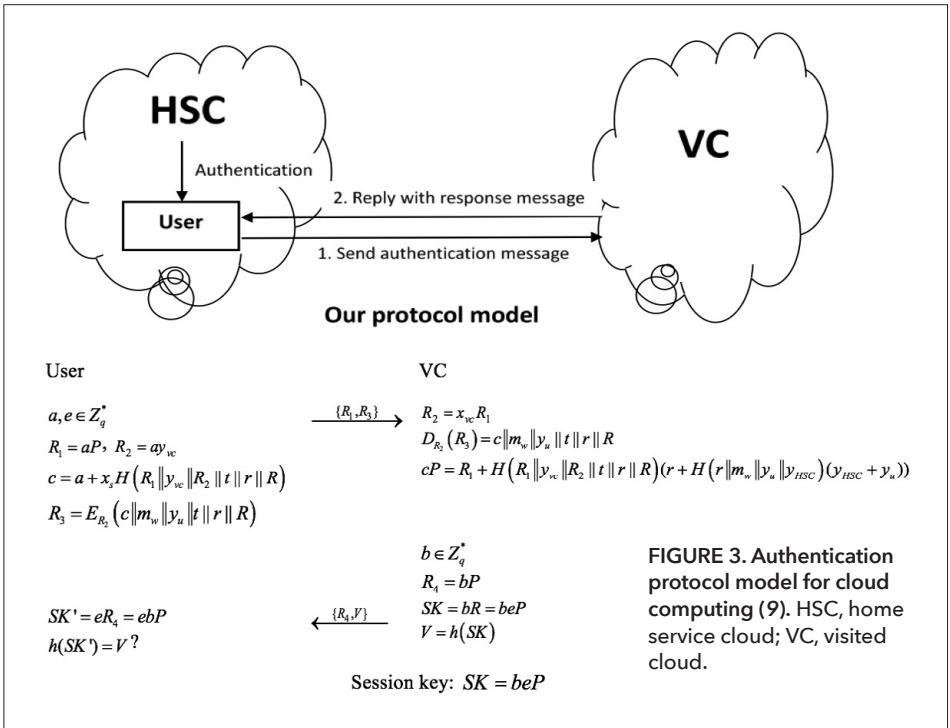
In cloud computing, the user can obtain many services provided by various public cloud servers; however, the user's registration is required for access. It is inconvenient to memorize multiple passwords; moreover, the clouds should be able to manage many accounts. Juang *et al.* presented an authentication scheme using the proxy signature to solve this problem (*10*). However, three limitations exist in their scheme: (1) an update of the user's public key is required to confer user anonymity; (2) HSC may become a bottleneck in the performance when many users in the same HSC need to simultaneously log onto different VCs; and (3) secret key sharing between HSC and VC must be required in advance.

Our proposed protocol overcomes the above limitations while also providing additional desirable features such as user privacy, nonrepudiation, freedom from the necessity for user public key updates, and no requirement for secret key sharing between HSCs and VCs. Moreover, the scheme is provably secure in the random oracle model with a higher efficiency than Juang *et al.*'s scheme (*10*).

Our one-round authentication protocol model for cloud computing scenarios is illustrated in Figure 3. The model is highly efficient because it requires only one round of communication and eliminates the need for HSC assistance.

Other advances

Our contributions to provably secure AKE protocols using enhanced security models include other security aspects. For instance, we developed schemes that are not based on ECC, but rather on chaotic maps (mathematical mappings or functions that exhibit some form of chaotic behavior). Farash and Attari (*11*) proposed the first chaotic-map–based 3PAKE (CM-3PAKE) protocol that does not use the server's public key, smart card, or symmetric cryptosystems. However, their protocol is vulnerable to impersonation attacks and offline password-guessing attacks. We therefore proposed an improved scheme that has the same advantages (*12*), but is more efficient with respect to communication and computation costs. We further proposed the first CM-3PAKE protocol without a timestamp that provides enhanced security measures and is also lightweight (*13*).

Lee *et al.* proposed the first password-free 3PAKE scheme and provided a well-organized Burrows–Abadi–Needham logic test to validate its security (*14*). However, their protocol is vulnerable to impersonation attacks. To address this weakness, we adopted biometrics technologies to construct an anonymous 3PAKE protocol based on chaotic maps (*15*). With a methodology similar to that outlined in (5), we use ProVerif to verify that our protocols (see *12*, *15*) achieve authentication, security, and anonymity.

References

1. Q. Xie *et al., IEEE Trans. Inf. Foren. Sec.* **12**,1382–1392 (2017).
2. P. Kocher *et al.,* in *Advances in Cryptology–CRYPTO '99, 19th International Cryptology Conference, Santa Barbara, California, USA, August 15–19, 1999, Proceedings,* pp. 388-397 (Springer, Berlin, 1999).
3. T. S. Messerges, E. A. Dabbish, R. H. Sloan, *IEEE Trans. Comput.* **51**, 541–552 (2002).
4. X. Li, Y. Zhang, *Secur. Commun. Netw.* **6**, 631–643 (2013).
5. Q. Xie, B. Hu, N. Dong, D. S. Wong, *PLOS ONE* **9**, e102747 (2014).
6. Q. Xie, J. Zhang, N. Dong, *J. Med. Syst.* **37**, 1–8 (2013).
7. M. Abadi, B. Blanchet, H. Comon-Lundh, in *Computer Aided Verification, 21st International Conference, CAV 2009, Grenoble, France, June 26-July 2, 2009, Proceedings,* pp. 35–49 (Springer-Verlag, Berlin, 2009).
8. M. Abadi, C. Fournet, in *Proceedings of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL 2001),* London, January 17-19, 2001, pp. 104–115 (ACM, New York, 2001).
9. Q. Xie *et al., Sci. Sin. Inform.* **42**, 303–313 (2012).
10. W. S. Juang *et al.,* in *1st Cross-Straits Conference on Information Security,* pp. 96–102 (Hangzhou, 2011).
11. M. S. Farash, M. A. Attari, *Nonlin. Dyn.* **77**, 399–411 (2014).
12. Q. Xie, B. Hu, T. Wu, *Nonlin. Dyn.* **79**, 2345–2358 (2015).
13. Q. Xie, J. Zhao, X. Yu, *Nonlin. Dyn.* **74**, 1021–1027 (2013).
14. C. C. Lee, C. T. Li, S. T. Chiu, Y. M. Lai, *Nonlin. Dyn.* **79**, 2485-2495 (2015).
15. Q. Xie *et al., Chin. Phys. B* **24**, 110505 (2015).
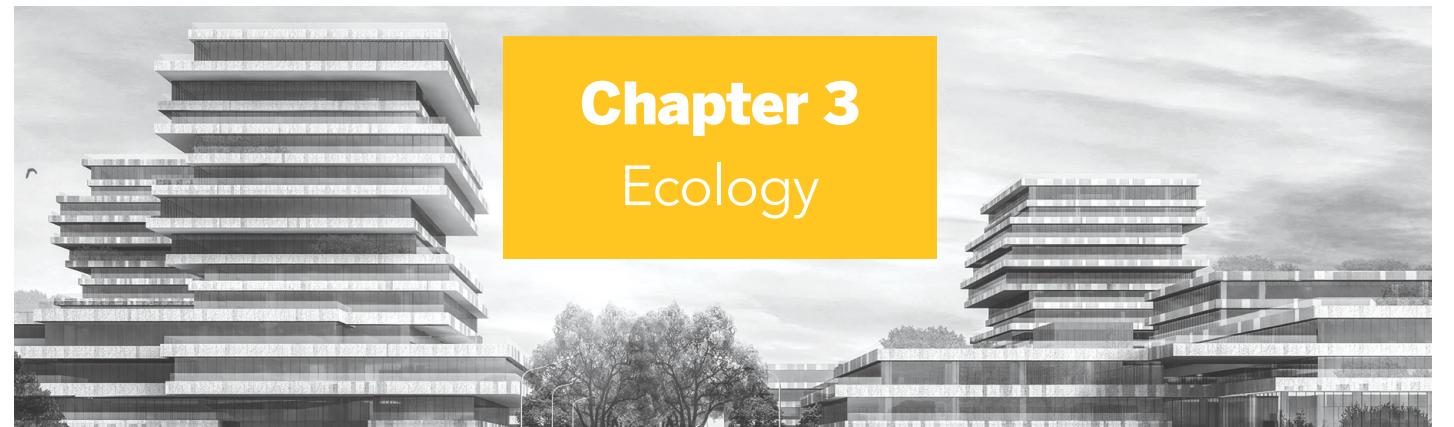
Chapter 3
Ecology

Living and after-life trait variation of plants in changing environments

Ming Dong[1]*, Yao-Bin Song[1], Guo-Fang Liu[2], Xu Pan[3], Yu-Kun Hu[3], Duo Ye[4], and Xiu-Fang Xie[5]

A plant functional trait is defined as any morphological, physiological, or phenological heritable feature measurable at the individual level that can respond to environmental changes and affect the environment (1–3). Trait-based approaches have shown strong predictability and generality in addressing many ecological topics, from the individual to ecosystem level, and from small scale to global scale, and have helped researchers in linking ecology to sociology. In this paper, we review our recent work on the responses and effects of plant functional traits in changing environments and, specifically, address the living traits of (clonal) plants and after-life traits related to leaf litter decomposability in the framework of trait-based ecology (4).

Intraspecific variation of plant functional traits

The basic assumption of current plant functional trait studies is that interspecific trait variations are more relevant than intraspecific trait variations (4). However, few studies have confirmed this assumption (5, 6); thus, intraspecific variation of plant functional traits across a large scale

cannot simply be ignored. *Phragmites australis*, a typical, widely distributed wetland perennial rhizomatous clonal plant, occurs in multiple climatic zones, ranging from temperate to tropic and from arid to humid regions in China and worldwide. This wide environmental range and inherent genetic variation cooperatively contribute to the large variation of functional traits in this species, making it a good model organism for assessing the variation and covariation of intraspecific functional traits on multiple spatial scales.

We used our field investigation data from 16 sites in undisturbed wetlands, covering most of the distribution range of *P. australis* in China, to examine the variation and covariation of leaf functional traits related to resource capture, also known as leaf economic traits, at three spatial scales in *P. australis* (7). *P. australis* showed considerable intraspecific variation in specific leaf area (SLA), leaf dry matter content, leaf nitrogen (N), and leaf phosphorus (P) at all tested spatial scales. Strong covariation among the four leaf economic traits was also found across the range of *P. australis*, resulting in a clear leaf economic spectrum (LES) at all spatial scales. The relationships between environmental variables and leaf economic trait variation were spatially specific. This study provided new empirical evidence of an LES within a species at multiple spatial scales, which implied that genetic and environmental variation, accompanying the resource trade-off, shaped intraspecific trait variation patterns.

Geographic patterns in leaf stoichiometry reflect plant adaptations to environments. We analyzed datasets of leaf N and P in *P. australis* and environmental (geographic, climate, and soil) variables from literature and field investigations in natural wetlands distributed in three climatic regions (subtropical, temperate, and highland) across China (8). We found no clear geographic patterns for leaf nutrient content in *P. australis* across China, except for leaf N/P ratio, which increased with altitude. Leaf N and N/P decreased with the mean annual temperature (MAT) and leaf N and P were closely related to soil pH, carbon (C)/N ratio, and available P. Climate and soil variables explained 62.1% of the total variation in leaf N, P, and N/P based on redun-

dancy analysis. These *P. australis* patterns differ from those of terrestrial plants, suggesting that changes in climate and soil properties can exert divergent effects on wetland and terrestrial ecosystems.

Variation of plant clonal traits in response to heterogeneous environments

Plant clonality is an adaptive strategy of plants by which clonal plants produce identical genetic individuals without meiosis (9). Clonal plants are widely distributed in many ecosystems (10, 11), especially in those with unfavorable physical conditions (10, 12). Clonal growth is dominant in monocotyledons; however, phylogenetic analyses reveal that clonality is in general widely distributed among plant species (10, 13). Functional clonal traits (groups) are defined as plant traits resulting from plant clonality that are related to whole genet fitness and/or ecosystem function (14).

To examine large-scale clonality–climate relationships, we compiled a clonality database with 4,015 vascular plant species in 545 terrestrial communities spanning tropic to temperate zones across China. Generally, plant clonality dominance increased with latitudinal gradient and the shift toward more stressful environments. However, the distribution of clonality in China was weakly correlated with latitude and climatic factors (MAT, mean annual precipitation, temperature seasonality, and precipitation seasonality). Woody clonal plants preferred wet or climatically stable environments, while herbaceous clonal plants were frequently found in cold, dry, or climatically unstable environments (15). When clonality–climate relationships were examined in different phylogenetic lineages, significant relationships were identified between clonality dominance and environmental factors for the predominant representatives of each of the taxonomic levels, although different taxonomic levels differed in the shape and strength of the clonality–climate relationships (16). These results highlight the importance of phylogenetic analysis in testing relationships between plant functional traits and environmental factors.

Clonal integration is the ability to share resources such as water, carbohydrates, and mineral nutrients between interconnected clone parts (ramets or groups of ramets) (17), which may be ecologically and evolutionarily advantageous. Using meta-analysis, we tested the generality of positive effects of clonal integration on plant fitness in terms of biomass and asexual reproduction (18). Generally, clonal integration increased the performance of whole clones and recipient parts, without decreasing that of donor parts. Interestingly, we also found that the effects of clonal integration on recipient parts were significantly positively related to clonal plant invasiveness. These results indicate that clonal integration could be one of the adaptive traits that explain clonal plant invasiveness.

Clonal plants could adjust their resource allocation to meet changing environments owing to functional trait plasticity. Using phylogenetic meta-analysis, we explored patterns of clonal plant biomass allocation in relation to resource availability. The results showed a greater plasticity

of vegetative growth traits compared with reproductive traits in clonal plants in response to changes in resource level. However, no evidence of trade-off patterns between clonal and sexual reproduction was found (19).

Plant after-life trait responses to multiple environments

Plant litter decomposition is an important ecosystem process in the terrestrial biogeochemical cycle. It is controlled mainly by litter quality, decomposers, and local environmental conditions (20). Many studies on litter decomposition have focused on dicotyledonous trees. However, in (sub)tropical areas, bamboo represents another main contributor to carbon cycling. Using 66 angiosperm species (including 31 bamboo species) arrayed across the phylogenetic tree, we set up a 1-year common garden experiment to test the effects of phylogeny, litter traits, and biotic factors on woody litter decomposition. The results showed higher tissue density and more recalcitrant wood chemistry in bamboo species wood, which contributed to its more rapid decomposition compared with dicotyledonous wood (21). The decomposition rate can be explained by the wood functional traits (density and dry matter content) of both bamboo and dicotyledons. Owing to their preference for deadwood with a lower density, termites had larger effects on the variance of wood decomposition. The deadwood's quality, diameter, and decomposer (termites) contributed differently to the decomposition of wood litter in woody monocotyledons and dicotyledons; these differences further our understanding of the driving mechanisms underlying (sub)tropical wood decomposition and its contribution to the global carbon cycle.

In arid ecosystems, many ecosystem processes are greatly affected by strong solar radiation (22). Using a factorial common garden experiment, we tested the effects of solar radiation and environmental moisture on leaf litter mass loss and decomposition rate in 13 typical arid species in Northern China (23) to quantify the effects of solar radiation on the leaf litter decomposition of different species. The effects of solar radiation on leaf litter decomposition were found to be species-specific. In species that decompose rapidly under shade, solar radiation can accelerate litter decomposition. Such interspecific variation can be explained by functional traits, especially SLA. This study was the first to quantify the effect of solar radiation on decomposition in multiple species, and its results also imply that plant community structure and composition in arid zones could also determine the effect of solar radiation on ecosystem processes.

Phylogenetic distances of coexisting species vary greatly within and among plant communities (24). However, little is known regarding their impact on decomposers and decomposition. Using a common garden experiment lasting 7 months, we conducted a litter mixture experiment with the leaf litter of 12 common coexisting temperate tree species from four families to test the effects of phylogenetic distances in litter mixtures on leaf litter decomposition (25). The average mass loss in different treatments was >50%. Phylogenetic distance did not appear to have any effect

[1]Key Laboratory of Hangzhou City for Ecosystem Protection and Restoration, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, China
[2]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China
[3]Institute of Wetland Research, Chinese Academy of Forestry, Beijing, China
[4]College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua, China
[5]School of Health Management and Education, Capital Medical University/Library of Capital Medical University, Beijing, China
*Corresponding author: dongming@hznu.edu.cn

on litter mass loss or on the abundance and diversity of invertebrate decomposers. However, litter mixtures with larger phylogenetic distance had lower microbial biomass and higher C/N ratios. Significant (and marginal) phylogenetic signals were found in leaf P, lignin concentrations, SLA, and nontannin phenols. Increased phylogenetic distance between litter mixtures may decrease the microbe digestibility of leaf litter species. This study is the first to examine the ecological impact of phylogenetic distance on decomposers and decomposition.

Summary

In the framework of trait-based ecology, our group addressed several interesting ecological questions related to intraspecific versus interspecific variation for clonal versus aclonal species from a local to regional scale and from individual to ecosystem levels. This work has contributed significantly to the development and advancement of our understanding of ecology at Hangzhou Normal University.

