Fighting the flu with nanobodies *pp. 511 & 598*

A U.S. quake warning system buys precious seconds *p. 514*

Environmental gene editing requires local engagement *p. 527*

Science

$15
2 NOVEMBER 2018
sciencemag.org

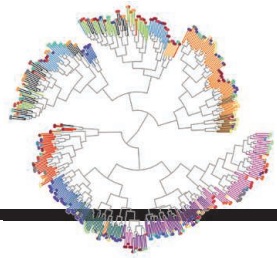
AAAS

SPECIAL ISSUE

COMPOSITE MATERIALS

Inspirations & applications

CONTENTS



524 & 577
Predicting viral origins

2 NOVEMBER 2018 • VOLUME 362 • ISSUE 6414

SPECIAL SECTION

COMPOSITE MATERIALS

ON THE COVER

This whimsical photo of a coconut turned into a car represents the increasing use of natural materials in commercial composites, particularly in the automotive sector. Coconut fibers (coir) and powdered coconut shells can be combined with a variety of polymers to make car parts, including door cladding, structural guards, and internal components. For more on the investigation and application of biological, hybrid, and carbon-based composite materials, see page 534.
Photo: Cary and Babs Wolinsky

NEWS

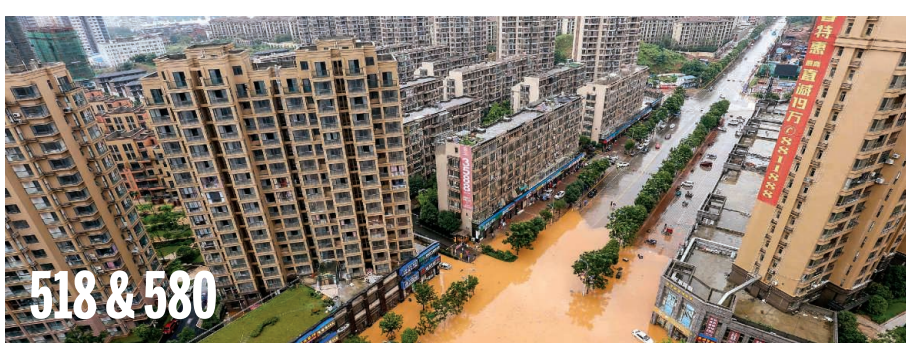
IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

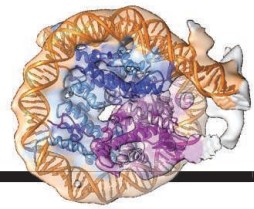
PERSPECTIVES



518 & 580

Published by AAAS

CONTENTS



595 Snapshot of nucleosomal transcription

2 NOVEMBER 2018 • VOLUME 362 • ISSUE 6414

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

CREDITS (FROM TOP): KUJIRAI *ET AL.*; ROBERT NEUBECKER

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Clarendon House, Clarendon Road, Cambridge, UK CB2 8FH

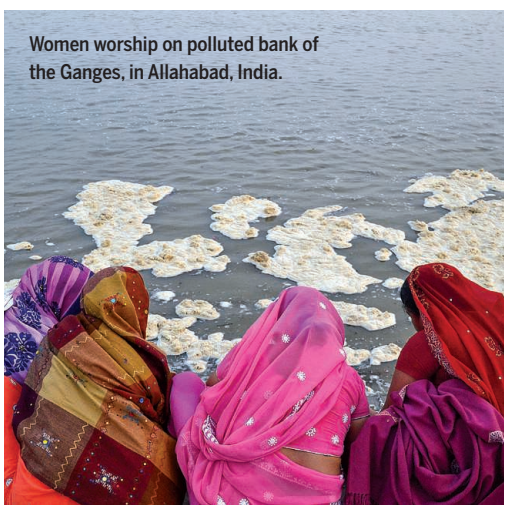
Published by AAAS

EDITORIAL

How to clean up the Ganges?

For millennia, the Ganges River, holy to Hindus, has provided livelihoods, food, and water for Nepal, India, and Bangladesh. Last month, one of India's leading environmental activists died after a 111-day hunger strike, failing to evoke changes to save India's most revered river (known as Ganga). After years of unrelenting abuse, Ganga is now one of the world's worst polluted rivers. India's Prime Minister Narendra Modi vowed in 2014 to clean Ganga by 2019, but despite increased funding and much lip service, the river is more polluted than before. Mr. Modi needs a new strategy.

Pressure on Ganga has been building for decades. With a tripling of human population since 1950 and rapid urbanization, 50 cities along Ganga daily release 6 billion liters of untreated sewage into the river, by far the largest source of pollution. Untreated industrial effluents compound the problem, together with run-off of chemical fertilizers and pesticides. Unfettered disposal of human and animal corpses into Ganga makes it unfit even for ritual bathing. Barrages and hydroelectric projects on the main stem of Ganga and its tributaries divert 60% of their waters, leaving little mainstream flow, which further concentrates pollutants in the river.



Women worship on polluted bank of the Ganges, in Allahabad, India.

"...despite increased funding and much lip service, the river is more polluted than before."

Earlier governments tried cleaning Ganga, investing some U.S.$14 billion, mostly in plants that treat sewage to an acceptable level of pollutants before discharge into waterways. Mr. Modi brought a new sense of urgency to the task, allocating U.S.$27 billion up to 2019. But money alone will not clean the river. Complex long-standing issues must be addressed, including increasing involvement of stakeholders, reducing corruption in pollution control agencies, increasing accountability and rule enforcement, and inciting behavioral change among citizens. This requires tackling three interrelated challenges, each involving different stakeholders. Municipalities must curtail discharge of untreated sewage and rapidly build sewerage infrastructures. Pollution control authorities must ensure treatment of municipal and industrial effluents to the needed standards. And water managers must enhance river flow so that secondary treated waste can be safely discharged into Ganga.

Since 1855, a profusion of barrages and dams has diverted water from Ganga and its tributaries for hydropower and canal irrigation. But over the past five decades, farmers have increasingly turned away from canal irrigation to shallow tube wells. Herein lies a big opportunity for cleaning Ganga. During the 1970s, modelers recommended cranking up the "Ganges water machine" to relieve flooding in the eastern parts of the river basin by promoting intensive groundwater-based irrigation during winter and summer months. The strategy was to draw down basin water tables in the copious alluvial aquifers. Shallow aquifers would then absorb monsoon floods and snowmelt, and protect Bihar, North Bengal, and Bangladesh from annual floods. Monsoons would replenish aquifers that shallow tube wells could then tap into for irrigation. This concept did not catch on back then, but today, the water machine is alive and kicking thanks to easy access to credit and cheaper drilling technology. With over 6.5 million shallow tube wells in Nepal, India, and Bangladesh, the Ganga basin is one of the most densely plumbed aquifer systems in the world. More than 80% of farmers depend on these wells. The water machine would be revved up even more but for the high cost of diesel that farmers must use to pump groundwater. But affordable electricity or solar pumps could wean farmers from canal irrigation, leaving more water to flow in Ganga and its tributaries, without adversely affecting hydropower.

The quickest, cheapest, and most effective way for Mr. Modi to show a less polluted Ganga by 2019 would be operating dams and barrages in the Ganga basin with the sole objective of augmenting river flows. This would be a start to controlling discharge of untreated sewage and industrial waste, which will take a long time.

–Tushaar Shah, Chittaranjan Ray, Uma Lele*

Tushaar Shah
is a senior fellow of the International Water Management Institute, Colombo, Sri Lanka, and is working in Gujarat, India. t.shah@cgiar.org

Chittaranjan Ray
is director of the Nebraska Water Center at the University of Nebraska, Lincoln, NE, USA. cray@nebraska.edu

Uma Lele
is president-elect of the International Association of Agricultural Economists, Milwaukee, WI, USA. umalele1@gmail.com

*The views expressed in this editorial are those of the authors and do not necessarily represent the opinions of their institutions.

10.1126/science.aav8261

PHOTO: RITESH SHUKLA/NURPHOTO/GETTY IMAGES

Corrected 1 November 2018. See full text.

NEWS

> “It’s the little spacecraft that could. It did everything we asked of it, and more.”
>
> **Jessie Dotson**, **a** NASA project scientist, announcing the demise this week of Kepler, the prolific exoplanet-hunting satellite, which ran out of fuel after 9 years in space.

IN BRIEF

Edited by **Jeffrey Brainard**

ASTRONOMY

Cosmic law renamed to expand credit

The change honors Georges Lemaître, an astronomer and priest, as co-discoverer of cosmic expansion.

Hubble’s Law, a cornerstone of cosmology that describes the expansion of the universe, should instead be called the Hubble-Lemaître Law, following a vote of more than 4000 members of the International Astronomical Union (IAU) announced this week. The change aims to credit Georges Lemaître, a Belgian astronomer and Catholic priest who in 1927 calculated that the universe was expanding, 2 years before U.S. astronomer Edwin Hubble found that galaxies farther from Earth were receding faster. Some astronomers criticized IAU for not allowing enough debate. “The IAU presented the issue as neat and tidy, but it is a much more murky and messy tale,” says Michael Merrifield of the University of Nottingham in the United Kingdom, because other researchers could have a claim. Others argue that IAU has no standing to get involved in the history of science and rename physical laws. The vote does not change the name of the Hubble constant, the actual expansion rate.

A bid to sequence all species

GENOMICS | The Earth BioGenome Project—an ambitious $4.7 billion, 10-year plan to sequence the genomes of all of Earth’s 1.5 million known species of animal, plant, fungus, and protozoan—officially began this week. In addition to providing insights into evolution, the genomic data are intended to help conserve and restore biodiversity as well as provide benefits for agricultural and biomedical research. The Wellcome Sanger Institute in Hinxton, U.K.; BGI in Shenzhen, China; and 15 other institutions have pledged to raise $600 million initially and agreed to coordinate their efforts and to work with other sequencing projects, to avoid overlap in species covered, for example. Most of these affiliated projects concentrate on taxonomic groups, such as insects or fungi, but some focus on the species of specific countries. The Wellcome Trust, for example, announced this week plans to sequence within 10 years all 66,000 known eukaryotic species in the United Kingdom.

Scrolls revealed as forged

BIBLICAL ARCHAEOLOGY | An independent analysis of their ink revealed that five fragments of text claimed to be pieces of the Dead Sea Scrolls are modern forgeries, the Museum of the Bible in Washington, D.C., announced last week. Scholars had long questioned the authenticity of these and the 11 other fragments in the museum’s collection. The texts were bought on the antiquities market by the Green family, the evangelical billionaires who have bankrolled the museum (*Science*, 20 October 2017, p. 295). Its administrators say they will replace the known fakes in the museum’s Dead Sea Scrolls display with other fragments in its collection, with labels saying that their authenticity, too, has been questioned.

Canada introduces carbon tax

CLIMATE CHANGE | Canada announced last week it will implement a tax on carbon emissions in the four provinces that do not already have either a tax or cap-and-trade system of their own—Manitoba, New

PHOTO: BETTMANN/GETTY IMAGES

Corrected 1 November 2018. See full text.

Brunswick, Ontario, and Saskatchewan. The federal levy will come into effect on 1 April 2019 at CA$20 per ton, rising to CA$50 per ton in 2022, Canadian Prime Minister Justin Trudeau said. The government expects the tax to raise CA$2.36 billion next year, but under a measure unique among countries that have implemented a carbon tax, it will return 90% of that revenue directly to taxpayers in the four provinces, providing them an incentive to reduce their personal emissions from burning fossil fuels. The size of the refund is based on each province's energy mix—ranging from an average of CA$248 per household in New Brunswick to CA$598 per household in Saskatchewan—and will increase with carbon prices. Ontario and Saskatchewan are challenging the federal tax in court.

Genomics lab leaders cleared

WORKPLACE | The Wellcome Sanger Institute, the world-renowned genomics center in Hinxton, U.K., last week cleared its director of allegations of bullying and gender discrimination after an investigation led by an outside attorney, Thomas Kibling. He concluded that director Michael Stratton had not bullied employees and that there was no evidence of gender discrimination, "wrongful exploitation of scientific work," or misuse of grant monies by Stratton and other senior institute managers. But the lawyer's report faulted the institute for a paucity of female scientists—seven of the institute's 33 faculty members are women—and lack of transparency in its processes for firing employees. The probe responded to a 170-page April complaint by a woman who was terminated; she also referred to a joint statement of concern from several other employees. In a statement, Stratton apologized for "failures in people management" and added he is "personally committed to" righting the gender imbalance among "our scientific leaders." Jeremy Farrar, director of the Wellcome Trust in London, the charity that funds the institute, also apologized in a statement that the institute failed to recognize and act on diversity and inclusion issues sooner.

Toughest TB cases routinely cured

PUBLIC HEALTH | A clinical trial from Belarus has provided what some tuberculosis (TB) experts say is the most compelling evidence that the drug bedaquiline can help cure most people who have an extensively drug resistant (XDR) form of the lung disease. Bedaquiline (sold as Sirturo) came to market in 2012 primarily to treat multidrug-resistant (MDR) TB—a form that doesn't respond to the most common TB drugs but usually remains treatable by other medications. About 600,000 people each year acquire MDR TB, and about 6% of those fall into the XDR category. Their TB doesn't respond to the second tier of drugs and typically produces a mortality rate of 60%. At the 49th Union World Conference on Lung Health held last week in The Hague, Netherlands, researchers reported that bedaquiline-containing regimens cured 80% of a group of 181 TB drug–resistant patients, 64% of whom had XDR. Only three people in the group died.

Hungary forces university move

HIGHER EDUCATION | The Central European University (CEU) in Budapest, under pressure from Hungary's right-wing government, will send incoming students to a new campus in Vienna starting next year. The private university, founded in 1991 by American-Hungarian billionaire philanthropist George Soros and based in Budapest since 1993, grants U.S.-accredited

Only a few scientists have squeezed into a South African cave rich in hominin bones.

PHOTO: ROBERT CLARK/CONTRIBUTOR/GETTY IMAGES

PALEOANTHROPOLOGY

Virtual tour reveals fossil cave

Only a handful of people have entered the Dinaledi chamber in the Rising Star cave in South Africa, where hundreds of bones of a mysterious early hominin called *Homo naledi* were discovered in 2015. To access it, researchers must squeeze through a chute only 18 centimeters wide. But now, anyone with a smartphone can explore the fossil chamber. The Perot Museum of Nature and Science in Dallas, Texas, and the University of the Witwatersrand in Johannesburg, South Africa, developed a virtual reality tour, which opened in May for visitors to the museum. To expand the exhibit's reach, the museum and the university last week released the tour in a freely downloadable app format, compatible with Google Cardboard, a virtual reality headset that can be made at home. The tour is available in English, Spanish, isiZulu, Setswana, and Sesotho.

Corrected 1 November 2018. See full text.

graduate degrees in environmental science, math, and other fields. In 2017, the government enacted a law requiring international universities to have a campus in the country where they are officially chartered. CEU says its collaboration with Bard College in Annandale-on-Hudson, New York, fulfills the requirements. Hungary's government, however, has declined to sign the required agreeement. The university said in its 25 October announcement that current students can remain in Budapest to finish their degrees and that it will "maintain as much research and educational activity in Budapest as possible."

Iran brings capital charges

INTERNATIONAL AFFAIRS | Prosecutors in Iran have charged four conservationists with "sowing corruption on Earth"—a crime punishable by death. The environmentalists, with the Persian Wildlife Heritage Foundation in Tehran, were arrested on suspicion of espionage in January. Iran's Revolutionary Guards accused them of using camera traps—intended for remotely monitoring rare Asiatic cheetahs and other wildlife—to eavesdrop on the nation's ballistic missile program. Two of the accused serve on International Union for Conservation of Nature (IUCN) panels that recommend whether to add or remove species from IUCN's Red List of Threatened Species. Taher Ghadirian and Houman Jowkar are members of the cat specialist group, and Ghadirian is also part of the bear specialist group. Many observers view the detainees as pawns in a power struggle between the hardline Revolutionary Guards and Iranian President Hassan Rouhani's relatively moderate administration, which in a review last spring determined that the spying accusation is baseless but has been powerless to secure the conservationists' release.

Court OKs Hawaiian telescope

ASTRONOMY | The Supreme Court of the State of Hawaii ruled this week that the Thirty Meter Telescope (TMT) project should be granted a permit for construction on Mauna Kea. The TMT, to be one of the world's largest optical telescopes, is opposed by some native Hawaiian groups because the mountain is a sacred site. The project's original permit was revoked in 2015 on procedural grounds because opponents had not been given opportunities to object. The state's supreme court dismissed another challenge, to the University of Hawaii's sublease of land for the project, in August. The TMT board now faces no legal obstacles to restarting construction, though it may still face protesters on the mountain.

Dinosaurs began colored eggs

PALEONTOLOGY | Birds were not the first to lay colored eggs: Dinosaur eggshells came in a panoply of hues and speckle patterns, researchers report in the 1 November issue of *Nature*. The authors argue that eggshell coloration likely had a single evolutionary origin, in the carnivorous dinosaurs called theropods that gave rise to modern birds. The researchers, led by Jasmina Wiemann of Yale University, used spectroscopic analysis of fossils to identify pigments in eggs of 15 species of dinosaurs and prehistoric birds, revealing their probable original colors. As with modern birds, the dinosaurs' tinted shells likely camouflaged their eggs from predators, while distinctive speckling patterns may have helped parents distinguish their own eggs from those of cuckoolike dinosaurian nest parasites, Wiemann says.

Tests of fossilized dinosaur eggs (top eggshell) show they had colors like those of modern bird eggs.

PHOTO: JASMINA WIEMANN

China allows rhino, tiger trade

WILDLIFE CONSERVATION | In a setback for conservationists, China this week said it will legalize the use of tiger and rhinoceros bones and tissues "in medical research and healing," as long as the material comes from farmed animals. The move ends a ban that had been in place since 1993. Many Chinese believe tiger and rhino parts have medicinal properties, despite a lack of scientific evidence. Conservationists fear the new policy will ease the ability of poachers to smuggle parts from wild animals into the Chinese market, further imperiling the endangered species.

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

The Swiss team tests a laser mapping submersible in Lake Annecy in France.

MARINE GEOLOGY

Seafloor mappers to compete for XPRIZE

Faster, cheaper autonomous systems could aid in resource extraction and science

By **Julia Rosen**

Next week, a small yellow and white–striped boat will slip out of port in Kalamata, Greece, and motor away from shore. The vessel won't carry a captain or crew, just an array of electronics that will tell it where to go, and when to drop the torpedo-shaped pod lodged in its stern. Once released, the sonar-equipped vehicle will descend several kilometers into the frigid abyss of the Hellenic Trench, the deepest part of the Mediterranean Sea, and map the sea floor with pinging pulses of sound. The team behind the effort is the first of eight competing over the next few months in the finals of the $7 million Shell Ocean Discovery XPRIZE. "I'm not sure if we are crazy or not, but we decided to go first," says Rochelle Wigley, a marine geologist at the University of New Hampshire in Durham, who leads the XPRIZE team of the Japanese Nippon Foundation and the General Bathymetric Chart of the Oceans (GEBCO), an international organization.

XPRIZE, a nonprofit based in Culver City, California, runs competitions to spur innovation, and in 2015, it turned to the problem of mapping the ocean floor, says contest director Jyotika Virmani. The catalyst was the disappearance of Malaysia Airlines Flight 370 somewhere over the Indian Ocean, and the stark realization that recovery teams knew little about what lay below the surface of the search area. "Instead of the airplane, unfortunately, they did find two new volcanoes, one of which is bigger than Mount Vesuvius," she says.

Sharper pictures of the ocean floor could help companies look for resources such as oil. (The energy company Shell is the prize's sponsor.) But researchers want a clearer view, too. For example, Dave Clague, a geologist at the Monterey Bay Aquarium Research Institute in Moss Landing, California, studies volcanic activity along midocean ridges—submarine mountain chains that generate new ocean crust—by identifying lava flows. But scientists have fine-scale maps for only a tiny fraction of the 65,000-kilometer-long system, limiting their understanding of how new crust forms and what happens to it as it moves away from the ridge.

Biologists also need better maps, to manage fisheries and identify deep-sea habitats. They have already discovered new colonies of cold-water corals just by looking for structures rising from the sea floor, says Craig Brown, a mapping expert at Nova Scotia Community College in Halifax, Canada. "They usually have quite dramatic topography," he says.

So far, just 9% of the sea floor has

Race to the bottom

The eight teams competing for the ocean mapping XPRIZE use a mix of uncrewed surface vehicles and autonomous underwater vehicles (AUVs).

TEAM NAME	COUNTRY	SURFACE OPS	NUMBER OF AUVS
Arggonauts	Germany	Five ships	Five
Blue Devil Ocean Engineering	United States	Two aerial drones	Two
CFIS	Switzerland	None	20
GEBCO-Nippon Foundation alumni	International	One ship	One
Kuroshio	Japan	One ship	Two
PISCES	Portugal	One ship, two acoustic beacons	One
Team Tao	United Kingdom	One ship	Five
Texas A&M	United States	One ship	One

CREDITS: (PHOTO) XPRIZE; (DATA) XPRIZE TEAMS

Corrected 1 November 2018. See full text.

Published by AAAS

been mapped in detail with modern sonar technology, Wigley says, and only 18% of the world's ocean bottom has been surveyed at all, often at resolution so coarse that jumbo jets—and volcanoes—would have no trouble hiding. The rest—four-fifths of the two-thirds of the planet covered by water—is virtually unknown. As usual, the limitations are money and time. The research vessels that do high-resolution mapping cost up to $100,000 a day to operate. And they move so slowly that it would take centuries for them to chart the world's oceans, Virmani says.

Satellites can also map the sea floor, by measuring slight variations in the ocean surface caused by the gravitational pull of massive seafloor features. But the resolution is crude. In recent years, researchers have turned to autonomous underwater vehicles (AUVs). They follow preprogrammed paths using inertial navigation systems that precisely track their speed and direction, and carry miniature multibeam sonars. By cruising close to the ocean bottom, they can detect contours in the seabed smaller than a meter—a vast improvement over the 50-meter resolution of a typical ship-based system working in the deep ocean, says Clague, who is not involved in the XPRIZE contest. But the AUVs are still slow. Efforts to add batteries and extend diving time only bulk up the AUV, requiring bigger ships to launch them, "which kind of defeats the purpose," Clague says.

XPRIZE hopes its competition will spark faster, cheaper autonomous systems. Starting from shore, the eight finalists must map between 250 and 500 square kilometers in 24 hours, at depths down to 4000 meters and resolutions of 5 meters or better. They must also carry instruments to collect images of 10 interesting features and find a trophy stashed on the sea floor. The technical challenges include building instruments to withstand enormous pressure, balancing battery life against speed, and making the robots smart enough to carry out the whole operation without human guidance. "Everything is hard," says Martin Brooke, an engineer at Duke University in Durham, North Carolina, and leader of its XPRIZE team.

Brooke's group—mostly engineering students—will try to gain time by using heavy-lift aerial drones to carry buoys that will lower tethered mapping pods into the ocean. Most teams use an autonomous surface vessel to save their AUV's precious power and to serve as a communication hub. The Swiss CFIS team, led by Toby Jackson, a financial trader–turned–inventor, will send 20 lightweight, 3D-printed AUVs directly from shore. Instead of sonar, they will use lasers, which can bounce light off the sea floor because they are at such close range.

Team Tao will also use a swarm approach, launching five custom-built AUVs from an autonomous catamaran it calls the "vending machine." Eventually, the system will carry two dozen subsea drones, says team leader Hua Khee Chan, an engineer at Newcastle University in the United Kingdom, allowing half to work while the others charge. Each AUV will follow a simple vertical path, enabling it to sample the temperature and salinity of the water column as it descends. Chan says it's "extra data that we get for free while it's traveling." Both Chan and Jackson say they aim to produce their AUVs for less than $25,000 a pop—a bargain compared with the sophisticated models used today, which can cost $1 million or more.

Cheaper, more flexible systems could help researchers rapidly fill the gaps in seafloor maps—and enable repeat surveys to monitor changes over time. Clague would like to measure how much lava is produced during a single eruption on a midocean ridge, which gives clues about magma generation in the mantle. Repeat mapping could also track movement along offshore faults that generate earthquakes, and in seafloor sediments after major weather events.

> ***"... you go into the ocean and things are not like math."***
> **Nuno Cruz**, University of Porto

As XPRIZE's sponsor, Shell reserves first rights to negotiate with each team for use of its technology, which it could use for oil and gas exploration or to monitor production wells and pipelines. Companies hoping to mine the sea floor for minerals are also eager to get a better look. But Wigley says mapping could also aid in marine protection. "If we understand the sea floor better, we can manage where it's happening better and understand the impacts better."

For now, that's a long way off, and most teams are just scrambling to prepare for the competition in Greece. A Portuguese team still hasn't tested its acoustic positioning system, which relies on a constellation of floating beacons, in deep water. "From the math, it should work," says team leader Nuno Cruz, an engineer at the University of Porto in Portugal. "But you go into the ocean and things are not like math." Some teams already know they won't win, but they are fine with that. Most entered for the challenge, not the purse, and XPRIZE is pleased with the progress they've made, Virmani says. "We've already shifted the field." ■

Julia Rosen is a journalist based in Portland, Oregon.

ENERGY

Advances in flow batteries promise cheap backup power

Upstart technology could enable widespread adoption of renewables

By **Robert F. Service**

Batteries already power electronics, tools, and cars; soon, they could help sustain the entire electric grid. With the rise of wind and solar power, energy companies are looking for ways to keep electrons flowing when the sun doesn't shine and the wind ebbs. Giant devices called flow batteries, using tanks of electrolytes capable of storing enough electricity to power thousands of homes for many hours, could be the answer. But most flow batteries rely on vanadium, a somewhat rare and expensive metal, and alternatives are short-lived and toxic.

Last week, researchers reported overcoming many of these drawbacks with a potentially cheap, long-lived, and safe flow battery. The work is part of a wave of advances generating optimism that a new generation of flow batteries will soon serve as a backstop for the deployment of wind and solar power on a grand scale. "There is lots of progress in this field right now," says Ulrich Schubert, a chemist at Friedrich Schiller University in Jena, Germany.

Lithium-ion batteries—the sort in laptops and Teslas—have a head start in grid-scale applications. Lithium batteries already bank backup power for hospitals, office parks, and even towns. But they don't scale up well to the larger sizes needed to provide backup power for cities, says Michael Perry, associate director for electrochemical energy systems at United Technologies Research Center in East Hartford, Connecticut.

That's where flow batteries come in. They store electrical charge in tanks of liquid electrolyte that is pumped through electrodes to extract the electrons; the spent electrolyte returns to the tank. When a solar panel or turbine provides electrons, the pumps push spent electrolyte back through the electrodes, where the electrolyte is recharged and returned to the holding tank

Corrected 1 November 2018. See full text.

(see graphic, right). Scaling up the batteries to store more power simply requires bigger tanks of electrolytes.

Vanadium has become a popular electrolyte component because the metal charges and discharges reliably for thousands of cycles. Rongke Power, in Dalian, China, for example, is building the world's largest vanadium flow battery, which should come online in 2020. The battery will store 800 megawatt-hours of energy, enough to power thousands of homes. The market for flow batteries—led by vanadium cells and zinc-bromine, another variety—could grow to nearly $1 billion annually over the next 5 years, according to the market research firm MarketsandMarkets.

But the price of vanadium has risen in recent years, and experts worry that if vanadium demand skyrockets, prices will, too. A leading alternative replaces vanadium with organic compounds that also grab and release electrons. Organic molecules can be precisely tailored to meet designers' needs, says Tianbiao Liu, a flow battery expert at Utah State University in Logan. But organics tend to degrade and need replacement after a few months, and some compounds work only with powerful acidic or basic electrolytes that can eat away at the pumps and prove dangerous if their tanks leak.

Researchers are now in the midst of "a second wave of progress" in organic flow batteries, Schubert says. In July, a group led by Harvard University materials scientist Michael Aziz reported in *Joule* that they had devised a long-lived organic molecule that loses only 3% of its charge-carrying capacity per year. Although that's still not stable enough, it was a big jump from previous organic flow cell batteries that lost a similar amount every day, Liu says.

Iron, which is cheap and good at grabbing and giving up electrons, is another promising alternative. A Portland, Oregon, company called EES, for example, sells such batteries. But EES's batteries require electrolytes operating at a pH between one and four, with acidity similar to vinegar's.

Now, Liu and his colleagues have come up with a flow battery that operates at neutral pH. They started with an iron-containing electrolyte, ferrocyanide, that has been studied in the past. But in previous ferrocyanide batteries, the electrolyte was dissolved in water containing sodium or potassium salts, which provide positively charged ions that move through the cell to balance the electron movement during charging and discharging. Ferrocyanide isn't very soluble in those salt solutions, limiting the electrical storage capacity of the battery.

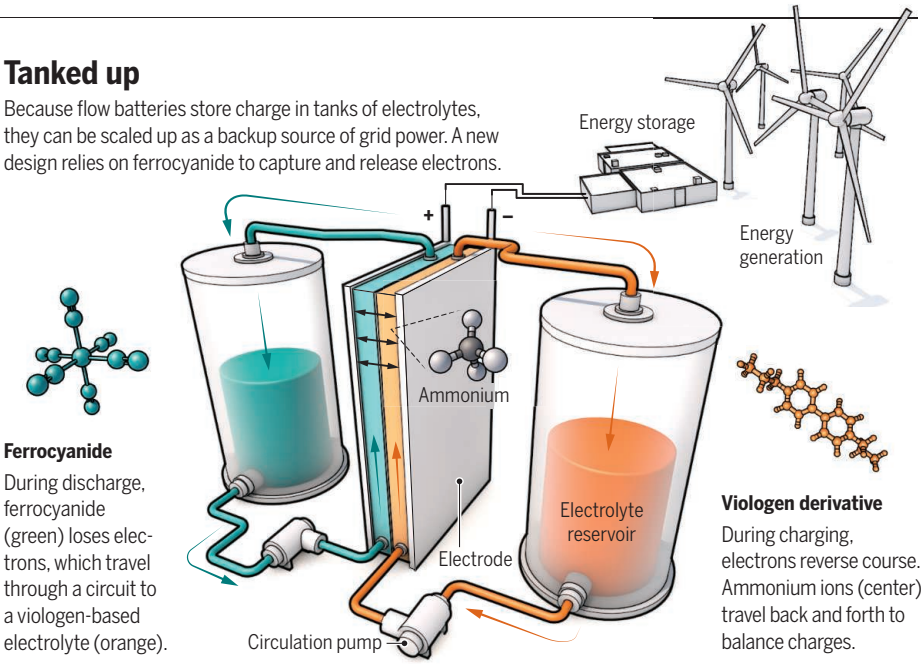
So Liu and his colleagues replaced the salts with a nitrogen-based compound called ammonium that allows at least twice as much ferrocyanide to dissolve, doubling the battery's capacity. The resulting battery is not as energy-dense as a vanadium flow battery. But in last week's issue of *Joule*, Liu and his colleagues reported that their iron-based organic flow battery shows no signs of degradation after 1000 charge-discharge cycles, equivalent to about 3 years of operation. And because the electrolytes are neutral pH and water-based, a leak likely wouldn't produce environmental damage.

"Overall, that's an excellent piece of work," says Qing Wang, a materials scientist at the National University of Singapore. Still, he and others caution that the battery is sluggish to charge and discharge. Liu says he and his colleagues plan to test other electrolyte additives, among other fixes, to boost conductivity.

It's too early to say which flow battery chemistry—if any—will support the renewable grid of the future. Another contender uses electrolytes made from metal-containing organic compounds called polyoxometalates, which store far more energy in the same volume than the competition. In the 10 October issue of *Nature Chemistry*, for example, researchers led by Leroy Cronin, a chemist at the University of Glasgow in the United Kingdom, reported a polyoxometalate flow battery that stores up to 40 times as much charge as vanadium cells of the same volume. The downside for now is that these electrolytes are highly viscous and thus more challenging to pump through the battery, Cronin says. "Today, no one flow battery fills all the needs," Schubert says. That means there's still plenty of room for innovation. ■

Tanked up

Because flow batteries store charge in tanks of electrolytes, they can be scaled up as a backup source of grid power. A new design relies on ferrocyanide to capture and release electrons.



Ferrocyanide

During discharge, ferrocyanide (green) loses electrons, which travel through a circuit to a viologen-based electrolyte (orange).

Viologen derivative

During charging, electrons reverse course. Ammonium ions (center) travel back and forth to balance charges.

Commercial flow batteries, such as this zinc-bromine system from Redflow, are helping back up renewables.

CREDITS: (PHOTO) REDFLOW LIMITED; (GRAPHIC) C. BICKEL/*SCIENCE*

Researchers distribute plastic bowls to collect fecal samples from Hadza people near Lake Eyasi in Tanzania.

MICROBIOLOGY

Microbiome conservancy stores global fecal samples

Research could help prevent or treat diseases of the gut

By **Tania Rabesandratana**

Whether in villages on the coast of Ghana or in the mountains of Rwanda, asking for people's poop is a good icebreaker, Mathieu Groussin says. "Everybody laughs," says Groussin, a microbiologist at the Massachusetts Institute of Technology (MIT) in Cambridge. "Especially when we stress that we need the whole fecal sample and show them the big bowl."

He's asking on behalf of the Global Microbiome Conservancy (GMC), an effort to identify and preserve gut bacteria from different peoples around the world. Most microbiome research has focused on Western, urban populations, which typically eat processed foods and use antibiotics. The few studies of traditional peoples have found a far more diverse gut microbiome that appears to be linked to the absence of certain diseases.

But as traditional societies change their lifestyles, that biodiversity is under threat, says Eric Alm, an MIT microbiologist who cofounded GMC in 2016 with Groussin and two other MIT postdocs. "Strains that coevolved with humans are currently disappearing," he says. Later this month, Groussin plans to expand the growing conservatory with samples from Nigerian communities affected by oil pollution near the Niger River delta.

Rescuing and preserving the microbes, Alm says, could pave the way for new treatments not just for gut ailments, but for other disorders linked to the microbiome—such as asthma, allergy, obesity, and diabetes. "I'm 100% confident that there are relevant medical applications for hundreds of strains we've screened and characterized," he says.

Gathering material from human subjects and importing it to the United States for posterity raises ethical and legal complications, however. And the project itself represents a pessimistic outlook, says Stephanie Schnorr, a biological anthropologist at the University of Nevada in Las Vegas who has studied the microbiome of the Hadza, a traditional people in Tanzania. "Keeping a biobank is already expecting that diversity loss is an inevitable process," she says.

Stool samples are put into small tubes and transported to the United States at –190°C.

GMC's biobank now houses about 11,000 strains, from about 40 people in seven countries. Its budget will support visits until 2021 to about 34 countries in total—from the Arctic to Africa, Asia, Oceania, and South America. After that, organizers hope to find several million dollars per year to expand the research and fund local science.

For now, the focus is on gathering strains. On collecting trips, Groussin retrieves the filled plastic bowls the morning after distributing them—or, in places with high fiber diets, within the hour. He then processes the samples in a makeshift lab in the back of a car. Portions of the stool are fixed and dried for DNA sequencing and lipid content measurement. The rest is divided into small tubes, preserved in glycerol, and shipped back to Cambridge in containers at –190°C. There, bacterial strains are isolated, using growth media that mimic the conditions of the gut, and preserved in perpetuity in large freezers.

The team is already uncovering novel strains. The 7000 strains in GMC's library that came from North American peoples include only five previously unknown genera. But the 4000 strains from Africa and the Arctic have already yielded 55 unknown genera.

Genomic data on the bacteria revealed another contrast between populations. In September, at a human microbiome symposium in Heidelberg, Germany, Groussin said GMC found preliminary evidence that so-called horizontal gene transfers between the strains living within one person are frequent enough to change the gut microbiome's function during a lifetime. These gene transfers are more frequent in industrialized populations, they found, possibly as a result of higher environmental pressures, such as antibiotic use.

Developing nations that have a history of exploitation can be wary of the effort. In Rwanda, for example, MIT researchers worked with John Rusine, director of the National Reference Library's biology lab in Kigali, to gain permission to transport stool samples to the United States. Rusine says he spent several months convincing the library's head to allow it. "Without his signature, we couldn't ship the samples out of the country." Each country keeps backup samples, and GMC trained local technicians to extract DNA. "Just keeping samples has no meaning if there is no further research here," Rusine says.

PHOTOS: CHRISTOPHER CORZETT

One way to ensure broad buy-in is to store samples in a territory perceived as "neutral, stable, safe, where their rights will be respected," says Maria Gloria Dominguez-Bello, a microbiologist at Rutgers University in New Brunswick, New Jersey. Dominguez-Bello leads an initiative to build an international storage facility modeled after the Svalbard Global Seed Vault, an underground cold storage building on a remote Norwegian island that safeguards plant diversity for future generations. Just as in the seed vault, researchers, institutions, or governments could make deposits in the microbiota vault, retrieve samples, and grant others access to them.

So far, the microbiota vault is just an idea, supported by a dozen volunteer scientists; it has no planned home yet, and Dominguez-Bello says she is seeking a few hundred million dollars to endow the project and get it started.

Who owns the preserved microbiome samples—and any scientific advances made using them—remains a legal puzzle. Microorganisms do fall under the Convention on Biological Diversity, a 1992 international treaty ratified by all United Nations member states except the United States. But it's not clear whether the convention applies to microbes that come from the human body.

The treaty intends to ensure that any R&D results or benefits arising from genetic resources are shared with the government or community that provided them in the first place. GMC is trying to abide by that spirit: In each country it visits, it has drawn up agreements stipulating that the stool samples and cultured bacteria strains remain the property of individual donors, and can only be used for noncommercial purposes.

But the very idea of preserving cells for future studies can be problematic, Schnorr cautions. For example, in her 2014 study of the Hadza, Schnorr used samples only in the exact way her study's consent forms described. She is skeptical that consent can be meaningful when scientists themselves don't know what questions they may ask in the future.

Kieran O'Doherty, a social scientist at the University of Guelph in Canada who has studied the ethics of microbiome research, thinks scientists should do more than archive the diversity of the human microbiome; they should look for ways to preserve it, by helping traditional peoples retain their sovereignty and natural resources. O'Doherty compares biobanking to small-scale efforts to fight climate change. "It's a good idea to use different light bulbs or drive a different car, but we need higher-level political action," he says. "To many scientists that's activism, and they're not comfortable with that." ■

IMAGE: XUEYONG ZHU AND IAN WILSON, SCRIPPS RESEARCH INSTITUTE

BIOMEDICINE

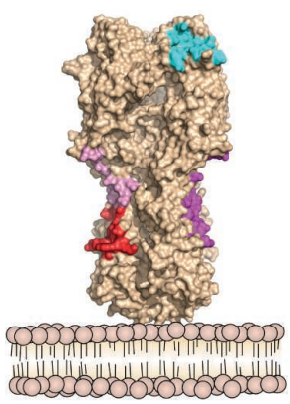
Llama antibodies inspire gene spray to prevent all flus

Strategy for mutable virus might outdo traditional vaccines

By **Jon Cohen**

Four llama antibodies and a harmless virus: This outlandish recipe could be the basis of a nasal spray designed to foil infection from all strains of influenza. The spray, containing a virus engineered to make a protein derived from the llama antibodies, has passed its first animal test, protecting mice from every known flu strain that infects humans, a research team reports on p. 598.

Although the strategy must go through more testing before human trials can begin, researchers who have struggled to develop a "universal" vaccine against the highly mutable flu virus say it merits serious attention. The nasal spray could prove a boon to the elderly, who typically suffer most from flu and get only weak protection from existing vaccines. And unlike traditional influenza vaccines, which are tailormade each flu season to match the viruses in circulation, it could be stockpiled as protection against a flu pandemic. "This is a great story and shows the power of antibody engineering," says immunologist Antonio Lanzavecchia, a leading flu vaccine researcher at the Institute for Research in Biomedicine in Bellinzona, Switzerland.



A flu-fighting antibody targets four sites (colored areas) on the virus's hemagglutinin surface protein.

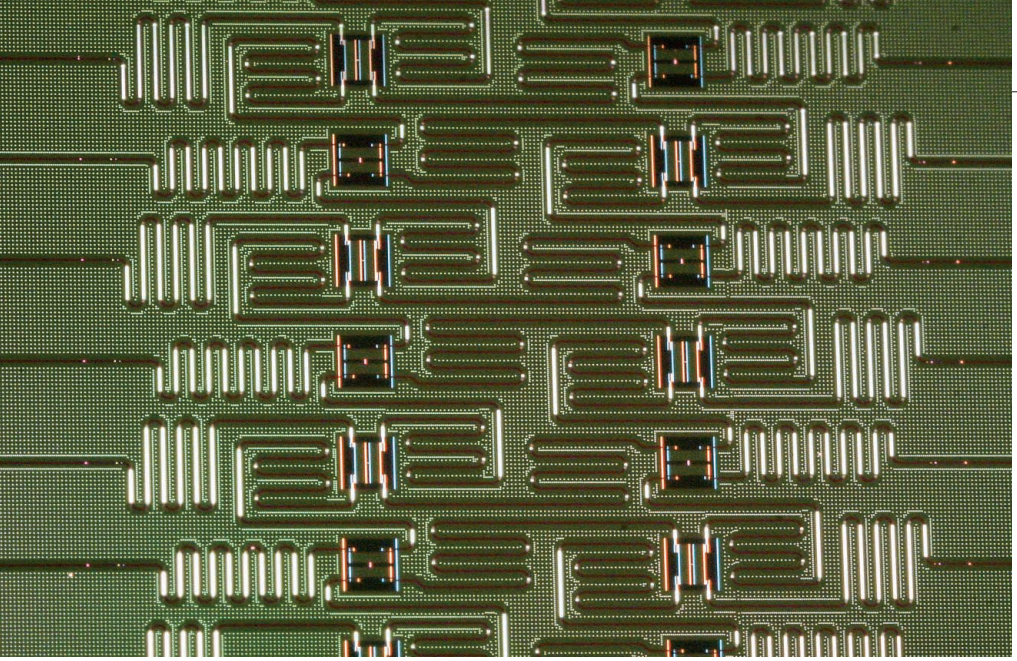
Antibody engineer Joost Kolkman at Janssen Infectious Diseases in Beerse, Belgium, and his colleagues thought an unusual class of antibodies made by llamas and their camel cousins might serve as a weapon against flu. These antibodies are unusually small because they lack the "light" peptide chain that normally bulks up each arm of the Y-shaped proteins. Researchers can further pare down the remaining "heavy" chains to create so-called nanobodies, able to reach into crevices of viruses that their full-size counterparts can't touch (*Science*, 11 May, p. 594).

To create nanobodies against the flu, the Janssen group injected llamas with a vaccine containing three different influenza viruses, as well as the viral surface protein, hemagglutinin, from two other flu strains. They then harvested four antibodies that each neutralized many flu strains. Ultimately, the team was able to engineer a gene that expressed a protein made up of nanobodies derived from all four antibodies. "It's very easy to link the domains together into one single molecule," Kolkman says. They spliced the gene into a benign adenovirus-associated virus (AAV) that's used in gene therapy experiments.

Test tube studies showed the four-in-one antibody prevented infection by 60 different influenza viruses from both the type A and B groups that infect people. "It's been quite hard to find an antibody that neutralizes both A and B," says Ian Wilson, a structural biologist at the Scripps Research Institute in San Diego, California, who helped work out how the nanobodies bound to the virus.

Mice given the synthetic antibody—delivered either by squirting the doctored virus into their noses of the mice or by infusing the protein directly into their circulation—had significantly higher survival rates than untreated rodents when injected with a variety of influenza strains. Wilson, who has published more than 50 papers on influenza antibodies, says he's never seen one with greater breadth and potency. Because AAVs can persist for months, the strategy could offer extended production. "Hopefully it would last the entire flu season in humans," Wilson says.

Immunologist James Crowe, an influenza antibody specialist and vaccine developer at Vanderbilt University in Nashville, cautions that human immune systems may see the llama-derived proteins as foreign and develop antibodies against them. He also notes that although AAV-based treatments are being tested for life-threatening diseases, giving the virus as a flu preventive would face more intense scrutiny from regulators. "The bar for putting AAV in a healthy individual is going to be very high," Crowe says. ■



Quantum computers made of superconducting circuits could outperform classical computers.

PHYSICS

Europe's €1 billion quantum flagship announces grants

Program aims to nudge quantum technology to market

By **Edwin Cartlidge**

The first phase of Europe's decadelong, billion-euro program to turn its quantum technology research into commercial products has come into focus. At an event held in Vienna on 29 October, the European Union announced that the first €132 million of its quantum flagship initiative will be split between 20 continent-wide consortia over the next 3 years to develop new kinds of quantum sensors, communications, and computers.

Backers hope the investment will keep Europe from being overtaken in a potent new area of technology. "It's important to start an applications sector to allow industry to grow in Europe," says Ian Walmsley, of the University of Oxford in the United Kingdom, who is also a member of the steering group that formulated the flagship. "No doubt it's growing elsewhere in the world." But it remains uncertain how the rest of the flagship will be paid for, and whether it will inject life into a fledgling European quantum industry.

Physicists have begun to find commercial applications for the strange laws of quantum mechanics, which allow a subatomic particle to be in two states at the same time and a measurement on one particle to instantly affect another, distant particle. For example, Swiss company ID Quantique, set up in 2001, sells equipment exploiting the quantum properties of photons to create uncrackable encryptions for banks and governments.

Basic research in quantum mechanics has flourished in Europe. But China is spending billions to commercialize quantum technology, including a satellite to send quantum-encrypted messages through space, launched in 2016—a first step toward a quantum internet. Meanwhile, the U.S. Congress is considering a $1.3 billion quantum initiative, and U.S. companies including Google, IBM, Intel, and Microsoft have already spent hundreds of millions of dollars to try to build a quantum computer that could outstrip conventional machines on certain tasks.

Such investment has been scarce in Europe, where companies without the huge cash reserves of U.S. tech firms have been reluctant to take risks. The quantum flagship—the third EU flagship research program after ones on graphene and the human brain—is intended to compensate. Without such support, says flagship spokesperson Tommaso Calarco of the Jülich research center in Germany, "the ideas that were developed and are still being developed in Europe could be converted into companies and jobs elsewhere."

The program was announced in 2016, and grant proposals from 140 consortia—each a mixture of academics and industrialists—were received earlier this year, then whittled down to the 20 winners across five categories. Seven winners will pursue basic science while many others will develop commercial prototypes. Four of the projects focus on quantum communication, among them an effort to develop a blueprint for a quantum internet. Two more will plunge into the race for quantum supremacy, which means executing a specific algorithm that the best classical computers can't handle.

Google aims to reach that milestone in the coming months using quantum bits, or qubits, made in superconducting circuits, says John Martinis, the company's head of quantum hardware in Santa Barbara, California. Yet Thomas Monz of the University of Innsbruck in Austria, who coordinates one of the European consortia, says his group's bid for quantum supremacy, which uses trapped ions as qubits, is based on an algorithm that will be more "meaningful" than Google's.

A full-scale quantum computer is decades off, however. Among the consortia developing more tangible quantum devices are a Dutch-led group that intends to make a portable and easy-to-use optical clock that could help telecom companies end their dependence on potentially unreliable GPS signals. A group led by the Fraunhofer Institute for Applied Solid State Physics in Freiburg, Germany, is working on a room-temperature device to supply the spin-polarized molecules needed for magnetic resonance imaging machines.

The grants amount to a fraction of the initiative's €1 billion commitment. Calarco says the next funding round could combine calls for fresh proposals with continued support for existing projects. But the source of the money is in question. Funding is supposed to be split 50-50 between the European Commission and member states, as in other flagships. But in the quantum initiative, member states can spend their share on national programs that merely share the same aims. To avoid this, Calarco is hoping that when the budget for the next EU research framework is decided next year, it will contain all of the remaining €850 million the flagship needs.

An additional uncertainty is how Brexit—the United Kingdom's departure from the European Union in 2019—will affect the flagship. It could remove a key funding source, although the United Kingdom could strike a deal like Switzerland, which pays to participate in EU research frameworks. But Brexit's effects on grantees will be delayed: The U.K. groups in the winning consortia will participate for the full 3-year initial period. "We don't know what form Brexit will take," Calarco says. "So we have 3 years to sort this out." ■

Edwin Cartlidge is a journalist in Rome.

CARDIOLOGY

Suspect science leads to pause in stem cell trial

Harvard University finds fraud in 31 papers, casting doubt on study rationale

By **Jocelyn Kaiser**

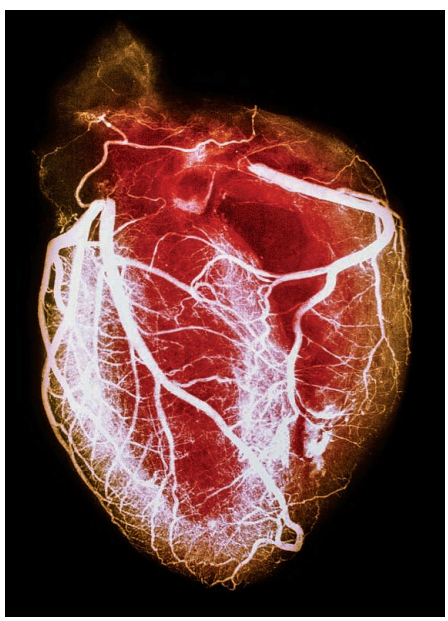
The saga of one of the biggest scientific fraud cases in the past decade has reached a turning point. In mid-October, Harvard University officials disclosed that they have called for the retraction of 31 papers from cardiac researcher Piero Anversa, who ran a Harvard Medical School lab studying the potential of stem cells to repair the heart. And this week, federal officials paused a related clinical study, explaining that the pending retractions "have raised concerns about the scientific foundations of this trial."

Anversa's research has long drawn controversy. Some critics of his work say studies like the now-halted trial should never have started, given long-standing questions about how—and whether—the cells Anversa studied might repair heart tissue. "The problem is that if you don't know how something works, then you don't really have a road map of what to address to make it better," says Deepak Srivastava, a pediatric cardiologist and president of the Gladstone Institutes in San Francisco, California.

But other researchers defend the push into the clinic. Many labs other than the Harvard group contributed to the basic research supporting the study, called CONCERT-HF, says David Goff, director of the Division of Cardiovascular Sciences at the National Heart, Lung, and Blood Institute (NHLBI) in Bethesda, Maryland. NHLBI, which funds the study at several U.S. sites, is pausing the trial to review it simply because "it's the prudent thing to do," he says.

Anversa, formerly of Harvard Medical School and Harvard-affiliated Brigham and Women's Hospital in Boston, rose to prominence by showing that, contrary to the received wisdom, the heart can rapidly regenerate muscle. He also reported that heart tissue contains stemlike cells known as c-kit+ cells that can rejuvenate heart muscle in mice. If true, the cells could form the basis of a treatment for heart failure, caused by weakened heart muscle.

But several labs could not replicate Anversa's work. One Anversa paper was retracted in 2014, and a 2011 paper in *The Lancet* drew an "expression of concern" from the journal. That paper reported on a small clinical trial that found modest benefits from c-kit+ cells as a heart disease treatment. Also in 2014, Harvard and Brigham and Women's revealed they had begun a scientific misconduct investigation of Anversa's work (*Science*, 18 July 2014, p. 254). Anversa has blamed problematic papers on another lab member. He lost a lawsuit claiming the investigation was mishandled and left Harvard and Brigham and Women's in 2015. Last year, Brigham and Women's agreed to pay $10 million to settle charges that Anversa and two colleagues fraudulently obtained federal funding.



The now-paused clinical trial explored using stem cells to repair damage to the heart.

IMAGE: SPL/SCIENCE SOURCE

The CONCERT-HF trial, led by cardiologist Roberto Bolli of the University of Louisville in Kentucky, began to enroll patients with heart failure 3 years ago. The study is testing four possible treatments: an injection in the heart of c-kit+ cells derived from a patient's own heart tissue, a combination of c-kit+ cells and mesenchymal stem cells from the patient's bone marrow, mesenchymal stem cells alone, or a placebo. The trial has recruited 125 of a planned 144 participants, Goff says. Blood stem cells and heart tissue have been collected from 117 of them, and 90 have received treatment. No safety issues related to the treatment have arisen, but one person died after a heart biopsy during early safety tests.

Goff says that when the CONCERT-HF trial was approved, reviewers were aware of concerns about Anversa's work. But eight other labs had published 11 studies supporting the c-kit+ cell treatment in animal models of heart disease. Scientists no longer think the cells engraft in the heart and differentiate into cardiac cells. Instead, many think the cells secrete some "paracrine factor" that promotes heart tissue growth.

Further concerns arose this month when Harvard and Brigham and Women's recommended that 31 papers from the Anversa lab be retracted because they "included falsified and/or fabricated data." (Harvard and Brigham and Women's also told Anversa in a 3 October letter that he "committed research misconduct" in eight published or submitted papers and a grant proposal, according to *The New York Times*.) The news prompted NHLBI to convene the trial's safety board, which recommended that the study pause for a review. NHLBI also expects to examine the list of 31 papers to ensure the trial is scientifically sound "out of an abundance of caution," Goff says.

He could not say how long the review will take, but promises it will be done "in an expeditious manner," particularly because patients have donated tissue and await treatment. About half of patients with chronic heart failure die within 5 years of diagnosis, he notes.

Despite the cloud over Anversa's research, some researchers think the trial should be completed. "There's plenty of reason to believe that there's still promise," says cardiologist Christopher Granger of Duke University in Durham, North Carolina. Bolli, a longtime Anversa collaborator, points to "overwhelming evidence" from animal studies that c-kit+ cells somehow promote heart tissue growth. "The controversy [over Anversa's work] does not really change the validity of using c-kit+ cells," Bolli says.

Another trial, based in Florida, planned to begin to treat infants born with a certain heart defect with c-kit+ cells this month; the status of the trial was unclear at press time.

Meanwhile, NHLBI hopes to clarify just what c-kit+ cells do and how much promise they hold. It is funding 14 preclinical grants to study the cells, Goff says. ■

With reporting by Jennifer Couzin-Frankel.

FEATURES

MOMENTS TO SPARE

California's new earthquake warning system delivers just seconds of notice. Even that is a victory

By **Paul Voosen**

The shaking woke Thomas Heaton on a quiet winter morning in Pasadena, California. The streets were empty, with sunrise hours away. As Heaton lay in bed next to his wife, waves vibrated through their house. Ten seconds. Fifteen. Twenty. As a seismologist, Heaton had spent his career studying seismic waves like these. By feel and duration, he guessed this quake was big, maybe a magnitude 6.5, and close, under west Los Angeles. Plenty dangerous. An aftershock rolled through. He was needed. "I've got to get to work," he told his wife.

At the time, in 1994, Heaton was the lead scientist at the earthquake field office of the U.S. Geological Survey (USGS) in Pasadena. He drove to the office in darkness, imagining the fires, collapsed bridges, and crumbled buildings closer to the epicenter. At the office, seismic readings partially validated his gut: "I was right about the magnitude and approximate distance," he says—though not the location. The quake had struck farther north, under the neighborhood of Reseda, on a previously unknown fault. The Northridge quake, as it came to be known, killed 57 and caused many billions of dollars in damage. There had been no warning, no sirens sending people into the streets. Heaton recalled how he had guessed the size of the earthquake when the first, gentle waves reached his bedroom. There must be a way, he thought, to translate his gut check into a short but useful warning.

After decades of work, Heaton's dreams have taken form. Last month, USGS unveiled ShakeAlert, the West Coast's earthquake early warning system. If all goes as planned, a dense network of seismometers in California, Oregon, and Washington will detect the first, weak waves of an earthquake and relay a rapid warning of ground shaking to come. To start, those warnings will go to first responders, power companies, and transit agencies. But in the next couple of years, alerts could roll out to the public to provide at least a few seconds of warning. Not much time, but enough to "drop, cover, and hold on," says Doug Given, a geophysicist in Pasadena who is leading the USGS effort.

For years, ShakeAlert was an academic side project of California seismologists, especially the gravelly voiced Heaton, now at the California Institute of Technology (Caltech) in Pasadena, and Richard Allen, his soft-spoken counterpart at the University of California (UC), Berkeley. They were inspired by warning systems in Mexico,

ILLUSTRATION: ROBERTO CIGNA

Published by AAAS

Japan, Taiwan, and Chile, among others, which emphasize detecting earthquakes at the source and warning distant cities before the seismic waves arrive. Many people thought such a system would be useless in fault-riddled California, where earthquakes seem to erupt underfoot anywhere. But Heaton and Allen persevered, deploying a pilot system in 2012.

Now, politicians are offering their support. Last year, some $13 million in annual funding flowed in from the federal government, along with $10 million more for sensor upgrades; California kicked in another $10 million. In his state-of-the-city speech last year, Los Angeles Mayor Eric Garcetti pledged: "By the end of 2018, we will deploy an earthquake early warning system to every corner of this city—in schools, at businesses, even on your smartphone."

This year's version doesn't quite measure up to that promise. Only half of the system's 1675 seismic stations have been installed. The technology to rapidly push alerts to mobile phones is not mature. And the public has yet to be trained in how to respond to such alerts, which are sure to include false alarms.

The system's scientific ambitions have also been humbled. The scientists developing ShakeAlert once promised it could warn of strong, violent shaking from a distant earthquake far in advance. That pitch stemmed especially from Allen, Heaton's friendly rival, who believed the final magnitude of an earthquake was determined by its first few seconds of rupture. If so, the system could catch an earthquake rupturing on a remote section of the San Andreas fault and give Los Angeles 1 minute or more of warning of severe shaking.

But over 15 years of development, reality has intruded: Faults fracture in complex, unpredictable ways. The current incarnation of ShakeAlert might offer 10 seconds of warning for a severe event—if you're lucky, Heaton says. "We're back to the simple ideas and just making the engineering part of this problem work," he says. "We're just trying to get it born."

THE SON OF A MATHEMATICIAN at Rutgers University in New Brunswick, New Jersey, Heaton, born in 1951, grew up on the stable ground of the East Coast. Dyslexia, which made molecular structures a jumble, pushed him out of chemistry into physics, but he didn't find his calling in the Cold War tasks of the time, either. "They had enough nuclear weapons to blow up the entire solar system and they didn't need any more," he says. Instead he was drawn to study Earth's own convulsions. As a graduate student at Caltech he experienced his first earthquake, an aftershock of the San Fernando quake in 1971. During that disaster, emergency workers took 3 hours to figure out where the heavy damage was. Seismologists were little help.

Heaton had three children, so he took a job with Exxon. He lasted less than a year, but while there he learned that Japan was already using early earthquake warnings to shut down bullet trains. "At that point, I got very excited," Heaton says in his Caltech office, where gas mains buckled by earthquakes serve as table stands.

He laid out his idea for a U.S. system in a 1985 paper in *Science*. Because seismic waves travel far more slowly than electrical signals, a "seismic computerized alert network" could detect an earthquake at its source and relay a warning of ground shaking to cities far from the epicenter. Automated systems could act immediately to prevent chemical spills, electrical fires, and other catastrophes. Such a system would do little to protect San Francisco from an earthquake like the one in 1906, which was centered near the city. But it could give minutes of warning for great quakes that start far from populated regions. It was a simple model with many assumptions—including, critically, an immediate detection of an earthquake's magnitude. "We can do it in 10 years," Heaton promised anyone who asked.

> ***"You're not going to get much time. If it's going to be dangerous, we won't know that till the last seconds."***
>
> **Thomas Heaton**, California Institute of Technology

It took longer. But as he climbed the ranks at USGS, Heaton updated Southern California's network of seismometers toward the always-connected compatibility needed for early warning. He also formed an alliance with Hiroo Kanamori, a decorated Caltech seismologist. Others in the field had spent years fruitlessly debating whether earthquakes can be predicted. Kanamori saw a better use for seismologists' talents: developing a warning system for earthquakes already underway. By the early 2000s, Allen, then an ambitious postdoc, had joined their effort.

Like Heaton, Allen was a transplant from stable terrain, namely, the United Kingdom. He, too, came to Caltech dissatisfied with sterile debates, in his case about Earth's internal structure. Early warning, it seemed, was the rare scientific discipline that could save lives. And at the time it appeared poised for a breakthrough.

At their most basic, earthquakes result when the strain built up between two locked chunks of Earth's crust becomes too much to bear and the slabs of rock slip past each other along a fault. The larger the slip area, the bigger the earthquake. As the rupture starts, it tosses off pressure (P) and shear (S) waves. P waves percuss the rock like a drumstick, traveling quickly through incompressible material. S waves, though more powerful, struggle through the rock because of their sashaying motion and lag well behind.

The classical view had been that nothing about the first waves from a rupture indicates how it will grow, reflecting an inherently chaotic, unpredictable system. But in the 1990s, lab-built models and some data on actual earthquakes suggested that a nucleation phase—a brief period of subtle slipping at the quake's start—could predict the size of the ensuing rupture. If that were true, forecasting the ultimate magnitude of an earthquake from only a few seconds of P waves might be possible. That ability could power a potent early warning system—a possibility that Yutaka Nakamura, an earthquake engineer at a private company in Japan, had already begun to pursue to improve bullet train warnings.

Allen and Kanamori built on Nakamura's work in a 2003 *Science* paper. In records from 53 California earthquakes, the largest a magnitude 7.3, they found a correlation between the time the initial P wave took to complete one cycle, called τ, and the resulting magnitude. That relationship became the core of an algorithm Allen developed called ElarmS. It led him to argue, in a 2005 *Nature* paper, that earthquakes are deterministic, their fate structured by their start, contrary to the conventional wisdom. "That paper," he notes, "was very controversial."

Heaton, though, doubted a chaotic system such as an earthquake would surrender its secrets to a simple equation. He recalled how his gut feeling and knowledge of past events had called out the Northridge quake. He started to develop code to re-create that intuition. As with ElarmS, the code relied on P waves from the first few seconds of a quake. But instead of using τ to leap to a final magnitude, the system compared the features of the initial waves with those of past quakes to create a digital gut check. Heaton called the project Virtual Seismologist.

Despite the ongoing debate, USGS began to finance Heaton, Allen, and other teams

to work on the algorithms that make up the core of ShakeAlert.

WHEN A MAGNITUDE-9.1 EARTHQUAKE struck 70 kilometers off Japan on 11 March 2011, the country's warning system was little help for people in the path of the torrential tsunami that swamped the coast; nearly 16,000 died. But the system did alert more than 50 million people and halt bullet trains and elevators in many regions before the shaking began. It also served as a wake-up call for U.S. researchers to push for their own system. "That was the tipping point," Allen says.

At a 2-day emergency summit at UC Berkeley a month later, the ShakeAlert team won a $6.5 million commitment from the Gordon and Betty Moore Foundation in Palo Alto, California, to build a prototype. USGS was sold, too, and agreed to run the system. The funders accepted that ShakeAlert need not be perfect; the Japanese public had appreciated the Tohoku warning despite its flaws. "They just had to make sure it worked reasonably well," Heaton says. And that meant solving the problems that Tohoku had exposed.

Despite the warning's success, it failed to alert Tokyo residents, far south of the quake, who were blindsided when the ground began to shake. The problem was that the system had located the earthquake to a single point. It then calculated how the shaking at that point, the hypocenter, would affect more distant locations. For earthquakes of magnitude 6.5 or smaller, which rupture for only a few seconds, that approach is reasonable. But the Tohoku fault rupture grew toward Tokyo, extending to some 400 kilometers over more than 3 minutes. "One thing we did not expect is that really long fault rupture," says Masumi Yamada, a seismologist at Kyoto University in Japan who studied with Heaton. As a result, the alert underestimated the quake's magnitude and extent.

The algorithms developed for ShakeAlert had the same shortcoming. "We realized really quickly that if there was a major earthquake along the southern San Andreas fault, we wouldn't expect shaking in Los Angeles because it was so far away," says Maren Böse, a seismologist at ETH Zurich in Switzerland who had also worked with Heaton. But while Yamada was at Caltech, Heaton worked with her to develop a way to track the growth of a rupture in real time, by measuring the shaking along its path. Böse, Heaton, and others then refined that technique. Virtual Seismologist yielded to an algorithm called the Finite-Fault Rupture Detector (FinDer), which updates its warnings as an earthquake progresses. FinDer, despite its late start, proved vital to showing that ShakeAlert could handle a Tohoku-size strike. "And really, if we can't do big earthquakes," Heaton says, "we're missing the point."

A KLAXON SOUNDED eight times, followed by an insistent robotic voice: "Earthquake. Earthquake." Heaton, seated at his desk, had a map of Southern California on screen. A simulated earthquake had just struck at the southern end of the state, by the Salton Sea, with an estimated magnitude of 7. The quake posed little threat to Pasadena, 250 kilometers to the north, and so ShakeAlert warned of only light shaking. But the rupture didn't stop there, and FinDer stayed on the case. A gray line began to extend toward Los Angeles, as did expanding rings of yellow and red: the warning P waves and damaging S waves. "Now it's getting closer," Heaton said. "And bigger as it goes."

> ***"After working on this for over a decade, here it was in action and I was on the receiving end."***
>
> **Richard Allen**, University of California, Berkeley

"Earthquake. Earthquake. Moderate shaking expected in 42 seconds," the voice warned. The estimated magnitude had gone up to 7.8—a 16-fold leap in energy. The rupture continued, and ShakeAlert upped its warning again: "Strong shaking expected in 23 seconds. Earthquake. Earthquake." Finally, 7 seconds before the damaging waves arrived, ShakeAlert gave its final warning: "Very strong shaking expected." The klaxon fired rapidly. And then silence. "You're not going to get much time," Heaton says. "If it's going to be dangerous, we won't know that till the last seconds."

Heaton says he still wishes that some signal buried in the first moments of an earthquake could reveal more. But even before Tohoku, the grand promise of predicting an earthquake's final magnitude from its first moments had begun to fall apart. In records for earthquakes with magnitudes above 7, "We started seeing a saturation effect," says Gilead Wurman, one of Allen's former students at UC Berkeley. "You'd start to underestimate the magnitude."

Heaton's recent work, conducted especially with Men-Andrin Meier, a seismology fellow at Caltech, has only solidified doubts. A 2016 comparison of P waves recorded within 25 kilometers of the hypocenter of earthquakes in the United States, Japan, and elsewhere showed that the small and large quakes looked identical at the start. The determinism of the nucleation phase, it seemed, was a ghost.

Meier and Heaton, along with Pablo Ampuero, another Caltech seismologist, have found that as an earthquake develops, it does drop a hint about its ultimate strength. In a database of 116 earthquakes greater than magnitude 7 created by a former postdoc, Lingling Ye, now at Sun Yat-sen University in Guangzhou, China, they found that once the rupture starts to slow, the median earthquake ends up no more than doubling in strength. "At some point you see it slowing down, and then you know after that it's all downhill," Ampuero says. Heaton believes that effect, which they call weak rupture predictability, is the only pattern they'll be able to tease out. But it emerges late and has little predictive value for individual earthquakes. There's no sign of a clear connecting thread from the start to the end of an earthquake.

Faced with the mounting evidence that determinism isn't holding up, Allen, too, is settling for weak predictability, something his own recent work has supported. That happens in science: Careers are made by staking out either side of a data-poor claim, and then a middle ground emerges. Yet even though most seismologists now agree that the start of an earthquake does not determine its end, many still think its early stages might somehow influence whether the rupture can grow by jumping faults or sections of locked rock. The earthquake's start may not drive all action, but it may still be a prologue that—in some way still not evident in the data—informs the rest of its story.

IN SEPTEMBER, while Allen was riding a Bay Area Rapid Transit train near San Francisco, his rail car ground to a halt. The conductor's voice came on the intercom. A magnitude-3.3 quake had struck 40 kilometers north of Berkeley, and the train system, following protocol, had stopped for safety. "I can't believe it: We have seen Yosemite, San Francisco, and now we have been in an earthquake!" one family of tourists said. The unplanned stop delighted Allen, too. "After working on this for over a decade, here it was in action and I was on the receiving end."

Over the past year, Given has pushed the ShakeAlert team to meld its unruly competing algorithms into a cohesive whole. First, a fast-estimating code called EPIC—

consisting primarily of Allen's ElarmS—generates an initial magnitude, treating the quake as a point source. But if EPIC sees a quake lasting more than a few seconds—and therefore larger than magnitude 6.5—FinDer takes the lead, tracking the rupture from there and updating the magnitude. The refinements will continue. "It's been kind of a closed club through this year," Given says, but the agency is now soliciting other researchers to improve the code.

New technologies will sharpen the warnings, too. GPS sensors, though slower than seismometers, can capture even shaking strong enough to max out conventional instruments, enabling the system to cope better with the biggest earthquakes. And Heaton expects artificial intelligence, especially neural networks, will in the next few years be able to discern P waves, an earthquake's first whisper, from seismic noise earlier than the existing algorithms. At first, Heaton was skeptical of the technology. "But then it dawned on me that this other neural network was in many ways more capable than this neural network," he says, pointing at his head.