References

1. C. Violle *et al*., *Oikos* **116**, 882–892 (2007).
2. K. N. Suding *et al*., *Glob. Change Biol*. **14**, 1125–1140 (2008).
3. E. Garnier, M.-L. Navas, K. Grigulis, *Plant Functional Diversity: Organism Traits, Community Structure, and Ecosystem Properties* (Oxford University Press, Oxford, 2015).
4. B. Shipley *et al*., *Oecologia* **180**, 923–931 (2016).
5. C. H. Albert *et al*., *J. Ecol*. **98**, 604–613 (2010).
6. N. Pérez-Harguindeguy *et al*., *Austr. J. Bot*. **61**, 137–234 (2013).
7. Y.-K. Hu *et al*., *Front. Plant Sci*. **6**, 901 (2015).
8. Y.-K. Hu *et al*., *Sci. Rep*. **7**, 43018 (2017).
9. H. de Kroon, J. van Groenendael, *The Ecology and Evolution of Clonal Plants* (Backhuys Publishers, Leiden, 1997).
10. L. Klimeš, J. Klimešová, R. Hendriks, J. van Groenendael, in *The Ecology and Evolution of Clonal Plants*, H. de Kroon, J. van Groenendael, Eds. (Backhuys Publishers, Leiden, 1997), pp. 1–29.
11. M. H. Song, M. Dong, *J. Veg. Sci*. **13**, 237–244 (2002).
12. C. Körner, *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems* (Springer Verlag, Berlin, 2003).
13. J. M. van Groenendael, L. Klimeš, J. Klimešová, R. J. J. Hendriks, *Phil. Trans. R. Soc. Lond. B* **351**, 1331–1339 (1996).
14. J. H. C. Cornelissen, Y.-B. Song, F.-H. Yu, M. Dong, *Ann. Bot*. **114**, 369–376 (2014).
15. D. Ye *et al*., *PLOS ONE* **9**, e94009 (2014).
16. D. Ye *et al*., *Sci. Rep*. **6**, 26850 (2016).
17. H. de Kroon, B. Fransen, J. W. A. Rheenen, A. Dijk, R. Kreulen, *Oecologia* **106**, 73–84 (1996).
18. Y.-B. Song *et al*., *Oecologia* **171**, 317–327 (2013).
19. X.-F. Xie *et al*., *Front. Plant Sci*. **7**, 603 (2016).
20. W. K. Cornwell *et al*., *Ecol. Lett*. **11**, 1065–1071 (2008).
21. G. Liu *et al*., *J. Ecol*. **103**, 1214–1223 (2015).
22. A. T. Austin, L. Vivanco, *Nature* **442**, 555–558 (2006).
23. X. Pan *et al*., *Sci. Rep*. **5**, 13217 (2015).
24. D. S. Srivastava, M. W. Cadotte, A. A. M. MacDonald, R. G. Marushia, N. Mirotchnick, *Ecol. Lett*. **15**, 637–648 (2012).
25. X. Pan *et al*., *Proc. R. Soc. Lond. B* **282**, 20150103 (2015).

Effects of engineered nanoparticles on the anammox process in wastewater treatment

Zheng-Zhe Zhang and Ren-Cun Jin*

Since the start of the 21st century, the rapid development of nanotechnology has revolutionized many industrial fields. Today, engineered nanoparticles (NPs) are

College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, China
*Corresponding author: jrczju@aliyun.com

mass-produced and, because of their distinctive properties, are widely used in many industrial and consumer products, such as sunscreens, cosmetics, electronics, clothing, tires, and medicines (*1*). As a result, NPs are inevitably released into the environment during production, storage, use, and disposal (*2*). Most released NPs flow into wastewater treatment plants at the end of their lifecycle (*3*). Therefore, their potential impact on biological wastewater treatment has become an increasingly serious public concern as well as an area of intensive research (*2*, *4*). However, limited information is available on the NP-induced impact on the anaerobic ammonium oxidation (anammox) process, in which anammox bacteria oxidize ammonium to gaseous nitrogen using nitrite as an electron acceptor. The anammox-based process is widely perceived as a novel, environmentally friendly, and cost-effective alternative to traditional biological nitrogen-removal technology, which consists of nitrification and denitrification via nitrate (*5*, *6*). According to a recent survey (*7*), the number of full-scale anammox installations worldwide exceeded 100 in early 2015. The release of NPs poses an emerging threat to the

nitrogen removal performance of the anammox process. Therefore, we review here our recent research progress on the impact of NPs on the anammox process (Figure 1).

CuNPs show a difference with ZnONPs and CuONPs in acute toxicity

The acute toxicity effects of copper NPs (CuNPs), zinc oxide NPs (ZnONPs), and Cu oxide NPs (CuONPs) on anammox sludge (dominated by *Candidatus Kuenenia stuttgartiensis*) were first investigated by a series of batch tests (*1*). The presence of CuONPs or ZnONPs in concentrations as high as 50 mg/g suspended solid (SS) hardly influences anammox activity, lactate dehydrogenase (LDH) production, or reactive oxygen species (ROS) production. However, CuNPs at 1.25 mg/g SS could greatly inhibit anammox activity and further induce a significant release of LDH at 12.5 mg/g SS. Further investigation indicated that anammox flocculant aggregates (flocs) with a lower aggregation degree were more sensitive to the stress of CuNPs than anammox granules. This difference may be attributed to their disparate structural characteristics. The contact between CuNPs and anammox cells in flocs with a larger specific surface area and looser structure formed more easily than that in granular aggregates with a compact, multilayered structure. Generally, anammox bacteria secrete massive amounts of extracellular polymeric substances (EPS) into the extracellular space (*8*); thus, EPS are always the first medium that anammox cells are exposed to in wastewater containing CuNPs. CuNPs at concentrations below 12.5 mg/g SS were adsorbed mainly onto the aggregate surface, in which EPS play a dominant role. The binding of EPS with CuNPs attenuate their toxicity. Thus, anammox granules with more EPS show a greater tolerance to CuNPs than the flocs. The Fourier-transform infrared spectroscopy and three-dimensional excitation–emission matrix fluorescence spectra of EPS show that CuNPs might interact with tyrosine- or tryptophan-containing proteins and the C–O–C and –OH functional groups (*1*).

Combined toxicity of NPs and detoxification strategy

As the final sinks of NPs, wastewater treatment plants are likely exposed to a variety of NPs, even though they
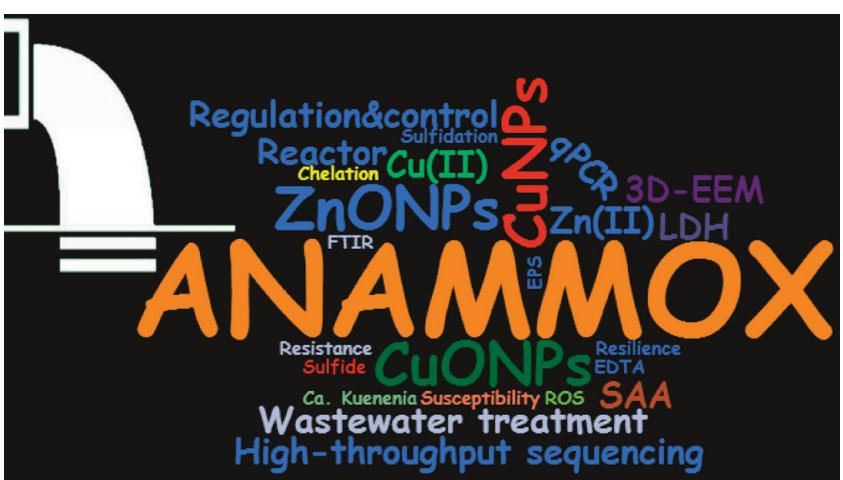
FIGURE 1. Keyword cloud representing recent research on the impact of nanoparticles on the anaerobic ammonium oxidation (anammox) process. Cu(II), copper (II); Zn(II), zinc (II); CuNPs, copper nanoparticles; CuONPs, copper oxide NPs; ZnONPs, zinc oxide NPs; qPCR, quantitative polymerase chain reaction; FTIR, Fourier-transform infrared spectroscopy; 3D-EEM, three-dimensional excitation–emission matrix; LDH, lactate dehydrogenase; ROS, reactive oxygen species; EDTA, ethylenediaminetetraacetic acid; Ca. Kuenenia, *Candidatus Kuenenia stuttgartiensis*; SAA, specific anammox activity; EPS, extracellular polymeric substances.

enter the waste stream at different locations. CuNP-induced cell membrane damage might enhance intracellular accumulation of NPs if other NPs are also present in the wastewater, such as CuONPs and ZnONPs, even if they themselves cannot damage the cell membrane. Therefore, it is also necessary to understand the combined impacts of such NPs on anammox consortia. Our recent study showed that the coexistence of CuONPs or ZnONPs at 5 mg/g SS did not enhance or weaken the toxicity of CuNPs (*9*).

The toxicity of NPs is widely considered to be mainly or at least partially attributed to the amount of toxic metal ions they release (*4*, *10*). Thus, the feasibility of two possible detoxification strategies—chelation and sulfidation—was also investigated. The addition of ethylenediaminetetraacetic acid (EDTA) alleviated the toxicity of CuNPs toward anammox granules. Although the concentration of Cu ions and the total Cu present in the wastewater increased, the level of Cu associated with cells significantly decreased in the presence of EDTA. In other words, the solubility of CuNPs was enhanced, but the bioavailability of Cu reduced, owing to chelation of Cu ions by EDTA. As mentioned above, the adsorption of NPs is generally the first step in wastewater treatment with anammox sludge. The complexation of Cu ions with EDTA is thermodynamically favorable, and exhibits greater stability than complexation with EPS. The fluorescence of proteins in EPS was enhanced following addition of EDTA, indicating that the interaction of Cu^{2+} with EPS was weakened. Consequently, the amount of Cu reaching the anammox cells was reduced. In turn, the toxicity of CuNPs depended on the amount of Cu arriving at the anammox cells rather than the amount of soluble Cu present in the wastewater. Similarly, the level of Cu ions in the wastewater solution was significantly reduced after the introduction of sulfide (S^{2-}) and remained at a stable and low level. However, the amount of Cu associated with cells, as well as that deposited at the lower part of the reactor, significantly increased. The presence of S^{2-} not only passivated the CuNPs by surface sulfidation, but also deactivated the Cu ions by forming CuS precipitates. Moreover, the interaction of Cu with

the functional groups of EPS weakened. As a result, the suppressive effect that CuNPs exert on anammox activity was attenuated owing to the presence of a shield of S^{2-}. Furthermore, CuNP-induced cell membrane damage became insignificant in the presence of EDTA or S^{2-}. This is probably because the deactivated Cu ions or passivated CuNPs could not damage the bacterial cell membrane. Although the amount of cell-associated Cu also significantly increased owing to the presence of CuONPs, the inhibition of anammox activity by CuNPs was not enhanced. Thus, it may be more precise to say that CuNP toxicity depends on the amount of active Cu reaching the anammox cells.

Susceptibility, resistance, and resilience of anammox consortia to long-term stress of CuNPs

CuNPs were added to the influent to monitor the long-term performance of the anammox reactor under the stress of CuNPs (*11*). CuNPs at 1.0 mg/L did not affect the performance; however, the reactor lost nearly all its nitrogen-removal capacity after being exposed to 5 mg/L of CuNPs for approximately 1 month; the relative abundance of anammox bacteria (*Ca. Kuenenia*) decreased greatly. After the withdrawal of CuNPs from the influent, the nitrogen-removal capacity of the anammox biomass completely recovered within 70 days. The above study, based on batch experiments, reported that the anammox activity was inhibited by 15.9% in the presence of 1.25 mg/g SS CuNPs (total load equivalent to 5 mg/L), and the inhibitory effect was aggravated when the exposure time was extended to 24 h (*1*). However, after the withdrawal of CuNPs from the solution and subsequent cultivation for 24 h, the activity of the anammox granules slightly recovered. The susceptibility and resilience of anammox bacteria to the stress of CuNPs were amplified in a continuous-flow experiment.

Notably, the enrichment of copper-resistance genes was detected in the anammox consortia when overwhelmed by the high stress of CuNPs (*11*). These genes belong to the *cus*, *cop*, and *pco* systems for eliminating excess intracellular Cu. This effect indicated the adaptive ability of anammox biomass during long-term stress. The Cu-resistance genes responsible for the removal of intracellular Cu could also have an important role in the subsequent recovery of the biomass. Overall, the anammox consortia showed susceptibility to the shock of CuNP introduction, resistance during long-term acclimatization to CuNPs, and resilience after the withdrawal of CuNPs.

Transient disturbance of ZnONPs enhances the resistance and resilience of the anammox process

Unlike CuNPs, ZnONPs showed no acute toxicity effect on anammox consortia (*1*). However, the actual effect of NPs may be not visible during just a few hours of exposure; hence, long-term observations are required to verify the effects. The chronic response of anammox consortia to ZnONP exposure was therefore investigated in terms of the reactor performance and community dynamics

(*12*). The presence of 1–5 mg/L ZnONPs did not affect the reactor performance, but nitrogen removal capacity was reduced by 90% when 10 mg/L ZnONPs was introduced over a 3-day period. In this case, anammox activity was significantly inhibited, but no significant stimulation in ROS production or extracellular LDH activity was observed. Therefore, the inhibition was presumably due to the accumulation of toxic Zn(II) ions in anammox cells. However, the resistance and resilience of this anammox reactor to ZnONPs were enhanced by a repeated "shock-recovery" mode. The upregulated abundance of Zn(II)-exporter ZntA might contribute to this enhanced resistance. In addition, these repeated transient disturbances improved the functional specificity of the anammox consortia despite the reduction of its diversity.

Conclusions and outlook

Overall, our studies indicated that the effects of NPs on the anammox process are complex and depend on the properties of NPs, the composition of wastewaters, and the condition of the bioreactors. This complexity indicates that the application of NPs should be extensively reviewed and revised. Their potential negative impacts on functional microorganisms in manmade ecological systems need to be weighed against the benefits of their application. Given the prevalence of NPs, their potential risks to the anammox-based process should be of great concern. While our current studies provide useful references for evaluating and controlling the risk of NPs to the anammox process, further investigations are still required. Although such NPs cannot oxidatively damage the anammox cells in the system without oxygen and light, their potential effects on the one-stage partial nitrification and anammox process may be disparate due to the presence of oxygen.

References

1. Z. Zhang *et al.*, *Bioresour. Technol.* **235**, 281–291 (2017).
2. D. Wang, Y. Chen, *Crit. Rev. Biotechnol.* **36**, 816–828 (2016).
3. S. K. Brar, M. Verma, R. D. Tyagi, R. Y. Surampalli, *Waste Manag.* **30**, 504–520 (2010).
4. S. Eduok *et al.*, *Ecotox. Environ. Safe.* **95**, 1–9 (2013).
5. M. C. M. van Loosdrecht, D. Brdjanovic, *Science* **344**, 1452–1453 (2014).
6. B. Kartal, J. G. Kuenen, M. C. M. van Loosdrecht, *Science* **328**, 702–703 (2010).
7. S. Lackner *et al.*, *Water Res.* **55**, 292–303 (2014).
8. Z. Zhang *et al.*, *J. Hazard. Mater.* **300**, 838–846 (2015).
9. Z. Zhang *et al.*, *J. Hazard. Mater.* **334**, 49–58 (2017).
10. Z. Wang, L. Zhang, J. Zhao, B. Xing, *Environ. Sci. Nano* **3**, 240–255 (2016).
11. Z. Zhang *et al.*, *Bioresour. Technol.* **241**, 35–43 (2017).
12. Z. Zhang *et al.*, *Sci. Total Environ.* **622**–**623,** 402–409 (2018).

Chapter 4
Life sciences
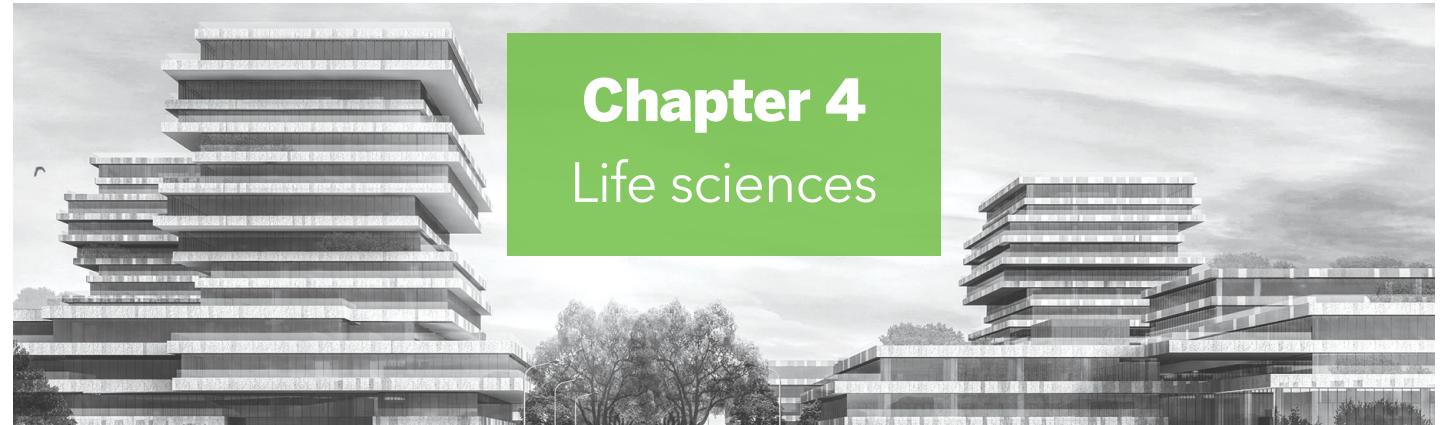
New prospects for targeting telomerase in cancer

Miao Wang, Jiang Liu, Deqiang Ding, Junzhi Zhou, Peng Xi, Ning Liu, Xiaoying Wu, Lu Xu, and Yu-Sheng Cong*

Telomeres are functional complexes at the termini of linear chromosomes, made up of tandem repeat DNA sequences and associated proteins (*1*). Mammalian telomeres comprise tandem TTAGGG repeats bound by a specialized protein complex. Apart from ensuring genome stability by protecting the ends of linear chromosomes from degradation and preventing cellular repair proteins from mistaking the chromosome ends as damaged DNA (*2*), telomeres are also critically involved in several cellular processes contributing to the pathology of aging and cancer (*3–5*).