Any warning system is only as good as its messaging, and how ShakeAlert will best reach the general public remains uncertain. "The technology for doing rapid massive alerting doesn't exist in the United States," Given says. The cellular messaging system that handles child abduction or severe weather alerts wasn't designed to relay warnings in seconds—more like minutes. Los Angeles will begin to test an alternative, using notifications on a smartphone app, but the fear is that such a system could easily overload.

USGS has set one important parameter: Instead of waiting until a risk is severe, ShakeAlert will skew toward more alerts, sounding an alarm once a location is at risk of "light shaking." That will increase the warning time—but it also will mean that, if the rupture grows, the prediction could change to severe shaking only seconds before hitting. And the public might grow complacent about those alarms and fail to respond to the rare mild threat that, in a moment, turns severe.

The faults riddling Heaton's adopted state guarantee that soon, ShakeAlert will get its first high-profile test. "It's a little terrifying," he says. "The world will be watching. Here's your chance to sing in front of everybody. You just hope you don't—" And Heaton's gravelly voice broke into a croak that echoed down the hall. ■

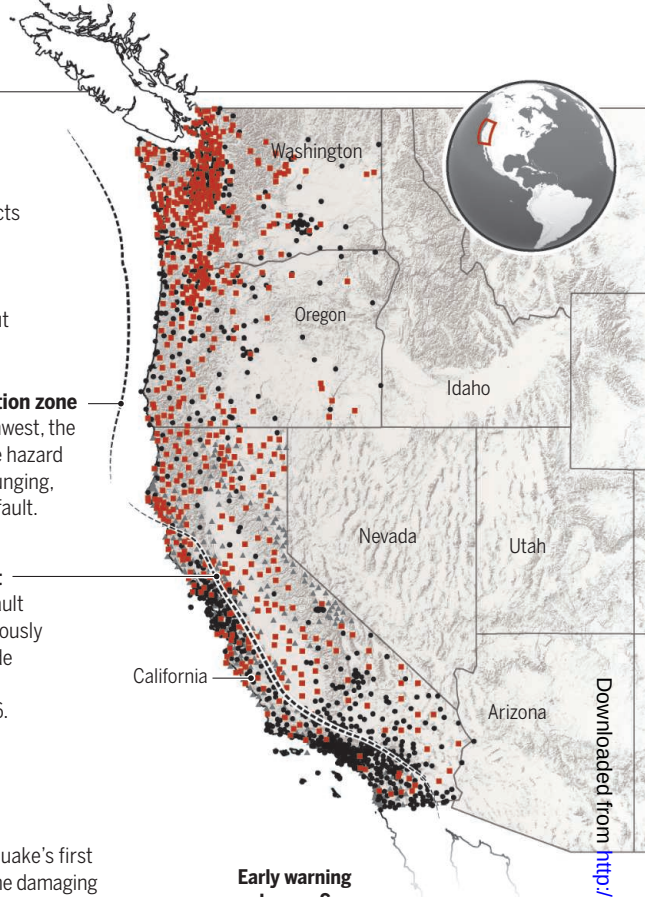
On the alert

ShakeAlert, the first U.S. earthquake early warning system, went live in October. It detects incipient earthquakes to provide seconds of warning to critical infrastructure and, eventually, the public. For now, the sensor network is densest in Southern California, but it will expand into the Pacific Northwest.

- Installed seismic stations
- Stations needing upgrade
- Planned seismic stations

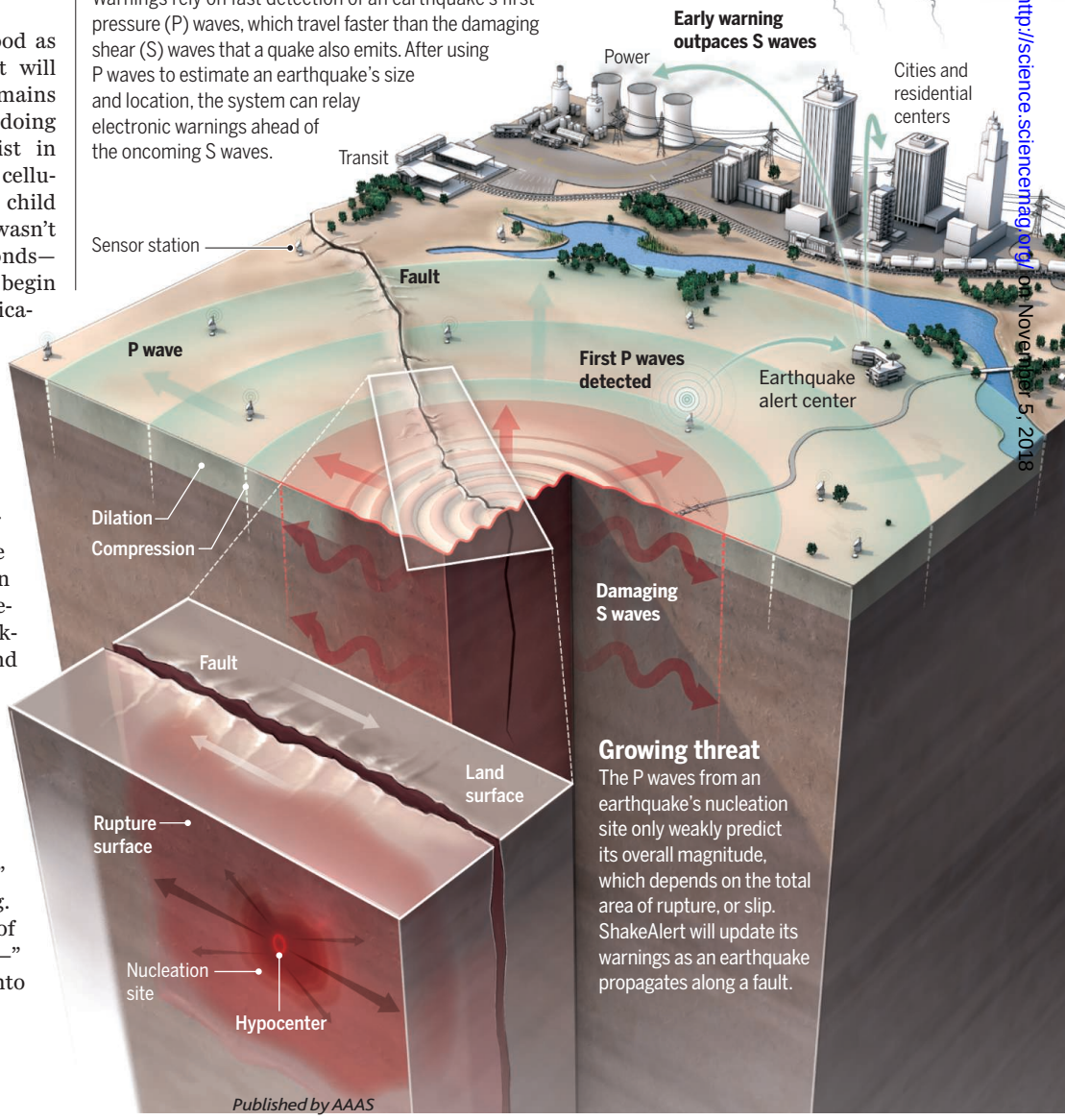
Cascadia subduction zone
In the Pacific Northwest, the biggest earthquake hazard comes from the plunging, offshore Cascadia fault.

San Andreas fault
The San Andreas fault ruptured most famously in the 7.8-magnitude San Francisco earthquake of 1906.



A race against time

Warnings rely on fast detection of an earthquake's first pressure (P) waves, which travel faster than the damaging shear (S) waves that a quake also emits. After using P waves to estimate an earthquake's size and location, the system can relay electronic warnings ahead of the oncoming S waves.



Growing threat

The P waves from an earthquake's nucleation site only weakly predict its overall magnitude, which depends on the total area of rupture, or slip. ShakeAlert will update its warnings as an earthquake propagates along a fault.

GRAPHIC: C. BICKEL/*SCIENCE*

Published by AAAS

CLIMATE

Shifting summer rains

Trace-element records in Chinese caves reveal the effects of climate change on Asian monsoons

By **David McGee**

Most of China's water supply depends on rainfall from the East Asian summer monsoon (EASM), a seasonal progression of rains that begins along the southern coast in spring, then sweeps north, reaching northeastern China in midsummer (see the photo). Projections of the EASM's response to future climate change are complicated by its complex interaction with the mid-latitude jet stream, which appears to govern the monsoon's northward march each spring and summer (*1*). To investigate the monsoon's sensitivity and dynamics, many scientists have turned to examining its past changes recorded in natural archives. Although past climates are not a direct analog of the 21st-century climate, they offer vital tests of the ability to describe monsoon behavior through theories and numerical models. On page 580 of this issue, Zhang *et al.* (*2*) provide new insight into the impact of past EASM and jet stream variations on rainfall patterns over China. Through examination of trace elements in Chinese stalagmites (a proxy for local precipitation amount) and climate modeling experiments, they show that cooling episodes in the North Atlantic shifted the summer jet stream south, delaying the onset of monsoon rains in northeastern China and increasing rainfall in central China. The finding demonstrates that local rainfall in the EASM regions can vary in opposition to monsoon strength, and it highlights the importance of future high-latitude warming in determining precipitation patterns in China.

Central to investigations of the EASM's history are records of the oxygen isotope

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. Email: davidmcg@mit.edu

PHOTO: HUMPHERY/SHUTTERSTOCK

In south-central China, monsoon flooding typically occurs in early summer before rains progress northward, as in this June 2016 flood in Jiujiang.

composition of rainfall recorded in stalagmites from Chinese caves (*3*). Because of their precise chronological control, continuous coverage of the last 640,000 years, and high temporal resolution, these isotope records excel on the "x axis" (time) in ways that are essentially unparalleled in paleoclimate research. A key limitation of these records, however, is in the "y axis" (oxygen isotope ratio) because the interpretation of oxygen isotope variations remains uncertain. It was initially argued that these variations (expressed as the $^{18}O/^{16}O$ ratio, or $\delta^{18}O$) reflect changes in the contribution of ^{18}O-depleted summer rains to total annual rainfall outside of the caves. This implies that low $\delta^{18}O$ values mean more summer rainfall at the cave site (*4*). At the opposite extreme, others have argued that $\delta^{18}O$ values in EASM rainfall have no relationship with local rainfall amounts and are instead set by changes in the composition or fraction of water vapor transported into the EASM region from India, where there is preferential rainout (removal from the atmosphere by precipitation) of water vapor with heavy oxygen isotopes (*5*). In this interpretation, $\delta^{18}O$ records from Chinese stalagmites provide information about remote atmospheric circulation, but not rainfall over China itself.

Key to testing both interpretations are independent reconstructions of local rainfall changes in the EASM region during periods with high-amplitude oxygen isotope changes. Zhang *et al.* do precisely that and provide important support for a newly emerging interpretive framework for the canonical Chinese cave records. The study presents measurements of trace impurities in the calcium carbonate lattice—substitutions of magnesium and strontium for calcium—in two stalagmites from central China that span the transition from the peak of the last ice age to the beginning of the current interglacial period, between 21,000 and 10,000 years ago. These trace-element variations are interpreted to reflect changes in the rate with which infiltrating waters passed through the rock above the cave, which should track local precipitation. Oxygen isotopes in these samples record the same pronounced variations as other Chinese stalagmites during the last deglaciation: high $\delta^{18}O$ values during periods of abrupt cooling of the Northern Hemisphere, interpreted as representing a "weak monsoon" and/or low summer rainfall over the cave site, and low $\delta^{18}O$ values when the Northern Hemisphere warmed.

Surprisingly, the trace-element records from these stalagmites indicate rapid infiltration of waters during cold periods when oxygen isotopes point to a "weak monsoon." The data suggest locally wet conditions at these times, and dry conditions during low-$\delta^{18}O$ "strong monsoon" intervals. Combining this new record with other recent precipitation reconstructions [e.g., (*6*)], Zhang *et al.* propose that strong monsoon intervals are marked by a tripole pattern of precipitation anomalies consisting of drying in central China and wet conditions to the south and northeast. Opposite responses—drying in the northeast and south, and wet conditions in central China—occur during high-$\delta^{18}O$, weak monsoon intervals.

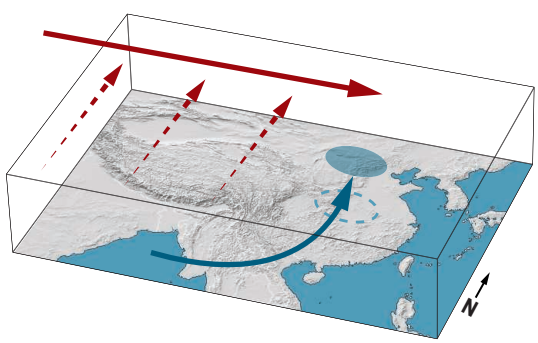
Zhang *et al.* investigate this tripole pattern with climate modeling experiments designed to mimic changes in Northern Hemisphere temperatures during the last deglaciation. The authors found precipitation changes consistent with the tripole pattern that appear to derive from the impact of high-latitude temperatures on the seasonal shifts of the jet stream (a narrow band of strong winds around 5 to 7 miles above Earth's surface that blow from west to east). In response to cooling of the Northern Hemisphere's high latitudes in the model, the steeper pole-to-equator temperature gradient slows the jet stream's seasonal migration north of the Tibetan plateau, delaying the summertime incursion of tropical moisture into northeastern China and increasing summer rainfall near the cave site in central China.

This finding of antiphased regional rainfall patterns in China, and of the key role of the jet stream's position in producing them, makes several important contributions to our understanding of the past and future of the EASM. That local precipitation varies in opposition to oxygen isotope values demonstrates that $\delta^{18}O$ values cannot be interpreted simply as a proxy for the abundance of summer rainfall at this and other sites in the EASM region. But it also appears incorrect to conclude that oxygen isotope records have no relationship to local rainfall in China and only reflect upstream processes: The data of Zhang *et al.* suggest that low-$\delta^{18}O$ intervals are marked by a consistent pattern of higher precipitation in both the south and northeast of China, and reduced precipitation in central China.

The results of Zhang *et al.* also demonstrate the importance of current and future warming of the Northern Hemisphere's high latitudes for regional rainfall patterns

East Asian summer monsoon rainfall

In response to warmer high-latitude temperatures, an earlier northward shift of the jet stream in summer (red arrows) draws more water vapor into northeastern China (blue arrow, blue ellipse) and shortens the rainy season in central China (blue dashed ellipse).



GRAPHIC: N. DESAI/*SCIENCE*

in China (see the figure). The tripole pattern evident in the paleoclimate data is very similar to the spatial pattern of interannual rainfall variability in China over recent decades (*7*), suggesting the robustness of this behavior across different climate states and forcings (physical factors that affect climate). In isolation, the polar-amplified Northern Hemisphere warming projected over the coming decades would suggest that droughts in central China may be a consistent feature of the 21st century.

Several mysteries remain for future investigation, however. The relationship between high-latitude temperatures, the jet stream's seasonal shifts, and the EASM proposed by Zhang *et al.* predicts that other processes that influence Earth's pole-to-equator temperature gradients, including ice ages and changes in the tilt of Earth's axis, should have pronounced effects on the EASM. Yet, oxygen isotope records from Chinese caves show almost no variability on the characteristic time scales of ice age cycles or changes in axial tilt. Longer records will be essential for testing whether the glacial-interglacial variability in EASM rainfall apparent in other paleoclimate records (*8*, *9*) is also present in infiltration proxies in stalagmites, but is somehow masked in oxygen isotope records. Indeed, the data of Zhang *et al.* intriguingly suggest an overall reduction in rainfall in central China over the deglaciation, a finding that requires further testing. Recently developed infiltration proxies involving calcium and strontium isotopes (*10*, *11*) offer valuable new tools for these efforts. In addition, modeling experiments that separate low- versus high-latitude influences on the monsoon, and experiments directly simulating the oxygen isotope composition of rainfall (*5*, *12*), will be necessary to fully unpack the processes driving monsoon variations and to decode the foundational records of past climate changes preserved in Chinese stalagmites. ■

REFERENCES AND NOTES

1. J. C. H. Chiang *et al.*, *Quat. Sci. Rev.* **108**, 111 (2015).
2. H. Zhang *et al.*, *Science* **362**, 580 (2018).
3. H. Cheng *et al.*, *Nature* **534**, 640 (2016).
4. Y. J. Wang *et al.*, *Science* **294**, 2345 (2001).
5. F. S. R. Pausata *et al.*, *Nat. Geosci.* **4**, 474 (2011).
6. Y. Goldsmith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1817 (2017).
7. J. C. H. Chiang *et al.*, *Geophys. Res. Lett.* **44**, 3788 (2017).
8. J. W. Beck *et al.*, *Science* **360**, 877 (2018).
9. S. C. Clemens *et al.*, *Nat. Commun.* **9**, 3364 (2018).
10. R. A. Owen *et al.*, *Earth Planet. Sci. Lett.* **443**, 129 (2016).
11. B. E. Wortham *et al.*, *Earth Planet. Sci. Lett.* **463**, 310 (2017).
12. Z. Liu *et al.*, *Quat. Sci. Rev.* **83**, 115 (2014).

ACKNOWLEDGMENTS

I thank G. Roe for helpful discussion and acknowledge funding related to this topic from NSF EAR-1434138.

10.1126/science.aav5280

NEUROPHYSIOLOGY

Identifying posture cells in the brain

The parietal cortex represents body posture and other factors in spatial awareness

By **Guifen Chen**

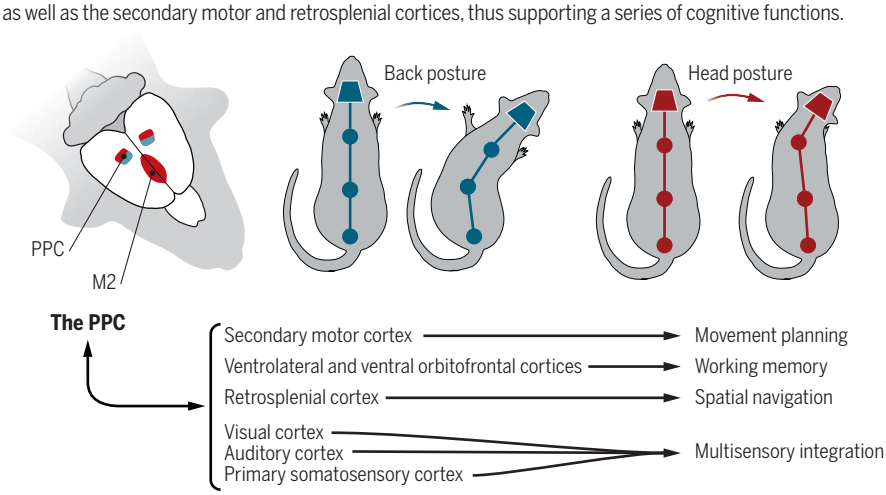
For an animal to successfully feed, mate, and avoid danger, its brain must integrate incoming information from many sensory modalities, combine the information with previously stored knowledge about the world, and then send appropriate output commands to the muscles. The information in this process is highly spatial in nature, but it is not anchored to any one coordinate reference frame. For example, sensory data from a fingertip tell the animal about a point in space, but exactly which point in space depends on the position of the finger relative to the wrist and arm it is attached to, as well as on the actual location of the animal in the world. Similarly, for the information on the retina, the point depends on depth of field, position of the eye within the socket, position of the head relative to the body, and location of the animal in the world. To integrate this highly spatial information, the brain needs to transform between coordinate systems. On page 584 of this issue, Mimica *et al.* (*1*) demonstrate how the posterior parietal cortex (PPC) in the brains of rats represents different aspects of the animal's posture, that is, the relative position of different parts of its body, implying that these posture cells could be one of the building blocks for coordinate transformation in the brain.

The parietal cortex has reciprocal connections to frontal motor areas as well as to almost all sensory areas in the mammalian brain (see the figure). It is thus expected to play a role in multiple cognitive functions such as multisensory integration, movement planning, working memory, and spatial navigation. Recordings from the parietal cortex of head-fixed monkeys revealed cells sensitive to light at specific positions on the retina, with responses further modulated by the position of the eyes, position of the head, and even by limb position (*2*–*5*). These experiments also revealed cells related to quick eye movement, grasping, and reaching. Such complex responses hinted that within the parietal cortex, some form of spatial transformation might occur between the various reference frames of the receptors (for example, eyes) and effectors (for example, hands) (*6*).

Building on this, Mimica *et al.* recorded single neurons from the PPC and frontal motor cortex (M2) of freely moving rats

Posture cells in the PPC-M2 network

Posture cells have been discovered in the PPC-M2 network. PPC neurons in rodents have reciprocal connections to various cortical areas, including visual, auditory, primary somatosensory, and prefrontal cortices, as well as the secondary motor and retrosplenial cortices, thus supporting a series of cognitive functions.



GRAPHIC: VERONICA FALCONIERI/*SCIENCE*

while tracking reflective markers attached to various parts of the rodent's body. They found that slightly more than half of the populations in PPC and M2 were tuned to postural features (including head, neck, and back positions), with some of the cells responding to the conjunctive postures of multiple body parts. The authors also demonstrated that they could reliably decode the activity of the neuron population in these regions so as to predict postures. A posture signal of this kind would be a key component of the transformation between reference frames because it would enable, for example, conversion from a head-centered coordinate frame to a body-centered coordinate frame.

Previously, PPC cells in rodents were shown to specifically encode simple locomotion behaviors, such as left and right turns and forward runs (*7*); and a more complex series of locomotion behaviors, for example, running straight followed by a right turn (*8*). Responses of these cells could be highly dependent on the given task, such as leftward movement in a maze versus in an open field (*9*), or they could be related to working memory, for example, remembering which of two paths to choose when approaching a T junction to reach a reward (*10*). However, it is possible that in all of these cases, the animal was simply exhibiting very different postures at different points in the activity being undertaken, for example, turning its head to the right prior to making a right turn, or turning in a very different manner during the maze task compared to the open-field task. Indeed, Mimica *et al.* disambiguated locomotion from posture and found posture to be far more correlated with PPC cellular activity, thus providing a more consistent and simpler explanation for the phenomenon of encoding complex locomotion behaviors.

However, even if posture is central to the PPC, there are many details left to uncover. Can it be driven by passive posture change, or is it dependent on actively updating posture only? And what happens when the animal is static: Does the PPC behave the same as when the animal is constantly shifting posture? Does the PPC represent current posture only, or does it also (or instead) handle the planning and execution of future postures? Mimica *et al.* present evidence that activity in PPC neurons precedes that in M2 neurons. These regions are then coupled to the superior colliculus, which contains cells tuned to, and likely driving, the rotational movement of the head in three-dimensional space (*11*). This suggests that PPC-M2 network activity could be the driving force for the downstream postural changes. If that is the case, how do the postural signals in the PPC and M2 become translated into rotation signals in the superior colliculus? Where does the transformation happen, and what is the nature of feedback between the different regions?

"The parietal cortex has reciprocal connections to... almost all sensory areas in the mammalian brain."

Further, can the posture cells support other complex cognitive functions in the PPC, such as spatial memory and navigation? To perform such functions, there must be a transformation from body-centric (egocentric) to world-centric (allocentric) coordinates, such that information is expressed in a manner invariant to the animal's location, and anchored to the external environment. Allocentric signals have been reliably observed in the hippocampus, which is connected to the PPC via the retrosplenial cortex. And indeed, cells in the PPC have been found to encode not just egocentric body movements (left or right turns), but also a direction signal expressed in both egocentric and allocentric reference frames (*12*). Thus, the transformation between egocentric and allocentric reference frames presumably happens in either the PPC or retrosplenial cortex (*13, 14*). But how? And what role do PPC posture cells play in the process? Future experiments will be needed to reconcile the newly found posture cells in the PPC-M2 network by Mimica *et al.* with previous work on the PPC in rodents and primates, so as to establish how these cells support the complex cognitive functions of the PPC. ■

Neuroscience, Physiology and Pharmacology, University College London, London, UK. Email: guifen.chen@ucl.ac.uk

PHOTOS: T. I. KAM *ET AL.* (2)

REFERENCES AND NOTES

1. B. Mimica *et al.*, *Science* **362**, 584 (2018).
2. R. A. Andersen *et al.*, *Science* **230**, 456 (1985).
3. R. A. Andersen *et al.*, *Annu. Rev. Neurosci.* **20**, 303 (1997).
4. R. A. Andersen, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **352**, 1421 (1997).
5. P. R. Brotchie *et al.*, *Nature* **375**, 232 (1995).
6. Y. E. Cohen, R. A. Andersen, *Nat. Rev. Neurosci.* **3**, 553 (2002).
7. B. L. McNaughton *et al.*, *Cereb. Cortex* **4**, 27 (1994).
8. D. A. Nitz, *Neuron* **49**, 747 (2006).
9. J. R. Whitlock *et al.*, *Neuron* **73**, 789 (2012).
10. C. D. Harvey *et al.*, *Nature* **484**, 62 (2012).
11. J. J. Wilson *et al.*, *Curr. Biol.* **28**, 1744 (2018).
12. A. A. Wilber *et al.*, *J. Neurosci.* **34**, 5431 (2014).
13. A. Bicanski, N. Burgess, *eLife* **7**, e33752 (2018).
14. L. H. Snyder *et al.*, *Nature* **394**, 887 (1998).

ACKNOWLEDGMENTS

G.C. thanks D. Manson for helpful discussion and comments.

10.1126/science.aav3819

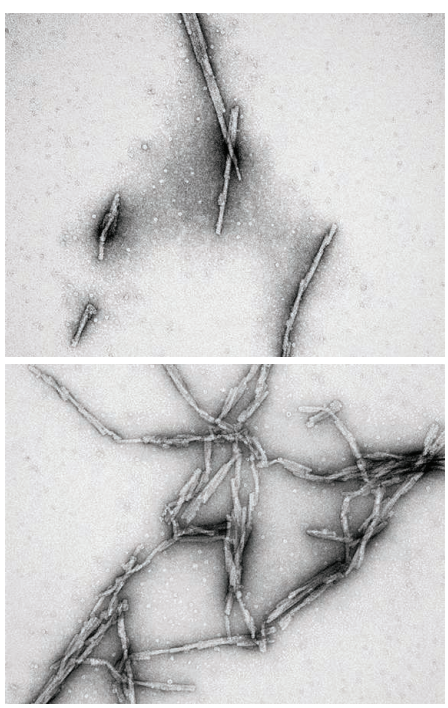
NEURODEGENERATION

Cancer enzyme affects Parkinson's disease

New insights identify a possible target for slowing neurodegeneration

By **Patrik Brundin**[1] *and* **Richard Wyse**[2]

Parkinson's disease (PD) is a progressive neurodegenerative disease characterized by rigidity, tremor, and slowness of movement, as well as a wide range of debilitating nonmotor symptoms (*1*). The neuropathology is characterized by loss of midbrain dopamine–producing neurons and the occurrence of intraneuronal protein aggregates, largely comprising the protein α-synuclein, that are distributed widely throughout the nervous system. Current treatments are focused on alleviating the impaired movement, and there is no therapy that slows PD progression. With almost 7 million patients worldwide, finding ways to prevent



The addition of PAR (bottom) accelerates pathogenic α-synuclein fibrillization.

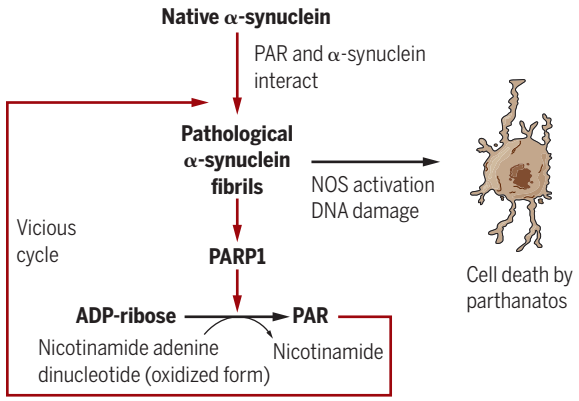
long-term neurodegeneration in PD and slow disease progression are vital goals. On page 557 of this issue, Kam *et al.* (*2*) describe a new molecular pathway that could explain why α-synuclein aggregates form in PD. Moreover, their findings suggest a pharmacological approach to slowing PD progression.

Kam *et al.* show that fibrillar aggregates of α-synuclein created in a test tube can activate the DNA damage repair protein, poly(ADP-ribose) (PAR) polymerase-1 (PARP1), which in turn triggers generation of PAR. Even in a test tube, PAR directly binds to and accelerates the fibrillization of recombinant α-synuclein into highly pathogenic fibrils (see the image), as part of a vicious cycle (described by the authors as "forward-feeding") (see the figure). When such α-synuclein fibrils are added to cell cultures or injected into the brains of mice, they are particularly potent at causing intracellular accumulation of α-synuclein and neuronal death, via activation of nitric oxide synthase (NOS) and induction of DNA damage, which activates PARP1. Kam *et al.* show that deletion of the *PARP1* gene or pharmacological inhibition of PARP1 activity, both in cell cultures and mice, prevents α-synuclein aggregates from propagating between neurons and causing neuronal death by parthanatos, which is PARP-dependent cell death, a form of programmed cell death that differs from both apoptosis and necrosis (*3*). The authors also present evidence for increased PAR levels in the cerebrospinal fluid of PD patients as well as in the region containing degenerating dopamine neurons.

PARP inhibitors have been developed as chemotherapeutic agents and are currently used to treat several forms of cancer, including BRCA-mutated breast cancer, as well as chronic lymphocytic leukemia. The rationale for using a PARP inhibitor is that some cancer cells are particularly vulnerable to loss of PARP1 activity because they rely heavily on the enzyme for DNA repair and thus to maintain the ability to divide. As a direct result of the current findings by Kam *et al.*, and supported by earlier studies in neurotoxin-induced and genetic animal models of PD (*4*–*6*), it should be possible to consider PARP inhibitors for repurposing as potential neuroprotective therapeutics in PD patients. Parthanatos potentially also contributes to other neurological diseases—for example, Alzheimer's disease, stroke, traumatic brain injury, and multiple sclerosis—and so PARP inhibitors have been suggested as potential therapeutics for these diseases (*7*, *8*).

Vicious cycle of PARP activity in PD

PARP1 produces PAR, which accelerates the fibrillization of α-synuclein and triggers parthanatos, a PARP-dependent form of cell death. This may contribute to the progressive neurodegeneration observed in PD.



Several drugs are currently being repositioned as possible disease-modifying agents in PD (*9*). A coordinated drug-repurposing initiative, the international Linked Clinical Trials initiative, has over the past 6 years initiated many clinical trials aimed at disease modification (*10*). Since 2012, PARP inhibitors (such as olaparib, ABT-888, BMN-673, rucaparib, and niraparib) have been considered by this initiative (*11*). Although there has been great interest in PARP1 as a therapeutic target, concerns about possible safety and tolerability issues, especially in conjunction with chronic treatment that would be necessary to treat PD, have previously weighed against moving into clinical trials. Notably, the study of Kam *et al.* suggests that relatively low doses of PARP inhibitors might be sufficient to attenuate elevated PAR levels in the brain to achieve clinical benefit. Furthermore, target engagement could be assessed by measuring PAR levels in cerebrospinal fluid, which could help guide appropriate dosing.

The first trial to be completed under the Linked Clinical Trials initiative was published in 2017 and suggested that exenatide, a glucagon-like peptide 1 receptor (GLP1R) agonist used to treat type 2 diabetes, slowed PD progression (*12*). Notably, several of the authors who contributed to the study by Kam *et al.* recently reported that GLP1R agonism inhibits α-synuclein toxicity. They demonstrated that a GLP1R agonist reduces the activation of brain-resident macrophages, called microglia, when they are exposed to α-synuclein aggregates. Consequently, microglia exposed to the GLP1R agonist are no longer able to trigger the transformation of neighboring supportive cells, a form of glial cell called astrocytes, into a neurotoxic form (*13*). Together, one can potentially consider simultaneously targeting pathological α-synuclein toxicity by means of two dissimilar approaches: PARP inhibition to reduce the assembly of α-synuclein into toxic aggregates, and concomitant activation of GLP1R to prevent microglia-mediated activation of astrocytes.

Therefore, in the future, combination therapies could be used in clinical neuroprotection trials for enhanced therapeutic gain beyond using a single neuroprotective agent. Clearly, further research is needed to define the interrelationship between PARP inhibitors and GLP1R agonists in their capacities to dampen α-synuclein responses and to unravel whether their effects are additive. In addition to these two approaches, several other strategies to target α-synuclein pathology—for example, by using immunotherapy against α-synuclein aggregates and small molecules to prevent α-synuclein aggregation—are either already in trials or about to enter the clinical arena (*14*). Identification of clinically effective neuroprotectants for PD has been a struggle for many years, but given the discovery of several potentially relevant molecular disease mechanisms, hopeful times lie ahead. ■

[1]Van Andel Research Institute, Center for Neurodegenerative Science, 333 Bostwick Avenue NE, Grand Rapids, MI 49503, USA. [2]The Cure Parkinson's Trust, 120 Baker Street, London W1U 6TU, UK. Email: patrik.brundin@vai.org

REFERENCES AND NOTES

1. W. Poewe *et al.*, *Nat. Rev. Dis. Primers*. **3**, 17013 (2017).
2. T.-I. Kam *et al.*, *Science* **362**, eaat8407 (2018).
3. S. A. Andrabi, T. M. Dawson, V. L. Dawson, *Ann. N. Y. Acad. Sci.* **1147**, 233 (2008).
4. Y. Lee *et al.*, *Nat. Neurosci.* **16**, 1392 (2013).
5. A. S. Mandir *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5774 (1999).
6. S. Lehmann, A. C. Costa, I. Celardo, S. H. Y. Loh, L. M. Martins, *Cell Death Dis.* **7**, e2166 (2016).
7. N. A. Berger *et al.*, *Br. J. Pharmacol.* **175**, 192 (2018).
8. S. Martire, L. Mosca, M. d'Erme, *Mech. Ageing Dev.* **146–148**, 53 (2015).
9. D. Athauda, T. Foltynie, *CNS Drugs* **32**, 747 (2018).
10. P. Brundin, R. K. Wyse, *Eur. J. Neurosci.* 10.1111/ejn.14175 (2018).
11. P. Brundin *et al.*, *J. Parkinsons Dis.* **3**, 231 (2013).
12. D. Athauda *et al.*, *Lancet*. **390**, 1664 (2017).
13. S. P. Yun *et al.*, *Nat. Med.* **24**, 931 (2018).
14. P. Brundin, K. D. Dave, J. H. Kordower, *Exp. Neurol.* **298**, 225 (2017).

ACKNOWLEDGMENTS

P.B. has received commercial support as a consultant for Renovo Neural, Roche, Teva, Lundbeck A/S, AbbVie, Neuroderm, Fujifilm Cellular Dynamics, Living Cell Technologies, ClearView Healthcare, FCB Health, IOS Press Partners, and Capital Technologies. P.B. has received commercial support from Renovo, Roche, Teva, and Lundbeck A/S. P.B. has ownership interests in Acousort AB, and P.B. and R.W. are on the steering committee of the NILO-PD trial.

10.1126/science.aav3986

GRAPHIC: N. DESAI/*SCIENCE*

MICROBIOLOGY

Evolving a pathogen to be protective

The deadly fungus *Candida albicans* can be pushed to protect its mammalian host

By **Christophe d'Enfert**

The yeast *Candida albicans* is a component of the human oro-gastrointestinal and genital microbiota where it can exist in a commensal state without causing pathological infection. Yet, it is one of the main fungal pathogens of humans, responsible for superficial infections as well as disseminated, often deadly, infections (*1*). As the focus of recent research has been on understanding what makes *C. albicans* such a harmful pathogen, we know little about the commensal state and whether the interaction of *C. albicans* with the healthy human host has some reciprocally beneficial (mutualistic) components. Exploring this aspect has been hampered, in part, by the observation that most laboratory animals (including mice) are not natural hosts of *C. albicans*, and persistence of *C. albicans* in the gastrointestinal (GI) tract of these animals requires antibiotic-induced dysregulation of the microbiota (dysbiosis). By contrast, when allowed to reach the bloodstream, *C. albicans* behaves as a highly virulent pathogen of mice. On page 589 of this issue, Tso *et al.* (*2*) report how they have taken advantage of this dichotomy in *C. albicans*–mouse interactions to evolve this fungal pathogen into a genuine commensal that protects the host from other pathogens and might pave the way to a universal vaccine.

Through serial passage of a *C. albicans* isolate in the GI tract of antibiotic-treated mice, Tso *et al.* obtained variants with increased fitness upon commensalism but decreased fitness upon blood-borne infection. Importantly, colonization of the GI tract or injection in the bloodstream of several of these variants protected mice against virulent *C. albicans* but also protected them from infection with the filamentous fungus *Aspergillus fumigatus* and the bacterial pathogens *Pseudomonas aeruginosa* and *Staphylococcus aureus* (see the figure).

One of the notable observations of Tso *et al.* is that GI-tract evolved variants display defects in one of the key attributes of *C. albicans* virulence: the yeast-to-hypha transition. The reversible switch between the yeast and hyphal (filamentous) forms is critical for disseminated infections (*3*). Yeasts contribute to *C. albicans* dissemination in the bloodstream and adhesion to tissues, whereas the filamentous form contributes to tissue invasion and escape from phagocytic cells in the host immune system. It was shown previously that strains with defects in the yeast-to-hypha transition have an increased ability to colonize the GI tract of antibiotic-treated mice (*4*). However, the observation that serial passage of *C. albicans* in the GI tract of such mice leads to the emergence of hypha-defective variants with increased fitness for this compartment was unknown and is somewhat puzzling.

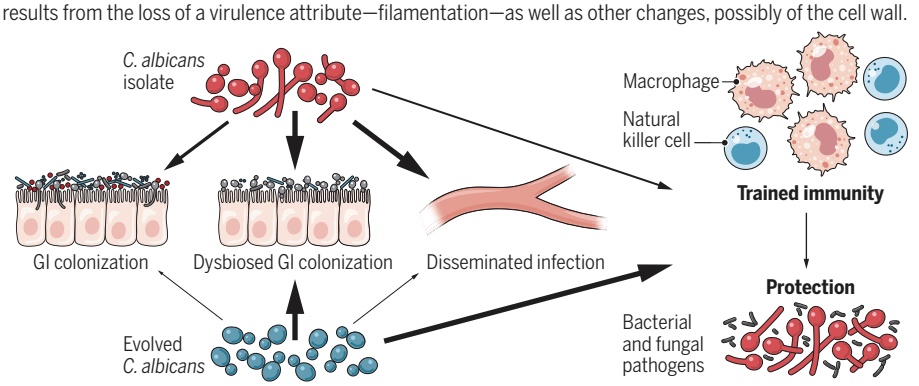
C. albicans is primarily a commensal, and its involvement in disease is occasional, with superficial infections resulting from local dysbiosis and disseminated infections usually associated with broad-spectrum antibiotic treatment, alteration of epithelial barriers, and induced or pathological immunosuppression. Notably, the ability of *C. albicans* to form hyphae is considered a hallmark of this species and is frequently used in clinics to distinguish *C. albicans* from other yeast species. Indeed, most *C. albicans* isolates form hyphae, albeit with different efficiency. Hence, one would expect that, if hyphal formation is disadvantageous for commensalism, this would have been lost during evolution. Similarly, most of the variants obtained by Tso *et al.* carry loss-of-function mutations in *FLO8*, a *C. albicans* gene that encodes a transcription factor key to the progression of the yeast-to-hypha switch. None of these mutations were observed in a collection of 182 genome-sequenced *C. albicans* isolates, including commensal isolates (*5*). Other loss-of-function *FLO8* mutations occurred rarely and were never found at homozygosity, suggesting that they are not selected.

A possible explanation for this paradox lies in the additional observation by Tso *et al.* that filamentation-defective variants are less fit than the wild type upon GI colonization of mice that have not received antibiotics and thus have an intact microbiota, such as in healthy human hosts. Perhaps the filamentation program is required for establishing a commensal equilibrium with other components of the microbiota, but this is disadvantageous when the microbiota is reduced, leading to the emergence of filamentation-defective variants. One important conclusion is that *C. albicans* has exploited an attribute of commensalism for pathogenesis.

What could be the hypha-specific factors that are important for *C. albicans* survival in the microbiota? Hypha-associated proteins that are expressed predominantly during hyphal growth have various functions, including cell wall proteins that are involved in the interaction with host cells and proteins, the cytolytic peptide toxin candidalysin, and secreted aspartyl proteases (*3*, *6*). Moreover, hypha adopt a metabolism that differs from those of yeasts (*7*). Future research should elucidate whether hyphal formation is indeed nec-

Evolving a fungal pathogen into a beneficial commensal

Evolved *C. albicans* differ from their parental *C. albicans* isolates by their ability to colonize the GI tract in the presence or absence of an intact microbiota, cause systemic infection, and induce trained immunity toward other bacterial and fungal pathogens (shown by arrow thickness). This shift toward a mutualistic interaction results from the loss of a virulence attribute—filamentation—as well as other changes, possibly of the cell wall.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, 25 rue du Docteur Roux, Paris, France. Email: christophe.denfert@pasteur.fr

essary for *C. albicans* commensalism in a host with an intact microbiota and, in this event, which of the above factors contribute to commensalism.

Another key observation of Tso *et al.* is that priming of mice by the *C. albicans* variants led to cross-protection against different fungal and bacterial pathogens. This is rapidly established, independent of adaptative immunity, relatively short-lived, and requires cytokine production. These are hallmarks of "trained immunity," a recently described property that allows enhanced innate immune cell responses when pathogens are reencountered (*8*). Colonization of the GI tract by *C. albicans* is known to protect against subsequent injury and infection (*9*), and a role of the gut fungal microbiome (mycobiome) in shaping immunity has started to be unraveled (*10*, *11*). Yet, the variants identified by Tso *et al.* showed increased cross-protection compared to the parental *C. albicans* isolate as well as *C. albicans* mutants defective in the yeast-to-hypha switch. This suggests that passage through the gut has led to changes, besides loss of filamentation, that contribute to triggering trained immunity. These changes may affect the fungal cell wall. Indeed, β-glucans, the main component of the cell wall skeleton, are strong inducers of trained immunity (*12*), and intestinal stimulation with mannans, the glycans that decorate cell wall proteins, protects against disease susceptibility, yet in a manner that may be distinct from trained immunity (*9*).

Understanding the exact nature of these changes represents a prerequisite to the development of the universal vaccine that Tso *et al.* propose based on the cross-protection properties of the *C. albicans* variants. Even though the prospects for a universal vaccine might seem hypothetical, one could anticipate that further exploration of the mechanisms by which the variants protect against different fungal infections may lead to new preventive approaches against these infections. Such approaches are greatly needed because fungal infections are responsible for up to a million deaths per year worldwide, and resistance to antifungals is rising (*1*, *13*). ■

REFERENCES

1. G. D. Brown *et al.*, *Sci. Transl. Med.* **4**, 165rv13 (2012).
2. G. H. W. Tso *et al.*, *Science* **362**, 589 (2018).
3. P. Sudbery, *Nat. Rev. Microbiol.* **9**, 737 (2011).
4. J. V. Pierce *et al.*, *Eukaryot. Cell* **12**, 37 (2013).
5. J. Ropars *et al.*, *Nat. Commun.* **9**, 2253 (2018).
6. D. L. Moyes *et al.*, *Nature* **532**, 64 (2016).
7. I. V. Ene *et al.*, *Cold Spring Harb. Perspect. Med.* **4**, a019695 (2014).
8. M. G. Netea *et al.*, *Science* **352**, aaf1098 (2016).
9. T. T. Jiang *et al.*, *Cell Host Microbe* **22**, 809 (2017).
10. M. L. Wheeler *et al.*, *Cell Host Microbe* **19**, 865 (2016).
11. I. D. Iliev *et al.*, *Science* **336**, 1314 (2012).
12. J. Quintin *et al.*, *Cell Host Microbe* **12**, 223 (2012).
13. M. C. Fisher *et al.*, *Science* **360**, 739 (2018).

10.1126/science.aav3374

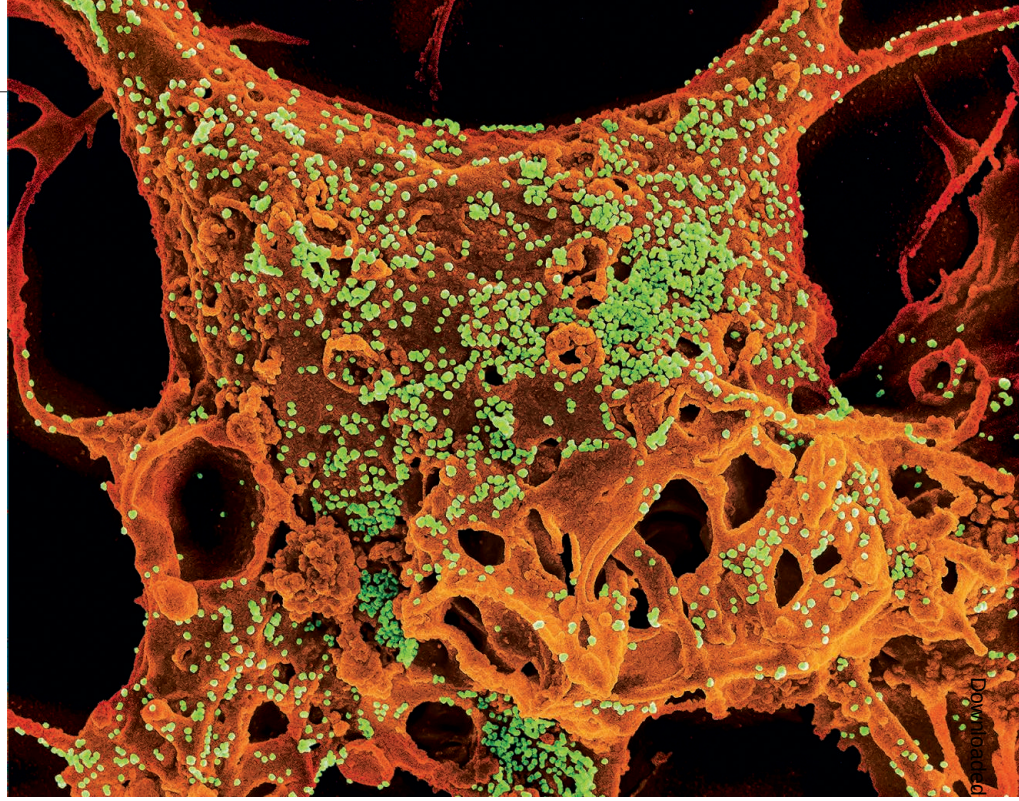


PHOTO: NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES, NATIONAL INSTITUTES OF HEALTH/SCIENCE SOURCE

INFECTIOUS DISEASE

Sources of human viruses

Machine learning predicts the hosts and vectors of RNA viruses that can infect humans

By **Mark Woolhouse**

Most emerging infectious diseases are caused by RNA viruses (*1*). Many of these that are newly found in humans have a natural mammal or bird reservoir; some are transmitted by arthropod vectors, such as mosquitos (*2*). If we do not know the reservoir host and/or vector, it is harder to identify individuals and populations at greatest risk of infection and to design an effective public health response. On page 577 of this issue, Babayan *et al.* (*3*) report their efforts to predict the reservoir hosts and vectors of human RNA viruses by applying machine-learning algorithms to virus genome sequence data.

More than 200 species of RNA virus are known to be capable of infecting humans, and two or three new species are discovered every year (*2*). Some of these, such as severe acute respiratory syndrome (SARS) coronavirus (*4*), posed an immediate epidemic threat; others, such as mammalian 2 bornavirus (*5*), remain restricted to small numbers of cases. Although we may know little of the basic biology of a new virus, it is now relatively straightforward to obtain its genome sequence (*6*). The study of Babayan *et al.* suggests that sequence data can give us a rapid prediction of reservoir host type, whether the virus is vector-borne, and, if so, the vector type.

Babayan *et al.* assembled a dataset of more than 500 genomes of single-stranded RNA viruses linked to information on known reservoir and vector species. Their first analysis was based on the "phylogenetic neighborhood," a cluster of closely related viruses. This approach predicted vector-borne transmission with 95% accuracy but correctly predicted the type of vector (mosquito, midge, sandfly, or tick) and the reservoir host type (as one of 11 broad taxonomic groups within mammals and birds) only 67% and 58% of the time, respectively. These results reflect the ways in which these traits are distributed among virus taxa. RNA viruses switch between hosts quite easily. By contrast, vector-borne transmission is a much more conserved trait; few genera have both vector-borne and non–vector-borne species, although there may be within-genus differences in the type of vector used.

Babayan *et al.* then used machine learning to improve their predictions. More than

Usher Institute at Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3FL, UK. Email: mark.woolhouse@ed.ac.uk

MERS (Middle East respiratory syndrome) coronavirus first emerged in 2012 and causes severe acute respiratory illness.

4000 genomic traits (such as amino acid bias or codon pair bias) were deployed in training, optimization, and validation algorithms. This markedly improved the predictions of vector type (to 91% accuracy) and host type (to 72% accuracy). Although the genomic traits themselves show strong phylogenetic signals, this result suggests that they also contain information not (or not easily) captured by analyses based on relatedness alone.

The algorithms can be used to predict unknown natural reservoirs or vectors. For example, they predict that Bas-Congo tibrovirus is transmitted by midges from an artiodactyl (even-toed ungulates) reservoir and that two species of ebolavirus—Bundibugyo and Taï Forest—have primate reservoirs. These assignations should be regarded as hypotheses to be tested empirically. Some predictions may be helpful in confirming or challenging current understanding. For example, they predict that Crimean-Congo hemorrhagic fever virus is not vector-borne, although it is thought to be transmitted by ticks.

Babayan *et al*. propose that their methodology could also be used for rapid assessment of emerging viruses of which we have no prior knowledge [a scenario that the World Health Organization calls "Disease X" (*7*)]. In the early stages of a public health emergency, knowledge of reservoir host and/or vector species could be vital for effective outbreak management. In this situation, the stakes are higher and greater caution is appropriate. It is helpful that the in silico predictions are assigned a confidence level and that the next best possibilities are identified: 95% confidence in a particular vector type is more useful than 60% confidence in a choice of three different host types. However, even if the reservoir host and/or vector are predicted with high confidence, the taxonomic resolution is currently insufficient; we need to identify species rather than broadly defined "types" such as sandflies or rodents. Even then, empirical confirmation would still be necessary.

Could confidence and resolution of the predictions be improved in the future? Methodological advances may help; Babayan *et al*. found variation in the performance of different machine-learning algorithms. However, the key requirement is for more and better-quality data, and here there is much still to do. It is likely that we currently know of only a small fraction of RNA viruses of mammals and birds (*8*). Finding and sequencing this hidden diversity of viruses would be an enormous undertaking (*9*), but the advent of metagenomics (techniques for sequencing all the genetic material in a sample) makes it considerably more feasible (*10*).

A collection of virus sequences from animal reservoirs would be especially valuable if it could be used to predict whether a virus is capable of infecting humans, even before any human cases are detected (*11*). Babayan *et al*. did not look for predictors of human infectivity, and it is not clear whether their approach would predict preadaptation to a potential new host as well as it predicts adaptation to an established reservoir host.

More focused approaches may be needed. For example, a key determinant of host range is cell receptor usage; host switching has been linked to a virus gaining cell entry via a phylogenetically conserved host receptor (*12*). Cell receptor usage also determines tissue tropism, and this is associated in turn with both pathogenicity and transmissibility, two traits that are highly relevant to the level of threat to human health. Predicting cell receptor usage directly from virus genome sequences is beyond our current capabilities, but undeniably the genome does contain the information. As our knowledge of viral cell receptors improves and the number and diversity of virus genome sequences accumulates, machine-learning approaches applied to protein-protein interactions (*13*) may provide a way forward.

Currently, our ability to predict human infectivity among RNA viruses is limited to (mostly weak) ecological correlates (*14*), and the value of massive investment in virus discovery and sequencing programs is hotly debated (*9*, *15*). This debate is timely, however, given that databases are growing fast and analytical methods are evolving rapidly. The study of Babayan *et al*. is a valuable step forward and hopefully presages further advances in our ability to extract information of public health value directly from virus genome sequences. ■

REFERENCES

1. S. S. Morse *et al*., *Lancet* **380**, 1956 (2012).
2. M. E. J. Woolhouse, L. Brierley, *Sci. Data* **5**, 180017 (2018).
3. S. A. Babayan *et al*., *Science* **362**, 577 (2018).
4. J. S. M. Peiris *et al*., *Lancet* **361**, 1319 (2003).
5. B. Hoffman *et al*., *N. Engl. J. Med.* **373**, 154 (2015).
6. M. E. J. Woolhouse *et al*., *Sci. Transl. Med.* **7**, 307rv5 (2015).
7. World Health Organization, *A Research and Development Blueprint for Action to Prevent Epidemics* (2018); www.who.int/blueprint/en/.
8. J. L. Geoghegan, E. C. Holmes, *R. Soc. Open Biol.* **7**, 170189 (2017).
9. D. Carroll *et al*., *Science* **359**, 872 (2018).
10. P. Simmonds *et al*., *Nat. Rev. Microbiol.* **15**, 161 (2014).
11. G. T. Keusch *et al*., *Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases* (National Academies Press, 2009).
12. M. E. J. Woolhouse, J. L. Ashworth, *Biochemist* **39**, 8 (2017).
13. B. Kim *et al*., *J. Bioinf. Comp. Biol.* **15**, 1650024 (2017).
14. K. J. Olival *et al*., *Nature* **546**, 646 (2017).
15. E. C. Holmes *et al*., *Nature* **558**, 180 (2018).

10.1126/science.aav4265

PHYSICS

Disorder at the border

The photoinduced phase transition in vanadium oxide involves ultrafast loss of coherence

By **Andrea Cavalleri**[1,2]

How does an ensemble of closely packed atoms rearrange when a solid undergoes a morphological change? Do the atoms follow a common, synchronized path, or do they move independently into their new positions? These elusively simple questions underpin many key problems in modern condensed-matter physics and also affect fields as far removed as soft-matter research and biological sciences. On page 572 of this issue, Wall *et al*. (*1*) use the Linac Coherent Light Source x-ray free-electron laser (*2*) to provide decisive new information on the nature of a structural phase transition. They study the time evolution of the monoclinic phase of crystalline vanadium dioxide (VO_2) after it is destabilized with light (*3*–*5*). Previous studies have assumed that this photoinduced structural transition proceeds coherently, that is, as a concerted structural rearrangement in which all atoms move at once (*6*, *7*). Wall *et al*. turn this notion on its head and find a prompt increase in disorder immediately after photoexcitation, long before the material heats.

"...immediately after photoexcitation, different dimers become uncorrelated at short distances..."

To understand the nature of the problem, it is helpful to recall the idea of crystal symmetry, which physicists define as a collection of mathematical operations that leave a unit cell unchanged. For example, a cubic crystal is invariant against all rotations

[1]*Department of Physics, University of Oxford, Oxford, UK.* [2]*Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany. Email: andrea.cavalleri@mpsd.mpg.de*

at right angles, because a cube is indistinguishable if looked at from the top, from the bottom, or from any one of its sides. Similarly, the cube is also invariant against reflections about a number of planes, for example, those that bisect the edges. However, if a phase transition takes place that distorts the unit cell—for example, one of the edges becomes longer—the material loses a subset of these symmetry elements.

A phase transition is in fact about acquiring or losing elements of symmetry, but how does this proceed microscopically? Instead of the geometric parameters of a unit cell, let us consider, for graphic simplicity, a magnetic material in which spin moments point either up or down at neighboring sites. The classical theory of phase transitions is based on a so-called mean-field approach. It is assumed that fluctuations of local order occur uniformly over long length scales. It follows that the order must be finite everywhere below the transition temperature T_c and zero everywhere above it. In this view, x-ray diffraction from an antiferromagnet would show a diffraction peak for all temperatures below T_c and no peak for all temperatures above it (see the figure, top). This type of transition is termed "displacive."

However, let us relax the constraint on the correlation length of the order and consider instead a situation in which the moments at each site do not necessarily disappear when crossing T_c, but rather become uncorrelated near the transition. That is, the size and angle of a moment at one site do not influence the size and angles of spins only a short distance away. Also in this limit, which is that of an "order-disorder" transition, the diffraction peak disappears when the transition temperature is crossed (see the figure, bottom). Hence, one cannot tell apart a displacive and an order-disorder transition by looking at the Bragg scattering alone. However, in the displacive case, the diffraction peak disappears because the individual spins are not there to scatter the light, whereas in the order-disorder case, it disappears because of a lack of constructive interference at the Bragg angle. In the order-disorder case, the x-rays are still scattered by the nonzero local moments, but in a more incoherent manner to other diffraction angles.

In the VO_2 structural transition discussed by Wall *et al.*, the increase in symmetry between the low- and high-temperature phases involves the disappearance of V^{4+} dimers, which in the low-temperature phase are arranged in chains with alternate elongated and contracted bonds. Such an antiferrodistortive transition can be thought of in the same terms as the magnetic case of the figure if the spin direction and size are substituted with the contracted or elongated bonds of different amplitudes. The authors differentiated the two possibilities of a displacive and of an order-disorder transition by measuring not only the dynamical changes in the Bragg peaks after photoexcitation, but also those x-ray photons scattered in the weak "halo" between peaks. As the cations acquired the high-temperature symmetry, a large increase in the diffuse scattering was also detected, indicating an ultrafast order-disorder transition.

The reason why this disordering happens so fast is that immediately after photoexcitation, different dimers become uncorrelated at short distances and start to move independently. This loss of correlation is a defining feature of the high-temperature state of VO_2, in which the frequency of the molecular vibration is independent of the wavelength at short length scales. A wavelength-independent frequency in a chain of atoms implies that if a single bond is "plucked" at one site in the crystal, no sound wave is launched, but the vibration remains localized. Thus, photoexcited low-temperature VO_2 acquires the essence of the high-temperature phase, that is, nonpropagating short-wavelength sound waves, even before the atoms have had time to rearrange completely. A photoinduced structural transition is then not only about atomic positions, but also about their oscillation frequencies and correlation lengths.

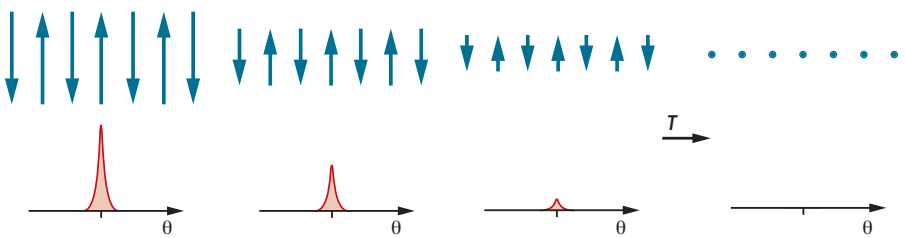
The most important implication of these measurements is that the much-publicized molecular movie, the idea that atomic positions as a function of time provide a complete picture of the microscopic physics, becomes a slippery concept of limited validity. The experiments reported here show the crucial importance of diffuse and inelastic x-ray scattering in dynamical experiments. No transition can be understood without a complete measurement of elastic and inelastic components of the scattered signal. More generally, the work of Wall *et al.* shows that x-ray free-electron lasers are opening up far more avenues of research than what was envisaged when these light sources were being planned, forcing us to reevaluate many old notions taken for granted up until now. ■

Symmetry changes can be in lockstep at all sites, or incoherent

Two ways in which a phase transition can proceed are illustrated for an antiferromagnetic spin system and its diffraction pattern (measured as angle θ) as a function of temperature *T*.

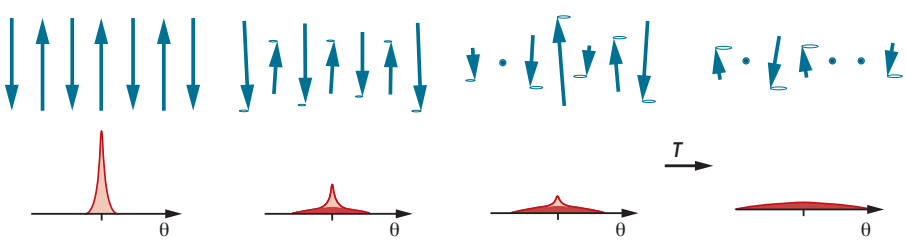
Displacive transitions

The ordering of spins uniformly decreases at each site with increasing temperature in a cooperative fashion, and diffraction peaks simply decrease in intensity.



Order-disorder transitions

The decrease in spin order occurs independently at each site. Diffraction peaks decrease in amplitude with temperature, but x-rays are also reflected incoherently into a diffuse background.



REFERENCES

1. S. Wall *et al.*, *Science* **362**, 572 (2018).
2. P. Emma *et al.*, *Nat. Photon.* **4**, 641 (2010).
3. A. Cavalleri *et al.*, *Phys. Rev. Lett.* **87**, 237401 (2001).
4. V. R. Morrison *et al.*, *Science* **346**, 445 (2014).
5. P. Baum *et al.*, *Science* **318**, 788 (2007).
6. A. Cavalleri *et al.*, *Phys. Rev. B* **70**, 161102 (2004).
7. C. Kübler *et al.*, *Phys. Rev. Lett.* **99**, 116401 (2007).

10.1126/science.aav2019

POLICY FORUM

TECHNOLOGY GOVERNANCE

Editing nature: Local roots of global governance

Environmental gene editing demands collective oversight

By **Natalie Kofler, James P. Collins, Jennifer Kuzma, Emma Marris, Kevin Esvelt, Michael Paul Nelson, Andrew Newhouse, Lynn J. Rothschild, Vivian S. Vigliotti, Misha Semenov, Rowan Jacobsen, James E. Dahlman, Shannon Prince, Adalgisa Caccone, Timothy Brown, Oswald J. Schmitz**

The end of malaria. Restored island habitats. Resiliency for species threatened by climate change. Many envisioned environmental applications of newly developed gene-editing techniques such as CRISPR might provide profound benefits for ecosystems and society. But depending on the type and scale of the edit, gene-edited organisms intentionally released into the environment could also deliver off-target mutations, evolutionary resistance, ecological disturbance, and extinctions. Hence, there are ongoing conversations about the responsible application of CRISPR, especially relative to the limitations of current global governance structures to safeguard its use [(*1*, *2*); see table S1]. Largely missing from these conversations is attention to local communities in decision-making. Most policy discussions are instead occurring at the national or international level (*3*, *4*), even though local communities will be the first to feel the context-dependent impacts of any release. To be fully representative, therefore, local inputs and perspectives must also be considered. As laboratories around the world develop and perfect gene-editing techniques with unprecedented capacity to alter wild species and, by extension, the ecological and cultural systems of which they are a part, we outline our vision for locally based, globally informed governance.

GENETICALLY ENGINEERING NATURE

CRISPR gene editing and other related genetic technologies are groundbreaking in their ability to precisely and inexpensively alter the genome of any species (*5*). CRISPR-based gene drives hold particular import because they are designed to rapidly spread genetic changes—including detrimental traits such as infertility—through populations of sexually reproducing organisms, to potentially reach every member of a species. Villages in Burkina Faso are weighing the release of gene drive–bearing mosquitoes that could suppress malaria. Nantucket Island residents in the United States are considering the release of genetically engineered white-footed mice to deplete Lyme disease reservoirs. New Zealand communities are discussing the possibility of using genetic methods to eliminate exotic predators.

But what if a gene drive designed to suppress an invasive species escaped its release site and spread to a native population? Or if a coral species gene edited to better adapt to environmental stressors dominated reef ecosystems at the expense of a diversity of naturally evolving coral species and the fish that depend on them (see the photo)? The gravity of these potential outcomes begs the question: Should humans even be meddling with the DNA of wild organisms? The absence of generally agreed on answers can be used to support calls for moratoria on developing and releasing genetically altered organisms, especially those with gene drives (*6*).

However, the promising benefits of environmental gene editing cannot be dismissed (*4*). Gene drives may provide a long-sought-after tool to control vectors of infectious disease and save millions of human lives. Projects to conserve ecosystems or promote species resilience are often intended to repair human-inflicted environmental damage. Put simply, either using this technology irresponsibly or not using it at all could prove damaging to humans, our welfare, and our planet.

National, regional, and international governmental agencies are working to clarify how existing research policy, field-testing frameworks, and risk-assessment guidelines apply to environmental gene editing, enacting some existing rules, and seeking to update and create new policies to address this technology. For example, the U.S. National Academies of Sciences, Engineering, and Medicine's report on gene drives deemed current ecological risk-assessment frameworks as adequate to predict potential ecological impacts of gene-drive release but recommended new guidelines to safeguard gene-drive research and encourage public discourse (*4*). In fact, several national and regional reports echo in their calls for improved fora that can support meaningful public debate (*7*, *8*); however, most frameworks for regulatory decision-making continue to largely preference science-based knowledge and technical risk assessments over ethical and societal considerations. At the international level, the Convention on Biological Diversity (CBD) has enlisted an expert technical panel to, in part, update its Cartagena Protocol (of which the United States is not a party) that oversees transboundary transport of living modified organisms to accommodate gene drive–bearing organisms (*9*). The International Union for the Conservation of Nature (IUCN) is also developing policy to address the release of gene-edited organisms (*3*). Although the CBD and the IUCN offer fora to engage diverse public feedback, a role largely fulfilled by civil society groups, none of these agencies currently use the broad and open deliberative process we advocate.

In the absence of widely agreed-upon governance guidelines or support for more optimal deliberative processes, the developers of a technology seeking consent to release a gene-edited organism may also serve as a community's source of expertise and information (*10*, *11*). Such an advice-and-consent relationship raises the possibility of a real or apparent conflict of interest. Ideally, in these cases, governance plans should incorporate expertise and perspectives that are independent, transparent, inclusive, and based on balanced deliberations.

Each decision to release a gene-edited organism has specific considerations that depend on the organism altered, the scope and intent of the alteration, the ecosystem(s) affected, consequences for human health, and the value systems of communities affected by such a decision. Underlying all of this are differing views about what is considered "natural" and to what degree humans should intervene in ecosystems (*12*). Different societal views about the human relationship to nature will therefore shape decision-making. Local community knowledge and perspectives must therefore be engaged to address these context-dependent, value-based considerations.

A special emphasis on local communities is also a matter of justice because the first and most closely affected individuals deserve a strong voice in the decision-making process. This is additionally a matter of urgency. Communities, technologists, and governments

See supplementary materials for author affiliations.
Email: natalie.kofler@yale.edu

will require methods to make responsible and informed decisions about environmental applications to keep pace with rapid progress in gene-editing technologies.

Compounding this challenge is that these decisions cannot be made in isolation. Organisms released into local environments may cross regional and even international borders. Hence, respect for and consideration of local knowledge and value systems are necessary, but insufficient, to anticipate the potentially ramifying global implications of environmental release of gene-edited organisms. What is needed is an approach that places great weight on local perspectives within a larger global vision.

INTEGRATED DELIBERATION

We propose a coordinating body that can convene communities, technology developers, and governmental and nongovernmental organizations in ways that ensure inclusive deliberations. Such a body would serve as a neutral third party to help inform decision-making that is free from conflicts of interest and locally based. This organization would (i) establish a deliberation framework that integrates diverse expertise and worldviews, including participants that represent impacted communities; (ii) facilitate deliberation to produce standardized reports and deliver recommendations; (iii) establish information-sharing protocols to connect deliberations around the world; and (iv) report on the outcomes of deliberation to inform global governance of gene editing.

Characterizing what defines an affected "local" community will be an important part of this process and will depend on the nature of the technology and how it is predicted to interact with the environment. For example, if a self-propagating gene drive is under deliberation to counter malaria transmission, then representatives from much of sub-Saharan Africa would deserve a voice. For cases where a technology is more limited in scope or designed to have more limited spread, predictive models could be used to define the communities most likely to be affected and relative riskiness of the alteration. In line with frameworks for responsible innovation, evaluation would ideally begin as early as possible, so that deliberation shapes the development of the application in question (*13*).

The integrated deliberation framework must incorporate expertise from fields such as genetics, philosophy, ecology, economics, law, and risk assessment, as well as representatives of diverse stakeholder groups, and members of affected communities. Historically marginalized voices (indigenous communities, ethnic minorities, women, and children) must be included. Network analysis could identify affected parties, especially those who have previously been overlooked; deliberative procedures should build on frameworks demonstrated to promote inclusive and democratic deliberation (*14*, *15*).

The needs of ecosystems could also be given voice to inform deliberative outcomes through custodial human proxies. Inspired by legislative precedent set by New Zealand, in which the Whanganui River was granted legal "personhood," human representatives, nominated by both an international body like the IUCN and the local community, would be responsible for upholding the health and interests of the ecosystems in question (*16*). Proposed gene-editing strategies would be placed in the larger context of alternative approaches to address the public health or environmental issue in question.