Neoplastic transformation must overcome two tumor suppression barriers–senescence and apoptosis–and progress through multistep tumorigenic processes before attaining a state of malignancy (*6*). During the process of replicative immortality and oncogenic transformation, telomerase activation plays a fundamental role in bypassing the cellular senescence and apoptosis barriers (*7*).

The telomerase enzyme comprises a catalytic telomerase reverse transcriptase (TERT) subunit and an RNA component (telomerase RNA) that serves as a template for telomeric DNA synthesis (*8*). Telomerase is usually inactive in normal human somatic cells, having limited proliferative capacity owing to transcriptional repression of the *TERT* gene during embryonic differentiation. However, the enzyme remains active in certain highly proliferative tissues including male germ cells, stem cell populations, and acti-

vated lymphocytes. Moreover, telomerase is reactivated in over 90% of cancerous cells during cellular immortalization and transformation (*9*).

Such telomerase activation is considered one of the six most common events of cancer development (*10*). Thus, gaining a better understanding of telomerase function and regulatory mechanisms may provide novel insights into the development of better therapeutic strategies and prognostic capabilities for cancer treatment. Recent progress from our and other groups suggests that beyond its canonical function in telomere maintenance, human telomerase additionally functions in a telomere-independent manner in cancer. Therefore, it is important to understand the actions of telomerase regarding both telomere-dependent and -independent mechanisms if we are to develop effective therapeutic strategies.

Regulation of human telomerase

The biological importance of telomerase in aging and cancer has attracted substantial attention in basic biological research and the biomedical industry. Telomerase activity is subject to complex regulation by diverse intracellular and extracellular signaling pathways and is dependent on cell type, tissue environment, and physiopathological conditions (*9*). Transcriptional regulation of the human *TERT* gene (h*TERT*) is the primary, rate-limiting step in telomerase activation in most cells. However, telomerase regulation under certain physiological and pathological conditions is more complex and less well understood.

We recently reported that endoplasmic reticulum (ER) stress transiently activates h*TERT* expression in human cancer cell lines and murine primary neural cells. Downregulation of h*TERT* sensitizes cells to undergo apoptosis under ER stress, whereas upregulation of h*TERT* protects against ER stress-induced cell death (*11*). These findings indicate a functional link between ER stress and telomerase in cancer. Since cancer cells frequently experience ER stress due to exposure to a tumor microenvironment characterized by hypoxia and nutrition deprivation, h*TERT* upregulation during tumorigenesis may confer cancer cells with resistance to apoptotic cell death resulting from a cytotoxic tumor microenvironment or in response to anticancer drugs.

Institute of Aging Research, Hangzhou Normal University School of Medicine, Hangzhou, China
*Corresponding author: yscong@hznu.edu.cn

Role of TERT in transcriptional regulation

In humans, telomerase is active early in embryonic development but is suppressed in most adult tissues; however, it is upregulated or reactivated in the majority of cancerous cells (*12*, *13*). Telomerase reactivation and telomere maintenance sets the foundation for cellular immortalization and the malignant conversion of cancer cells (*7*).

Several recent studies from our and other groups suggest that telomerase may also promote cancer progression independent of its telomere-lengthening enzymatic activity (*3*, *14*). For example, the nuclear factor-kappa beta (NF-κB) pathway is critical for inflammation and cancer progression, and telomerase can directly regulate NF-κB-dependent transcription by binding to the NF-κB p65 subunit, which is recruited to the promoters of NF-κB target genes, including proinflammatory cytokines interleukin-6 and tumor necrosis factor α (*15*, *16*). In addition, we reported that hTERT protein activates expression of the vascular epithelial growth factor (*VEGF*) gene by interacting with the *VEGF* promoter and the transcription factor Sp1 (*17*). Moreover, hTERT binds to Sp1 in vitro and in vivo, and stimulates angiogenesis in an Sp1-dependent manner. Significantly, ablation of the mouse *TERT* (m*TERT*) gene in the first generation of *Tert*-null mice attenuated tumor growth accompanied by VEGF downregulation (*17*). Thus, h*TERT* upregulation in cancer may enhance tumor angiogenesis by upregulating *VEGF* expression.

Together, these observations indicate that hTERT has dual functions. On the one hand, it maintains telomere length and unlimited proliferation capacity; on the other, it functions as a transcription regulator to facilitate oncogenic transformation by regulating the expression of cancer-related genes.

Roles of TERT in cancer progression

Cancer progression is an integrated, multistep process that involves cellular immortalization, cell polarity changes, migration, angiogenesis, and metastasis, all of which require the coordinated regulation of gene expression programs (*10*).

h*TERT* expression increases during cancer progression

FIGURE 1. Telomere-dependent and -independent functions of hTERT in tumorigenesis. Human telomerase reverse transcriptase (hTERT) interacts with telomerase RNA to synthesize telomeric DNA and to maintain telomere homeostasis, which confers unlimited replicative capacity to the transformed cells. However, under various physiological and pathological conditions, hTERT interacts with different cellular factors to alter gene expression profiles and confer transformed cells with cancer-specific capacities during tumorigenesis. TERT, telomerase reverse transcriptase; BRG1, Brahma-related gene 1; TERC, telomerase RNA component; VEGF, vascular endothelial growth factor; MMPs, matrix metalloproteinases; IL-6, interleukin 6; TNF-α, tumor necrosis factor alpha; EMT, epithelial–mesenchymal transition; ROS, reactive oxygen species.

and is associated with an unfavorable prognosis in certain human cancers (*18–20*). In mouse models, telomerase is upregulated in several tumor types despite the long telomere reserves in mouse cells (*21*). Exogenous h*TERT* expression, along with oncogenic RAS and the SV40 early region, facilitates the tumorigenic transformation of fibroblasts, kidney epithelial cells, and mammary epithelial cells (*22–24*). These findings suggest other telomere-independent roles for telomerase in cancer progression. However, the detailed molecular mechanisms underlying telomerase action in cancer, beyond telomere maintenance, remain elusive.

In addition to regulating the Sp1- and NF-κB–dependent transcription (as discussed in the previous section), another reported function of hTERT, independent of its telomere-lengthening and telomerase activity, is to promote epithelial–mesenchymal transition (EMT). Using a mouse gastric cancer model, Xu and colleagues (*25*) reported that ectopic h*TERT* expression promotes EMT and the acquisition of pluripotency of gastric cancer cells.

Since hTERT regulates several important cancer-related genes (matrix metalloproteinases, cytokines, *VEGF*, and EMT-related genes) (*15–17*, *25*), it is conceivable that activation of telomerase during tumorigenesis not only provides infinite proliferative potential by maintaining telomere length but may also influence the tumor microenvironment to favor the invasive growth and metastasis of cancer cells (Figure 1).

Concluding remarks

Telomerase plays a critical role in the initiation and development of cancer by maintaining telomere length and promoting the proliferative capacity of cancer cells. Evidence from basic and clinical studies indicates the additional involvement of TERT in several essential cellular functions, including transcriptional regulation, mitochondria-mediated apoptotic cell death, cell survival under stress, oncogenic transformation, and EMT. These activities may provide transformed cells with additional cancer-specific capacities beyond immortality during tumorigenesis. In particular, the roles of telomerase as a transcriptional regulator of the Wnt/β-catenin, Sp1/VEGF, and NF-κB signaling pathways provide important insights into the regulatory function and biochemical properties of TERT at the molecular level. A future challenge is to identify the targets and downstream signaling molecules of telomerase, which will provide important insights into its role in human cancer. Such mechanistic studies would provide a novel perspective for developing specific strategies for the therapeutic manipulation of telomerase in cancer.

References

1. E. H. Blackburn, *Nature* **408**, 53–56 (2000).
2. T. de Lange, *Science* **326**, 948–952 (2009).
3. P. Martinez, M. A. Blasco, *Nat. Rev. Cancer* **11**, 161–176 (2011).
4. E. Sahin *et al.*, *Nature* **470**, 359–365 (2011).
5. R. Benetti, M. Garcia-Cao, M. A. Blasco, *Nat. Genet.* **39**, 243–250 (2007).
6. W. E. Wright, J. W. Shay, *Exp. Gerontol.* **27**, 383–389 (1992).
7. L. Xu, S. Li, B. A. Stohr, *Annu. Rev. Pathol.* **8**, 49–78 (2013).
8. K. Collins, J. R. Mitchell, *Oncogene* **21**, 564–579 (2002).
9. Y. S. Cong, W. E. Wright, J. W. Shay, *Microbiol. Mol. Biol. Rev.* **66**, 407–425 (2002).
10. D. Hanahan, R. A. Weinberg, *Cell* **144**, 646–674 (2011).
11. J. Zhou *et al.*, *Aging Cell* **13**, 197–200 (2014).
12. W. E. Wright, M. A. Piatyszek, W. E. Rainey, W. Byrd, J. W. Shay, *Dev. Genet.* **18**, 173–179 (1996).
13. J. W. Shay, S. Bacchetti, *Eur. J. Cancer* **33**, 787–791(1997).
14. Y. S. Cong, J. W. Shay, *Cell Res.* **18**, 725–732 (2008).
15. A. Ghosh *et al.*, *Nat. Cell Biol.* **14**, 1270–1281 (2012).
16. D. Ding, P. Xi, J. Zhou, M. Wang, Y. S. Cong, *FASEB J.* **27**, 4375–4383 (2013).
17. N. Liu *et al.*, *Nucleic Acids Res.* **44**, 8693–8703 (2016).
18. L. Terrin *et al.*, *Clin. Cancer Res.* **14**, 7444–7451 (2008).
19. N. W. Kim *et al.*, *Science* **266**, 2011–2015 (1994).
20. B. Ivan *et al.*, *Clin. Cancer Res.* **6**, 452–459 (2000).
21. E. Gonzalez-Suarez *et al.*, *EMBO J.* **20**, 2619–2630 (2001).
22. B. Elenbaas *et al.*, *Genes Dev.* **15**, 50–65 (2001).
23. W. C. Hahn *et al.*, *Nature* **400**, 464–468 (1999).
24. W. C. Hahn *et al.*, *Mol. Cell Biol.* **22**, 2111–2123 (2002).
25. Z. Liu *et al.*, *Oncogene* **32**, 4203–4213 (2012).

The development of recombinant hNeuritin and its therapeutic potential in nerve regeneration

Haiyan Wang[1], Yu Hong[1], Yuqin Niu[2], Rui Gao[2], Rong Chen[1], Jingling Zhu[2], Xiaoming Song[3], Yunhua Zhang[2], Guijie Liu[3], Yi Yang[1], Jin Huang[2]*, and Lei Yang[1]*

Neurological disorders are an important cause of disability and death worldwide (*1*). Recently, much effort has been devoted to developing treatments for these ailments. Although their causes vary, they share a common pathogenesis associated with neuronal cell death or dysfunction (*2*). Thus, the key to neuroprotection and therapeutic intervention is sustaining neuronal survival and promoting nerve regeneration.

Neurotrophic factors are proteins that mediate neuronal survival, nervous system development, and many other neuronal physiological functions. Given their high potency, as demonstrated by many animal studies, neurotrophic factors are regarded as candidate therapeutic agents for use in neurological disorders (*3*).

Neuritin, a neurotrophic factor involved in nervous system development and plasticity (*4*), promotes neurite growth, regulates the formation and maturation of synapses, and inhibits neuronal apoptosis (*5*, *6*). Neuritin is upregulated in structures demonstrating neural plasticity (*7*), and its upregulation correlates with functional recovery after central and peripheral nervous system injury (*8*, *9*). Given that changes in neuritin gene expression coincide with functional recovery after neurological disorders, it has been hypothesized that neuritin may be involved in endogenous recovery processes following injury. However, neuritin expression is largely restricted to nervous system development and plasticity-related changes, and is low under normal physical conditions.

To study neuritin function, we performed a series of experiments that included development of a recombinant neuritin protein, exploration of its therapeutic potential, and elucidation of the mechanisms involved in regeneration. Recombinant hNeuritin was developed using the *Pichia pastoris* secretory expression system, and effectively promoted neurite outgrowth of undifferentiated PC12 cells in vitro (*10*). Importantly, recombinant hNeuritin displayed potent neuroprotective properties, as illustrated by the accelerated functional recovery and axonal regeneration of rat sciatic

[1]Preventive Medicine, School of Medicine, Hangzhou Normal University, Hangzhou, Zhejiang, China
[2]The Key Laboratory of Xinjiang Endemic and Ethnic Diseases and Department of Biochemistry, Shihezi University School of Medicine, Shihezi, China
[3]Laboratory Animal Center, Hangzhou Normal University, Hangzhou, Zhejiang, China
*Corresponding authors: huangjin623@163.com (J.H.), yanglei62@hznu.edu.cn (L.Y.)

FIGURE 1. Development and bioactivity analysis of recombinant hNeuritin. (A) Construction of the recombinant pPIC9K-neuritin plasmid. **(B)** Purification of recombinant hNeuritin. Peaks 1 and 2 indicate the effluent and elution protein (recombinant hNeuritin) peaks, respectively. **(C)** Separation and visualization of purified neuritin by SDS-PAGE and Coomassie staining. Lanes 1 through 4 indicate the protein molecular weight marker, the culture supernatant, the protein effluent, and the eluted protein, respectively. **(D)** Western blot analysis showing purified hNeuritin protein. Lane 1: protein marker; Lane 2: recombinant neuritin (approximately 11 kDa). **(E)** MALDI-TOF mass spectrometry was used to determine the precise molecular weight of purified recombinant hNeuritin. **(F)** Effects of recombinant hNeuritin on neurite outgrowth in undifferentiated PC12 cells (magnification: ×200). Cells were treated with PBS and 1 μg/mL recombinant neuritin for control and neuritin groups, respectively. MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; PBS, phosphate buffered saline.

FIGURE 2. Recombinant hNeuritin promotes structural and functional recovery of rat sciatic nerve injury. (A) Sciatic Function Index of sciatic nerves in different rat groups at the indicated time points (n = 5, $^*p < 0.05$, $^{**}p < 0.01$ vs. NS group). **(B)** Recovery of nerve conductive velocity in different groups at the indicated time points (n = 5, $^*p < 0.05$, $^{**}p < 0.01$ vs. NS group). **(C)** Sevier staining of nerve fibers in the NS and N5 (hNeuritin 5 μg) groups 4 weeks after injury. **(D)** GAP43 and NF200 expression in the NS and N5 groups 4 weeks after injury, and the quantification of the OD ratio for each (n = 3, $^*p < 0.05$ vs. NS group). NS, normal saline; N3μg, hNeuritin 3 μg; N5μg, hNeuritin 5 μg; OD, optical density.
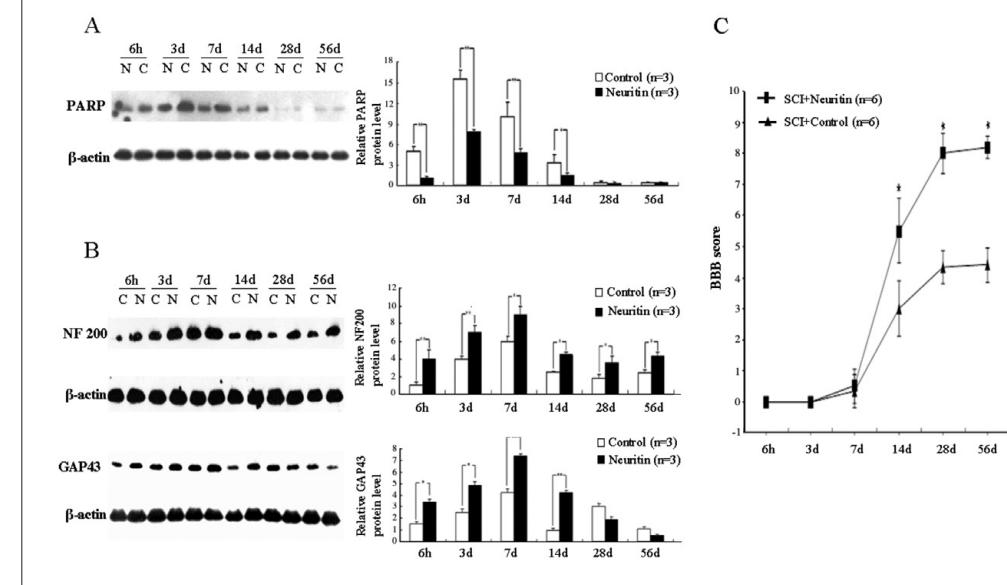
FIGURE 3. Recombinant hNeuritin accelerates nerve regeneration after rat spinal cord injury. (A) Western blot analysis of PARP expression at the indicated time points and quantification of its expression normalized to β-actin (n = 3, $^*p < 0.05$, $^{**}p < 0.01$ vs. control group). N, hNeuritin treated; C, control. **(B)** Western blot analysis of NF200 and GAP43 expression at the indicated time points and quantification of their expression normalized to β-actin (n = 3, $^*p < 0.05$, $^{**}p < 0.01$ vs. control group). N, hNeuritin treated; C, control. **(C)** Basso, Beattie, and Bresnahan (BBB) scores of locomotor function in control and recombinant hNeuritin groups at the indicated time points (n = 6, $^*p < 0.05$ vs. control groups).

nerve (*11*) and repair of spinal cord injury (SCI) (*12*) in vivo. Furthermore, we have revealed the mechanisms by which neuritin enhances regeneration: Neuritin promoted neurite growth by inhibiting Notch signaling through interaction with NEURL1 (*13*). The link between neuritin and Notch signaling illustrated their joint participation in neurite regeneration when used to treat neurological disorders.