Coral reef ecosystems are estimated to support over 25% of all marine fish worldwide and to contribute over US$1 trillion in economic, social, and cultural value globally. CRISPR-based strategies have been proposed as a means to protect coral from bleaching. Corals genetically engineered to be more resilient to heat stress and ocean acidification could be used to help conserve the Great Barrier Reef (shown above), a UNESCO world heritage site.

To promote equitable representation, neutral and informed facilitators would bring history to the table, reveal existing power structures, and foster relationships between groups that hold disparate ideological stances. Deliberants would be encouraged to observe and reflect on their values—how they value nature, how they perceive risk, their level of trust in technology, and their motives and agendas—with full transparency. Questioning what is "natural" and to what degree ecosystems should be restored will also require exploration to ensure an appropriate ethical basis for decision-making.

In following with recent proposals for more meaningful deliberation over socially complex environmental issues (*17*), facilitators would also strive to cultivate certain virtues in participants. These decisions involve complexity and uncertainty and are motivated by concern for both human and nonhuman well-being; thus, deliberants must be encouraged to think in interdisciplinary ways, act with humility, and be mindful of their membership in an interdependent, planetary community. To incentivize virtue-based participation, the coordinating body must function in a manner that upholds extreme transparency and trustworthiness with deliberative outcomes that carry authority. Accordingly, an outright refusal to either participate in or heed the outcomes of the deliberative process could mark certain agents as untrustworthy, an outcome likely not in line with their strategy for long-term success.

Each deliberative process would yield a standardized report that summarizes concerns raised, areas where consensus was reached, and recommendations as to whether or how a gene-edited organism should be developed for environmental release. Reports could also accommodate a recommendation of "maybe, but not yet" and stipulate ecological, technical, and ethical considerations that require further study, reflection, and consideration. Ultimate control over the deliberative process would be shared by nominated local leaders, but the coordinating body would provide frameworks for deliberation and provide support throughout the process.

To allow deliberative outcomes to have immediacy, while not being encumbered by what will likely be a long-drawn-out process

PHOTO: DANIELA DIRSCHERL/GETTY IMAGES

to update national and international regulatory policy, a certification model for integrated deliberation could be one way to lend immediate authority and impact. In this vein, any project that successfully passes through the deliberative framework could be given a seal of completion to alert regulatory bodies and the general public that informed and inclusive deliberation steered the development of that specific technological application. This is not to deny that new regulatory policies need to be in place to cope with this technology, but a certification approach could address concerns about environmental applications in a timely way, while providing incentive for developers to participate.

FROM VISION TO ACTION

Our proposed environmental gene editing coordinating body could be jointly supported by several intergovernmental organizations. Similar to how the Intergovernmental Panel on Climate Change falls under the joint sponsorship of the World Meteorological Organization and the United Nations Environment Programme (UNEP), this new initiative could receive shared support from several concerned organizations, such as the World Health Organization (WHO), UNEP, or the IUCN. Falling under the auspices of existing intergovernmental organizations would serve to hasten the development of the coordinating body, so it can meet the rapid development of gene-editing technologies. Joint support would also lend immediate accountability and authority, while ensuring that environmental and human health concerns are equally upheld in deliberative procedures.

A trust fund built on contributions by concerned governments, nongovernmental organizations, and intergovernmental organizations and managed in a manner consistent with International Public Sector Accounting Standards could provide financial support. Our hope is that governments from around the world and global health and environmental organizations will financially support this organization, because in the absence of global guidelines, improper use of this technology could prove costlier than any up-front investments that ensure its responsible use.

An interdisciplinary leadership committee diverse in gender, age, geography, and worldviews and whose members are not direct beneficiaries of any decision-making process, would need to oversee the following essential next steps. First, a task force must be established to design the integrated deliberative framework. This task force would also be charged with defining the scope and type of edits to be covered, convening deliberative processes, and overseeing iterative improvements to deliberative design. Organizational procedures could be built on models used by existing international organizations. For example, the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services has developed protocols to integrate normative inputs into science-based policy (*18*), and effective measures for global coordination of diverse stakeholders and expertise could be garnered from the IUCN.

Second, an online registry for all projects intending to release genetically engineered organisms into the environment must be created. Currently, no central database exists for environmental gene-editing applications or for decision-making outcomes associated with their deployment, and this potentially puts the global community at risk. Third, a communications task force needs to create an online space that allows communities, technology developers, and policy-makers from around the world to share information resources, discuss issues faced, and provide expertise. The communications team would also oversee an annual summit and the publication of annual reports to share lessons learned and promote continued conversation on concerns raised.

"Put simply, either using this technology irresponsibly or not using it at all could prove damaging..."

Finally, leveraging the experience and infrastructure of its support organizations (for example, perhaps the WHO and UNEP), a global coordination task force would be charged with coordinating multiple communities, nations, and regions to ensure successful deliberative outcomes. As a hypothetical example, genetic strategies to eliminate invasive possums from New Zealand must include representatives from Australia, the country likely to be affected should animals be transported outside the intended range. Similarly, the African Union is currently deliberating appropriate governance of gene drive–bearing mosquitoes to combat malaria on a regional scale. The global coordination team would establish mechanisms to provide integrated deliberation services for regional and national decision-making. Coordination would serve to minimize geopolitical threats and ensure that the rights of affected communities are upheld at all levels of decision-making. Moreover, the new avenues forged for open communication by the coordinating body will enable deliberative outcomes to shape gene-editing governance on a global scale.

The success of this approach will depend on inputs and expertise from diverse worldviews and disciplines. Important questions remain to be answered: How can deliberative procedures effectively weigh local benefits with more-widespread global risks? How would transfer of control for the deliberations to local leaders take place? What structures are in place to guarantee historically marginalized voices are heeded in deliberation? What institutional procedures and evaluation mechanisms are needed to ensure accountability? Through the collective creation of this new governance model, the first that proposes a connection between local needs and global frameworks and expertise, our world may realize this technology's most profound benefit—the opportunity to inspire a more healthy and just future for all who share our planet. ■

REFERENCES AND NOTES

1. S. Jasanoff, J. B. Hurlbut, *Nature* **555**, 435 (2018).
2. L. Montoliu *et al.*, *CRISPR J.* **1**, 128 (2018).
3. IUCN, Development of an IUCN policy on synthetic biology; https://www.iucn.org/theme/science-and-knowledge/our-work/culture-science-and-knowledge/synthetic-biology-and-biodiversity-conservation/development-iucn-policy-synthetic-biology.
4. The National Academies of Sciences, Engineering, and Medicine, *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values* (National Academy Press, Washington, DC, 2016).
5. A. J. Piaggio *et al.*, *Trends Ecol. Evol.* **32**, 97 (2017).
6. ETC Group, Reckless driving: Gene drives and the end of nature (2016); www.etcgroup.org/content/reckless-driving-gene-drives-and-end-nature.
7. Australian Academy of Science, Synthetic gene drives in Australia: Implications of emerging technologies (2017); www.science.org.au/support/analysis/reports/synthetic-gene-drives-australia-implications-emerging-technologies.
8. New Partnership for Africa's Development, Gene drives for malaria control and elimination in Africa (2018); www.nepad.org/publication/gene-drives-malaria-control-and-elimination-africa.
9. CBD, Portal on synthetic biology; https://bch.cbd.int/synbio/.
10. C. Emerson, S. James, K. Littler, F. Randazzo, *Science* **358**, 1135 (2017).
11. I. Swetlitz, "In a remote West African villiage, a revolutionary genetic experiment is on its way—if residents agree to it," *STAT News*, 14 March 2017; www.statnews.com/2017/03/14/malaria-mosquitoes-burkina-faso/.
12. G. E. Kaebnick, *Hastings Cent. Rep.* **47**, S60 (2017).
13. R. Owen *et al.*, in *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, R. Owen, J. Bessant, M. Heintz, Eds. (Wiley, 2013), pp. 27–50.
14. S. Burall, *Nature* **555**, 438 (2018).
15. C. P. Neuhaus, *Hastings Cent. Rep.* **48**, 25 (2018).
16. E. L. O'Donnell, J. Talbot-Jones, *Ecol. Soc.* **23**, 7 (2018).
17. M. Ferkany, K. P. Whyte, *J. Agric. Environ. Ethics* **25**, 419 (2012).
18. S. Díaz *et al.*, *Science* **359**, 270 (2018).

ACKNOWLEDGMENTS

Authors thank the Yale Institute for Biospheric Studies, the Yale Interdisciplinary Center for Bioethics, and the Franke Program in Science and the Humanities for funding the 2017 Editing Nature Summit, which informed this manuscript. We are grateful for the help and support of J. and M. R. Vigliotti, I. Papangeli, G. Kaebnick, M. E. Tucker, L. Stevenson, C. A. Kroop, A. Miller, and our dearly remembered friend and colleague Calestous Juma.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/suppl/DC1

10.1126/science.aat4612

An Ebola treatment center sits empty in Freetown, Sierra Leone, in November 2016.

PUBLIC HEALTH

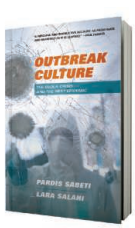
Improving outbreak response

Learning from Ebola failures is key for crafting better plans for public health emergencies

By **William F. Pewen**

"Knowing is not enough; we must apply. Willing is not enough; we must do." This observation, made many years ago by Johann von Goethe, might well have been said about the West African Ebola outbreak of 2014–2016, when neither advanced technology nor humanitarian intent was sufficient to avert a major health crisis.

In *Outbreak Culture*, geneticist Pardis Sabeti joins journalist Lara Salahi to examine the epidemic from the perspective of the clinical workers, researchers, and health agency personnel who experienced it firsthand. They note in particular three factors that foster a culture that can compromise outbreak response: (i) fear; (ii) the instinct to protect oneself, others, or institutions; and (iii) the desire to exploit. The book focuses on what happens when these motives collide, indicting poor policies, ineffective decision-making, and bad actors.



Outbreak Culture
Pardis Sabeti and Lara Salahi
Harvard University Press, 2018. 275 pp.

An illustrative prologue depicts the tragic death of Sheikh Humarr Khan, the physician director of the Kenema Government Hospital who succumbed to Ebola infection in July of 2014. Before his death, physicians struggled to reconcile ethical questions: Should they use a new experimental treatment? If he subsequently perished, would the attempt be viewed by the public as a noble failure or the exploitation of yet another African life? Would offering him a new treatment be equitable? Could a patient without Khan's training provide truly informed consent to receive such a therapy?

The authors deftly weave surveys, interviews, and retrospectives on previous outbreaks into a concise depiction of the Ebola epidemic, with particular focus on human behaviors that exacerbated the course of the outbreak. They describe the negative impacts of the egregious quarantine in Liberia and acknowledge the contribution of local burial practices to viral transmission, but they also remind readers how poorly health care workers returning to the United States were treated and how inadequately the country responded to a mere handful of cases.

Accountability proved a critical problem in the Ebola response. Funds often flowed directly to individuals, fostering corruption: Less than 2% were directed to the pressing needs of local responders.

Sabeti and Salahi present a wealth of evidence supporting the imperative that outbreak response must operate in a coordinated, real-time manner. The construction of 11 U.S.-built Ebola treatment units is an emblematic example: Just two units were ever occupied because they were no longer needed by the time they were completed.

Lags in detection and mobilization and delays in the establishment of best clinical practice proved costly. Failures synergized: When the Sierra Leone Ministry of Health and Sanitation counted only laboratory-confirmed Ebola deaths, yet had just two phlebotomists in the field collecting blood samples, cases were underreported. Meanwhile, response was compromised by a deficit of collaboration: 43% of the authors' survey respondents reported data hoarding.

Sabeti and Salahi provide a set of principles to improve outbreak response. They propose a foundation of universal ethical standards coupled with increased accountability. They note that cooperation could be incentivized, from eliminating stigma and discomfort in implementing quarantine to reforming data sharing and patent systems in order to promote collaboration.

Most provocative is the authors' vision of a system of global readiness—one that combines a military-styled response and centralized governance. Given that substantial compliance issues exist even under existing World Health Organization health regulations, could such an integrated authority gain the support of national governments and the cooperation of local populations? In an era of increasing nationalism and waning support of international efforts, the bar to achieve such an ambitious vision appears high.

Yet in a time when it can be easier to sequence a virus than to obtain a clinical sample or share genomic data, changing outbreak culture demands systems innovations and policy commitments. As Sabeti's colleague Nathan Yozwiak notes, "The limitations are human, they're not technical anymore." ■

10.1126/science.aav2473

The reviewer is visiting associate professor and interim chair of the Department of Health, Nursing and Nutrition at the University of the District of Columbia, Washington, DC, USA. Email: william.pewen@udc.edu

PALEONTOLOGY

Collector's item or cultural heritage?

A riveting tale of a smuggled dinosaur illuminates an enduring tension in paleontology

By **Victoria M. Arbour**

When a dinosaur is named as a defendant in a court case, it is all but ensured that headlines will follow. So it was in 2013, when the U.S. government sought to seize a smuggled dinosaur skeleton in a case entitled *United States v. One Tyrannosaurus Bataar Skeleton*. In *The Dinosaur Artist,* journalist Paige Williams tells the story of Eric Prokopi—the man who purchased, prepared, and attempted to sell the *Tarbosaurus* in question—and reveals how the skeleton made its way out of, and ultimately back to, Mongolia. Throughout the book, she folds together multiple stories to illuminate the history of fossil hunting, revealing how fossils wind up as precious objects of national and cultural heritage or coveted collector's items.

Tarbosaurus bataar was a close relative of *Tyrannosaurus rex* that roamed the humid floodplains of Mongolia during the Late Cretaceous, toward the end of the age of dinosaurs. Some *Tarbosaurus* skeletons have been excavated by teams of paleontologists and can be found in museums, where they are used for research and public display. Others enter a not-so-secret underground trade in fossils illegally exported from Central Asia and are sold piecemeal at gem and mineral shows, on eBay, or through private dealers.

Fossils sold to private hands are effectively lost to science. Specimens in the public trust, such as accredited museums, can be examined over and over again by paleontologists now and in the future; private collections do not have the same mandates to care for fossils in perpetuity. For this reason, many countries, including Mongolia, make it illegal to excavate or export scientifically important fossils without a permit.

Prokopi built a living by selling fossils he scoured from beaches and waterways in Florida, starting when he was just a teenager. The *Tarbosaurus* was his biggest acquisition yet, shipped to America via a dealer based in England in crates labeled "broken fossil bones" and "fossil reptiles." After more than a year of work cleaning the bones and fabricating an armature to mount them together, the skeleton went up for auction in 2011. When the Mongolian government raised concerns over the sale of the skeleton, it kickstarted a multiyear investigation that would ultimately see the *Tarbosaurus* seized and returned to Ulaanbaatar. Prokopi would lose his house and face time in jail for smuggling fossils.

I must admit that I came away from *The Dinosaur Artist* with respect for Prokopi's passion for fossils, if not for his methods. In a world where full-time positions for academic paleontologists are extremely rare, and where obtaining such a position requires that one have the time and money to pursue a graduate degree, he forged his own path.

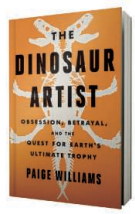
The smuggled *Tarbosaurus* was eventually returned to Mongolia.

PHOTO: BYAMBASUREN BYAMBA-OCHIR/AFP/GETTY IMAGES

The Dinosaur Artist puts into stark relief the way wealth, privilege, and nationality shape who gets to participate in paleontology and reveals how shifting norms complicate the legacy of revered paleontologists of the past. Mary Anning, a 19th-century English fossil hunter who struggled to be taken seriously as a contributor to science (despite making numerous important discoveries), is now lauded as an unsung hero of our discipline, for example. But, as Williams reveals, she also sold her fossils to make a living. Similarly, Roy Chapman Andrews, who led the first American paleontological expedition to Mongolia, inspired many of today's prominent paleontologists. But, like Anning, he also sold some of his finds.

The reviewer is curator of paleontology, Royal BC Museum, Victoria, BC, Canada. Email: varbour@royalbcmuseum.bc.ca

The Dinosaur Artist
Obsession, Betrayal, and the Quest for Earth's Ultimate Trophy
Paige Williams
Hachette Books, 2018.
432 pp.



Mongolian fossils have contributed greatly to our understanding of dinosaur evolution; however, most have been excavated by foreign teams of paleontologists who receive the credit and fame for these discoveries. As such, the tale of the smuggled *Tarbosaurus* raises some uncomfortable questions for academic paleontologists who, like myself, take part in foreign expeditions. Are our efforts just another form of western colonialism? Have those of us who study Mongolian fossils truly done all we can to help grow capacity for research and public engagement within that country?

Although the heart of Williams's tale is Prokopi's story, I found inspiration in Mongolian paleontologist Bolor Minjin's quest to bring the *Tarbosaurus* home and to reboot paleontological research and education in her home country. She raised the alarm over the sale of the *Tarbosaurus* and, with the nonprofit Institute for the Study of Mongolian Dinosaurs that she founded, has helped repatriate numerous Mongolian skeletons that had been smuggled out of the country. Each summer, she leads a team that travels across the Mongolian countryside in a 40-foot mobile museum, bringing dinosaurs directly to her fellow Mongolians. But the resources for building education and research from the ground up in Mongolia pale in comparison with the sums involved in the sale of smuggled dinosaurs, and Minjin faces a constant uphill battle.

Ultimately, *The Dinosaur Artist* is a compelling, nuanced look at the surprisingly complex politics that surround fossil collecting. It should be required reading for both professional paleontologists and fossil enthusiasts and is a gripping read for anyone interested in the interplay between culture, history, and science. ■

10.1126/science.aav2020

LETTERS

Edited by **Jennifer Sills**

Beyond hierarchical one-on-one mentoring

The NextGen Voices section "Quality mentoring" (5 October, p. 22) demonstrates how traditional hierarchical mentoring relationships, when they work, can be sources of incredible psychosocial and practical support. However, when these relationships are not strong, they can hinder or even harm mentees (*1*, *2*). The unequal power dynamic of a senior mentor (typically one who is male and white) and junior mentee can be especially problematic for individuals belonging to systematically marginalized identity groups (such as women, people of color, and individuals with disabilities) and can exacerbate a sense of isolation for the mentee (*3*). Furthermore, mentees, more than mentors, say that mentoring relationships should directly address cultural diversity (*4*). A mentoring network, including peer mentoring, can address the shortfalls of traditional one-on-one mentoring.

A mentoring network with multiple modes of mentoring (*5*) dismantles the guru mentor myth, which suggests that one senior mentor is a necessary and sufficient source of mentoring. Instead, a mentoring network framework centers on the mentees and what they need and desire to thrive in their career; it then meets their varied needs through a host of mentoring relationships (*6*, *7*). Peer mentoring can serve as an important node in an individual's broader mentoring network and reduce the reliance on hierarchical relationships (*8*).

Peer mentoring is a truly horizontal mentoring experience (*9*) that offers participants access to resources, support, and accountability in a regular group meeting setting. Evidence suggests that peer mentoring is most effective with groups of five to eight participants who are all at a similar career stage, have complementary fields of expertise, and share social identities (such as gender, race, ethnicity, or ability status) (*10*). There is no senior mentor, and thus the model asserts that each peer mentoring participant has useful wisdom and perspectives to share as well as areas in which they need advice. Peer mentoring provides an opportunity to collaboratively problem solve, share ideas and perspectives, and develop community and thus serves as a mechanism for developing independence and career self-efficacy (*11*).

Peer mentoring becomes especially important as mentees mature and develop into independent scientists. Through peer mentoring, individuals participate in reciprocal and interactive relationships in which they have the opportunity to develop not only their own problem-solving skills and career self-efficacy but also their confidence and skills as mentors. Peer mentoring can be a component of a professional development program (*12*) or a stand-alone activity (*10*). Those looking for the mentoring so valued by contributors to the NextGen Voices survey might consider giving peer mentoring a try.

M. Claire Horner-Devine,[1,2*] Torie Gonsalves,[1] Cara Margherio,[3] Sheri J. Mizumori,[4] Joyce W. Yen[1]
[1]ADVANCE Center for Institutional Change, University of Washington, Seattle, WA 98195, USA. [2]Counterspace Consulting, Seattle, WA 98103, USA. [3]Center for Evaluation and Research for STEM Equity, University of Washington, Seattle, WA 98195, USA. [4]Department of Psychology, University of Washington, Seattle, WA 98195, USA.
*Corresponding author. Email: mchd@uw.edu

REFERENCES

1. B. R. Ragins, J. L. Cotton, J. S. Miller, *Acad. Manage. J.* **43**, 6 (2000).
2. S. E. Straus, M. O. Johnson, C. Marquez, M. D. Feldman, *Acad. Med. J. Assoc. Am. Med. Coll.* **88**, 1 (2013).
3. D. J. Davis, *Mentor. Tutor. Partner. Learn.* **16**, 3 (2008).
4. A. Byars-Winston, "Race matters: Building the 21st century clinician, educator and scientist" (2016); https://academicaffairs.ucsf.edu/ccfl/media/lectures/Race%20Matters%20Slides.pdf.
5. K. A. Rockquemore, "When it comes to mentoring, the more the merrier" *Chron. High. Educ.* (2014); https://chroniclevitae.com/news/326-when-it-comes-to-mentoring-the-more-the-merrier.
6. K. E. Kram, *Mentoring at Work: Developmental Relationships in Organizational Life* (Scott Foresman, Glenview, IL, 1985).
7. R. DeCastro, D. Sambuco, P. A. Ubel, A. Stewart, R. Jagsi, *Acad. Med.* **88**, 4 (2013).
8. A. Darwin, E. Palmer, *Higher Educ. Res. Dev.* **28**, 2 (2009).
9. M. C. Horner-Devine, C. Margherio, S. J. Mizumori, J. W. Yen, "Peer Mentoring Circles: A strategy for thriving in science," *BioMed Central Blog* (2017); https://blogs.biomedcentral.com/bmcblog/2017/05/18/peer-mentoring-circles-a-strategy-for-thriving-in-science/.
10. E. Daniell, *Every Other Thursday: Stories and Strategies from Successful Women Scientists* (Yale, Cambridge, MA, 2006).
11. L. R. Fraga, "Comment: Mentoring for institutional gain" in *A Truly Diverse Faculty: New Dialogues in American Higher Education*, S. A. Fryberg, E. J. Martinez, Eds. (Palgrave Macmillan, New York, 2014), pp. 257–264.
12. J. W. Yen, M. C. Horner-Devine, C. Margherio, S. J. Mizumori, *Neuron* **94**, 3 (2017).

10.1126/science.aav7656

Engaging community with humility

In his Policy Forum "Building an evidence base for stakeholder engagement" (10 August, p. 554), J. V. Lavery rightly proposes additional reporting and evidence collection to understand best practices for community and stakeholder engagement. However, we are concerned that he framed stakeholder engagement too narrowly.

Lavery's proposed consumer model for engagement replicates the individual

PHOTO: KLAUS VEDFELT/GETTY IMAGES

Peer mentoring can provide support, facilitate collaborative problem solving, and build confidence.

focus of human subject research in an arena where stakeholders' interests and power are often collective. His reference to the human-centered design that informed Apple tech products for purchase by individual consumers fails to capture the complexity and challenges that communities and stakeholders face when confronted with large-scale technical interventions that affect the shared environment. Lavery's avoidance of the term "public" in favor of this consumer model contrasts with important research and practice in the pursuit of more democratic visions for the governance of science and technology (*1*, *2*).

Although Lavery cites the potential for community and stakeholder engagement to increase the success of science programs, he does not explore stakeholders' role in defining what would count as success. For community and stakeholder engagement to matter and serve as more than just window dressing, scientists and funders need to relinquish some of their power and authority by allowing stakeholders—including historically marginalized groups that are traditionally excluded from governance spaces—to influence problem framings, program goals, and other key decisions (*3*, *4*). Communities and stakeholders should help identify the desired goals of a science program instead of serving as a means to achieve the program sponsor's predetermined ends (*5*, *6*). For community and stakeholder engagement to take a more central role in science programs, scientists and funders will have to embrace a new kind of humility.

Jason Delborne,* Adam Kokotovich, S. Kathleen Barnhill-Dilling
North Carolina State University, Raleigh, NC 27695, USA.
*Corresponding author. Email: jadelbor@ncsu.edu

REFERENCES

1. A. Stirling, K. R. Hayes, J. Delborne, *BMC Proc.* **12**, 43 (2018).
2. A. Delgado, H. Åm, *PLOS Biol.* **16**, e2003921 (2018).
3. S. Jasanoff, J. B. Hurlbut, *Nature* **555**, 435 (2018).
4. R. A. Salmon, R. K. Priestley, J. Goven, *J. Environ. Stud. Sci.* **7**, 53 (2017).
5. D. Tomblin *et al.*, *Astropolitics* **15**, 141 (2017).
6. P. Macnaghten *et al.*, *J. Responsible Innovation* **1**, 191 (2014).

10.1126/science.aav4987

Evolution of teeth in South America

In his Perspective "Marsupial responses to global aridification" (5 October, p. 25) on the Report "Rapid Pliocene adaptive radiation of modern kangaroos" (A. M. C. Couzens and G. J. Prideaux, 5 October, p. 72), P. D. Polly discusses the evolution of high molar teeth in Australian kangaroos and wallabies. He cites evidence that, in South America, mammals developed these high-crowned teeth in the context of humid forests. However, this has not been determined with certainty.

Opinions have oscillated in recent years regarding the context and causes of the high-crowned molars [so-called "precocious hypsodonty" (*1*)] exhibited by late Eocene and early Oligocene South American mammals. Several lines of evidence, including fossilized plant tissue, mammals, and stable isotopes, suggest that high-crowned teeth developed in the context of relatively arid and open habitats (*2*–*5*). Curiously, these habitats seem to have been dominated by palms rather than grasses (*4*) and thus were neither true grasslands nor analogous to most modern South American open habitats. The discrepancy is reminiscent of the ancient South American mammal communities themselves, which were not analogous to those found there today (*6*).

As Polly explains, geographically isolated Cenozoic mammal species, such as those of Australia and South America, are well suited for independently documenting how mammalian herbivores have responded to changing climates and other factors over millions of years. In South America, unlike in Australia, the relationship between aridity and the evolution of high molars broadly parallels that seen in the Northern Hemisphere (*4*, *7*); it just happened to take place some 15 to 20 million years earlier (*2*).

Darin A. Croft
Department of Anatomy, Case Western Reserve University School of Medicine, Cleveland, OH 44106-4930, USA. Email: dcroft@case.edu

REFERENCES

1. B. Patterson, R. Pascual, *Q. Rev. Biol.* **43**, 409 (1968).
2. J. J. Flynn, A. R. Wyss, D. A. Croft, R. Charrier, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **195**, 229 (2003).
3. D. A. Croft, J. J. Flynn, A. R. Wyss, *Arq. Mus. Nac. Rio de Janeiro* **66**, 191 (2008).
4. R. E. Dunn, C. A. E. Strömberg, R. H. Madden, M. J. Kohn, A. A. Carlini, *Science* **347**, 258 (2015).
5. M. J. Kohn *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **435**, 24 (2015).
6. D. A. Croft, R. K. Engelman, T. Dolgushina, G. Wesley, *Proc. R. Soc. Ser. B Biol. Sci.* **285**, 20172012 (2018).
7. R. Pascual, E. Ortiz-Jaureguizar, *J. Hum. Evol.* **19**, 23 (1990).

10.1126/science.aav6745

ERRATA

Erratum for the Research Article "Saturn's magnetic field revealed by the Cassini Grand Finale" by M. K. Dougherty *et al.*, *Science* **362**, eaav6732 (2018). Published online 12 October 2018; 10.1126/science.aav6732

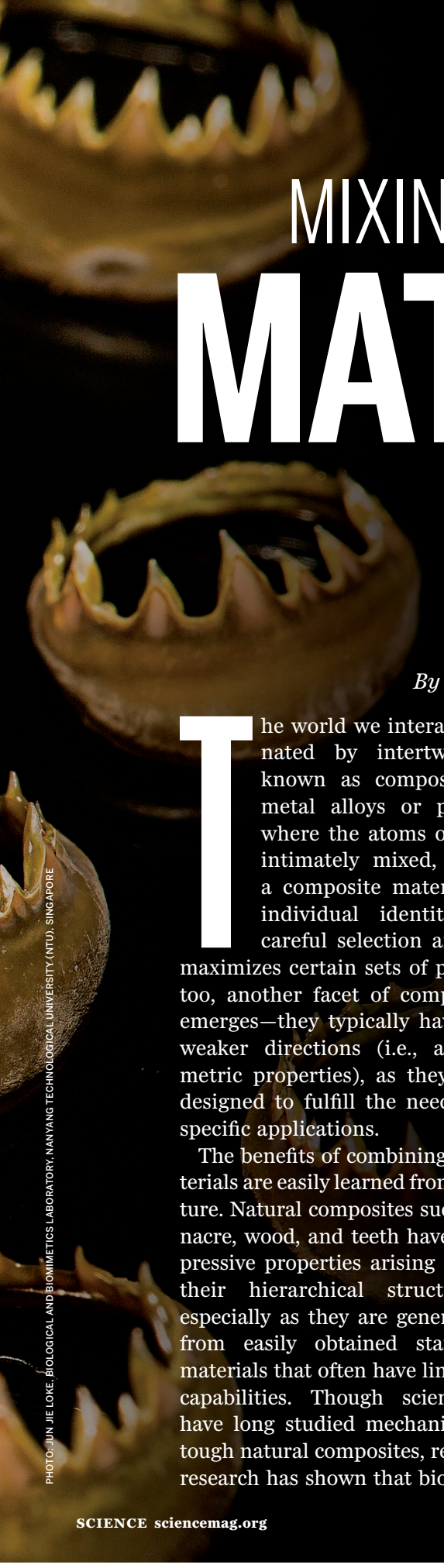


PHOTO: JUN JIE LOKE, BIOLOGICAL AND BIOMIMETICS LABORATORY, NANYANG TECHNOLOGICAL UNIVERSITY (NTU), SINGAPORE

MIXING AND MATCHING MATERIALS

By **Marc Lavine** *and* **Brent Grocholski**

The world we interact with is dominated by intertwined materials known as composites. Unlike in metal alloys or polymer blends, where the atoms or molecules are intimately mixed, components of a composite material retain their individual identities, and their careful selection and combination maximizes certain sets of properties. Here, too, another facet of composite materials emerges—they typically have stronger and weaker directions (i.e., asymmetric properties), as they are designed to fulfill the needs of specific applications.

The benefits of combining materials are easily learned from nature. Natural composites such as nacre, wood, and teeth have impressive properties arising from their hierarchical structures, especially as they are generated from easily obtained starting materials that often have limited capabilities. Though scientists have long studied mechanically tough natural composites, recent research has shown that biological sponges transform materials as simple as silica into components that are both tough and exceptionally efficient at manipulating light. Further, natural materials, many of which end up as waste in the process of harvesting food, may have properties that can compete with those of synthetic materials and have the advantage of originating from renewable resources.

Carbon-based composites such as those fabricated from carbon fiber offer the unique combination of high strength and low density. Replacing carbon fiber with graphene or carbon nanotubes opens a wide range of new properties for optimization. The potential to develop composites with exceptional mechanical, thermal, and electrical properties provides great incentive to overcome processing hurdles to efficient production scale-up.

In all good composites, the whole is better than the sum of the parts. The same may be true of research on composites, as looking beyond one's traditional sources of materials may lead to better composite materials.

INSIDE

REVIEWS

RELATED ITEM

▶ PODCAST

Squid sucker teeth are strong enough to pierce and anchor prey but are composed solely of protein building blocks that can be reshaped upon heating.

REVIEW

Composites from renewable and sustainable resources: Challenges and innovations

Amar K. Mohanty[1,2]*, Singaravelu Vivekanandhan[3], Jean-Mathieu Pin[1], Manjusri Misra[1,2]

Interest in constructing composite materials from biosourced, recycled materials; waste resources; and their combinations is growing. Biocomposites have attracted the attention of automakers for the design of lightweight parts. Hybrid biocomposites made of petrochemical-based and bioresourced materials have led to technological advances in manufacturing. Greener biocomposites from plant-derived fiber and crop-derived plastics with higher biobased content are continuously being developed. Biodegradable composites have shown potential for major uses in sustainable packaging. Recycled plastic materials originally destined for landfills can be redirected and repurposed for blending in composite applications, thus leading to reduced dependence on virgin petro-based materials. Studies on compatibility of recycled and waste materials with other components in composite structure for improved interface and better mechanical performance pose major scientific challenges. This research holds the promise of advancing a key global sustainability goal.

The era of natural fiber composites currently known as biocomposites dates back to 1908 with the introduction of cellulose fiber–reinforced phenolic composites. This innovation was followed by synthetic glass fiber–reinforced polyester composites, which obtained commodity status in the 1940s. The use of biobased green polymers to manufacture auto parts began in 1941, when Henry Ford made fenders and deck lids from soy protein–based bioplastic. The use of composite materials, made with renewable and sustainable resources, has become one of the vital components of the next generation of industrial practice. Their expanding use is driven by a variety of factors, including the need for sustainable growth, energy security, lower carbon footprint, and effective resource management, while functional properties of the materials are simultaneously being improved. Innovative sustainable resources such as biosourced materials, as well as wastes, coproducts, and recycled materials, can be used as both the matrix and reinforcement in composites to minimize the use of nonrenewable resources and to make better use of waste streams.

Composite materials find a wide range of potential applications in construction and auto-parts structures, electronic components, civil structures, and biomedical implants. Traditionally, industrial sectors that require materials with superior mechanical properties use composites made from glass, aramid, and carbon fibers to reinforce thermoplastics such as polyamide (PA), polypropylene (PP), and poly(vinyl chloride) (PVC), as well as thermoset resins such as unsaturated polyester (UPE) and epoxy resin. In addition to fiber, mineral fillers such as talc, clay, and calcium carbonate are being used in composite manufacturing. Such hybrids of fiber and mineral fillers play a major role in industrial automotive, housing, and even packaging applications. Carbon black plays a vital role as a reinforcement, especially in rubber-based composites. The key environmental concern with regard to composite materials is the difficulty of removing individual components from their structures to enable recycling at the end of a material's service life. At this point, most composite materials are either sent to a landfill or incinerated. Wood and other natural fibers (e.g., flax, jute, sisal, and cotton), collectively called "biofibers," can be used to reinforce fossil fuel–based plastic, thus resulting in biocomposite materials. Synthetic glass fiber–reinforced biobased plastics such as polylactides (PLAs) are a type of biocomposite. Biofiber-PP and biofiber-UPE composites have reached commodity status in many auto parts, as well as decking, furniture, and housing applications. Hybrid biocomposites of natural and synthetic fibers as well as mixed matrix systems also represent a key strategy in engineering new classes of biobased composites. As part of feedstock selection, a wide range of renewable products that includes agricultural and forestry residues, wheat straw, rice straw, and waste wood, as well as undervalued industrial coproducts including biofuel coproducts such as lignin, bagasse, and clean municipal solid wastes, is currently being explored to derive chemicals and materials. Recent advancements in biorefinery concepts create new opportunities with side-stream product feedstock that can be valorized in the fabrication of a diverse array of biocomposites.