Discovery of human neuritin

A human neuritin complementary DNA (cDNA) of 1618 bp was identified in a human fetal brain cDNA library, as reported in 2001 by Jin Huang, Lei Yang, and colleagues (*14*). The cDNA fragment contained an open reading frame encoding a 142-aa protein with a calculated mass of 15.3 kDa. Database searches and comparisons revealed that the cDNA had 98% homology with the rat neuritin gene.

Development of bioactive recombinant hNeuritin

The fragment encoding the active human neuritin sequence (nucleotides 652–1077), which lacks the signal peptide and the glycosylphosphatidylinositol domain was amplified from a previously constructed pcDNA3.1-His-neuritin vector using primers containing EcoRI–BamHI restriction sites (*10*). The polymerase chain reaction (PCR) product was inserted into the shuttle plasmid pPIC9K after EcoRI and NotI digestion (Figure 1A); this pPIC9K-neuritin plasmid was transformed into DH5α, positive colonies were extracted, and these were transformed into *P. pastoris* GS115 by electroporation. Positive recombinant clones were screened using G418 and were verified by PCR screening (*10*).

The recombinant colonies were cultured at 30°C in yeast medium containing methanol, at a final concentration of 1%, for induction. Three days later, the medium was collected and centrifuged; recombinant hNeuritin was purified from the supernatant through nickel iron affinity chromatography, identified by Western blot and mass spectrometry, and freeze-dried. The recombinant hNeuritin was secreted into the supernatant and was effectively purified, as confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Western blotting (Figure 1B–D). The precise molecular weight of the peptide was then analyzed using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry as 10.97 kDa (Figure 1E). This soluble product significantly promoted neurite outgrowth of undifferentiated PC12 cells in vitro (Figure 1F) (*10*).

Therapeutic potential of recombinant hNeuritin in neurological disorders

Given the differences in recovery between peripheral and central nervous system injuries, we explored the efficacy of using recombinant hNeuritin as treatment for both sciatic nerve injury and SCI, using nerve fiber and neuronal cells, respectively.

Therapeutic potential of recombinant hNeuritin in rat sciatic nerve injury

A rat sciatic nerve injury model was established using the clamp injury method: The left sciatic nerve (injured side) was compressed 0.5 cm below the piriformis to one ratchet of a gearless mosquito forceps for 30 s. In the sham group, the sciatic nerve was exposed, without injury. In the experimental groups, the sciatic nerves were clamp-injured and randomly assigned to the recombinant hNeuritin treatment groups or the normal saline (NS) control group. The recombinant hNeuritin treatment groups received gelatin sponges infiltrated with 3 μg or 5 μg hNeuritin in 1 mL NS (1.5 μg and 2.5 μg neuritin per 100 g body weight), and the NS control group received sponges infiltrated with 1 mL NS, placed at the injury site. Repeated intramuscular injections of recombinant hNeuritin or NS, at the same doses, were given at the injury site for 5 weeks (*11*).

At 2, 4, 6, and 8 weeks after injury, all groups underwent sciatic nerve function analysis, after which the rats were sacrificed and the entire sciatic nerve from both sides were collected and fixed for subsequent morphological analysis. Delivery of a high concentration of exogenous recombinant hNeuritin (5 μg) to the injured nerves promoted the expression of GAP43 and NF200 (Figure 2D) from the surviving neurons. Under hNeuritin stimulation, axons and dendrites sprouted and the surface area of the nerve terminals increased, contributing to the establishment of new synaptic contacts and stabilization of the cytoskeleton, providing a milieu for the ordered arrangement of the newly formed nerve fibers (Figure 2C). This arrangement enhanced the conduction velocity of the nerve fibers (Figure 2B) and finally accelerated their functional recovery (Figure 2A). These findings demonstrated that exogenously applied recombinant hNeuritin stimulated nerve regeneration and functional recovery and thus promoted the repair of injured sciatic nerves (*11*).

Therapeutic potential of recombinant hNeuritin in rat SCI

A rat SCI model was established using the weight-drop method (*12*): The T10 spinal cord (SC) segment was hit by a 15-g weight dropped from a height of 20 cm. Thirty minutes later, a 0.1-mm long polyethylene catheter was inserted into the subarachnoid space of the T12 SC segments of the experimental group rats, toward the T10 SC, as a sterile delivery channel through which gelatin sponges were placed locally on the injured tissue. Solutions of hNeuritin at a concentration of 30, 60, and 90 μg in 1 mL NS were prepared. Subsequently, rats were injected daily with 100 μL of these hNeuritin solutions or NS through the catheter tube for 1 week; in this way, the gelatin sponges were soaked, and released the solutions to the injured tissue. The sham group underwent SC exposure, but were not injured or catheterized.

At 6 h and 3, 7, 14, 28, and 56 days after SCI, the hind

FIGURE 4. Neuritin promotes neurite growth by inhibiting Notch signaling through interaction with NEURL1. (A) Endocytosis and degradation of NEURL1-mediated JAG1 in 293T cells expressing HA-JAG1, with or without neuritin treatment (top). Western blot analysis (bottom) of HA-JAG1, Flag-NEURL1, and His-Neuritin, *p < 0.01, scale bar = 10 μm. (B) Western blot analysis of Notch intracellular domain (NICD) and HES1 in Flag-NEURL1-transfected 293T cells, with or without His-Neuritin treatment, *p < 0.01. (C) Representative images (top) of neurites in pcDNA3.1-HA-JAG1 or pcDNA3.1 empty vector-transfected SH-SY5Y cells treated with recombinant hNeuritin or His peptide as negative control; scale bar = 50 μm. The positive cells were recorded as having one or more neurites (bottom left), while the mean length of neurites (bottom right) was calculated from over 100 cells in each group, *p < 0.01.
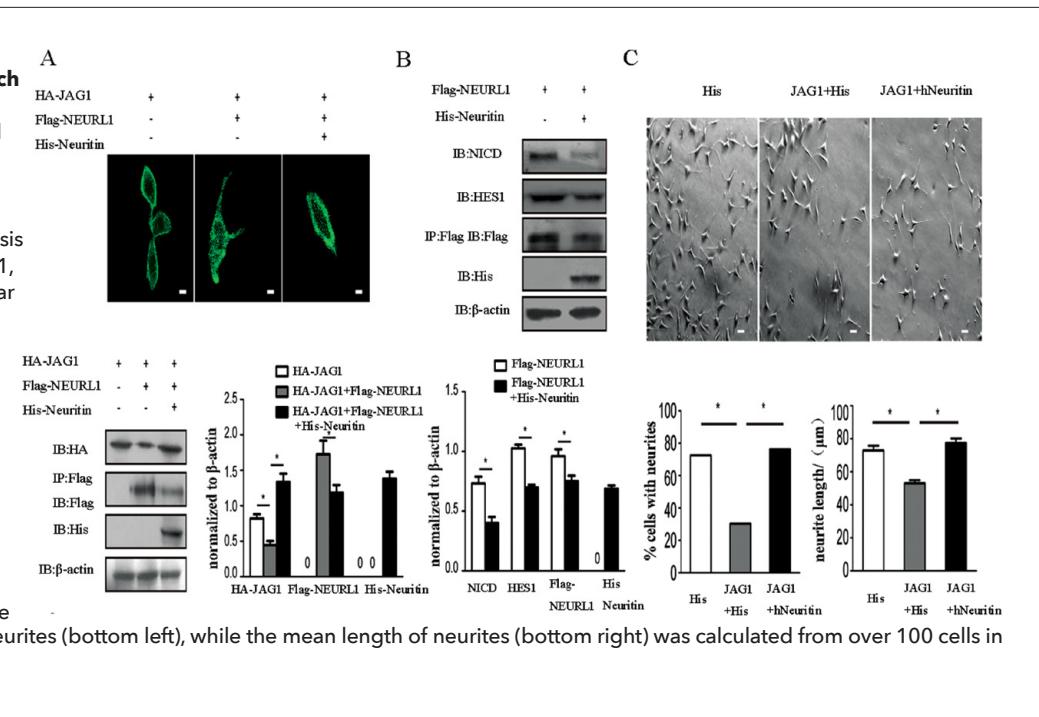
limb motor function was analyzed using the Basso, Beattie, and Bresnahan (BBB) scale, after which rats were sacrificed and tissues collected from 0.5 cm above and below the injury site at the T10 SC level (12). Tissues were fixed or placed in liquid nitrogen for subsequent morphological or Western blot analysis, respectively. The recombinant hNeuritin decreased neuronal apoptosis by preventing degradation of PARP protein (Figure 3A), resulting in increased expression of NF200 and GAP43 (Figure 3B). The latter may promote axonal and dendritic sprouting, leading to establishment of new contacts through the increased surface area of nerve endings, and contributing to motor function recovery after SCI (Figure 3C). These findings suggested that recombinant hNeuritin maintained neuronal survival and stimulated axonal regeneration, contributing to nerve regeneration and functional recovery after SCI (12).

Mechanism of neuritin action

NEURL1 is one of 57 genes encoding putative neuritin interactors, as determined by a yeast two-hybrid screen of a rat embryonic brain cDNA pool, in which neuritin was used as bait (13). Subsequent experiments confirmed the interaction between NEURL1 and neuritin and the inhibitory effect of neuritin on NEURL1. NEURL1 is a molecule that promotes endocytosis of Delta/Jagged 1 (JAG1), activating Notch signaling. Our findings indicated that the NEURL1-promoted endocytosis of Notch ligand JAG1, activation of the Notch receptor intracellular domain, and the expression of the hairy and enhancer of split-1 gene, a downstream target gene of Notch signaling, were significantly suppressed by neuritin overexpression (Figure 4A, B)(13). It has also been reported that Notch inhibits neurite outgrowth (15). We found that recombinant hNeuritin rescued Notch-mediated neurite retraction (Figure 4C). Thus, neuritin is an upstream and negative regulator of NEURL1, promoting neurite growth by inhibiting Notch signaling (13).

In summary, using recombinant hNeuritin, we have found that neuritin promotes functional recovery and regeneration of injured nerve fibers, and accelerates axonal regeneration and functional recovery of injured neurons, via a mechanism that involves Notch signaling. Furthermore, we have established protein expression and purification technologies for future investigations of recombinant hNeuritin. Although it is unclear whether the results from these animal studies will translate to human disorders, our findings suggest that hNeuritin is of clinical interest for the treatment of nerve injuries.

References

1. GBD 2015 Neurological Disorders Collaborator Group, *Lancet Neurol.* **16**, 877–897 (2017).
2. L. Berliocchi *et al.*, *Curr. Opin. Pharmacol.* **12**, 49–54 (2012).
3. Y. S. Levy, Y. Gilgun-Sherki, E. Melamed, D. Offen, *BioDrugs* **19**, 97–127 (2005).
4. E. Nedivi, D. Hevroni, D. Naot, D. Israeli, Y. Citri, *Nature* **363**, 718–722 (1993).
5. U. Putz, C. Harwell, E. Nedivi, *Nat. Neurosci.* **8**, 322–331 (2005).
6. E. Nedivi, G.Y. Wu, H.T. Cline, *Science* **281**, 1863–1866 (1998).
7. E. Nedivi, S. Fieldust, L. E. Theill, D. Hevron, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2048–2053 (1996).
8. S. Di Giovanni *et al.*, *FASEB J.* **19**, 153–154 (2005).
9. K. N. Fargo, T. D. Alexander, L. Tanzer, A. Poletti, K. J. Jones, *J. Neurotrauma* **25**, 561–566 (2008).
10. Y. Zhang *et al.*, *Appl. Microbiol. Biotechnol.* **99**, 8035–8043 (2015).
11. H. Wang *et al.*, *Front. Neurosci.* **10**, 589 (2016).
12. R. Gao *et al.*, *Hum. Gene Ther.* **27**, 544–554 (2016).
13. P. Zhang *et al.*, *Front. Mol. Neurosci.* **10**, 179 (2017).
14. J. Huang *et al.*, *Nat. Sci. Fudan University* **40**, 521–524 (2001).
15. N. Sestan, S. Artavanis-Tsakonas, P. Rakic, *Science* **286**, 741–746 (1999).

Chapter 5
Mathematics
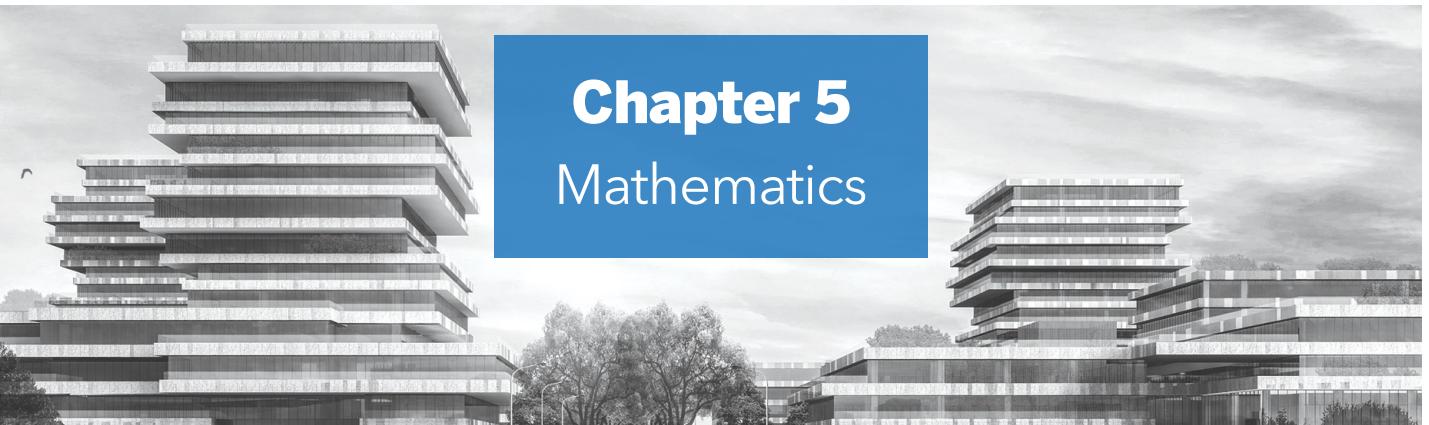
Iterative algorithms and an overview of recent results for fixed-point problems

Feng Gu

Let E be a metric space and $f : D(f) \subset E \to E$ be a function. Then, $x \in E$ is called a fixed point of f if $f(x) = x$. In many types of mathematical problems, examining the existence of a solution is equivalent to finding a fixed point of a suitable function. Therefore, finding a fixed point of a function is of utmost importance; this has applications in many fields, including biology, chemistry, economics, engineering, game theory, and physics. Fixed-point theorems provide conditions under which the functions have solutions. Fixed-point theory seamlessly combines topology, analysis, and geometry. Recently, many important (common) fixed-point theorems have been established for diverse spaces, such as partially ordered, multiplicative, and quasi-partial metric spaces, and G-, b-, and S-metric spaces [see (1–14) and references therein].

In 2006, Mustafa and Sims (1) first extended the definition of a metric space to the concept of generalized metric spaces, which can simply be called G-metric spaces. Subsequently, Mustafa and colleagues (2) obtained some fixed-point results for mappings that satisfied different contractive conditions in G-metric spaces. Shatanawi (3) provided proofs for some fixed-point theorems for Φ-mappings in G-metric spaces. Chugh and colleagues (4) obtained fixed-point results for mappings satisfying property P in G-metric spaces. Abbas and Rhoades (5) performed the first studies on common fixed-point problems in G-metric spaces; subsequently, many authors extended their ideas and obtained many common fixed-point theorems for the mappings satisfying different contractive conditions. Among these authors, Gu and Shatanawi (6) for the first time used common (E.A) properties to discuss common fixed-point problems of contractive mappings for three pairs of weakly compatible self-maps satisfying a generalized weakly G-contraction condition in generalized metric spaces, obtaining the following result:

Theorem GS. Suppose (E, G) is a G-metric space and $A, B, C, f, g, h : E \to E$ are six mappings such that (A, B, C) is a generalized weak G-contraction mapping of type A with respect to (f, g, h). If one of the following conditions is satisfied, then the following pairs (A, f), (B, g), and (C, h) have a common point of coincidence in E: (i) the subspace fE is closed in E, $AE \subset gE$, $BE \subset hE$, and two pairs (A, f) and (B, g) satisfy the common (E.A) property; (ii) the subspace gE is closed in E, $BE \subset hE$, $CE \subset fE$, and two pairs (B, g) and (C, h) satisfy the common (E.A) property; (iii) the subspace hE is closed in E, $AE \subset gE$, $CE \subset fE$, and two pairs (A, f) and (C, h) satisfy the common (E.A) property.

Moreover, if pairs (A, f), (B, g), and (C, h) are weakly compatible, then A, B, C, f, g, and f have a unique common fixed point in E.

In 2008, Bashirov and colleagues (8) introduced the definition of multiplicative metric spaces, studied multiplicative calculus, and derived a fundamental theorem of multiplicative calculus. In 2012, Florack and Assen (9) provided an example of multiplicative calculus application in biomedical image analysis. In 2011, Bashirov and colleagues (11) exploited the efficiency of multiplicative calculus as compared to that of Newtonian calculus. They showed that multiplicative differential equations are more suitable than ordinary differential equations to describe certain practical problems. Moreover, Bashirov and colleagues (8) demonstrated some interesting applications for multiplicative calculus. By using a multiplicative absolute value function, they also defined the multiplicative distance between two nonnegative real numbers as well as that between two positive square matrices. Recently, Abbas, Ali, and Suleiman (12) obtained results for a common fixed point of quasi-weak commutative mappings on a closed ball in

Department of Mathematics, Hangzhou Normal University, Hangzhou, Zhejiang, China
Corresponding author: mathgufeng@163.com (F. G.)

multiplicative metric spaces. Moreover, they also presented sufficient conditions for the existence of a common solution for the multiplicative boundary value problem.

In 2015, Gu and Cho (13) introduced the following two concepts and established a new common fixed-point theorem.

Definition GC1. The self-maps A and B of a multiplicative metric space (E,d) are said to be compatible if $\lim_{n\to\infty} d(ABy_n, BAy_n)=1$ when $\{y_n\}$ is a sequence in E such that $\lim_{n\to\infty} Ay_n = \lim_{n\to\infty} By_n = u$ for some $u \in E$.

Definition GC2. Let (E,d) be a multiplicative metric space. Let A and B be two self-maps of (E,d). The pair (A,B) is called a weakly compatible mapping if $Au=Bu$. Additionally, $u \in X$ implies $ABu = BAu$, which means that $d(Au,Bu)=1$; thus, this implies that $d(ABu,BAu)=1$.

Theorem GC3. Suppose (E,d) is a complete multiplicative metric space, and A,B,f, and g are four mappings of E into itself. Suppose that there exists $\tau \in \left(0,\frac{1}{2}\right)$ such that

$$d(Ax,By) \leq \phi \left(\begin{array}{c} d^\tau(fx,gy), d^\tau(fx,Ax), d^\tau(gy,By), \\ d^\tau(Ax,gy), d^\tau(fx,By) \end{array} \right) \text{ for all }$$

$x,y \in E$. Here, $\phi:[0,\infty)^5 \to [0,\infty)$ is a five-variable function that satisfies certain appropriate conditions. If one of the following conditions is satisfied, (a) Either A or f is continuous, (A,f) is compatible, and (B,g) is weakly compatible; or (b) Either B or g is continuous; (B,g) is compatible, and (A,f) is weakly compatible, then A,B,f, and g have a unique common fixed point.

Most recently, Liu and Gu (15) proved the following common fixed-point theorem in the b-complete b-metric spaces.

Theorem LG. Let f,g,A,B,S, and T be six self-mappings on a b-complete b-metric space (E,d), then the following conditions hold:

(i) $fE \subset BTE$, $gE \subset ASE$;

(ii) $fS=Sf, AS=SA, gT=Tg, BT=TB$;

(iii) $\forall x,y \in E$,

$$d(fx,gy) \leq \frac{1}{s^4}\phi \left(\begin{array}{c} d(fx,ASx), d(gy,BTy), d(ASx,BTy), \\ d(fx,BTy), d(gy,ASx) \end{array} \right)$$

where $s \geq 1$ is the coefficient of (E,d), and function $\phi:[0,\infty)^5 \to [0,\infty)$ satisfies the following conditions:

(a) ϕ is nondecreasing and upper semicontinuous about each variable.

(b) For $t>0$,
$$\max\{\phi(0,0,0,t,t),\phi(t,t,t,0,2t),\phi(t,t,t,2t,0)\}<t.$$

Assuming that one of the following conditions is satisfied, then f,g,A,B,S, and T have a unique common fixed point z in E. Moreover, z is also a unique common fixed point of each of the pairs $\{f,AS\}$ and $\{g,TB\}$.
(1) Either f or AS is continuous; $\{f,AS\}$ is compatible, and $\{g,BT\}$ is weakly compatible;
(2) Either g or BT is continuous; $\{g,BT\}$ is compatible, and $\{f,AS\}$ is weakly compatible;
(3) Either AS or BT is a surjection, and $\{f,AS\}$ and $\{g,BT\}$ are weakly compatible.

It is important to discuss the existence of fixed points of nonlinear mappings; however, it is more important to design iterative schemes to approximate their fixed points for real-world applications. Fixed-point computation is essential to the study of many practical problems, including utility-based bandwidth allocation, intensity-modulated radiation therapy, image-guided radiotherapy, computed tomography, and magnetic resonance imaging.

The Picard iterative scheme may be the oldest iterative scheme, and was first used to solve the initial value problems of differential equations before being applied to solve fixed-point problems of nonlinear operators.

The well-known application of the Picard iterative scheme is the Banach contraction mapping principle (BCMP). The proof of BCMP is established via a constructive method. BCMP yields critically remarkable results in mathematics because it is notably existent, unique, and constructive. Moreover, BCMP is also the most widely used theorem for solving nonlinear equations.

Nonexpansive mappings, which are more general than contraction mappings, have recently been extensively investigated. In 1965, Browder (16) and Kirk (17) proved the existence of fixed points in real uniformly convex Banach spaces. The Picard iterative scheme cannot be used to prove the existence of fixed points of nonexpansive mappings. Even if the fixed-point set of nonexpansive mappings is nonempty, we cannot guarantee that the sequence generated via the Picard iterative scheme converges to their fixed points. In 1953, Mann (18) introduced a more complicated iterative scheme called the normal Mann iterative scheme. Another efficient mean-valued iterative process is the Ishikawa iterative process, which was first introduced by Ishikawa (19) in 1974. Both schemes have been extensively studied [see (20–24) and references therein].

It is well known that only weak convergences of the normal Mann iterative scheme and Ishikawa iterative scheme are obtained in infinite-dimensional Hilbert spaces. Several problems in fields such as economics, image recovery, quantum physics, and control theory arise in infinite-dimensional spaces. For such problems, we often require a strong convergence (or norm convergence) rather than a weak convergence, because it illustrates the physically tangible property that the energy $\|y_n - y\|$ of the error between the iterative sequence $\{y_n\}$ and the solution y is arbitrarily small. The necessity for the discussion on strong convergence was also analyzed in (25); here, a convex function f was minimized via the proximal-point scheme. It was shown that the velocity of convergence of the sequence $\{f(y_n)\}$ is more desirable when $\{y_n\}$ strongly converges than when it weakly converges. Such properties have a direct impact when the process is directly executed in the underlying infinite-dimensional space.

Recently, mathematicians have attempted to modify the normal Mann iterative scheme and Ishikawa iterative scheme such that strong convergence is guaranteed without any further compactness restriction [e.g., see (26–30).]

The most famous regularization method that modifies the normal Mann iterative scheme has been applied to contraction mappings. This method combines the ideas of two iterative methods: the Halpern iterative algorithm (explicit algorithm) and the Browder iterative algorithm (implicit algorithm). In 2001, Moudafi (31) introduced a viscosity algorithm. The iterative sequence generated by the viscosity iterative scheme strongly converges to a fixed point of nonlinear mappings. It is worth mentioning that the fixed point is also a unique solution of a certain monotone variational inequality [for more details, see (32, 33)]. The viscosity algorithm connects nonlinear analysis with optimization theory. In 2007, Gu (34) established a common fixed-point theorem for a finite family of asymptotically pseudocontractive mappings in the framework of Banach spaces [see (34) for more details]. As such, the projection method, which was first proposed by Haugazeau (35) in 1968, has been rapidly developed in recent years.

In 2012, Gu and Lu (36) investigated zero-point problems of the sum of two monotone operators as based on the hybrid projection. Specifically, they obtained the following result:

Theorem GL. Suppose C is a nonempty, closed, and convex subset of Hilbert space, H. Suppose $T:C \to C$ is a nonexpansive mapping, and its fixed-point set is nonempty; $A:C \to H$ is an a-Lipschitz continuous and monotone mapping, and $B:C \to H$ is a b-inverse-strongly monotone mapping. Suppose $M:H \rightrightarrows H$ is a maximal monotone operator such that $D(M) \subset C$. Assume that $F := F(T) \cap (B+M)^{-1}(0) \cap VI(C,A)$ is nonempty. Let $\{x_n\}$ be a sequence generated via the following iterative process:

$$\begin{cases} x_1 \in C, \\ C_1 = C, \\ z_n = \mathrm{Pr}oj_C(J_{s_n}(x_n - s_nBx_n) - r_nAJ_{s_n}(x_n - s_nBx_n)), \\ y_n = \alpha_n x_n + (1-\alpha_n)T\mathrm{Pr}oj_C(J_{s_n}(x_n - s_nBx_n) - r_nAz_n), \\ C_{n+1} = \{u \in C_n : \|y_n - u\| \leq \|x_n - u\|\}, \\ x_{n+1} = \mathrm{Pr}oj_{C_{n+1}} x_1, n \geq 0. \end{cases}$$

where $J_{s_n} = (I + s_nM)^{-1}$, $\{r_n\} \subset \left(0,\frac{1}{a}\right)$, $\{s_n\} \subset (0,2b)$, and $\{\alpha_n\} \subset (0,1)$. Assume that the following conditions are satisfied:

(a) $0 < \alpha \leq r_n \leq \beta < \frac{1}{a}$;

(b) $0 < \gamma \leq s_n \leq \delta < 2b$;

(c) $0 \leq \alpha_n \leq \sigma < 1$,

where $\alpha, \beta, \gamma, \delta$, and σ are real constants. Then, the sequence $\{x_n\}$ strongly converges to $\mathrm{Pr}oj_{\mathcal{F}} x_1$.

Recently, many authors have investigated zero-point problems, variational inequality problems, and fixed-point problems based on the hybrid projection techniques [see (37–41) and references therein]. An emerging method is to use the hybrid projection technique to modify the forward–backward splitting algorithm, which results in each iterative step only managing the individual operators, and not the sum operator. However, most results on this topic are presented in Hilbert spaces. Even within the framework of Banach spaces, uniform smoothness and uniform convexity are essential. If the above techniques are required to provide better service for practical problems in the real world, the restrictions on both spaces and control sequences should be reduced. Solving these problems will enrich fixed-point theory and related convex optimization problems in the future.

References
1. Z. Mustafa, B. Sims, *J. Nonlinear Convex Anal.* **7**, 289–297 (2006).
2. Z. Mustafa, H. Obiedat, F. Awawdeh, *Fixed Point Theory Appl.* **2008**, Article ID 189870 (2008).
3. W. Shatanawi, *Fixed Point Theory Appl.* **2010**, Article ID 181650 (2010).
4. R. Chugh, T. Kadian, A. Rani, B. E. Rhoades, *Fixed Point Theory Appl.* **2010**, Article ID 401684 (2010).
5. M. Abbas, B. E. Rhoades, *Appl. Math. Comput.* **215**, 262–269 (2009).
6. F. Gu, W. Shatanawi, *Fixed Point Theory Appl.* **2013**, Article ID 309 (2013).
7. H.-C. Wu, *J. Nonlinear Var. Anal.* **1**, 175–199 (2017).
8. A. E. Bashirov, E. M. Kurpinara, A. Ozyapici, *J. Math. Anal. Appl.* **337**, 36–48 (2008).
9. L. Florack, H. V. Assen, *J. Math. Imaging Vis.* **42**, 64–75 (2012).
10. S. K. Malhotra, J. B. Sharma, S. Shukla, *J. Nonlinear Funct. Anal.* **2017**, Article ID 9 (2017).
11. A. E. Bashirov, E. Misirli, Y. Tandogdu, A. Ozyapici, *Appl. Math. J. Chinese Univ. Ser. B* **26**, 425–438 (2011).
12. M. Abbas, B. Ali, Y. I. Suleiman, *Int. J. Math. Math. Sci.* **2015**, Article ID 218683 (2015).
13. F. Gu, Y. J. Cho., *Fixed Point Theory Appl.* **2015**, Article ID 165 (2015).
14. K. Zoto, S. Radenović, J. Dine, I. Vardhami, *Commun. Optim. Theory* **2017**, Article ID 22 (2017).
15. L. Liu, F. Gu, *J. Nonlinear Sci. Appl.* **9**, 5909–5930 (2016).
16. F. E. Browder, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 1041–1044 (1965).
17. W. A. Kirk, *Amer. Math. Monthly* **72**, 1004–1006 (1965).
18. W. R. Mann, *Proc. Amer. Math. Soc.* **4**, 506–510 (1953).
19. S. Ishikawa, *Proc. Am. Math. Soc.* **44**, 147–150 (1974).
20. F. Gu, *Proc. Amer. Math. Soc.* **129**, 2293–2300 (2000).
21. F. Gu, *J. Math. Anal. Appl.* **329**, 766–776 (2007).
22. F. Gu, *Positivity* **12**, 503–509 (2008).
23. X. Qin, J. C. Yao, *J. Inequal. Appl.* **2016**, Article ID 232 (2016).
24. S. Y. Cho, B. A. Bin Dehaish, X. Qin, *J. Appl. Anal. Comput.* **7**, 427–438 (2017).
25. O. Guler, *SIAM J. Control Optim.* **29**, 403–419 (1991).
26. F. Gu, *Acta Math. Sinica, Chinese Ser.* **53**, 1209–1216 (2010).
27. F. Gu, *Bull. Malays. Math. Sci. Soc.* **34**, 591–599 (2011).
28. H. Zhou, Y. Jia, *Proc. Amer. Math. Soc.* **125**, 1705–1709 (1997).
29. X. Qin, Y. Su, *J. Math. Anal. Appl.* **329**, 415–424 (2007).
30. S. S. Chang, *Comput. Math. Appl.* **37**, 17–24 (1999).
31. A. Moudafi, *J. Math. Anal. Appl.* **241**, 46–55 (2000).
32. H. K. Xu, *J. Math. Anal. Appl.* **298**, 279–291 (2004).
33. X. Qin, S. Y. Cho, L. Wang, *Fixed Point Theory Appl.* **2013**, Article ID 148 (2013).
34. F. Gu, *Comput. Math. Appl.* **53**, 1792–1799 (2007).
35. Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes*, Ph.D. thesis, Université de Paris, France, 1968.
36. F. Gu, J. Lu, *Fixed Point Theory Appl.* **2012**, Article ID 162 (2012).
37. K. Nakajo, W. Takahashi, *J. Math. Anal. Appl.* **279**, 372–379 (2003).
38. X. Qin, Y. Su, *Nonlinear Anal.* **67**, 1958–1965 (2007).
39. X. Qin et al., *J. Comput. Appl. Math.* **225**, 20–30 (2009).
40. H. Zhou, *Nonlinear Anal.* **71**, 120–125 (2009).
41. S. Y. Matsushita, W. Takahashi, *J. Approx. Theory* **134**, 257–266 (2005).
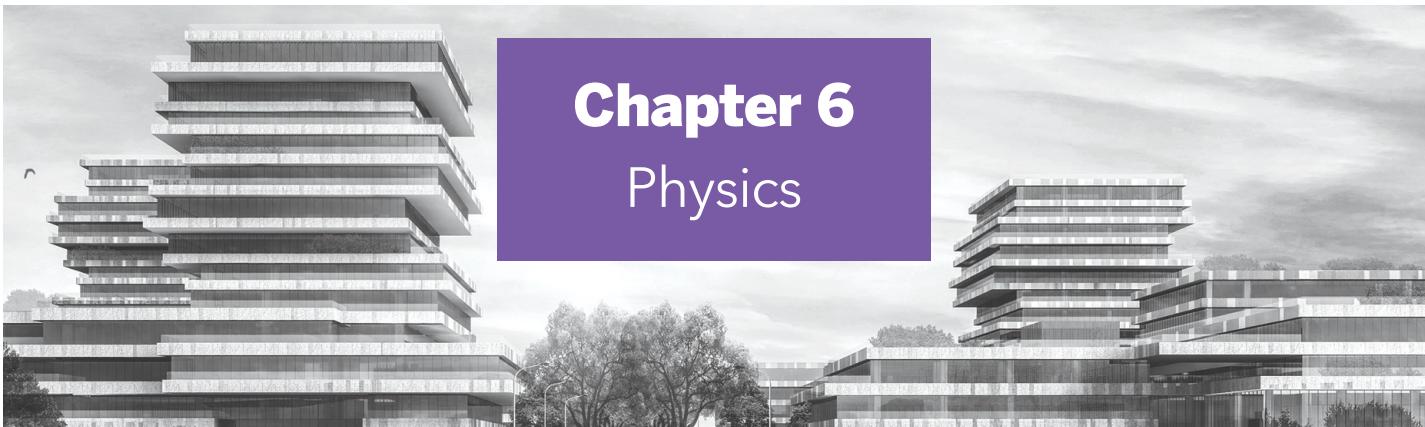
Chapter 6
Physics

Exploring frontiers of quantum materials and low-dimensional structures using the computational approach

Xiao-Yong Feng[§], Jianhui Dai[§], Yi Ding*, and Chao Cao*

The complexity of both quantum mechanics and many-body effects makes understanding emergent physical phenomena in realistic materials extremely difficult. For decades, theoreticians in the condensed matter physics community relied on semi-empirical models and tunable parameters to fit experimental observations and make limited predictions. This scenario began to change with the recent development of the computational approach. Numerical simulations based on quantum mechanics can now provide quantitatively comparable results in many cases, and can be used to predict the behavior of material. Moreover, these simulations provide useful insights and *ab initio* parameters, leading to reliable first-principles models with which theoreticians can explore unconventional physical properties and underlying mechanisms. This paper shares some intriguing investigations utilizing the computational approach by the Condensed Matter Group at Hangzhou Normal University.