Materials scientists can help in advancing sustainable alternatives by quantifying the environmental burden of a material through its product life-cycle analysis (*1*, *2*). The exponential growth of population and modernization of our society will lead to a threefold increase in the demand for global resources if the current resource-intensive path is continued (*3*). According to the United Nations, a truckload of plastic waste is poured into the sea every minute. By 2050, at current rates, the amount of plastic in the ocean will exceed the number of fish. The benefit of diverting plastic packaging material is estimated at around $80 billion to $120 billion, which is currently lost to the economy (*4*). If diverted for composite use, the recycled and waste plastic currently destined for landfills and incineration would be used for sustainable development, thereby reducing dependence on nonrenewable resources such as petroleum. Postindustrial food processing wastes are being explored as biofiller in biodegradable plastics for the development of compostable biocomposites. Low-value biomass and waste resources can be pyrolyzed to provide biocabon (biochar) as sustainable filler for biocomposite uses. The increased sustainability in composite industries requires basic and transformative research toward the design of entirely green composites. Renewable resourced–based sustainable polymers and bioplastics, as well as advanced green fibers such as lignin-based carbon fiber and nanocellulose, have great potential for sustainable composites. Biobased nonbiodegradable composites show promising applications in auto parts and other manufacturing applications that require durability. Biodegradable composites also show promise in sustainable packaging. This comprehensive Review on composites from sustainable and renewable resources aims to summarize their

"Wood and other natural fibers (e.g., flax, jute, sisal, and cotton), collectively called 'biofibers,' can be used to reinforce fossil fuel–based plastic, thus resulting in biocomposite materials... Biofiber-PP and biofiber-UPE composites have reached commodity status in many auto parts, as well as decking, furniture, and housing applications."

[1]Bioproducts Discovery and Development Centre, Department of Plant Agriculture, University of Guelph, Guelph, ON N1G 2W1, Canada. [2]School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada. [3]Sustainable Materials and Nanotechnology Lab, Department of Physics, V.H.N.S.N. College (Autonomous), Virudhunagar, Tamilnadu 626 001, India.
*Corresponding author. Email: mohanty@uoguelph.ca

current status, constraints on wider adoption, and future opportunities. In keeping with the broad focus of this article, we analyze the current development of such composites and discuss various fibers and fillers for reinforcements, current trends in polymer matrix systems, and integration of recycled and waste coproducts into composite systems to outline future research trends.

Fibers and fillers from renewable and sustainable resources

In polymer composites, plastic resins remain as continuous phase, whereas fibers and fillers stay in discontinuous phase to provide reinforcement effects. The composite performance is governed through the interface between the fibers and the polymer matrix. In composite science, the key target is the interface and related interfacial bonding as the stress transfer between the fiber and polymer matrix dictates the overall mechanical performance (*5*). When deciding on the appropriate fiber and filler system for sustainable composite uses in industrial sectors, it is necessary to compare the cost and availability, consistency of properties, and environmental advantages of sustainable fibers with those of their traditional synthetic counterparts. Figure 1 depicts examples of different types of fibers and fillers for biocomposites.

Lignocellulosic plant fibers

This category, well known as natural fiber or biofiber, is broadly classified as wood and nonwood fibers. These include various types such as bast, leaf, seed or fruit, straw, grass, and wood (*6*). The mechanical performance of various plant fibers depends on their cell wall structure, composition, and morphology. Cellulose content, lumen size, and microfibrillar angle are other key factors governing the stiffness of plant fiber reinforcements. The advantages of biofibers over traditional glass fibers and mineral fillers are their lower density, reduced cost, eco-friendly nature, and enhanced performance in certain applications. Among the available natural fillers that can be used for composite applications, wood is the most commonly used. Cotton fibers are also widely available. Other agricultural natural fibers such as flax, jute, kenaf, industrial hemp, and sisal are used as well. Because of their desirable structural properties, the construction field is the major arena for the use of natural fiber composites. Unconventional natural fibers such as agro-residues (e.g., wheat and rice straws, ground coconut shells) and grasses (e.g., miscanthus, switch grass, and bamboo) find application in biocomposites.

Biofuel coproducts, food processing wastes, and other postindustrial wastes

Value-added uses of coproducts and by-products from biorefinery and biomass conversion processes are beneficial for sustainable development. Distillers' dried grains with solubles (DDGS) from the corn ethanol industry, lignin coproduct from lignocellulosic ethanol industry as well as the pulp and paper industry, and bagasse from the sugarcane ethanol industry are being used in biocomposites (*7*). Lignin, which is polyphenolic in composition, has found value-added uses in sustainable biobased composite materials. Lignin, with its many functional surface –OH groups as well as with surface modification, has shown improved reinforcing effect on resulting biocomposite performance. Additionally, fruit and vegetable pomace, coffee chaff, and grain hulls have been explored in composites. Most of these biofillers are used for waste valorization, whereas the resulting biocomposites can be used for nonstructural applications. Chicken feathers, currently treated as waste in the global poultry industry, can find application as renewable fiber reinforcement in lightweight biocomposites. Recycled carbon fiber from the aerospace and sporting goods industries is another sustainable source that can be used to generate hybrid composites with biofibers.

Biocarbon, a new sustainable filler and functional material

Biocarbon, also known as biochar, has emerged as a new sustainable material for many applications. Biochar is not limited to filler and reinforcement uses for biocomposites (*8*); it is also beneficial for the development of next-generation functional carbon materials (*9*) for potential applications in energy storage and filtration devices. The thermochemical conversion of biomass when oxygen is absent or in limited supply (also known as pyrolysis) generates liquid bio-oil, solid biochar, and syngas. Depending on temperature, time, and the nature of biomass in the pyrolysis process, the amount of oil, char, and gas may vary. Biochar or biocarbon (BioC) is an amorphous carbon-rich material that can be tunable in terms of chemical structure, porosity, size, and intrinsic modulus. Two other amorphous carbon–based materials are activated carbon (AC) and carbon black (CB). The main distinctions among such carbon-rich materials are based on their origin, formation process, and structure. The carbon content of BioC varies from 40 to 90%, as compared with 80 to 95% for AC and >95% for CB. Regarding origin of formation, BioC is generated from biomass; AC from asphalt, coal, and biomass; and CB from petroleum and coal tar.

Matrix polymer from renewable and sustainable resources

The majority of plastic resins in biocomposites are predominantly concentrated among petro-based

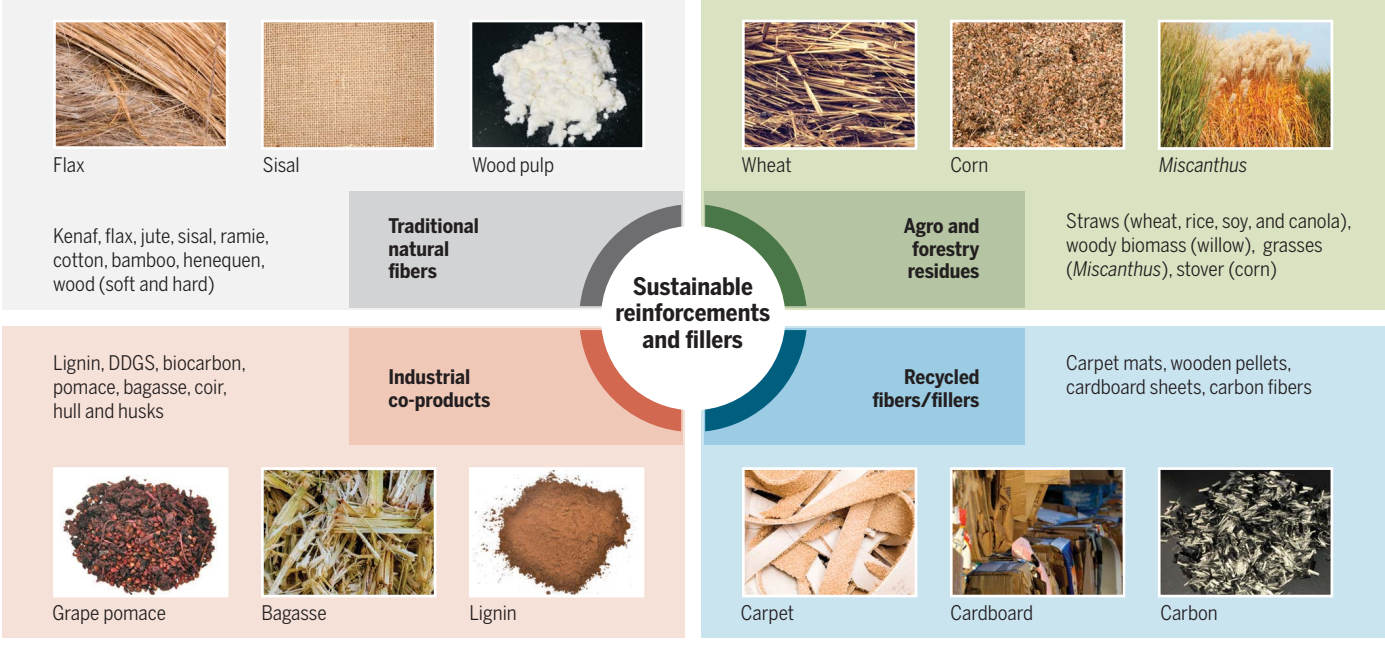


Fig. 1. Fibers and fillers from renewable and sustainable resources.

PHOTOS: (CLOCKWISE FROM TOP LEFT) ISTOCK/LIOR2, ISTOCK/LUCATO, A. K. MOHANTY; ISTOCK/IVAN4ES, ISTOCK/RDONAR; ISTOCK/NP-E07, A. K. MOHANTY; ISTOCK/OCEANBOUNDDB, ISTOCK/BANKSPHOTOS, A. K. MOHANTY; ISTOCK/JULIO RICCO, A. K. MOHANTY

thermoplastics [e.g., PP, polyethylene (PE), and PVC] and thermosets (e.g., UPE and epoxy). Global annual production of plastics was estimated as 407 million metric tons (MT) in 2015 (*10*), and production related to polyolefin with PP and PE exceeds this amount by >50%. Technically, the majority of all plastics can be made from renewable resources (*11*). Figure 2 depicts various types of plastics for use in composites. Plastics that are biobased, biodegradable, or both are known as bioplastics. The current trend of biobased polymer development is definitely growing: from a total production capacity of ~2.05 million MT in 2017 to an expected 2.44 million MT in 2022 (*12*). Bioplastics are generally categorized in three groups: (i) biobased and biodegradable plastics [i.e., polyhydroxyalkanoates (PHAs) and PLAs]; (ii) entirely or partly biobased and nonbiodegradable plastics [i.e., bio-based PE (bio-PE), biobased polyamide (bio-PA), biobased polyethylene terephthalate (bio-PET), and bio-thermosets such as polyfurfuryl alcohol (PFA) or bio-polyurethane (bio-PU)]; and (iii) petroleum-based biodegradable plastics [i.e., polybutylene adipate terephthalate (PBAT), polycaprolactone (PCL), and polybutylene succinate (PBS)]. PBS can be fully biobased, and partially biobased PBS is commercially available at present. Fully biobased polyethylene furanate, a new nonbiodegradable material, shows immense promise in industrial applications. Degradable plastics are attracting increased attention for disposable plastic applications, and nondegradable bio-PE and bio-PET have also demonstrated potential for packaging applications. Poly(trimethylene terephthalate) (PTT) is a semicrystalline biobased plastic generated from condensation of biobased propanediol and petrochemical terephthalic acid. Bio-PAs can be used in the automotive and consumer products sectors, although further large-scale production will be needed to achieve costs comparable to those of their petro-based counterparts. Biobased plastics from biological sources show promise for reducing environmental concerns caused by petrochemical plastics. Bioethanol is produced from starch and sugarcane fermentation and commercialized for fuel applications. The dehydration of bioethanol can produce ethylene, which, upon polymerization, makes bio-PE, a material that has already been commercialized. Besides this complete deoxidation strategy, catalytic redox approach of carbohydrates by biological or chemical processes can lead to high-value monomers (*13*). Terephtalic acid can be obtained from furfural and may lead to elaboration of fully biobased PET (*14*). Recently, efforts have been initiated to convert greenhouse gases such as CO_2 into bioplastics like polypropylene carbonate (PPC). Renewable resource–based polymeric materials should be safe in production and can be recyclable, and their disposal after use should be harmless to the environment. The success of the polymer industry will depend on the development of such sustainable polymers (*15*). The development of polymers from sustainable renewable feedstocks poses key scientific challenges from a combined societal, economic, environmental, and human health perspective. The sustainability measurement is difficult to quantify. Although more energy is required to produce biobased low-density polyethylene (bio-LDPE) in comparison with its petroleum-based counterpart, its overall global warming potential from the life-cycle assessment perspective is relatively lower (*16*).

Renewable and sustainable resource-based biocomposites: Scientific challenges

Fiber-matrix adhesion, matrix and fiber modification, hybrid strategy, and the desired processing approach are key factors in making high-performance biocomposites for specific end-use applications. Across the wide spectrum of possible matrix and fiber/filler systems, hybrid synergistic assembly for improved compatibilization is a key scientific challenge. Biofibers are hydrophilic and thus have reduced compatibility with hydrophobic polymer matrices. A myriad of treatments—including chemical (e.g., silane maleic anhydride), mechanical (e.g., cutting, carding), physical (thermal, plasma, irradiation), and combinations of chemical, mechanical, or biological techniques—have been developed to tackle the biofiber drawbacks and improve compatibility with the matrix. PP and PLA are two examples of thermoplastic resin in plant fiber–based biocomposites uses. The former results in a partial biobased composite and the latter a fully biobased composite. In contrast to hydrophobic polymers such as PP, lignocellulosic natural fiber shows good compatibility with biodegradable polyester–type bioplastics such as PLA and PCL (*17*). Because the natural fiber adheres poorly to hydrophobic polymers such as PP and PE, it is necessary to treat the fiber and/or to use a coupling agent during reactive extrusion processing. The most commonly used coupling agent is maleic anhydride–grafted polypropylene (MAPP) for PP-based biocomposites. The optimal content of coupling agents on specific natural fiber–PP system governs the overall performance. In natural fiber thermoplastic composites, short fibers (<10 mm) are preferred. In twin screw extruders, the reaction takes place among fiber, plastic, and coupling agent, with processing and materials combination optimized for development of compatibilized composite pellets formulated for additional processing. Further processing may include injection molding into the appropriate shape for structure-property-processing tests or the final shape for industrial applications. Besides the correct material combination, the final properties of composites also depend on the processing method. Conditions such as screw profile and speed are among the vital determinants of final properties in a reactive extrusion process. In polymer composites, cost-performance targets are affected by both the matrix and the reinforcement options selected. A hybrid composite system requires a blend of two or more plastics and mixed fibers. Inclusion of recycled or undervalued waste materials further heightens the scientific challenges. Figure 3 highlights the key aspects of sustainable hybrid composite design. The successful compatibilization of polymer matrices that allows a good balance of properties can be achieved by using block copolymers as additives to act as physical compatibilizers. The illustrated example related to the PE/*i*PP (polyethylene/isotactic polypropylene) compatibilized tough blends (*18*) is particularly relevant because it allows the direct recycling of those materials from municipal wastes (Fig. 3A, i). The compatibilizer can also react, permitting the grafting into the polymer backbone or terminal group, as exemplified in PLA/PP blends with epoxy reactions (Fig. 3A, ii) (*19*). The synergistic combination of the formulation components is maximized by increasing their compatibility via ex situ modifications or in situ process engineering (Fig. 3B). Although the filler aspect ratio can be mechanically tailored, chemical modifications through silane treatment and maleation of the matrix and/or fiber are classical industrial strategies to improve their intrinsic reactivity. All of these strategic concepts are combined in the design of hybrid composites (Fig. 3C). A multifunctional compatibilizer is typically required because it can act physically and chemically at the interphase, linking the different modified matrix-filler and fiber phases.

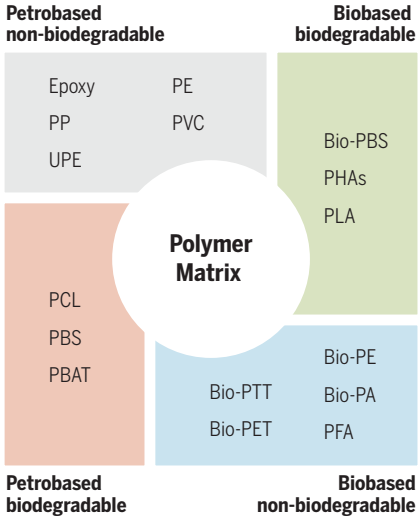


Fig. 2. Matrix polymers for biocomposites.

Renewable and sustainable resource-based biocomposites: Progress, recent developments, and future perspectives

Structure-property-processing correlation is important in research and development to find specific uses of composites. Biocomposites can be designed and engineered with various combinations of polymer matrix and reinforcing fibers and fillers from renewable and sustainable resources. The reinforcement could be in the form

of short fibers, knitted or woven fabrics or nonwoven fabrics, and braids. Various methods may be chosen for composite manufacturing, depending on materials and the targeted end-use application. For instance, short-fiber, particulate-type filler–reinforced thermoplastic biocomposites are manufactured through melt mixing with the use of extrusion and injection molding processes. Fabric thermoplastic–type reinforcements undergo thermoforming and compression molding–type processing. The resin infusion process is commonly adopted for thermoset resins. Hand-layup, vacuum-assisted resin transfer molding, and sheet molding compound techniques are some of the key processing methods for fabricating thermoset biocomposites. Depending on various combinations chosen from the range of fibers and resins systems available, the biocomposites can be classified

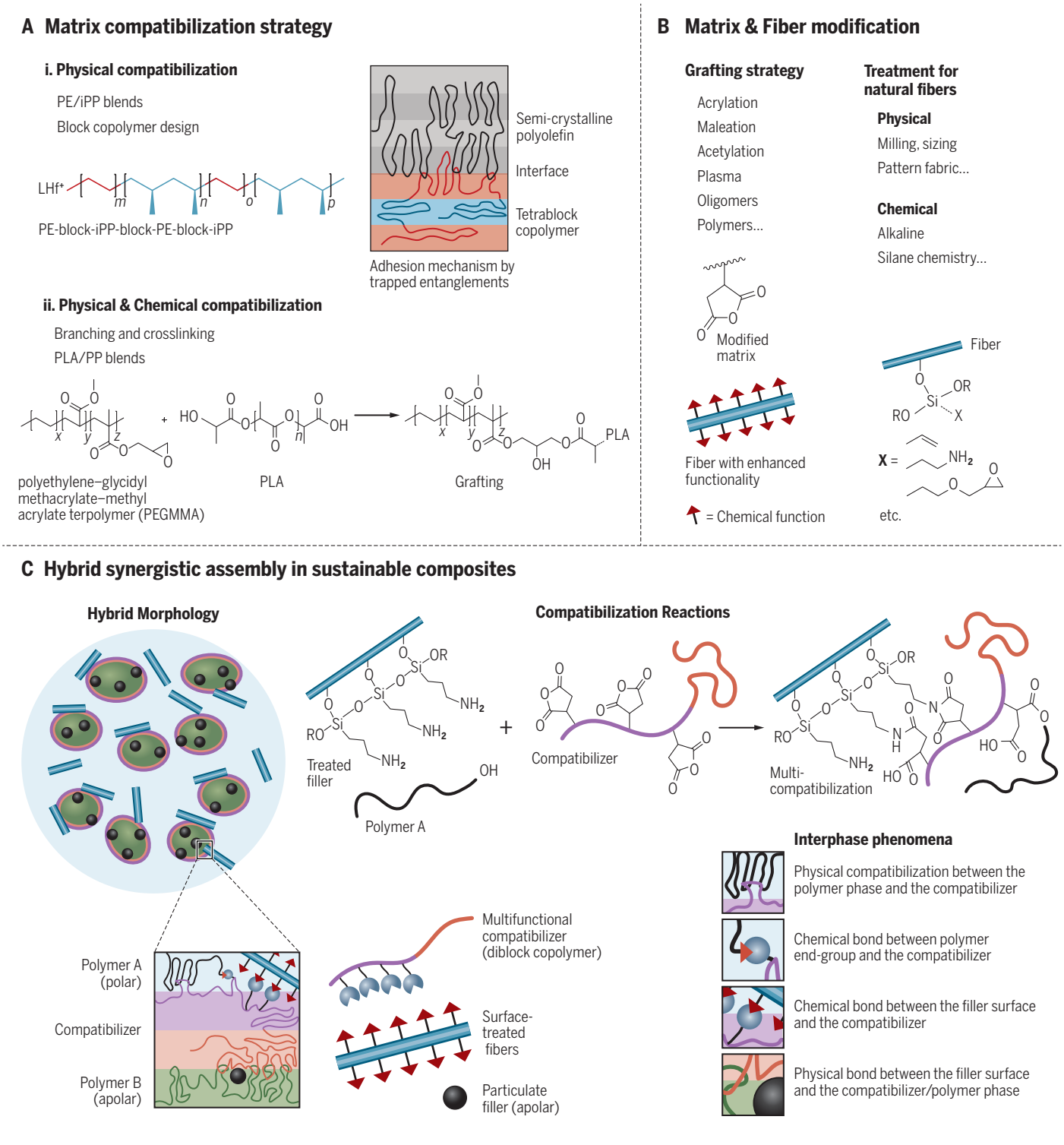
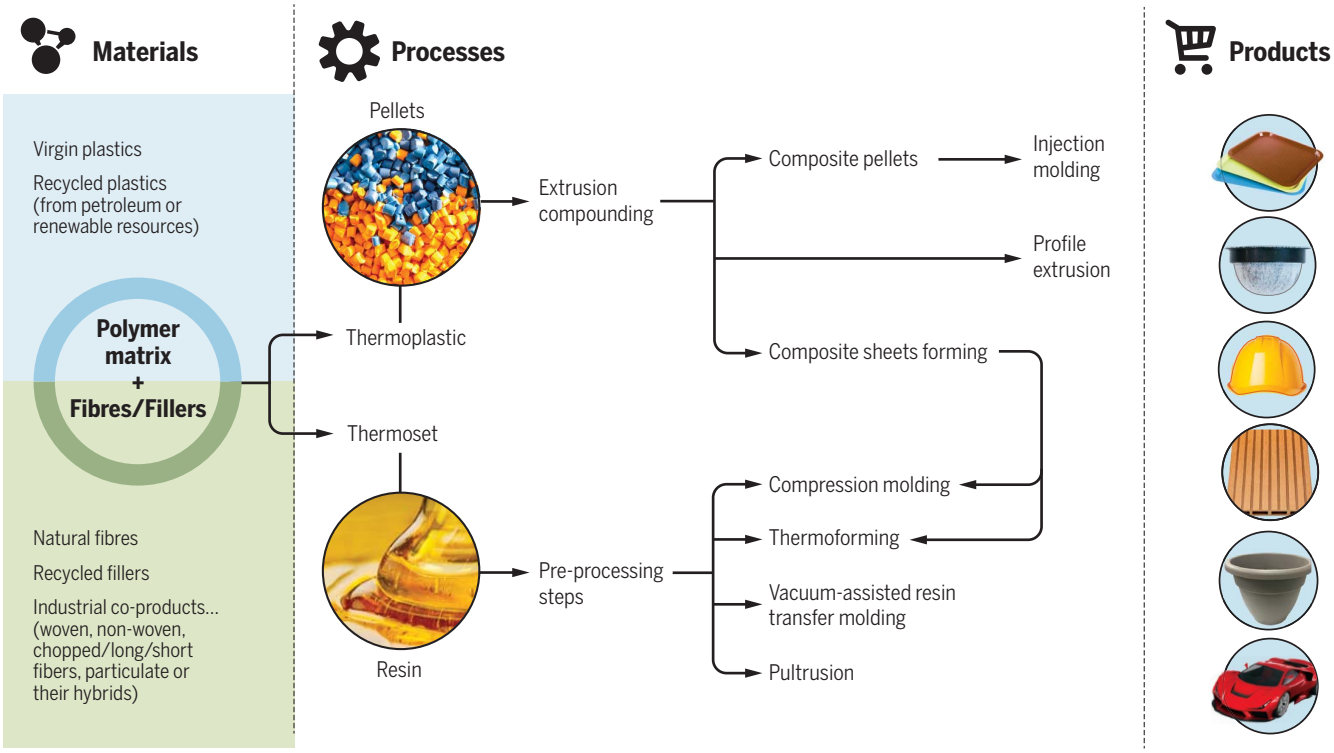


Fig. 3. Compatibilization strategies. The scheme in section i of (A) was redrawn from (*18*) with permission. The scheme in section ii of (A) was reprinted and adapted from (*19*) with permission; copyright (2015) American Chemical Society.

into three types: partially biobased, fully biobased, and hybrid biocomposites. In accordance with the selected polymer resin, the resulting composites may be thermoplastic or thermoset in nature. Figure 4 shows a schematic of biocomposite production from some representative raw materials and their processing in manufacturing, packaging, and consumer goods areas. Natural fiber composites are more environmentally friendly, economic, and lightweight than traditional glass or aramid fibers and talc-filled composites in both thermoplastic and thermoset platforms. They have many industrial uses, including applications in construction, automotive parts, and sporting goods. Biocomposites also have potential usage in electronics and specialty niche markets. A wide array of agro-forestry biofibers (predominantly wood, as well as flax, kenaf, and sisal) have been explored as natural reinforcement or fillers for composite fabrication. For auto manufacturers, one direct benefit of using sustainable composites is greater assurance of long-term price stability. Thus, many automakers continue to seek low-cost alternatives using nontraditional fibers and fillers and matrix systems including "waste" agro-forestry coproducts, waste rubber, cork, and recycling waste (*20*). Another example is the substitution of ground coconut shells (the major by-product of the coconut processing industry) for mineral-based talc. Natural fiber composites in auto parts have been thoroughly explored (*21*). Both reinforcement and matrices are being substituted with sustainable materials. Plastics originating from renewable resources (e.g., PLA and Bio-PA) are being studied for automotive applications. More attention is now being devoted to the use of recycled content with virgin plastics for the fabrication of composite materials. However, the broader acceptance of these biocomposite materials in automotive applications depends on many factors, including class A finishing, moisture repellence, structural stability, and flame-retardant properties. Sustainable composites are used in auto parts such as trim panels, seat backs, packaging trays, spare wheel covers, headliners, dashboards, and air-baffle components. Apart from automotive components, sustainable composites also receive considerable attention in construction and packaging applications. Tables 1 and 2 summarize a few representative biocomposites with their key properties and industrial uses. Needle-punch flax/PP mats on compression molding result in biocomposites with very high impact properties that may result from the aligned nature of the fibers, along with high fiber loading (50%) (*22*). The construction and municipal waste plastic-wood fiber–based biocomposites exhibited undesirable properties compared with virgin plastic LDPE-based wood-plastic composites (*23*). Multicomponent systems using such mixed plastic wastes (mostly PP and PE) deliver inferior performance due to incompatibility. Food wastes provide another potential sustainable resource for composite manufacture. The use of biofuel coproduct and food processing wastes as filler in biocomposites may help improve modulus, as compared with virgin plastic, while sacrificing toughness. The composite matrix can also be made from renewable resources. Bio-PU can be prepared from rubber seed oil monoglyceride reacting with diisocyanate (*24*). Combination of PLA and PBS, and, in a larger sense, biodegradable aliphatic polyesters with natural fibers, can lead to biodegradable formulations (*25*). Because of their woodlike aesthetic, hybrid biocomposite formulations containing natural resins with lignin and natural fibers were commercially successful when used in jewelry and musical instruments (*26*). Coffee chaff, a lignocellulosic waste from the coffee roasting industry, has been used to develop compostable biocomposites for a disposable food packaging application in coffee pods (*27*). Kiwifruit skin waste biomass and PLA, as well as grape pomace—biodegradable polyester-based biocomposites—resulted in compostable knives and clips, respectively, for industrial applications (*28*). DDGS, a corn ethanol coproduct, has been used as filler to increase the thermomechanical properties of PHA bioplastic and the biodegradation of the resulting composites, showing potential for disposable agricultural applications (*29*). Chicken feather fiber (CFF) waste from the poultry industry, with glass fiber and epoxy resin, resulted in lightweight hybrid biocomposites (*30*). As compared with glass fiber–epoxy composites, the CFF or CFF–glass fiber hybrid composites exhibited 30 to 40% density reduction because of internal void and lower aspect ratio of CFF as compared with glass fiber. The biocarbon from pyrolyzed biomass and waste resources has created opportunities to advance sustainable composites, ranging from their uses in commodity thermoplastic such as PP (*31*) to high-melting



IMAGES: (CLOCKWISE FROM TOP LEFT) ISTOCK/RAEVA. ISTOCK/HEINTEH. CLUBCOFFEE LP. ISTOCK/T_KIMURA. ISTOCK/BEN-SCHONEWILLE. A. K. MOHANTY. ISTOCK/THEBIAN. ISTOCK/ANTHIACUMMING

Fig. 4. Biocomposite product examples: From raw materials to manufacturing.

engineering plastic such as PET (*32*). Limitations of plant fiber in current biocomposite uses include unwanted odor, hydrophilicity, and low thermal stability. Because biofibers degrade around 200°C, their widespread use in engineering plastics is restricted. In the automotive industry, tires and polyurethane-based products contain CB filler up to 50 weight %. Again, most of the interior auto parts are made in black with the help of petro-based CB. Along with these components, the electronic housing products are also made with CB as a colorant. The pyrolysis process also results in bio-oil and syngas, which have been explored as raw materials for various grades of fuels and chemicals. In comparison with traditional natural fibers, biocarbon is thermally stable and is particularly suitable to reinforce or fill thermoplastic engineering plastics, like PET, for biocomposite uses. Traditionally associated with high strength and stiffness, the definition of advanced composites is evolving toward expression of multifunctionality that may include a combination of sophisticated mechanical properties with a high concentration of biobased content, light weight, and electrical and thermal conductivity. Appropriate functionalization of bioresin with the intended reinforcement can result in marked performance improvements. Highly functionalized resin derived from vegetable oil reinforced with fiberglass fabric resulted in advanced biocomposites (*33*). Lignin-based carbon fibers and nanocelluloses are considered the next generation of biobased reinforcements for sustainable composite applications. Carbon fiber is attracting attention for lightweight composite uses. Synthetic carbon fiber is made mostly from acrylonitrile. Lignin, being a renewable resource, has been heavily researched for the development of lignin-based carbon fibers (LCFs) (*34*). The diversity of lignin sources poses challenges in the manufacturing process. The pretreatment process and mechanical performance, among other aspects, must be further developed to bring more cost-competitive LCFs to commercialization. Nanocelluloses from agricultural and industrial waste have also attracted attention as sustainable materials (*35*). Nanofibrillated cellulose and cellulose nanocrystals are two key types of nanocelluloses. Polymer- and nanocellulose-based composites are under constant development (*36*). Melt processing of nanocellulose and thermoplastic has drawn more attention, and the major scientific challenges pertain to process development for improved dispersion of nanocellulose in the polymer matrix. Surface modification of nanocellulose is critical for improved fiber-matrix adhesion. Treatment of microfibrillated cellulose (MFC), a nanosized cellulose fiber, with silane changed its surface characteristics from hydrophilic to hydrophobic without affecting the crystalline structure (*37*), and such modification improved fiber-matrix adhesion in MFC-epoxy composites. Green composites have been developed using MFC-modified waxy maize starch–based bioresin and modified liquid crystalline cellulose through hand lay-up and compression molding. Such advanced composites exhibit very high strength (~800 MPa), thus creating new possibilities for structural applications (*38*). 3D printing, also known as additive manufacturing, can be widely used in sustainable composite manufacturing, especially in biomedical, automotive, and construction industries. 3D printing enables the creation of complex structures that cannot be created by traditional composite manufacturing technologies. A 3D printing technique has been developed for processing unidirectional continuous fiber–reinforced PLA composites with very high modulus and strength (*39*).

Challenges in adopting sustainable composites widely

More eco-friendly composites with enhanced sustainability face challenges to their wide-scale application. Measuring the sustainability of plastic and reinforcement/fillers is a complex task affected by factors such as the nature of the feedstock, energy input during production, durability, health impacts, and after-life recycling or disposal (*40*). Biomass supply chains, which address types of biomass, harvesting and collecting strategies, transport and storage mechanisms, as well as processing methodologies, are complex in nature and often vary with biomass type. It is necessary to establish a unified protocol for the effective utilization of bioresources, including waste resources. The sustainable method of expanding purpose-grown biomass, for example, requires the use of marginal agricultural land. Such an approach is essential for meeting the emerging massive requirement for biomass in the future. Durability is a critical test for any biocomposite material proposed for replacing traditional synthetic composite materials. To achieve functionality, biocomposite materials for automotive, construction, and other structural applications need to deliver the required service life and long-term durability. Inclusion of bioplastic and recycled materials in sustainable composite uses poses major scientific challenges. Designing and engineering new classes of biocomposite materials that can exhibit high tolerance against various external factors is essential. The classification of biodegradable and nonbiodegradable composites is also important from an application perspective. In addition to durable applications,

Table 1. Properties of representative biocomposites and their hybrids. –, not determined; MAPE, maleic anhydride–grafted polyethylene.

Resin	Filler	Impact strength	Tensile strength	Tensile modulus	Comments	Reference
Plastic waste (PE and PP)	Wood flour	2.9–6.2 kJ/m^2 Unnotch	6–13 MPa	2.3–3.9 GPa	MAPE compatibilization and lubricant utilization	(*23*)
PP	Wood, poultry litter biochar	8.1 kJ/m^2 Notch	27 MPa	4.3 GPa	Hybrid biocomposites–MAPP compatibilization	(*31*)
PP	Flax fiber	751 J/m Unnotch	40 MPa	6.5 GPa	Needle-punch fiber mat composite	(*22*)
Waxy maize starch	Neat and modified liquid crystalline cellulose, microcrystalline cellulose	–	505–790 MPa	22–32 GPa	Starch/cellulose hybrid biocomposites	(*38*)
Epoxy/acrylate	Glass fiber	237 kJ/m^2 Notch	532 MPa	37 GPa	Methacrylated epoxidized sucrose soyate resin/glass fiber	(*33*)
Bio-polyurethane (Bio-PU)	Sisal fiber	–	57–119 MPa	1.2–2.2 GPa	Rubber seed oil polyurethane	(*24*)
PBS/PLA	Flax fiber	9.1-17.8 kJ/m^2 Notch	39–55 MPa	3.6–7.4 GPa	Fully biodegradable composite	(*25*)
PLA	Carbon fibers, twisted yarns of jute fibers	–	57–185 MPa	5.1–19.5 GPa	Continuous fiber reinforcement probed by 3D printing	(*39*)

Table 2. Biocomposites in automotive, packaging, and other applications. TPE, thermoplastic elastomer; TPO, thermoplastic olefin; ABS, acrylonitrile butadiene styrene.