Computational exploration of unconventional behaviors in FeSe-/CrAs-based materials

Iron-based superconductors are a hot topic in the field of quantum materials. Among them, the vacancy-ordered iron selenides, which possess a general chemical formula of $A_yFe_{2-x}Se_2$, exhibit peculiar electronic and magnetic properties. Through extensive first-principles calculations, we predicted, independent of a neutron experiment reported at the same time, a novel block-spin magnetic state in $(K,Tl)_yFe_{1.6}Se_2$ (*1*). This state, which we found is insulating when $y = 0.8$, has a checkerboard antiferromagnetic configuration, as shown in Figure 1, with an unusually large magnetic moment in each Fe-square block. We modeled the magnetic structure with an extended J_1–J_2 Heisenberg model and revealed the hidden role played by the next-nearest-neighbor coupling, a magnetic interaction now recognized as the dominant coupling linked to the superconductivity in most iron-based superconductors. Our computation also shows that the electronic and magnetic structures are very sensitive to the concentration and symmetry of Fe vacancies (*2, 3*). For example, the $TlFe_{1.5}Se_2$ system, in which Fe vacancies form ordered orthorhombic superstructures, has a stripe-like antiferromagnetic ground state. Furthermore, with the variation of magnetic states, the band structures and Fermi surfaces of bulk $A_yFe_{2-x}Se_2$ are also changed. It is worth noting that there is increasing evidence indicating the coexistence of these states in this class of superconducting ternary iron chalcogenides.

In addition to iron-based superconductors, in 2015, alkali chromium arsenides $A_2Cr_3As_3$ (A = K, Rb, Cs) were found to be superconducting (*4*). These materials are of great interest not only because Cr was one of the two remaining 3*d* transition-metal (TM) elements (the other being manganese) with unknown superconducting phenomena, but also because the main building blocks, $[Cr_3As_3]^\infty$, form quasi-one-dimensional (Q1D) nanotubes. First-principles calculations provided insight into the electronic structure of the Q1D system (*5*), revealing three relevant molecular orbital bands that cross the Fermi energy and disperse strongly only along the tube direction. Owing to the relatively large electron–electron correlations in the Cr 3*d* orbitals, these peculiar band-structure features motivated us to model
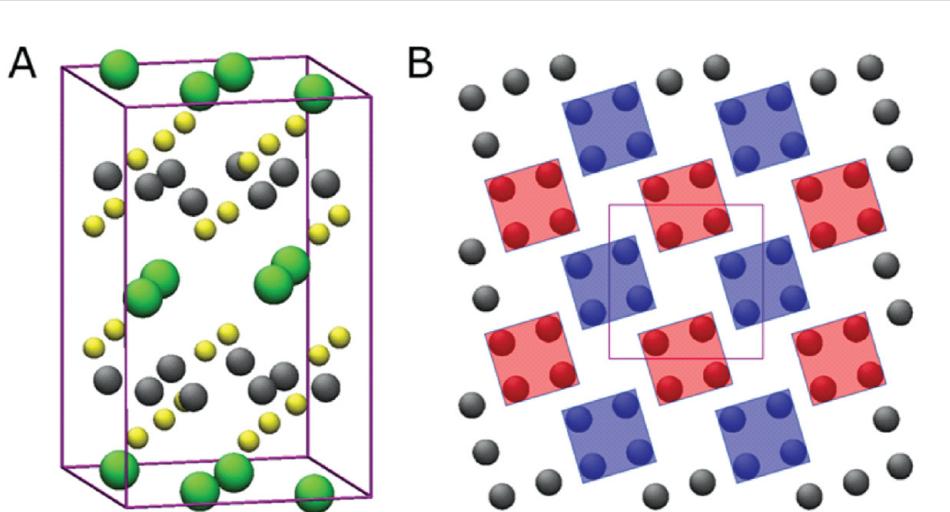
FIGURE 1. (A) Geometrical structure and (B) checkerboard antiferromagnetic state of $(K,Tl)_yFe_{1.6}Se_2$.
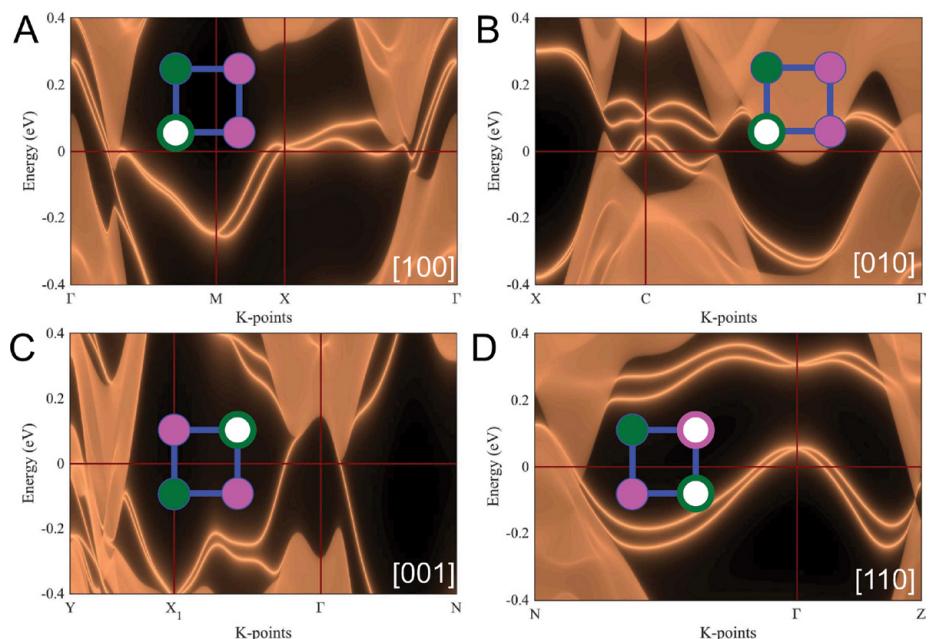
FIGURE 2. Possible surface states in $TaSb_2$. Surface spectrums for 2D semi-infinite slabs with the (A) [100], (B) [010], (C) [001], and (D) [110] surfaces shown. Here, M, X, C, N, Z, and Γ are the respective high symmetry points for each surface; topological surface states exist only between the solid circles of different colors.

the $[Cr_3As_3]^\infty$ structure by a twisted Hubbard tube (*6*), which revealed a Q1D Luttinger liquid characteristic and possible spin-triplet superconducting instabilities in this CrAs-based material. Our calculations also predicted an enhanced Q1D feature and associated magnetic frustration in a related material, KCr_3As_3, in which the concentration of cations is reduced (*7*). Because of the

enhanced Q1D feature, a new superconducting phase may emerge once the system is driven to the quantum disorder region under physical or chemical pressure.

Computational design of tunable physical properties in bulk and 2D materials

Computational modeling can help not only in understanding novel physical behaviors in materials, but also in designing useful physical properties in bulk and two-dimensional (2D) materials. For example, through first-principles calculations, the unique electronic structures and topological properties of a series of TM dipnictides, XPn_2 (X = Ta, Nb; Pn = P, As, Sb), have been systematically investigated (*8*). These compounds are weakly topological, as shown in Figure 2, meaning that they could be viewed as 2D topological insulators stacked together with nearly the same electron-hole compensation character as in $NbAs_2$, $NbSb_2$, and $TaAs_2$.

For a 2D graphene sheet, a typical zero-band-gap semimetal, our computational investigation has shown that the adsorptions of TM adatoms and dimers will effectively induce magnetization in the system (*9*). In particular, Fe/Co-adatom- and Fe-dimer-decorated graphene can turn to a half-metal, in which one spin channel is insulating while the other is metallic.

Atom adsorptions also alter the electronic and magnetic properties of black and blue phosphorene (P) systems (*10*). A theoretical study on 20 different adatoms has shown that both black and blue P sheets exhibit good adsorption capability to foreign atoms, and the original semiconducting behaviors are modified by adatoms. These doped P sheets exhibit p-type, n-type, and mid-gap states; and diverse spintronic features, such as half-metals, bipolar semiconductors, and spin-gapless semiconductors, can be realized in them. With defects,

Department of Physics, Hangzhou Normal University, China
[§]These authors contributed equally to this work.
*Corresponding authors: ydinghz@hznu.edu.cn (Y.D.), ccao@hznu.edu.cn (C.C.)

the binding strengths of adatoms are enhanced on the boron-nitride surface, where 1D TM chains will form along line defects (11). Defects also significantly change the intrinsic semiconducting properties: Band gaps are reduced, and versatile magnetic properties are exhibited depending on the adatoms.

Computational prediction of new low-dimensional structures

Using the computational approach, we can further predict new low-dimensional materials before experimental synthesis. One example is monolayer graphene-like transition-metal dichalcogenides (12). Through comprehensive calculations, the detailed structural, vibrational, and electronic properties of MX_2 (M = Mo, Nb, W, Ta; X = S, Se, Te) sheets have been revealed. The calculated thickness, Raman and infrared modes, and band features were validated in subsequent experiments. Another example is the computational discovery of silicene-related nanostructures (13, 14). The pristine silicene sheet is a buckled honeycomb, while binary SiX and XSi_3 (X = B, C, N, Al, P) sheets have flat, chair-buckled, or washboard-buckled hexagonal structures depending on the elemental constituents. Interestingly, we found that the CSi_3 sheet maintains zero-band-gap semimetallicity as in the silicene sheet, for which the p_z orbitals of Si and C atoms contribute equally to the linear Dirac-like bands near the Fermi level. In contrast, for the SiC_3 sheet, unbiased structural search calculations revealed three nearly degenerate lowest-energy structures, which can be regarded as different Si-graphene hybrid honeycomb lattices, and two of them exhibit the regular and distorted Dirac band features, as shown in Figure 3.

Computational studies are also conducted to explore new properties and applications in known nanomaterials. For example, although the graphene-like Be_3C_2 sheet was proposed as a new Dirac material, our latest study found an antiferromagnetic ground state (15). Noticeable band gaps are opened in Be_3C_2 nanosheets and nanoribbons, the valence and conduction band edges of which are suitable for application to photocatalytic water splitting. For the MoN_2 sheet, the experimentally proposed MoS_2 structure is unstable in the pristine case, while surface decoration by hydrogen, halogen, and alkali atoms will stabilize the MoN_2 monolayer (16, 17). Anisotropic Dirac cones were discovered in the hydrogenated structures, while partially occupied nearly-free-electron bands were found in the alkali-decorated structures, in which the delocalized electrons further form an intrinsic
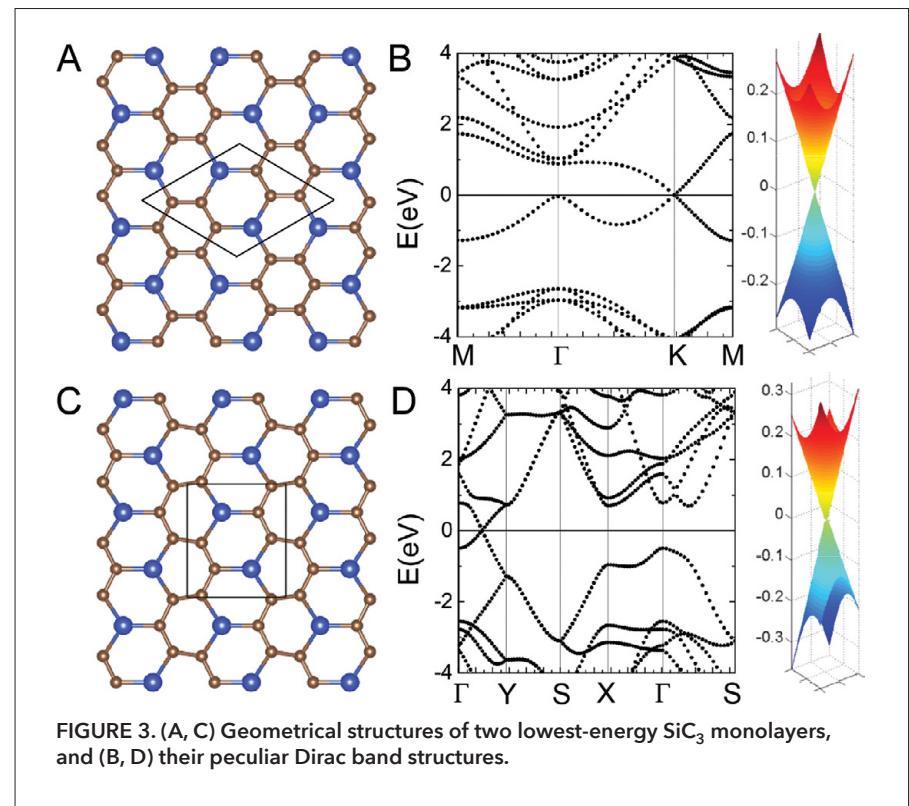
FIGURE 3. (A, C) Geometrical structures of two lowest-energy SiC_3 monolayers, and (B, D) their peculiar Dirac band structures.

2D electron gas in the free-space state for electron transport without ion scatterings.

Conclusions and outlook

The aforementioned examples illustrate how the computational approach can help theoreticians understand physical phenomena in new superconducting materials, find tunable electronic and magnetic properties in bulk and 2D materials, and predict novel stable structures for nanoscale systems. Nevertheless, many challenges remain for state-of-art computational methods. For example, the density-functional based first-principles method fails for so-called "strongly correlated materials." Even for "weakly correlated materials," it is never easy to model a large-scale system (e.g., glassy materials) efficiently or predict complex structures. New theories and algorithms, or even new hardware, are required to overcome these challenges. Researchers at Hangzhou Normal University are actively addressing these challenges and contributing to scientific knowledge.

References

1. C. Cao, J. Dai, *Phys. Rev. Lett.* **107**, 056401 (2011).
2. C. Cao, J. Dai, *Phys. Rev. B* **83**, 193104 (2011).
3. C. Cao, F. Zhang, *Phys. Rev. B* **87**, 161105 (2013).
4. J.-K. Bao *et al.*, *Phys. Rev. X* **5**, 011013 (2015).
5. H. Jiang, G. Cao, C. Cao, *Sci. Rep.* **5**, 16054 (2015).
6. H. Zhong, X.-Y. Feng, H. Chen, J. Dai, *Phys. Rev. Lett.* **115**, 227001 (2015).
7. C. Cao *et al.*, *Phys. Rev. B* **92**, 235107 (2015).
8. C. Xu *et al.*, *Phys. Rev. B* **93**, 195106 (2016).
9. C. Cao *et al.*, *Phys. Rev. B* **81**, 205424 (2010).
10. Y. Ding, Y. Wang, *J. Phys. Chem. C* **119**, 10610–10622 (2015).
11. Y. Wang, Y. Ding, *Ann. Phys. (Berl.)* **526**, 415–422 (2014).
12. Y. Ding *et al.*, *Physica B* **406**, 2254–2260 (2011).
13. Y. Ding, Y. Wang, *J. Phys. Chem. C* **117**, 18266–18278 (2013).
14. Y. Ding, Y. Wang, *J. Phys. Chem. C* **118**, 4509–4515 (2014).
15. Y. Ding, Y. Wang, *J. Mater. Chem. C* **5**, 10728–10736 (2017).
16. Y. Wang, Y Ding, *J. Mater. Chem. C* **4**, 7485–7493 (2016).
17. Y. Ding, Y Wang, *J. Mater. Chem. C* **5**, 683–689 (2017).

Graph mining from a physics perspective: Prediction, identification, and recommendations

Linyuan Lü[1,2*], Tianlong Fan[1], Hao Liao[3], Jianlin Zhang[1*], and Yi-Cheng Zhang[1,2*]

With significant developments in web techniques, information can be easily created and accessed by individuals. The explosive growth of data has driven scientific research into the big data era. Various data from real socioeconomic systems—such as online social networks, transportation networks, and the stock market—have significantly expanded physics research and promoted the formation of many interdisciplinary research fields such as econophysics, sociophysics, and biophysics. In the past few years, we have focused primarily on graph mining problems, especially from a physics perspective. Our studies have promoted the development of a new research field called "information physics," an interdisciplinary domain related to both statistical physics and information science.

A network is a type of graph where nodes represent entities and links represent their relationships or interactions. Given that a network has become a unified way to represent different systems, such as biological, social, and information systems, we can obtain their common characteristics by analyzing their corresponding networks. We have applied the concepts, theories, and methods of statistical physics to address many critical issues in information science, including link prediction, vital nodes identification, and personalized recommendation. In the following sections, we introduce some representative results obtained in these three areas.

[1]Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou, China
[2]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China
[3]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
*Corresponding authors: linyuan.lv@gmail.com (L.L.), zhangjohn@vip.sina.com (J.Z.), and zinppy@gmail.com (Y.-C.Z.)

Link prediction

The study of link prediction is an important branch of research in the field of network sciences. It determines whether a nonobserved link (1) should exist or (2) will exist (1). The first type of prediction uncovers missing links, while the second predicts future links and is related to the evolution of networks. A better understanding of the mechanisms of network formation or evolution enables us to design increasingly effective link prediction methods for these networks. However, for an arbitrary network, high-accuracy algorithms can, to a certain extent, reveal the network organization or evolutionary mechanisms.

Existing link prediction methods can be divided into four types: probabilistic models, similarity-based algorithms, machine learning–based methods, and methods based on likelihood analysis. Similarity-based methods are the most frequently used. Each pair of nodes is assigned a similarity score given by an algorithm. A higher score indicates a higher probability that the link exists. However, accurately defining the similarity is a challenge [see methods reviewed in (1)].

In general, the precision of an algorithm is an indication of the ability of the algorithm to explain link formation. However, even in the same network, different algorithms may yield different values of precision. Therefore, link predictability reflected by precision is associated with specific algorithms, and not the inherent characteristics of the network itself.

To measure the inherent link predictability of networks that are not dependent on specific algorithms, we recently proposed a "structural consistency" index based on the first-order perturbation on the network adjacency matrix (2). The basic assumption is that if the addition of links leads to significant structural changes, these links are difficult to predict. This implies that if the network structural features are not significantly changed when multiple randomly chosen links are removed or added, then the network has high predictability.

Specifically, for a given network with a set of links denoted by E, we randomly select a small number of links to constitute the perturbation set ΔE, and the remaining links constitute a set E^R. Their corresponding adjacency matrices can be represented by ΔA and A^R, where $A = A^R + \Delta A$. Matrix A^R can be written as $A^R = \sum_{k=1}^{N} \lambda_k x_k x_k^T$ for its symmetry property, where λ_k and x_k are the eigenvalue and eigenvector for A^R, respectively. We perturb network A^R with ΔE and use first-order approximation to constitute

FIGURE 1. An example of calculating structural consistency. In (**A**), the dashed links constitute the perturbation set $\Delta E=\{(5,8),(6,9)\}$, while the set E^R is formed by the solid links. (**B**) is the adjacency matrix of the given network, where the blue and black squares correspond to the links in ΔE and E^R, respectively. (**C**) shows the perturbed matrix. (**D**) is the perturbed network, where the outcome links $E^L=\{(3,8),(6,9)\}$ are represented by the dashed red lines. The structural consistency, σ_c, equals 0.5 because only one of the two blue links is recovered. Figure modified from (2).
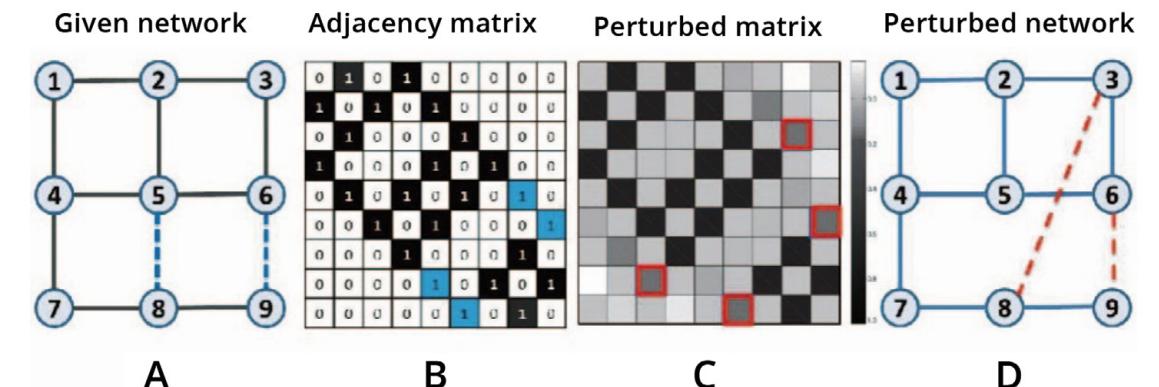
FIGURE 2. (**A**) An example of a network with three shells. (**B**) The $h_i^{(1)}$, $h_i^{(2)}$, and $h_i^{(3)}$ of node 5 in (A). (**C**) The H-sequence of a router network. Panel C reprinted from (7); distributed under a Creative Commons Attribution License 4.0 (CC BY); see link at creativecommons.org/licenses/by/4.0.
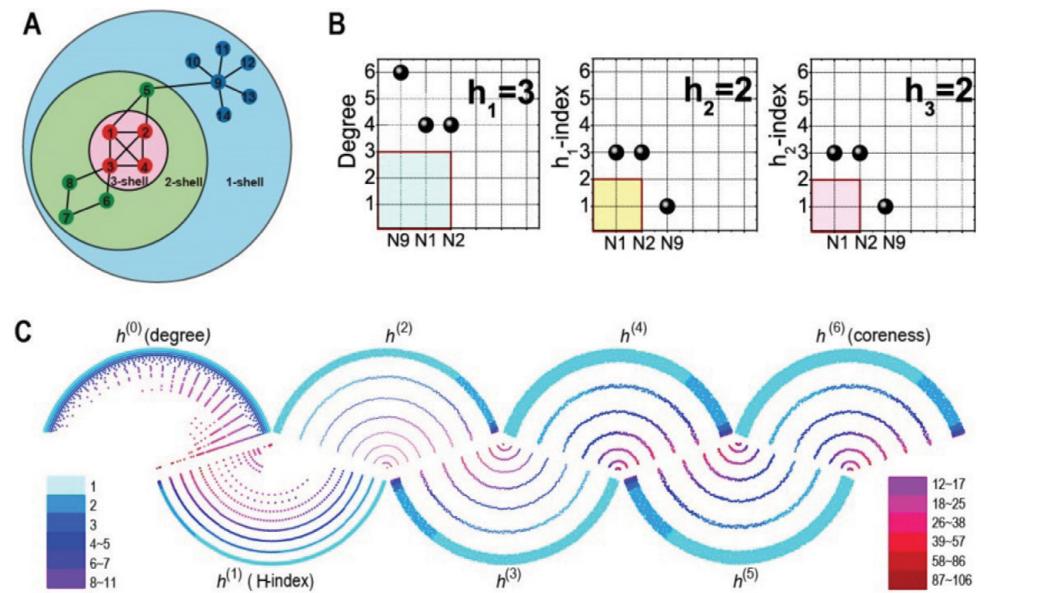
the perturbed matrix. After perturbation, we obtain the modified eigenvalue $\lambda_k + \Delta\lambda_k$ and its corresponding eigenvector $x_k + \Delta x_k$. Left-multiplying the eigenfunction by x_k^T and neglecting all second-order terms, we obtain $\Delta\lambda_k \approx x_k^T \Delta A x_k / x_k^T x_k$. Next, we obtain the perturbed matrix $\tilde{A} = \sum_{k=1}^N (\lambda_k + \Delta\lambda_k) x_k x_k^T$ by fixing the eigenvectors. According to our assumption, if deleting ΔE does not significantly change the structural features, the network is highly predictable, meaning that A and $\tilde{A}$ are almost the same. For quantitative measurement, we rank the nonobserved links (i.e., links that are not in the set E^R) in descending order according to their corresponding values in $\tilde{A}$. The set of top-L ranked links is represented by E^L, where L is the number of links in the perturbation set, and $L=|\Delta E|$. Then, the links in $E^R \cup E^L$ construct the perturbed network. The structural consistency is $\sigma_c = |\Delta E \cap E^L|/L$. Figure 1 shows an example of calculating the structural consistency of a simple network.

We further devised a link prediction method, called a structural perturbation method, to predict missing links. It outperforms many benchmark methods, including the stochastic block model (3) and the hierarchical structure model (4). This algorithm can also be used to detect spurious links (5).

Vital nodes identification

Owing to the heterogeneity of real networks, the nodes usually play significantly different roles in network structure and function. The criteria for identifying vital nodes are diverse owing to different structural or functional objectives. For example, important nodes for promoting the dissemination of information may not be the key nodes for maintaining network connectivity. We have previously reviewed significant developments (6) and also introduced state-of-the-art methodologies based on

physics principles, such as percolation-based methods and message-passing theory. In addition, we compared many methods on different real networks and provided an outlook on future research directions (6).

Among many node centrality indices, those of degree, H-index, and coreness are well-known and widely used; however, they were considered unrelated based on published literature. Our recent work uncovered their intrinsic relationship: Degree, H-index, and coreness are respectively the initial, intermediate, and steady state of a sequence constructed by an operator H (7). We call their relation the "DHC theorem" on networks. The operator H acts on a real-number sequence $(x_1, x_2, …, x_n)$, and returns an integer $y = H(x_1, x_2, …, x_n) > 0$. There are at most y elements in the sequence, each of which is no less than y. If we assign the sequence as the degree of node i's neighbors, then we have $y=h_i^{(1)}$, or the H-index of node i. If the operator further acts on the sequence of the H-indices of node i's neighbors, we have $y=h_i^{(2)}$, or the second-order H-index. Finally, the sequence $(h_i^{(0)}, h_i^{(1)}, h_i^{(2)}, …, h_i^{(n)}, …)$ converges to the steady states, which is the coreness of node i. We also proved that when the updating process is asynchronous, namely that when only a few nodes update in each time step, the convergence can also be guaranteed. In this sense, the DHC theorem provides an effective way to calculate a node's coreness in large-scale evolving networks in a decentralized manner (8). In directed and weighted networks, the DHC theorem still holds (6). The theorem is a useful tool that can be utilized in real applications, such as identifying important nodes in functional brain networks, assessing the social influence that cities have online (9), and quantifying the relative importance of countries in the World Trade Web.

Figure 2A shows an example of a network where the nodes can be divided into three shells. Take node 5 as an example. Figure 2B shows the step-by-step calculation of the n-order H-index of node 5. Clearly, the sequence converges after two steps, and the coreness of node 5 is two. Figure 2C shows the H-sequence of a router network. The iterative process starts from the left side and converges to coreness in six steps. The color represents the node degree. The location of a node is determined by its $h^{(n)}$-index. The higher the $h^{(n)}$-value, the closer the node is to the center. Nodes with the same $h^{(n)}$-values are in the same layer.

Personalized recommendation

Search engines and recommender systems are both useful tools to help users find relevant information. Recommender systems uncover users' personal interests based on their previous activities and suggest new items to them. This differs from search engines, which require users to provide keywords. Recommending items for target users is essentially a problem of link prediction on bipartite networks (10). Most of the existing methods are based on user or item similarity. Although the accuracy may be high, such methods are risky and could yield poor diverse results. A recommender system is similar to either a concave or a convex lens: It may broaden or narrow a user's field of vision (11). Therefore, how to address the accuracy-diversity dilemma becomes the main challenge. Some physical pro-

cesses have been applied to solve this problem, including mass diffusion (MD) and heat conduction (HC) on bipartite networks. Representative methods include the hybrid method, which combines MD and HC (12), the preferential diffusion method (11), and weighted HC with the addition of a ground user who selects all the items (13).

Another challenge is how to avoid recommendation congestion. Sometimes, not all user requirements can be satisfied owing to the constraint of the recommended items, and congestion may occur. Inspired by the Gini index, we proposed a metric to quantitatively measure an algorithm's ability to avoid recommendation congestion. Considering the HC process on a bipartite network with different thermal conductivities, we developed a new method called directed weighted conduction (DWC) (14). Compared with the state-of-the-art methods, the DWC algorithm has lower congestion and can significantly enhance both novelty and diversity, while maintaining a certain precision.

In summary, we have reviewed here some representative examples of progress on graph mining from a physics perspective, including link prediction, vital node identification, and personalized recommendations. Although these three problems appear different, they are all essentially ranking problems and can complement each other to improve accuracy. For example, we can apply link prediction methods before identifying the important nodes of a network with missing links (15). The success of many physics-based approaches to graph mining demonstrates that physics can also contribute to the development of information science.

References
1. L. Lü, T. Zhou, *Physica A* **390**, 1150–1170 (2011).
2. L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2325–2330 (2015).
3. R. Guimerà, M. Sales-Pardo, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22073–22078 (2009).
4. A. Clauset, C. Moore, M. E. J. Newman, *Nature* **453**, 98–101 (2008).
5. P. Pan, T. Zhou, L. Lü, C. K. Hu, *Sci. Rep.* **6**, 22955 (2016).
6. L. Lü et al., *Phys. Rep.* **650**, 1–63 (2016).
7. L. Lü, T. Zhou, Q.-M. Zhang, H. E. Stanley, *Nat. Commun.* **7**, 10168 (2016).
8. H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, M.-Y. Zhou, *Phys. Rep.* **689**, 1–54 (2017).
9. T. Fan, Y.-Y. Zhu, L.-L. Wu, X.-L. Ren, L. Lü, *J. Univ. Electr. Sci. Technol. China* **46**, 766–776 (2017).
10. L. Lü et al., *Phys. Rep.* **519**, 1–49 (2012).
11. L. Lü, W. Liu, *Phys. Rev. E* **83**, 066119 (2011).
12. T. Zhou et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4511–4515 (2010).
13. Y. Zhou, L. Lü, W. Liu, J. Zhang, *PLOS ONE* **8**, e70094 (2013).
14. X. Ren, L. Lü, R. Liu, J. Zhang, *New J. Phys.* **16**, 063057 (2014).
15. S.-Y. Tan, J. Wu, L. Lü, M.-J. Li, X. Lu, *Sci. Rep.* **6**, 22916 (2016).
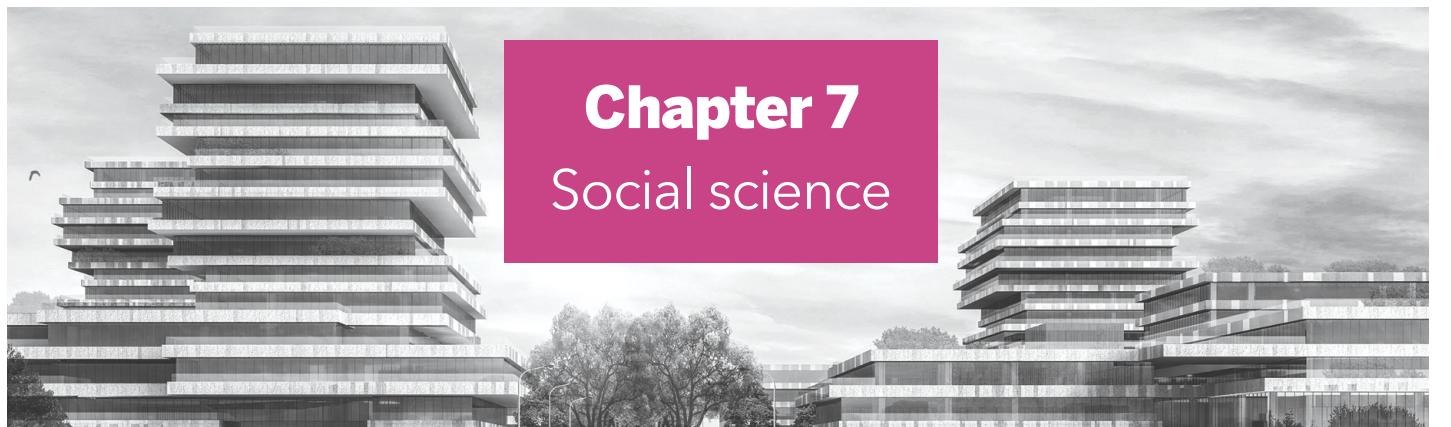
Marginocentric Beijing: Multicultural cartography and alternative modernity

Qingben Li

"**M**arginocentric cities" is a term coined by Marcel Cornis-Pope and John Neubauer to describe those multiethnic nodal cities "that, at favorable historical conjunctions, have rewritten the national cultural paradigm from the margin, ascribing to it a dialogic dimension, both internally (in dialogue with other ethnic traditions) and externally (in dialogue with larger geocultural paradigms)" (1). Whereas Cornis-Pope and Neubauer's map of marginocentric cities is limited to East-Central Europe, this paper employs a journalistic memoir by Michael Meyer, entitled *The Last Days of Old Beijing: Life in the Vanishing Backstreets of a City Transformed*, to argue that the concept of a "marginocentric city" fits Beijing as a nodal space of cultural exchange where boundaries might be elusive and national geographies dislocated. I propose that the constitutive dialectics of being simultaneously central and marginal should be considered relative to their complex relationships with the projects of modernity as seen through the eyes of Chinese citizens of Beijing, a city whose modernity differs from Western modernity. Therefore, Meyer's neutral text, which does not merely criticize the transformation of old Beijing, but also combines tradition and modernity internally and connects East to West externally through narrative, is a cross-cultural writing that offers a more acceptable view of what has happened to the city.