Manufacturer	Resin	Filler	Applications	Reference
Ford	PP	Coir fiber	Load floor, package shelf	(*20*)
	PP	Cellulose fibers	Console armrests	
	TPE	Powdered coconut shells, shredded tires	Structural guards	
	TPO	Powdered coconut shells, shredded battery cases, magnesium silica fibers	Rear deck lid applique brackets and side-door cladding	
Hyundai	ABS	Cork	Veneer	
Mercedes Benz	Epoxy	Flax/sisal fiber mats	Interior door panels	(*21*)
Toyota	Bio-nylon 6,10	Short glass fiber	Radiator end tank	
Volkswagen	Polyurethane	Flax/sisal hybrid mats	Door trim panels	
Mitsubishi Motors and Fiat SpA.	Bio-PBS	Bamboo fiber	Interior vehicle components	
Toyota - Lexus	Bio-PET	No information	Luggage compartment liner	
Club Coffee	Biodegradable polymer blend	Coffee chaff	Coffee pods	(*27*)
Competitive Green Technologies	PP	BioC	Additive/coloring agent	(*8*)
Zespri	PLA	Kiwifruit skin	Biodegradable spoon-knife	(*28*)
Scion	Aliphatic polyesters	Grape pomace	Biodegradable net clip for vineyard	
Tecnaro	Natural resins	Lignin and natural fibers	Construction, electronics, furniture, headphones	(*26*)

certain biocomposites are targeted for short life cycles. These materials must adhere to international standards for biodegradability and compostability. One challenge is having the required composting facility at the disposal site of such materials.

Outlook

Fossil fuel–based traditional composite structures persist in the environment. Because they are minimally recycled, these materials often end up being incinerated or placed in a landfill. In addition to fiber-reinforced composites, minerals such as talc and calcium carbonate–filled polymer composites and/or their hybrids with fibers are being used widely in composite industries. Hybrid biocomposites with petro- and biobased combinations, which are neither 100% fossil fuel–based nor 100% biobased, have achieved some commercial success. Wood and other agricultural natural fibers (flax, jute, etc.) used with petro-based plastics (PP, PE, epoxy, etc.) are more eco-friendly than 100% fossil fuel–based composites and have found use in housing structures, decking industries with wood plastic composites, other natural fiber–based hybrid biocomposites in automotive parts, and consumer products. Biocomposites from recycled fibers and natural fibers have also entered into consumer product applications. Currently, all-green (i.e., 100% bio-based) composites exhibit limited success because of their cost and durability restrictions in automotive and/or housing structures. All-green composites and biodegradable plastics are gaining momentum in sustainable packaging. The use of sustainable biocarbon fillers derived from waste biomass, industrial waste, and food waste demonstrates enormous potential for lightweight sustainable composites in auto parts and other growing demands from the manufacturing sector. Achieving increased utilization of wastes and undervalued industrial coproducts depends on creating a strong value proposition across the entire value chain. The economic and functional merits of composites made from renewable and sustainable resources must be coupled with leadership from industry executives and senior government officials to drive global growth in this innovative class of materials for positive societal, environmental, and economic impacts.

REFERENCES AND NOTES

1. *Nat. Mater.* **16**, 691 (2017).
2. *Nat. Mater.* **15**, 113 (2016).
3. United Nations Environment Programme, International Resource Panel, www.unep.org/resourcepanel (2018).
4. "The new plastics economy: Rethinking the future of plastics & catalysing action" (Ellen McArthur Foundation, 2017).
5. K. L. Pickering, M. A. Efendy, T. M. Le, *Compos. A* **83**, 98–112 (2016).
6. A. Bourmaud, J. Beaugrand, D. U. Shah, V. Placet, C. Baley, *Prog. Mater. Sci.* **97**, 347–408 (2018).
7. S. Vivekanandhan, N. Zarrinbakhsh, M. Misra, A. Mohanty, in *Liquid, Gaseous and Solid Biofuels-Conversion Techniques* (Intech, 2013), chap. 17.
8. A. Bali, M. Tiessan, *Bioplastic Mag.* **13**, 14 (2018).
9. W.-J. Liu, H. Jiang, H.-Q. Yu, *Chem. Rev.* **115**, 12251–12285 (2015).
10. R. Geyer, J. R. Jambeck, K. L. Law, *Sci. Adv.* **3**, e1700782 (2017).
11. G.-Q. Chen, M. K. Patel, *Chem. Rev.* **112**, 2082–2099 (2012).
12. European Bioplastics, "Bioplastics: Facts and figures" https://docs.european-bioplastics.org/publications/EUBP_Facts_and_figures.pdf (2018).
13. R. A. Sheldon, *Green Chem.* **16**, 950–963 (2014).
14. Y. Tachibana, S. Kimura, K. Kasuya, *Sci. Rep.* **5**, 8249 (2015).
15. D. K. Schneiderman, M. A. Hillmyer, *Macromolecules* **50**, 3733–3749 (2017).
16. C. Liptow, A.-M. Tillman, *J. Ind. Ecol.* **16**, 420–435 (2012).
17. A. Gallos, G. Paes, F. Allais, J. Beaugrand, *RSC Advances* **7**, 34638–34654 (2017).
18. J. M. Eagan *et al*., *Science* **355**, 814–816 (2017).
19. Y. Xu *et al*., *Ind. Eng. Chem. Res.* **54**, 6108–6114 (2015).
20. P. Malnati, "Recycled waste products get new life as lightweight, cost-effective auto parts," *Plastics Engineering* (2018); http://read.nxtbook.com/wiley/plastics_engineering/june_2018/index.html#recycled_waste_products_get_n.
21. O. Akampumuza, P. M. Wambua, A. Ahmed, W. Li, X. H. Qin, *Polym. Compos.* **38**, 2553–2569 (2017).
22. K. Oksman, *Appl. Compos. Mater.* **7**, 403–414 (2000).
23. I. Turku, A. Keskisaari, T. Kärki, A. Puurtinen, P. Marttila, *Compos. Struct.* **161**, 469–476 (2017).
24. I. O. Bakare, F. E. Okieimen, C. Pavithran, H. P. S. Abdul Khalil, M. Brahmakumar, *Mater. Des.* **31**, 4274–4280 (2010).
25. A. Bourmaud, Y.-M. Corre, C. Baley, *Ind. Crops Prod.* **64**, 251–257 (2015).
26. D. Kun, B. Pukánszky, *Eur. Polym. J.* **93**, 618–641 (2017).
27. S. Ackrill, "The simple solution that consumers want for single serve waste," *CoffeeTalk* (2018); http://coffeetalk.com/ctmagazine/05-2018/57083/.
28. F. H. Graichen *et al*., *Ind. Crops Prod.* **106**, 74–85 (2017).
29. S. A. Madbouly *et al*., *Green Chem.* **16**, 1911–1920 (2014).
30. M. Zhan, R. P. Wool, *J. Appl. Polym. Sci.* **133**, 44013 (2016).
31. O. Das, A. K. Sarmah, D. Bhattacharyya, *Waste Manag.* **49**, 560–570 (2016).
32. M. Idrees, S. Jeelani, V. K. Rangari, *ACS Sustainable Chem. Eng.* 10.1021/acssuschemeng.8b02283 (2018).
33. N. Hosseini, D. C. Webster, C. Ulven, *Eur. Polym. J.* **79**, 63–71 (2016).
34. W. Fang, S. Yang, X.-L. Wang, T.-Q. Yuan, R.-C. Sun, *Green Chem.* **19**, 1794–1827 (2017).
35. M. Rajinipriya, M. Nagalakshmaiah, M. Robert, S. Elkoun, *ACS Sustain. Chem. Eng.* **6**, 2807–2828 (2018).
36. K. Oksman *et al*., *Compos. A* **83**, 2–18 (2016).
37. J. Lu, P. Askeland, L. T. Drzal, *Polymer* **49**, 1285–1296 (2008).
38. M. M. Rahman, A. N. Netravali, *Compos. Sci. Technol.* **162**, 110–116 (2018).
39. R. Matsuzaki *et al*., *Sci. Rep.* **6**, 23058 (2016).
40. C. R. Álvarez-Chávez, S. Edwards, R. Moure-Eraso, K. Geiser, *J. Clean. Prod.* **23**, 47–56 (2012).

ACKNOWLEDGMENTS

Funding: This work is financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC)–Discovery Grants; the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), Canada/University of Guelph Bioeconomy for Industrial Uses Research Program; the Agriculture and Agri-Food Canada and Competitive Green Technologies through AgriInnovation Program; and the Ontario Research Fund, Research Excellence Program, Round-7 (ORF-RE07) from the Ontario Ministry of Research, Innovation and Science (MRIS), Canada. **Competing interests:** The authors have no competing interests.

10.1126/science.aat9072

REVIEW

Biological composites–complex structures for functional diversity

Michaela Eder, Shahrouz Amini, Peter Fratzl*

The bulk of Earth's biological materials consist of few base substances—essentially proteins, polysaccharides, and minerals—that assemble into large varieties of structures. Multifunctionality arises naturally from this structural complexity: An example is the combination of rigidity and flexibility in protein-based teeth of the squid sucker ring. Other examples are time-delayed actuation in plant seed pods triggered by environmental signals, such as fire and water, and surface nanostructures that combine light manipulation with mechanical protection or water repellency. Bioinspired engineering transfers some of these structural principles into technically more relevant base materials to obtain new, often unexpected combinations of material properties. Less appreciated is the huge potential of using bioinspired structural complexity to avoid unnecessary chemical diversity, enabling easier recycling and, thus, a more sustainable materials economy.

The challenges of life—from feeding to locomotion, communication, and protection—require materials and systems to fulfill a large variety of functions. This has been the driving force for the evolution of teeth, skeletons, claws, and sensory systems. Although natural materials have long been a prime source of inspiration for engineered ones, with our modern tools from nanoscience that allow inspection and construction at the scale of molecules, research in this direction is rapidly expanding (*1*–*5*). This may seem surprising at first glance, given human development of steel construction, copper wiring, and silicon chips. By contrast, the major parts of biological materials are based on just a few substances, polysaccharides (such as cellulose and chitin), proteins, and a few minerals (Fig. 1), which have considerably poorer properties than engineering materials. But the renewed interest in biological materials is precisely due to this fact: Natural organisms cannot rely on materials selection (*6*) to design a system; they have to generate it from the available basket of base materials. Hence, the diversity of properties and functions of biological materials is less due to a diversity of compositions than to a diversity of structures (Fig. 1). For example, the protein keratin is used in gripping and snapping tools (such as nails or beaks) and also for thermal insulation in wool; composites of protein and mineral build skeletons and teeth; and chitin in insects and arthropods is used to build carapaces and sensory systems and to give structural color.

A study of these materials leads to a number of questions: How does one get valuable functionality based on cheap base materials? How does one build a complex system based on little variation in constituents so as to facilitate recycling? Can we use structure to make materials active and adaptive? How does nature combine several functions in one material?

Answers to such questions require considering the function of biological materials in the natural environment to which they adapt. An evolutionary perspective can be particularly useful because it may reveal how evolutionary pressures led to the adaptation of material structure. As a consequence, bioinspired materials research is a relatively new field at the interface between materials science and various branches of biology, including functional morphology, evolutionary and developmental biology, and sensory physiology.

Multifunctionality of biological materials

Biological materials may need to simultaneously fulfill different functions to serve not only the needs of the (living) organism but also the needs of populations, such as growth, locomotion (*7*), signaling (*8*), repair, mechanical stability (*9*), resistance against light irradiation (*10*) or against low temperatures (*11*), and the possibility for functional adaptation (*12*). Because of this, Torquato *et al.* stated in 2003 that "the ultimate multifunctional materials are provided by nature" (*13*).

The potential for multifunctionality is inherently contained in the fact that properties are tuned through the internal structure. In particular, hierarchical structuring that is omnipresent in biological materials opens the possibility of adopting different physical properties in different size ranges (*14*). A typical example would be structural colors that require features at the submicron range corresponding to the wavelengths of visible light, whereas mechanical properties may be controlled at different scales. Most biological materials are based on fibers as primary motif. Bone is an excellent example in which collagen fibers and mineral are assembled into a great variety of structures that have different mechanical performances (*15*). Bone is

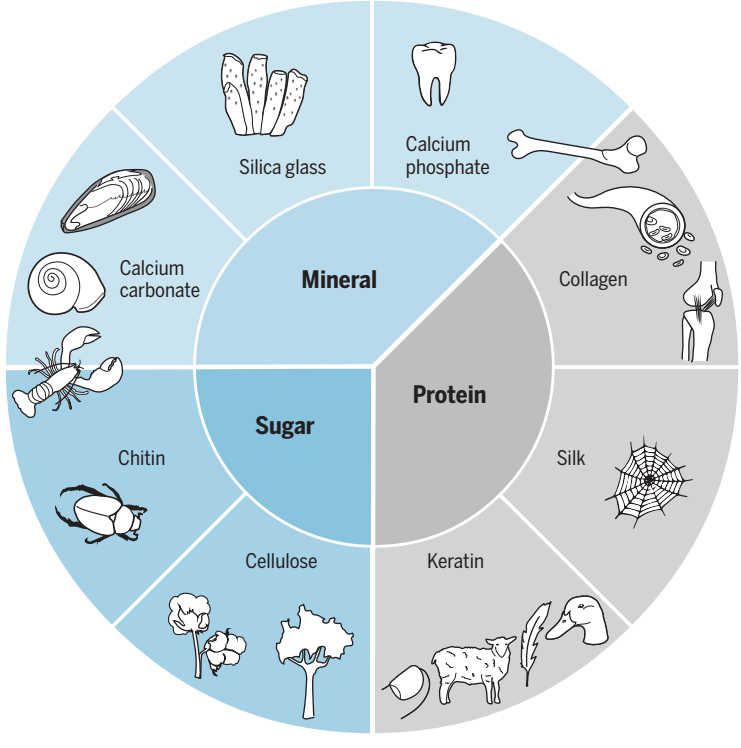


Fig. 1. Biological materials are built with a limited number of building blocks, based on polysaccharides, proteins, and minerals. A diversity of structures leads to a diversity of functions in tooth, bone, artery wall, tendon, spider web, beak, feather, wool, fingernail, tree, cotton, beetle carapace, lobster shell, snail shell, mussel shell, and the skeleton of the glass sponge (clockwise from top).

Max Planck Institute of Colloids and Interfaces, Department of Biomaterials, Research Campus Golm, 14424 Potsdam, Germany.
*Corresponding author. Email: fratzl@mpikg.mpg.de

not only the mechanical support for our body but also the reservoir for some of the most important ions for the functioning of cells: calcium and phosphate. More generally, hard-soft hybrid materials in the form of multilayers or tessellations allow the tuning of a wide range of properties without a change in composition (*16*).

The combination of toughness and strength is one aspect of mechanical functionality that has been studied extensively in biomaterials such as nacre and bone (*2*, *17*, *18*). However, biodiversity is huge, and many more examples of multifunctional biological materials have been and will be studied. Here we focus on three examples. The first is a thermoplastic protein that is manufactured into a graded porous material in the squid sucker ring. The second describes an example of serotiny in plants, in which a seed capsule stays closed for many years only to open and shed seeds during the first rain after a bush fire—all this by a cellulose-based composite without the help of living cells once the material is synthesized. The third discusses a range of organisms for which light manipulation is performed by the same material needed for mechanical protection or for water repellency.

Thermoplastic material with graded properties and strong grip—squid sucker ring teeth

Despite a fully proteinaceous structure, the toothed sucker rings of squid tentacles present remarkable physicochemical and thermomechanical characteristics, which, by far, transcend the properties of their building blocks (*19*, *20*). These gripping tools, which are used for piercing and anchoring of prey, attain their multifunctionality through one type of protein building block called "suckerins" and through their interactions into complex structural arrangements and architectures (*19*). The macroscopic geometrical arrangement of the sucker ring provides excellent stability during the lateral bending of these formidable teeth to facilitate an effective puncturing action actuated by their surrounding tentacle musculature (*21*). As shown in Fig. 2, sucker ring teeth (SRT) consist of an oriented tubular structure with graded pore size and pore volume fraction. Given the homogeneous molecular structure and in the absence of cross-linking or metal ions coordination (*22*), it is this graded distribution of porosity that provides the graded mechanical properties of SRT, which is a hallmark of many biological materials such as crustacean exoskeleton (*5*). In peripheral regions, a smaller diameter and lower pore fraction promote higher elastic modulus and hardness (*23*), which is required for gripping. A higher flexibility is obtained in the core zones by wider nanotubes and a higher pore fraction.

The SRT proteins self-assemble to form a block copolymer structure, whereby nanoconfined β sheets (Fig. 2) promote rigidity, and an amorphous matrix provides flexibility (*24*). Moreover, the isotropic orientation of the β sheets, which differentiate them from the aligned β sheets in spider silk, tune their load-bearing characteristic for compression and shear purposes. By contrast, the aligned β sheets of spider silk provide a superior elasticity for uniaxial tension (*25*). In the absence of covalent bonds, the suckerins self-assemble through hydrogen-bond interactions to form a robust but sufficiently dynamic supramolecular network (*24*). Consequently, this distinctive characteristic offers the possibility to provide self-healing, adhesive properties and high solubility and solution stability in weak acidic solvents (e.g., acetic acid), as well as thermoplastic properties (*19*, *20*), and nominates SRT as a model structure for a wide range of biomedical applications.

Because of this reversible network, the SRT protein structure can be thermally processed, molded, and reshaped with no or minimal changes in the intrinsic properties (*26*). This reversible protein network distinguishes SRT from related materials in the squid beak (*27*) or in insect exoskeletons (*28*). In such biological materials, the presence of a chitin-protein composite and interchain covalent cross-links prevent the proteinaceous matrix from any structural recovery, whereas the simplicity of the structural organizations at the molecular scale promote self-assembly and recovery in SRT. Such properties are attractive for many applications (Fig. 2), including biomedical applications such as drug delivery and bioadhesives (*19*, *20*).

Durable and multiresponsive encapsulation—*Banksia* seed pods

Many plants do not release their seeds straight after maturation but rather wait for conditions that will provide better chances for distribution and germination. The Australian plant genus *Banksia* stores seeds for up to 15 years in metabolically inactive closed pods (follicles; Fig. 3, A and B) (*29*) before increased temperatures, as caused by bush-fires, trigger seed release. The required opening temperatures are not only species dependent (*30*) but can also vary within species and adapt to geographic location and climatic

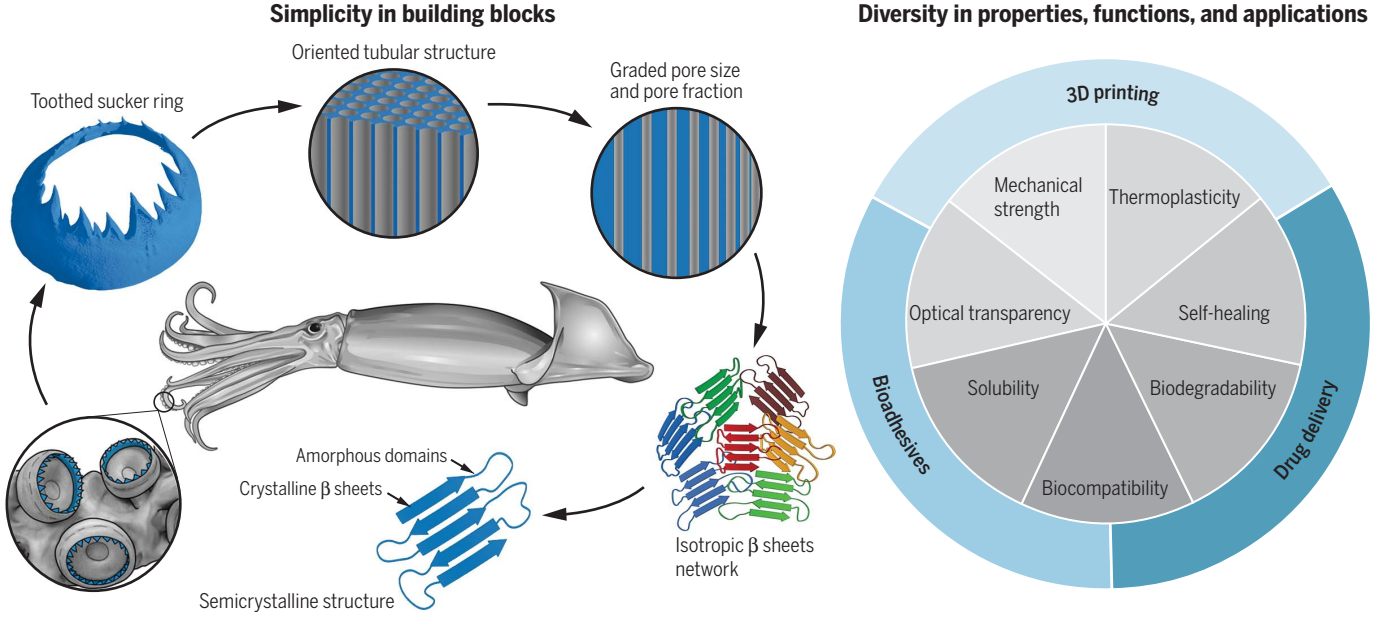


Fig. 2. Squid sucker ring teeth possesses a semicrystalline supramolecular network with variant physicochemical and thermomechanical characteristics. Although SRT are entirely made of a protein family called "suckerins," the different length scales in their tuned networks and structural arrangement result in a diversity of properties and applications. [Image credit: The schematic image of the squid is reproduced from (*21*) with permission from Mary Ann Liebert, Inc.]

conditions (*31*). Opening is initiated by increased temperatures, but the seeds are not released immediately (Fig. 3, C and G). Instead, they are held back until a species-dependent number of rainfalls triggers seed release (Fig. 3, D and H) in an environment conducive for germination (*32*). The seed pods, therefore, consist of a biological composite that allows for decade-long durability followed by a predefined deformation triggered and powered by two consecutive signals—heat and humidity—making it a small robotic device that can perform a simple task without an external energy source, triggered by a rare event. Thus, these seed pods may provide additional inspiration in the current quest for actuating soft robotic devices (*33*–*36*).

Durability for long-term seed storage and multiresponsiveness for opening and seed release are achieved by a structured functional composite made of cellulose, hemicelluloses, lignin, wax, and tannins. Each follicle consists of two pericarp valves in the centimeter range. The valves are multilayers with endo-, meso-, and exocarp, which have different cellulose orientations (*37*) that shrink by different amounts upon ripening and, thus, store internal stresses. They are connected by the junction zone (Fig. 3, B, C, and E)—a tissue with a high surface area of interdigitating cells—that is sealed with wax that melts at about 45 to 50°C (*31*, *38*), temperatures that can be reached on summer days in the field. These waxes are not related to seed pod opening but instead contribute to follicle integrity in creating a barrier against water loss and an antiadhesive film against wetting, insects, and microorganisms. The low melting temperatures may furthermore seal microcracks, which probably arise during long-term exposure to environmental challenges such as microbes, strong bird beaks, or weathering. In addition to the self-sealing waxes, condensed tannins in the mechanically weak parenchymatic tissue contribute to integrity by a high antioxidative capacity and increased water retention ability (*39*, *40*).

In *Banksia attenuata*, the initial opening temperatures and the dimensional stability depend on internal follicle geometry, in particular, on the radii of the biaxially curved follicle interiors (*31*), indicated by the white line in Fig. 3F. Opening at increased temperatures occurs by the softening of the endocarp (green layer in Fig. 3, F and G), which changes the internal force balance and allows the stored prestresses to be released (*31*) by the formation of a crack and an initial opening of the junction zone (Fig. 3C). Further opening for seed release (Fig. 3D) requires wetting and drying cycles, which activate bending of the endo-mesocarp bilayer (*32*) (Fig. 3H).

The *B. attenuata* example illustrates how adaptation strategies and (multi)functionalities in biological materials can be better understood by including ecological aspects in biomaterials research. Furthermore, it highlights the importance of sampling biological tissues in their natural environments and of relating their properties to the environmental challenges in those habitats. Comparative studies between species also increase our knowledge of biological material functionalities and possibly evolutionary processes. For example, a recent study of structural color in marine algae found similar optical structures for light harvesting and radiation protection in distantly related species (*10*). This example shows how biological materials science can help reveal new examples of convergent evolution, in which structures with similar form and function develop independently from each other. On the other hand, known examples of convergent evolution might help in understanding and revealing functionalities of biological materials.

Light-manipulating, superhydrophobic, antimicrobial, damage-tolerant mechanical devices—from insect wings to seashells

Numerous organisms manipulate light for harvesting and protection, communication, sensing, camouflage, or mate attraction in sunny uplands, dark forests, or marine ecosystems (*41*, *42*). To achieve and tune optical functionalities, nature has devised photonic structures which are tailored to manipulate light through optical phenomena such as specular reflection, diffraction, transmission, collection, or a combination of these effects (*41*, *42*). Photonic structures can interact with light to acquire structural colors by reflection (*43*), antireflectivity by diffraction (*44*), or optical transparency and light sensation mechanisms by collection and transmission (*45*, *46*). Probably the best-known examples of structural colors are insect cuticles and wings (*44*). The two-dimensional (2D) arrangements of multilayered cuticle in *Papilio palinurus* (*43*) and *Morpho* butterflies (*47*) are responsible for their iridescent green and blue colors, respectively. The iridescent green is achieved by continuous multilayered concavities that reflect a yellow light by a normal incident and a polarized blue light through a double reflection (*43*). By contrast, the iridescent blue in *Morpho* butterflies is formed by discrete multilayers of the cuticle (Fig. 4) (*47*). In addition to the coloration purposes, biological photonic structures have also been used for antireflective properties and optical transparency. The optical transparency of cicada wings and moth eyes attained by hexagon arrays of cone-shaped nanoprotuberances that provide a graded reflex index (*44*) aid in reducing reflection and camouflage and in efficient light harvesting.

Optical properties generated by structures offer the potential for combination with other functionalities. Numerous examples of this are found in nature (Fig. 4). The geometrical features and arrangement of the nanopillars in butterfly and cicada wings increase the surface roughness and enhance antiwetting properties. Consequently, the reduced wettability promotes superhydrophobic characteristics (*48*), resulting in self-cleaning and antibiofouling properties (*44*) such as those based on nanoprotuberance arrays of the cicada wing (Fig. 4). Furthermore, the presence of these nanopillar features on the wing surface prevents bacterial fouling, because the physical shape of the pillars helps rupture the cell membrane of adhering bacteria (*44*). Applications of the multifunctional photonic structures have not been limited to surface characteristics. Beyond the reflection-diffraction effects, the bulk photonic structures have been tuned to serve collection and transmission applications while integrating optical characteristics with mechanical properties. Such structures can be found in the integrated visualization system of chiton (*46*) or

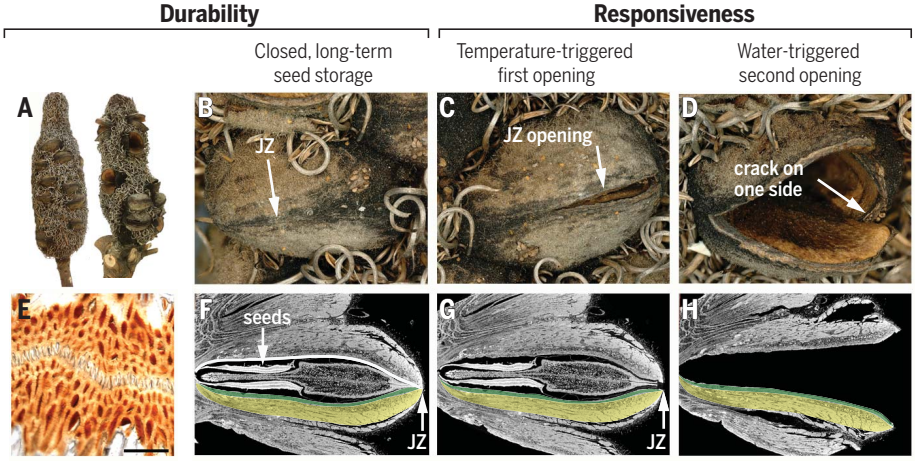


Fig. 3. Opening of *B. attenuata* infructescences (cones) collected in Western Australia. (**A** to **D**) Cones from the North [(A), left side] contain mainly closed follicles (B). Half-open (C) and open follicles (D) are frequently found on cones in the South, where opening temperatures are lower [right infructescence in (A)]. (**E**) Light micrograph of the junction zone (JZ) sealed with wax (scale bar 100 μm). (**F** to **H**) Virtual cuts through micro-tomographic reconstructions of closed (F), half-open (G), and open follicles (H) showing the seeds with the separator in between [(F) and (G)] and the endocarp-mesocarp bilayer (colored in green and yellow in one of the two pericarp valves). The white line in (F) indicates the internal valve curvature, which changes with geographic location and climate.

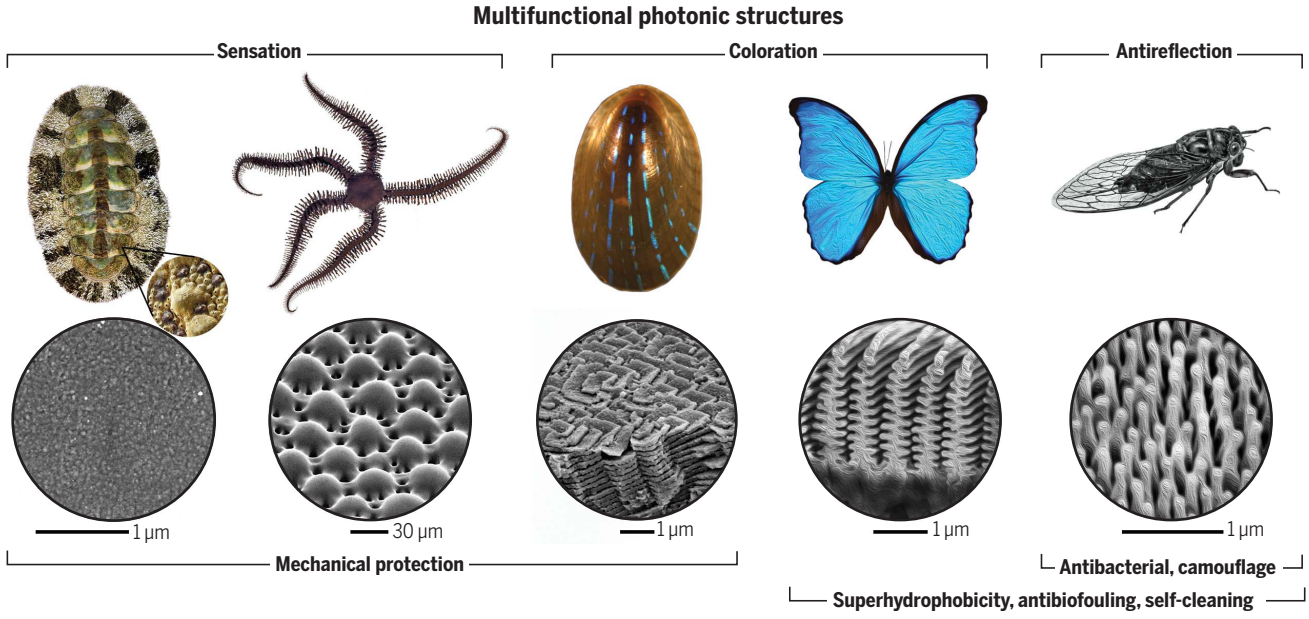


Fig. 4. Biological photonic structures with additional integrated functionalities. These photonic structures are honed to manipulate light for photosensation, coloration, or optical transparency purposes while achieving additional characteristics through the architectural arrangement of the same structural materials. The mechanical robustness of the chiton and brittlestar lenses (left) have been attained along with light transparency for photosensation mechanisms. The surficial architectures of the butterfly and cicada wings (middle and right) provide superhydrophobic surfaces and consequently promote other characteristics such as antibiofouling, self-cleaning, and antibacterial properties. [Image credits: From left to right, chiton, adapted from (*46*) and reprinted with permission from the American Association for the Advancement of Science; brittlestar, adapted with permission from Springer Nature (*45*); blue-rayed limpet, adapted from (*50*); morpho butterfly wing, adapted and reformatted from a micrograph by Shinya Yoshioka (Osaka University) and (*47*) by permission from Springer Nature; and cicada wing, adapted and reformatted from (*44*) and (*52*) by permissions from Annual Review, Inc., and IOP Publishing, respectively.

the light sensory system of the brittlestar (Fig. 4) (*45*, *49*). The chiton dorsal shell, which is used as a biomineralized armor for protection, contains translucent lenses dispersed across the shell plate. Similar to the shell structure, these microlenses are made of aragonite; however, the highly aligned crystalline structure of the lenses, as well as a lower amount of intercrystalline organic domains, minimizes light scattering (*46*). Similarly, the dorsal arm plates of brittlestars are covered with a 3D network of microlenses (Fig. 4), which are composed of single calcite crystals (*45*, *49*). These photosensory organs can collect and transfer the light to the dermal receptors while providing mechanical protection for the animal.

A prominent example of composite photonic architectures is the mineralized blue-rayed limpet shell (Fig. 4), which is tuned to attain mechanical robustness while achieving three differentiated optical properties: transparency, selective reflection, and absorption (*50*). In this shell, which is decorated with bright blue strips, the translucency is achieved through an irregular lamellar structure at the outer layer. By contrast, the selective reflection of blue-green light is created by means of a co-oriented calcite lamellae structure. Beneath this multilayer zigzag structure (Fig. 4), a disordered layer composed of amorphous colloidal particles absorbs the light to provide better contrast (*50*). These examples illustrate how nature develops site-specific design strategies to achieve the required functionalities and variant characteristics through manipulation of the architecture using available materials.