Institute for Arts Education, Hangzhou Normal University, China
Corresponding author: liqingben1@sina.com

Beijing's multicultural space and urban cartography

The Last Days of Old Beijing was published in June 2008, slightly more than one month before the Summer Olympic Games were held in Beijing on August 8. Michael Meyer, an American volunteer in Beijing, was living as no other Westerner had lived—in a shared courtyard home in Beijing's oldest neighborhood, *Dazhalan* [although he misspelled it; the correct name is *Dashilan* (大栅栏)]. The house was in one of Beijing's famed *hutong* (lanes), where Meyer taught English for two years at a local elementary school. His book is neither a novel nor an academic treatise, but describes with affection the life stories of the local residents, such as the widow who shares the courtyard, student Little Liu, coteacher Miss Zhu, migrant Recycler Wang, soldier Liu, and many others whom he integrates with Beijing's history of transformation over thousands of years. What was impressive about this book was the description of Beijing's recent transformation—traditional ways of life increasingly fading away, and many ancient houses being replaced by shopping malls, skyscrapers, the capital's first Walmart department store, and widened streets, all aiming to realize the goal of hosting the 2008 Summer Olympics. In this way, an "old" Beijing was replaced by a "new" Beijing, and life in the vanishing backstreets of the city was transformed, as the book's subtitle suggests.

I agree with the author's position; however, I am more interested in his book's narrative on the history of Beijing's transformation, which created an urban cartography of multicultural places combining tradition and modernity, while connecting East to West. The narrative offers opportunities to reflect on and reconsider these relationships from a spatial perspective, which is important because, as Edward Soja points out, "It is space, more than time, that now hides consequences from us" (2). Clearly, we cannot entirely separate space from time, and both dimensions are generally required to interpret all the consequences of a city's development, along with, in Beijing's case, multicultural space.

After a close reading, it is clear that Meyer did more than simply criticize the demolition and transformation of old Beijing. His attitude is bipartisan and, sometimes,

indeterminate, which challenges traditional linear models, historiographic generalizations, and literary writings. After living in the countryside of Sichuan Province in southwest China for two years, Meyer's first encounter with Beijing was very positive when he arrived there in 1997. His impression was that "Beijing felt cosmopolitan. It also looked unlike any other Chinese city. Here, the center was not a hollow cavern of wide boulevards and monolithic apartment blocks, but a chain of central lakes surrounded by a compact mixture of architecture built on a human scale" (3). Beijing's flat configuration, open sky, and climate reminded him of his Minnesota home.

However, when he found a house in a *hutong* in 2005, the picture was quite different. The most serious problem of living in a *hutong* might be the difficulty of being observed at the public toilet, which was "four slits in the floor facing one another without dividers. A squatting man hacks up a wad of phlegm. Another, wearing pajamas, lights a cigarette" (3). Because most of the houses had no washrooms, the residents were forced every morning to use the public toilet located far from their houses, which was neither convenient nor private. This situation suggests that the transformation of the old city was reasonable in some ways. In fact, many of the residents wanted to move away when they began earning sufficient compensation to live in convenient, modern apartments, particularly younger adults. Some of them found that it was useless to resist moving away, while others did not want to leave their old homes, filled with memories.

Despite its transformation, new Beijing still has many traditional buildings with an old-world ambience and style, and it retains many prehistoric and historic relics. Along with the Forbidden City, the Summer Palace, the Temple of Heaven, and the Great Wall, some *hutong* remain, although they have been greatly reduced in number. Of the many historical sites and cultural relics still found in new Beijing, you can visit the first Paleolithic site, which was unearthed in the basement of the Oriental Plaza. Beijing's most modern, upscale shopping mall is located on the Avenue of Eternal Peace, the city's main thoroughfare, which runs past Tiananmen Square and Wangfujing Street, the pedestrian-only commercial street that Meyer mentions in his book.

We might tend to imagine old Beijing as a place isolated from foreign influences, but that would be an illusion. Its more than 3,000 years of history and 800 years as China's capital have given the city a history rich in international and intercultural exchanges. All these exchanges have left cultural trails that remain in the multicultural space of modern Beijing.

Beijing's multicultural heritage can be understood through Itamar Even-Zohar's "polysystem" theory, which "recognizes that there is a regular process in attitude change in culture: We do not understand or accept anything new except in the context of the old" (4). Consequently, no city can cut off its traditions at the roots, and, therefore, we should understand new Beijing in the context of its old self. Moreover, although new Beijing

has lost many of its historical and cultural relics, a fact that deserves criticism, the remains of the historical sites named above attest to the richness and openness of Beijing's urban history.

The alternative modernity of marginocentric Beijing and cross-cultural writings

Beijing experienced the rise and fall of imperial dynasties and was often devastated during those transitions, which Meyer describes at length. For example, in 1153, the Jurchen conquered Beijing (the secondary capital of the Liao dynasty), destroyed it, and rebuilt it as the central capital of the Jin (Golden) dynasty. Less than a century later, in 1215, the city was destroyed again, this time by the Mongols under Genghis Khan. As Meyer describes it, they conquered the city with a "glorious slaughter," and the city fell in "a barrage of flaming arrows. Every citizen was marked for death. When the siege ended, the capital's streets were slippery with melted flesh, and the landscape showed only vultures and ash" (3).

Beijing was also humiliated by foreign invasions. For example, in 1860, Anglo-French forces assaulted Beijing and burned the Old Summer Palace, whose ruins are still visible. Citing a historical record, Meyer describes (with a humorous tone) a Westerner's impressions of Beijing at that time: "The British attaché recorded that in 1865, Lord Stanley sneered, 'Peking's a giant failure, isn't it? Not a two-story house in the whole place, eh?'" (3). This statement is typical of most Westerners' contemptuous attitudes toward China at the time, and it testifies to the decline of China's national forces and to its identity transformation from the center to the margin of world attention. Cornis-Pope and Neubauer propose that, for marginocentric cities, "it is their marginality, as well as their multiethnic composition, that has allowed the cities to look simultaneously to both East and West, establishing a fertile nexus between cultural traditions" (1).

As they grew aware of China's backwardness and marginalization, the Chinese population began to learn from the West. In the late nineteenth century, Beijing was the first Chinese city to experience Westernization, which created a new, more complex relationship between China and the West. "Learning from Western science and technology is the way to boycott Western invasions" was the slogan at that time. The process of economic and military reforms began in 1898, followed by the political revolution of 1911 and the May Fourth cultural movement of 1919. All these events occurred in Beijing, or, at least, were closely tied to the city. Thus, Beijing was a central witness to the process of China's modernization.

Of course, this multifaceted modernity clearly sacrificed the city's historical sites in the process. More than 7,000 thousand *hutong* were recorded in a survey in 1949; during the late 1990s, on average 600 of them were destroyed each year. In 2005, the state's news media reported that only 1,300 *hutong* remained (3).

Meyer's statements do not just criticize the demolition and transformation of old Beijing, but also reveal the history of Beijing's transformation that created multiculturalism by combining tradition and modernity internally while connecting East to West externally. Furthermore, we should remember that the transformation of the city was compelled by modernity.

Currently, Beijing leads the arts in China, with the largest number of high-quality cultural institutions, and it is known for its impressive cultural infrastructure. The city is a modern cosmopolitan metropolis that welcomes many international and domestic immigrants, including students and workers, and offers a unique, rich cultural atmosphere. Beijing's culture has a growing international influence that supports healthy East–West relations, which are expected to grow stronger because of the 2022 Winter Olympics to be held there.

Marginocentric cities and multicultural space

We conclude that Beijing is typical of a marginocentric city because it functions as a national center while being at the international margins. Precisely because of its marginocentricity, Beijing has encouraged the breaching of both traditional and modern boundaries, and has pioneered its own long journey toward Westernization, which has reshaped the city's cartography. Beijing's long history has developed its multicultural space of dialogue and negotiation between East and West and between tradition and modernity. I agree with Cornis-Pope's notion of the marginocentric city and its multicultural space, but we also should recognize that the multicultural space of the marginocentric city does not neutralize social differences by evaluating all cultural practices as valid per se. "Space is a product … and the space thus produced also serves as a tool of thought and of action … In addition to being a means of production, it is also a means of control, and, hence, of domination, of power" (5). In other words, although there are different levels of multicultural space, the production of space is still directed, regulated, and defined by the power accorded to the dominant ideology. I refer to Meyer's book as cross-cultural, not only because he is an American and has an American cultural perspective, but also because of the cross-cultural content of his writing. His book

seems to always try to interpret Beijing's transformation from a cross-cultural perspective, such as when he discusses other countries that have experienced the reshaping of their old cities.

When we clearly perceive that other cities have made mistakes like Beijing's, our sense of guilt about the transformation of old Beijing might be somewhat assuaged, because that perception might rationalize the direction taken by the city's modernization. Even so, we cannot forgive the excessive modernization of Beijing, and we must hope to hold on to the memories of its history as we enjoy its modern conveniences. As a Chinese citizen, I often complain about my overdeveloped city and its attendant excessive transportation and overcrowding, along with its heavily polluted air. However, I would be unhappy if a foreigner were to make similar comments about its shortcomings. I confess that this is an unfair attitude toward foreigners, but I fortunately feel more forgiving when I read Meyer's neutral impressions of Beijing. *The Last Days of Old Beijing* is an obviously cross-cultural work that can gain wide acceptance.

References

1. M. Cornis-Pope, J. Neubauer, *ACLS Occasional Paper* **52**, 26 (2002); available at https://acls.org/uploadedfiles/publications/op/52_literary_cultures_in_east_central_europe.pdf.
2. M. Sorkin, Ed., *Variations on a Theme Park*: *The New American City and the End of Public Space* (Noonday Press, New York, 1992), pp. 94–122.
3. M. Meyer, *The Last Days of Old Beijing: Life in the Vanishing Backstreets of a City Transformed* (Walker & Company, New York, 2008), pp. 6–143.
4. I. Even-Zohar, *Poetics Today* **11**, 1–6 (1990); available at http://www.tau.ac.il/~itamarez/works/books/Even-Zohar_1990--Polysystem%20studies.pdf.
5. H. Lefebvre, *The Production of Space* (Wiley-Blackwell, Toronto, 1992), p. 59; available at https://monoskop.org/images/7/75/Lefebvre_Henri_The_Production_of_Space.pdf.

Hangzhou Normal University's Overseas Talent Recruitment Scheme

Key disciplines or research areas

Key disciplines:

Public management, chemistry, biology, mathematics, foreign languages and literature, psychology, art theory, Chinese language and literature, nursing, fine arts, pedagogy, Chinese history, law, and ecology.

Key platforms:

- Institutes of Psychological Sciences (IPS): psychology, pedagogy, and clinical neuroscience
- Holistic Integrative Pharmacy Institutes (Key Laboratory of Elemene Class Anti-cancer Chinese Medicine of Zhejiang Province): pharmacy, oncology
- Institute of Aging Research: aging biology, molecular basis of aging-associated diseases

Recruitment plan

National, provincial, and municipal recruitment plan for high-level overseas talent:

1. National One Thousand Talents Project
2. Provincial Thousand Talents Program
3. Global leads to "521" plan

Hangzhou Normal University Outstanding Talent Plan:

Category 1 - Distinguished Scholars ("Jie" Talents), including two subcategories: Tier I and II scholars
Category 2 - Top Young Scholars ("Ying" Talents), including two subcategories: Tier III and IV scholars
Category 3 - Outstanding Young Scholars ("Jun" Talents), including two subcategories: Tier V and VI scholars

Remuneration

National, provincial, and municipal recruitment plan for high-level overseas talents:

One-off subsidy and research start-up funding from the Chinese government; university offers the corresponding position and remuneration

Hangzhou Normal University Outstanding Talent Plan:

- Basic salary plus performance-based pay
- Research support
- Relocation allowance
- Other benefits: medical insurance, assistance with spousal employment, and schooling for children

Contact information:
Submit applications to Mrs. Li or Mrs. Lu
Telephone: +860571-28868831, +86 0571-28869127
E-mail: hsdzp@hznu.edu.cn

For additional information regarding applications, such as the number of openings, please visit:
rsc.hznu.edu.cn and www.hznu.edu.cn

ENERGY
CHILWEE

www.cnchaowei.com
MADE IN CHINA

MANUFACTURER / OPERATOR / SERVICE PROVIDER

超威
CHILWEE

ADVOCATING GREEN ENERGY PERFECTING HUMAN LIFE

SANOFI - INSTITUT PASTEUR 2018 AWARDS
€ 350 000
FOR BIOMEDICAL RESEARCH

INSPIRED BY INSTITUT PASTEUR
SUPPORTED BY SANOFI

CALL FOR NOMINATIONS

SPECIAL FOCUS FOR GLOBAL HEALTH
1 – IMMUNOLOGY
2 - MICROBIOLOGY & INFECTION

2 AWARDS
1 International award - € 200 000
1 International Mid-Career award - € 150 000

DEADLINE FOR APPLICATION : JUNE, 25th
Awards Ceremony : Thursday, 15th November, 2018

CONTACT :
2018awards@pasteur.fr

MORE INFORMATION ON :
www.sanofi-institutpasteur-awards.com

A distinguished international Jury will choose the Awardees:
Dr. Elizabeth H. Blackburn, Prof. Pascale Cossart, Prof. Anne O'Garra, Dr. Yasmine Belkaïd, Prof. Alain Fischer, Prof. Jörg H. Hacker, Prof. Jules A. Hoffmann, Prof. Antonio Lanzavecchia, Dr Gary J. Nabel, Prof. Staffan Normark, Prof. Michel C. Nussenzweig, Prof. Jeffrey V. Ravetch, Prof. Philippe Sansonetti.

Scientific Coordinator of Awards: Prof. Pascale Cossart

SANOFI

Institut Pasteur

By **Carly Phillips**

My second coming out

told my dad using a "rip off the Band-Aid" strategy. I took a deep breath, dialed his cellphone, and said, "Hey Dad, it's Carly, I just, um, wanted to tell you that I'm, um, dating a woman. OK, talk to you later!" And just like that, I had both hung up on and come out to my dad. Fast forward 9 years: I'm an out and proud gay woman, accepted and loved by my family. I'm also in the depths of a Ph.D. program—and I've realized that I no longer want to be an academic. As I wrestled with whether and how to disclose this to my adviser, I was struck by how familiar the whole process felt, from a journey of self-discovery to a risk of rejection and loss of support. And I realized that I could learn from my queer experience to make this second coming out as smooth as possible.

When I started grad school, I was sure I would become a professor. So I focused on my research—but I also dabbled in the nonacademic realm, participating in outreach events and attending workshops about alternative careers. I convinced both my adviser and myself that this was just a phase, like the period when I wore a "Legalize Love" shirt and carefully proclaimed that I was an LGBTQ ally but certainly not queer myself.

Yet soon I was sneaking away to science communication conferences and arranging informational interviews at government organizations, reminiscent of my younger self wistfully wandering around West Hollywood, sneaking into gay bars, and watching *Ellen* whenever I could. I even quietly told other grad students and a handful of faculty members about my changing career interests, just as I had tried out my "Hi, I'm Carly and I'm a homosexual" line on friends and acquaintances before telling my family.

But keeping secrets and lying by omission ate away at me. As risky as disclosing my queerness had felt, I owed my family the chance to embrace a full version of me. My research mentor deserved the same.

As I thought about telling her, I reflected on the Band-Aid strategy I had used with my dad. My abrupt approach set up a one-way flow of information and created a tacit understanding that my gayness existed but was not to be discussed, which led me to avoid talking with my family about large aspects of my life for years. We have since reached a point where we can discuss most things freely. But if I had come out differently, we likely would have gotten here more quickly and with considerably less angst.

So, using the wisdom of my 28 years, I approached my

"I have been reminded that, in any coming out, the community catches you."

second coming out with a clarity, poise, and confidence that my frantic 19-year-old self could only dream of. I reminded myself that pulling off the Band-Aid is just the first step; the discussion that follows is often more important. Despite my racing heart and clammy hands, I initiated a calm, professional face-to-face conversation with my mentor so that we could both share our thoughts. She was surprised but not disappointed, and she began helping me network and plan for a new career track. We now talk openly about training opportunities, positions I might be interested in, and whether Chaco sandals are considered professional attire outside universities.

In this process I have been reminded that, in any coming out, the community catches you. Whether that means hosting a newly out and isolated friend for Christmas with your family or reading the eighth draft of a colleague's blog post, queers and scientists show up for each other. Since my second coming out, I have been overwhelmed by the support I have received from the many nonacademic scientists I've spoken with, whether it's offering to read cover letters, helping me rework a CV into a resume, or gently reminding me to not talk about my research too much during interviews.

Most important, I have been reminded that coming out does not change who I am at my core, as a daughter, a student, or a scientist. Even if I leave academia, I can still contribute to our understanding of the natural world, use that understanding to improve people's lives, and—hopefully—continue wearing Chacos to work. ■

Carly Phillips is a grad student at the University of Georgia in Athens. Send your story to SciCareerEditor@aaas.org.