Outlook

These examples of biological materials systems mostly function without direct action of living cells. Hence, they are composite materials in the engineering sense, synthesized for specific functions that are important for the organism. These materials based on proteins and polysaccharides are not meant to last forever but rather to be reused, degraded, or even digested. For this to be possible, it is a great advantage that all materials of a given organism are of the same general type with additional functionality being introduced by structure rather than by the addition of materials with different chemical composition. Instead of durability, which is a hallmark of many engineering materials, the strategy pursued by nature is a material lifetime adapted to the function, followed by degradation and reuse (by the same or by different organisms). This is greatly facilitated by the fact that there is not too much diversity in the combination of materials (*51*), so that enzymes for degradation are readily available, for example. Hence, nature's paucity of base materials could actually be an advantage when we use this to inspire a new materials economy with fewer materials that are structured specifically for the function we need.

Multifunctionality blurs the distinction between material and device. A tree stem is clearly a material that can be used in building construction, but it is also a device that fulfills several functions for the plant, from mechanical support to water and nutrient transport. In similar ways, most of our organs, such as muscle or bone and even liver, might be considered functional devices or materials. When translating this into engineering, the same comment applies: Multifunctional materials need to incorporate knowledge about the desired function to be useful. Whereas nature grows materials, organs, and whole organisms from the base substances it can synthesize, today's engineering processes are based on a separation of tasks and knowledge fields, from chemistry, materials science, production engineering, and device development all the way to product design, with various selection and assembly processes along this chain. The advantages of a more integrated approach have been recognized in many areas, for example, in biomaterials development in which the compatibility and the communication of materials with cells and organs is an essential part of the materials research itself. Research on materials and on functional devices are traditionally close in the textile industry or architecture. In general, an increased interaction of materials scientists with product designers could reduce the distinction between material and functional device in all areas of engineering and, thus, provide a growing need for the development of multifunctional materials.

REFERENCES AND NOTES

1. J. Aizenberg, P. Fratzl, *Adv. Funct. Mater.* **23**, 4398–4399 (2013).
2. U. G. K. Wegst, H. Bai, E. Saiz, A. P. Tomsia, R. O. Ritchie, *Nat. Mater.* **14**, 23–36 (2015).
3. A. R. Studart, *Chem. Soc. Rev.* **45**, 359–376 (2016).
4. F. Barthelat, Z. Yin, M. J. Buehler, *Nat. Rev. Mater.* **1**, 16007 (2016).
5. Z. Liu, M. A. Meyers, Z. Zhang, R. O. Ritchie, *Prog. Mater. Sci.* **88**, 467–498 (2017).
6. M. F. Ashby, Y. J. M. Bréchet, D. Cebon, L. Salvo, *Mater. Des.* **25**, 51–67 (2004).
7. T. N. Sullivan, B. Wang, H. D. Espinosa, M. A. Meyers, *Mater. Today* **20**, 377–391 (2017).
8. E. Moyroud *et al*., *Nature* **550**, 469–474 (2017).
9. M. F. Ashby, L. J. Gibson, U. Wegst, R. Olive, *Proc. R. Soc. London, Math. Phys. Sci.* **450**, 123–140 (1995).
10. C. J. Chandler, B. D. Wilts, J. Brodie, S. Vignolini, *Adv. Opt. Mater.* **5**, 1600646 (2017).
11. E. Kuprian *et al*., *Plant Cell Environ.* **40**, 3101–3112 (2017).
12. R. Weinkamer, P. Fratzl, *Mater. Sci. Eng. C Mater. Biol. Appl.* **31**, 1164–1173 (2011).
13. S. Torquato, S. Hyun, A. Donev, *J. Appl. Phys.* **94**, 5748–5755 (2003).
14. R. Weinkamer, P. Fratzl, *MRS Bull.* **41**, 667–671 (2016).
15. N. Reznikov, J. A. M. Steele, P. Fratzl, M. M. Stevens, *Nat. Rev. Mater.* **1**, 16041 (2016).
16. P. Fratzl, O. Kolednik, F. D. Fischer, M. N. Dean, *Chem. Soc. Rev.* **45**, 252–267 (2016).
17. A. R. Studart, *Nat. Mater.* **13**, 433–435 (2014).
18. H.-L. Gao *et al*., *Nat. Commun.* **8**, 287 (2017).
19. S. H. Hiew, A. Miserez, *ACS Biomater. Sci. Eng.* **3**, 680–693 (2017).
20. A. Pena-Francesch *et al*., *APL Mater.* **6**, 010701 (2018).
21. K. Kumar *et al*., *Soft Robot.* **4**, 317–323 (2017).
22. E. Degtyar, M. J. Harrington, Y. Politi, P. Fratzl, *Angew. Chem. Int. Ed.* **53**, 12026–12044 (2014).
23. A. Miserez *et al*., *Adv. Mater.* **21**, 401–406 (2009).
24. P. A. Guerette *et al*., *ACS Nano* **8**, 7170–7179 (2014).
25. S. Keten, Z. Xu, B. Ihle, M. J. Buehler, *Nat. Mater.* **9**, 359–367 (2010).
26. V. Latza *et al*., *Nat. Commun.* **6**, 8313 (2015).
27. A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite, *Science* **319**, 1816–1819 (2008).
28. S. O. Andersen, *Insect Biochem. Mol. Biol.* **40**, 166–178 (2010).
29. B. B. Lamont, D. C. Lemaitre, R. M. Cowling, N. J. Enright, *Bot. Rev.* **57**, 277–317 (1991).
30. A. S. George, *Nuytsia* **3**, 239–473 (1981).
31. J. C. Huss *et al*., *Adv. Sci.* **5**, 1700572 (2017).
32. R. M. Cowling, B. B. Lamont, *Aust. J. Ecol.* **10**, 169–171 (1985).
33. B. Mazzolai, L. Beccai, V. Mattoli, *Front. Bioeng. Biotechnol.* **2**, 2 (2014).
34. W. Wang *et al*., *Nature* **559**, 77–82 (2018).
35. W. Hu, G. Z. Lum, M. Mastrangeli, M. Sitti, *Nature* **554**, 81–85 (2018).
36. S. Poppinga *et al*., *Adv. Mater.* **30**, e1703653 (2018).
37. A. B. Wardrop, *Aust. J. Bot.* **31**, 485–500 (1983).
38. J. C. Huss *et al*., *J. R. Soc. Interface* **15**, 20180190 (2018).
39. S. Quideau, D. Deffieux, C. Douat-Casassus, L. Pouységu, *Angew. Chem. Int. Ed.* **50**, 586–621 (2011).
40. W. Vermerris, R. M. Nicholson, *Phenolic Compound Biochemistry* (Springer, 2008).
41. F. P. Barrows, M. H. Bartl, *Nanomater. Nanotechnol.* **4**, 1 (2014).
42. S. Tadepalli, J. M. Slocik, M. K. Gupta, R. R. Naik, S. Singamaneni, *Chem. Rev.* **117**, 12705–12763 (2017).
43. P. Vukusic, J. R. Sambles, C. R. Lawrence, *Nature* **404**, 457 (2000).
44. G. S. Watson, J. A. Watson, B. W. Cribb, *Annu. Rev. Entomol.* **62**, 185–205 (2017).
45. J. Aizenberg, A. Tkachenko, S. Weiner, L. Addadi, G. Hendler, *Nature* **412**, 819–822 (2001).
46. L. Li *et al*., *Science* **350**, 952–956 (2015).
47. P. Vukusic, J. R. Sambles, *Nature* **424**, 852–855 (2003).
48. T. Darmanin, F. Guittard, *Mater. Today* **18**, 273–285 (2015).
49. I. Polishchuk *et al*., *Science* **358**, 1294–1298 (2017).
50. L. Li *et al*., *Nat. Commun.* **6**, 6322 (2015).
51. M. F. Ashby, *Materials and Sustainable Development* (Butterworth-Heinemann, Oxford, 2015).
52. G. Xie *et al*., *Nanotechnology* **19**, 095605 (2008).

ACKNOWLEDGMENTS

The authors are grateful for many discussions about active materials, particularly with W. Schäffner, K. Krauthausen, and M. Friedman from Humboldt-Universität zu Berlin and with J. Dunlop from the University of Salzburg, Austria. We thank J. Blumenthal, Weissensee School of Arts and Design, for drawings in Fig. 1. **Funding:** Partial support was provided by the DFG Cluster of Excellence "Image-Knowledge-Gestaltung" (DFG-EXC 1027). P.F. was also supported by a Leibniz Award of the DFG. **Competing interests:** None declared.

10.1126/science.aat8297

REVIEW

Composites with carbon nanotubes and graphene: An outlook

Ian A. Kinloch[1], Jonghwan Suhr[2], Jun Lou[3], Robert J. Young[1], Pulickel M. Ajayan[3]*

Composite materials with carbon nanotube and graphene additives have long been considered as exciting prospects among nanotechnology applications. However, after nearly two decades of work in the area, questions remain about the practical impact of nanotube and graphene composites. This uncertainty stems from factors that include poor load transfer, interfacial engineering, dispersion, and viscosity-related issues that lead to processing challenges in such nanocomposites. Moreover, there has been little effort to identify selection rules for the use of nanotubes or graphene in composite matrices for specific applications. This review is a critical look at the status of composites for developing high-strength, low-density, high-conductivity materials with nanotubes or graphene. An outlook of the different approaches that can lead to practically useful nanotube and graphene composites is presented, pointing out the challenges and opportunities that exist in the field.

Carbon nanomaterials have had an unprecedented impact over the past three decades in defining the reach and applications of nanotechnology. Starting with the discovery of fullerenes and moving through the carbon nanotube (CNT) era to graphene and other two-dimensional (2D) materials, the academic world has been flush with new ideas, inventions, and innumerable attempts to find the killer applications for these remarkable nanostructures. Here we will discuss one such application, which when first introduced seemed close to embracing these materials but for various reasons has not met expectations: that of composite materials. In particular, the question of whether structures such as carbon nanotubes and graphene, touted as ideal reinforcements in composite matrices because of their mechanical properties, really are the right choices for mechanical reinforcement still remains largely unanswered. Moreover, the realm of functionalities that can be accessed by introducing nanotubes versus graphene in a matrix needs careful evaluation and thought. Although both are sp^2 allotropes, the structure, morphology, and dimensionality of these two profoundly interesting carbon nanostructures are indeed quite different, and so is the nature of their interactions with the adjacent matrix (*1*). Hence, the overall composite mechanical behavior provided by these two reinforcement units could be distinct. It would indeed be useful to have selection rules that choose one over the other in composite applications, but no such rational approaches have emerged in the design of nanocomposites with CNT or graphene phases.

[1]National Graphene Institute and School of Materials, University of Manchester, Manchester M13 9PL, UK. [2]School of Mechanical Engineering and Department of Energy Science, Sungkyunkwan University, Suwon, 16419, South Korea. [3]Department of Materials Science and NanoEngineering, Rice University, Houston, TX 77005, USA.
*Corresponding author: Email: ajayan@rice.edu

We will consider three important questions that may broadly define the fate of this field. Why should these unique structures be of any interest in composites, why have they not produced substantial progress after much effort, and what can be done to make them work as good reinforcements in composites? The discussion will go beyond just that of mechanical properties and also consider their excellent electrical and thermal properties.

Graphene represents the thinnest possible atomically flat layer, made of a planar hexagonal honeycomb lattice of strong C-C bonds that build up graphite when stacked via weak van der Waals interlayer forces. Graphene has a Young's modulus near 1 TPa in the plane, reflecting the intrinsic carbon bond stiffness, but being very thin, it can be flexible in bending, twist, and other deformation modes (*2*). The in-plane electrical and thermal conductivities are the highest among known materials, but the through-thickness properties of stacked graphene are very poor. CNTs are rolled-up graphenes with the in-plane properties translated to axial properties, making them among the stiffest axial fibers ever created. Similar to graphene, they can also be easily bent, twisted, and buckled (*3*). As graphene stacks up to make multilayer structures, CNTs can also have a nested structure of tubes inside tubes [single-walled (SWNTs) to multiwalled nanotubes (MWNTs)], which has a notable effect in determining their mechanical properties. In general, although the local stiffness that can be measured is extremely high, owing to a near defect-free structure of graphene and nanotubes, a couple of major issues make reinforcement of these structures in matrices challenging. First, both graphene and nanotubes are particulate fillers, with their larger dimensions (lateral size of graphene or lengths of nanotubes) reaching several hundred micrometers at the most or, in exemplary cases, millimeters. Short fibers are generally poor load carriers in fiber composites, and this effect is clearly seen when composites are made with CNT or graphene dispersions. Second, the surfaces of both graphene and nanotube are atomically smooth, devoid of any dangling bonds or defects (except at the edges in graphene or tips in nanotubes), which means that strong matrix-filler bonds are hard to accomplish, leading to poor interfacial load transfer during mechanical deformation (*1*) and high electron and phonon scatter, compromising electrical and thermal properties. This interfacial problem was a major roadblock in carbon fiber composites before industry figured out the sizing of fibers via chemical modification, but for nanotubes and graphene, this problem is severe, and attempts to chemically functionalize CNT or graphene surfaces may substantially compromise their intrinsic properties (*4*). This compromise also relates to a third issue, which is the inhomogeneous dispersion of nanotubes and graphene in the matrix. Without proper surface treatments, CNTs or graphene tend to aggregate easily, owing to strong van der Waals interactions between them, to form poorly dispersed bundles or agglomerates in the matrix, which often leads to poor interfacial connectivity and formation of mechanical stress concentration or other functionally singular sites, resulting in severely affected composite performances. Noncovalent functionalization methods (*5*) could be used to partially overcome the dispersion challenge, but this approach is ineffective in solving the interface problem. Therefore, a systematic and careful engineering approach is needed to design CNT or graphene composites with optimal performances (Fig. 1). We can show the potential for using CNT or graphene composites for applications by plotting the hypothetical position of different CNT or graphene composites in a modified Ashby plot (Fig. 1B) and the position of future continuous CNT or graphene fiber composites. Finally, the ideal dispersion of material throughout the composite is very application dependent; typically for mechanical properties, one wants as high a loading of the filler as possible that is aligned in the direction of the load, whereas for electrical percolation, one aims for a random percolated network with as low a concentration as possible. For multifunctional applications, the required microstructures may therefore even be contradictory.

Carbon nanotube and graphene fillers, interfaces, and load transfer

A great deal of effort has been made to develop lightweight, strong composite materials with CNTs and graphene as reinforcement, and although these are considered to be discontinuous short fillers, they possess outstanding mechanical properties. The extremely high Young's modulus of CNTs and graphene, their nanoscale dimensions, along with particular geometries that offer high specific surface area, present unprecedented opportunities to efficiently tailor the interface properties between the reinforcements and composite matrices. CNT or graphene nanocomposites may not be as strong or stiff as a continuously reinforced composite such as typical carbon fiber laminates that are currently

used in primary load-carrying structural applications. However, if the extraordinary potential and multifunctionality of CNTs and graphene are fully realized and their nanocomposites properly designed, they might still become game-changing composite materials. In addition to the challenges of dispersion, viscosity control, and sizing of the CNTs and graphene without compromising intrinsic properties, challenges still exist in accurately characterizing the reinforcement's mechanical properties. These properties include Young's modulus and strength, and also the interfacial shear strength at the filler-matrix interface, which will determine the critical length of the fillers required for the most efficient load-transfer capability. Although these factors are critically important in designing short-fiber composite systems, unfortunately most attention seems to be focused on improving practical issues such as dispersion in matrix materials and increasing the loading fractions of these fillers without suffering viscosity-related processing issues.

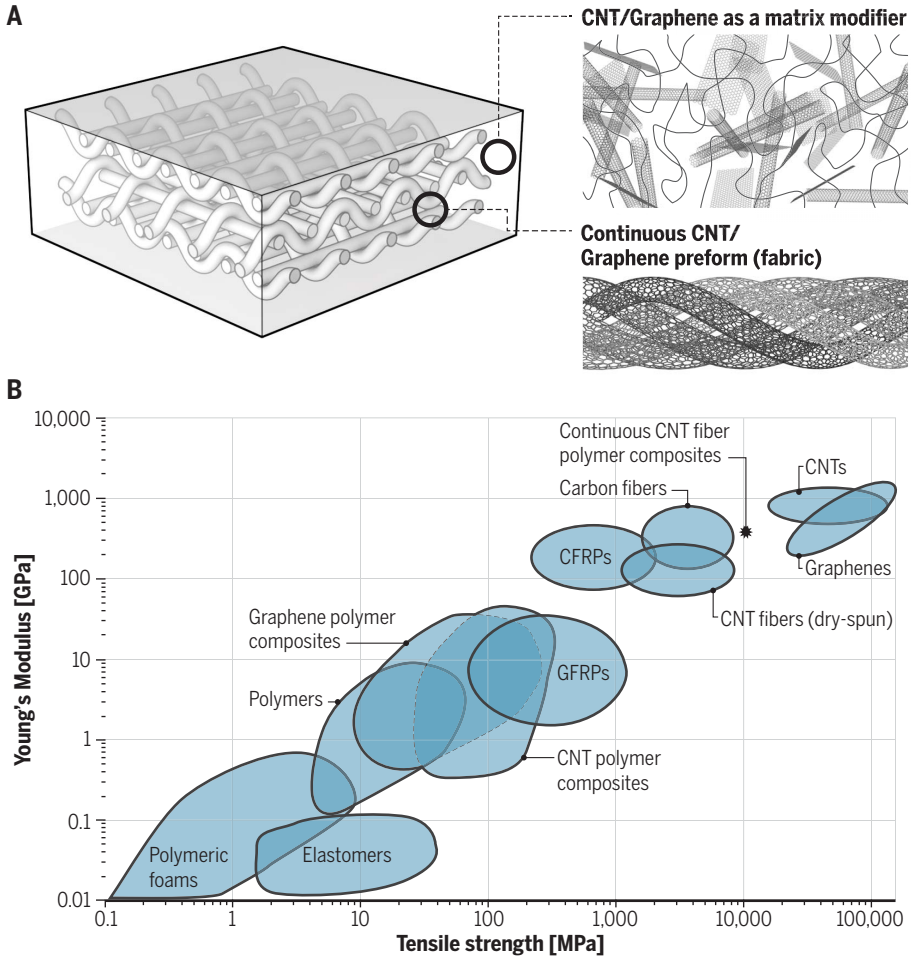


Fig. 1. (A) A schematic illustration of a futuristic CNT or graphene polymer composite that consists of continuous CNT fiber preform (fabric) in a polymer matrix and chemically modified CNT or graphene as matrix modifiers. The continuous CNT or graphene spun fibers can exceed the mechanical properties of carbon fibers that are used for state-of-the-art structural reinforcements in composites. The best load-carrying capability ever reached yet could be achieved by engineering the continuous CNT or graphene fiber preforms in composite design. It can be envisioned that the multifunctional CNT or graphene composites are realized by marrying the best features of the continuous CNT or graphene fibers and the modifiable polymer matrix with chemically functionalized CNT or graphene particulates. (**B**) Ashby plot of Young's modulus plotted against tensile strength comparing the mechanical properties of conventional polymer composites, including glass fiber–reinforced plastic (GFRP) and carbon fiber–reinforced plastic (CFRP), with CNT or graphene–based polymer composites. Thick outlines represent families of each material. In this projected Ashby plot, both mechanical properties of CNT or graphene polymer composites, which are considered to be particulate composites, are scattered around in between those of GFRPs and polymers. Those properties are shown to be below the ones of CFRPs that are continuously reinforced composites conventionally used for load-carrying structures. However, the mechanical properties of CNT spun fibers can indicate that CNT or graphene fibers can replace carbon fiber and, if properly designed and manufactured, the continuously reinforced CNT or graphene spun fiber composites might be used in making ultralight yet superstrong structures in the near future. The Ashby plot was made by taking many reported values from the literature and also our estimated scenario (e.g., for the CNT fiber composites).

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

A substantial challenge has been the proper characterization of fillers and filler-matrix interfaces during loading and assessing the efficiency of load transfer in composites. Raman spectroscopy has been the most powerful technique for both the characterization of carbon-based nanomaterials and the assessment of their mechanical properties (*6*). Raman spectroscopy can also be used to evaluate their mechanical properties by observing the shift of the Raman bands when the materials are subjected to deformation (see Box 1 and Fig. 2).

In polymer composites, we are normally concerned with the matrix-reinforcement interface. However, in layered reinforcements, it is also necessary to consider the van der Waals bonding between the walls of carbon nanotubes and the individual layers in multilayer graphene. This bonding is relatively weak compared with the strong covalent bonding within the graphene layers. When MWNTs and multilayer graphene are used in composites, their ability to reinforce is therefore limited by easy shear between the walls or layers, respectively. It is possible to track internal stress transfer between the walls of carbon nanotubes and between the layers of graphene from stress-induced Raman band shifts. Imperfect stress transfer is manifested as Raman band broadening during deformation and a lower Raman band shift rate compared with the single-walled or monolayer material. Comparing Raman band shifts under stress for SWNTs and MWNTs in epoxy nanocomposites shows that interwall stress transfer efficiency is only around 70% for MWNTs (*7*). Lower Raman band shift rate in multilayer graphene obtained with increasing number of graphene layers shows a similar trend to that in MWNTs. Raman spectroscopy has been effective in evaluating the efficiency of stress transfer at the interface before and after chemical modification of CNT or graphene surfaces. Several chemical modification schemes have been proposed (*1*, *8*, *9*) (Fig. 3), which improves the interfacial strength between CNT or graphene fillers and the polymer matrices.

Properties of carbon nanotube and graphene composites

The simplest way to look at the effect of nanofiller addition to a matrix is from the resulting mechanical properties of the nanocomposites, most commonly assessed through stress-strain curves (Fig. 4A). A telling example is natural rubber, which is a relatively strong elastomer with an extension to failure of ~1000%, where the addition of both the MWNTs and graphene leads to a major change in the stress-strain curves and a large increase in the stiffness even

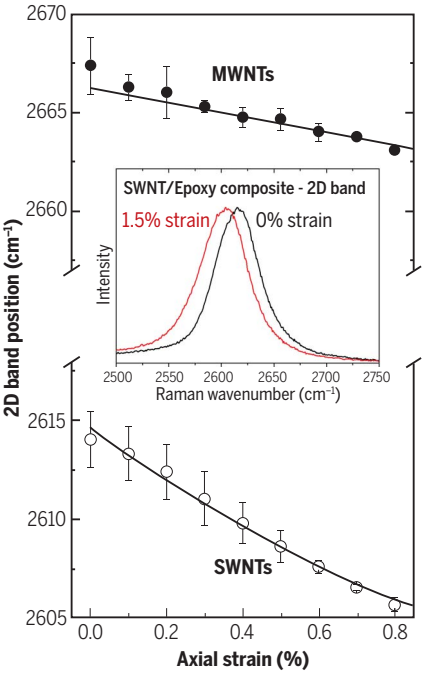


Fig. 2. Dependence of the 2D Raman band position upon strain during the deformation of CNTs in an epoxy resin matrix composite (0.1 wt %). The higher slope for the SWNTs reflects their higher Young's modulus. (**Inset**) The shift of the 2D band for the SWNTs. [Reprinted from (*7*) with permission from Springer Nature]

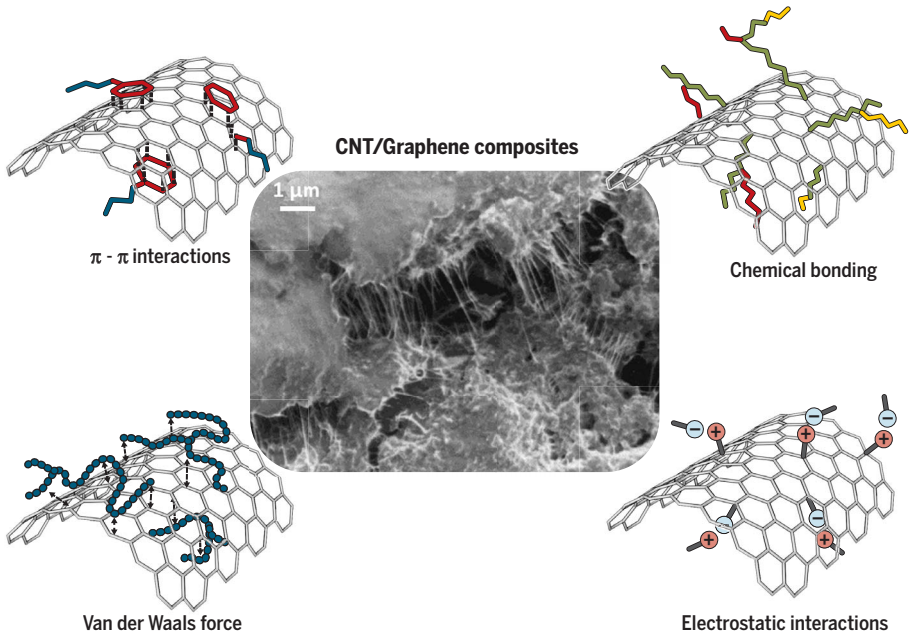


Fig. 3. A schematic representation of four different chemical modification schemes for CNT or graphene interacting with polymer matrix. Shown are π-π interactions (i.e., noncovalent interactions), chemical bonding, van der Waals force, and electrostatic interactions. (**Inset**) A scanning electron micrograph (SEM) image of the fracture surface of SWNT epoxy composites showing good dispersion and pullout of the nanotube bundles from the matrix (*9*). CNT or graphene surfaces with high specific surface area are favorable for chemical interactions with other molecules, which can allow for several chemical modification schemes. The ends of CNTs and edges of graphene facilitate additional chemical interactions with the polymer matrix. The chemically modified CNT or graphene surfaces can promote more uniform dispersion, control loading fractions of CNT or graphene without suffering high viscosity, and improve interfacial strength between the fillers and polymer matrix in the composites. Although the nanoscale dimensions present a formidable change, it is vital to understand the interactions between the CNT or graphene fillers and matrix toward optimum performance in composite design. In particular, the characterization and quantification of the interfacial shear strength will be essential.

Box 1. Raman spectroscopy and CNT or graphene composites.

sp^2 allotropes of carbon have electronic structures that lead them to undergo very strong resonance Raman scattering so that well-defined-spectra can be obtained even from individual CNTs or single layers of graphene (*16*, *38*). These structures show characteristic G and 2D bands in their Raman spectra, and where there are internal or edge defects, a D band may also be seen. Subtle differences between the spectra of SWNTs, DWNTs (double-walled carbon nanotubes), and MWNTs and between those of mono-, bi- and multilayer graphene enable the spectra to be used as fingerprints of the materials. The Raman bands shift during deformation for both CNTs and graphene (*6*). Such downshifts during tensile deformation are large and relatively easy to measure and enable one to monitor the deformation of CNT or graphene in nanocomposites inside a polymer matrix (*6*). The shift rates (in cm^{-1}/% strain) are found to scale with the Young's modulus of the nanocarbons, which enables such band shifts to be used as a universal stress sensor for composite mechanics (*39*).

at relatively low loading of the fillers (*10*, *11*). In both cases, however, the strength of the rubber is not increased, and the strain to failure is even reduced. The effectiveness of the reinforcements can be best assessed by plotting Young's modulus of the nanocomposite, E_c (as the stress at 100% strain for rubbers with nonlinear stress-strain curves), against the filler loading, and an approximately linear increase in E_c with filler volume fraction is seen in both cases. The effect of adding carbon black (commonly used nanofiller) shows a much smaller increase in the modulus of rubber (about a factor of 3 lower than for graphene).

The mechanics of deformation of fiber-reinforced composites are now well established, but this is not the case for nanocomposites. An obvious way forward is to adapt the continuum mechanics theories developed for macroscopic composites, such as those reinforced by carbon or glass fibers, to composites at the nanoscale. Some researchers have taken an opposing viewpoint and have suggested that polymer nanocomposites are actually quasi-homogeneous molecular blends that should be regarded as molecular composites or self-reinforced composites (*12*, *13*) and that the properties are controlled by interactions on the molecular scale between the nanoparticles and the polymer matrix.

One issue that arises is the difficulty of realizing the promised high stiffness of the nanotubes or graphene in polymer-based nanocomposites. The effective stiffness of the fillers can be assessed by using the rule of mixtures, and the experimentally derived stiffness of CNT or graphene fillers in composites is almost always observed to be far below the anticipated ~1 TPa modulus. In the case of CNTs, realization of the high stiffness only occurs when the CNTs are extended and aligned in polymer fibers or tapes (*14*). Values approaching ~500 GPa have been found, and this can be rationalized by using a modified rule of mixtures that takes into account filler size (length of CNTs and diameter for graphene flakes) and orientation (*15*):

$$E_c = E_{eff}\eta_o\eta_l V_p + E_m(1 - V_p) \quad (1)$$

where η_o is the Krenchel orientation factor and η_l is the length factor. The parameter E_{eff} is the effective Young's modulus of the filler that depends only on its structure. For example, it will

be ~ 1 TPa for monolayer graphene but is lower for few-layer graphene and graphene oxide (*16*). A similar reduction in E_{eff} occurs when going from SWNTs to MWNTs as the result of inter-wall sliding. The orientation factor η_o is 1 for aligned CNTs and flakes but much smaller for randomly oriented CNTs. Hence, the mechanical properties of CNTs are more sensitive to tube orientation, and randomly oriented graphene flakes offer better reinforcement than randomly oriented CNTs. The length factor η_l ($0 \leq \eta_l \leq 1$) reflects the efficiency of stress transfer from the matrix to the filler that is controlled by both the geometry (e.g., aspect ratio) of the filler and the strength of the filler-matrix interface. It approaches unity for long CNTs or large-diameter flakes with strong filler-matrix interfaces. The high values of E_{eff} found for aligned CNTs in fibers and tapes are therefore the result of long, extended CNTs with a high degree of alignment.

The very high strengths reported for CNTs and graphene are also not realized in nanocomposites. The celebrated experiment (*17*) that measured a strength of 130 GPa for graphene ("200 times stronger than steel"!) has been confirmed theoretically, but the experiment involved the nano-indentation of a monolayer graphene membrane and the deformation of less than a few thousand carbon atoms. Single-nanotube deformation experiments involve similar numbers of atoms. The strength of all materials falls as their size increases owing to the presence of defects, which is one reason why high strengths are not found in macroscopic nanocomposite specimens. Backing out the effective strength of the filler from the fracture strength of the nanocomposite is possible, using a rule of mixtures equation for fracture analogous to that for stiffness. Modest values of strength are generally found in randomly aligned nanocomposites, but values as large as 88 GPa have been reported for SWNTs in highly aligned polymer tapes (*18*). Strength values of ~10 GPa extracted from fracture data for graphene nanocomposites (*19*), although being well below the quoted 130 GPa graphene strength value, compare favorably with that of carbon fibers. Dispersion of the nanofillers also plays a major role in their effective strengths in nanocomposites, with high values only found for uniform dispersions.

In addition to being processed into extended structures for potential high-strength and high-stiffness composites, CNTs and graphene may best be applied in their particulate state as dispersed matrix modifiers or fiber modifiers in composites. The CNTs or graphene dispersed polymer composites exhibit time-dependent viscoelastic behavior under mechanical loadings, owing to the nature of polymers experiencing molecular rearrangements to release induced local stress or strain as a time-dependent behavior. The presence of CNTs or graphene can affect the polymer chains' mobility particularly in the vicinity of the reinforcements in composites. The viscoelastic behavior of nanotube and graphene composites has been reported, leading to strong matrix-modifying effects of these fillers. With only 0.1 weight % (wt %) of graphene platelets (GPLs), epoxy-GPL composites creep substantially less than SWNT-epoxy and MWNT-epoxy composites at the same loading fraction, mainly because of stronger interaction between graphene and epoxy. Functionalization of graphene (e.g., oxidation) can further promote better interaction with polymer matrix (*20*), leading to a higher glass transition

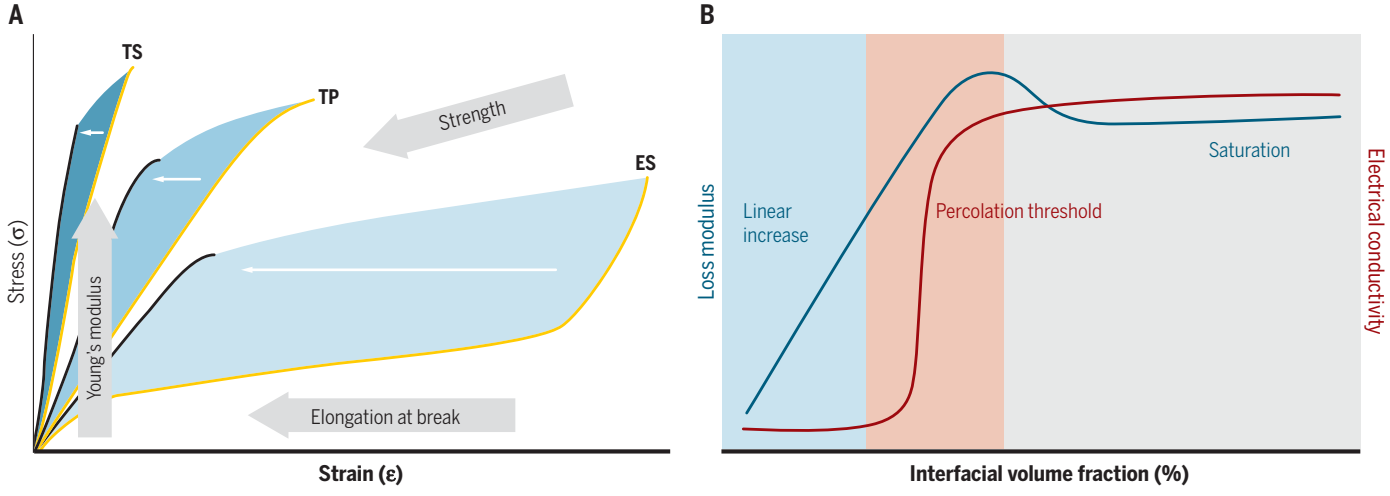


Fig. 4. Schematic illustrations of representative mechanical and electrical properties of CNT or graphene nanocomposites. (**A**) A schematic representation of tensile stress-strain behaviors of three different polymers as CNT or graphene loading fraction increases: thermoset polymer (TS, dark blue region), thermoplastic polymer (TP, blue region), and elastomer (ES, light blue region). Schematic of representative tensile stress-strain curves are compared for pristine thermoset, thermoplastic polymers and elastomer, and its nanocomposite with CNT or graphene reinforcements, respectively. For the pristine polymers, the stress-strain curves are depicted with yellow lines, while for the corresponding CNT or graphene nanocomposites, the curves are illustrated with black lines. As the CNT or graphene loading increase (depicted as horizontal white arrows), the stress-strain curve of each pristine polymer shifts from the yellow line to the black line. This shift indicates that with the increase of CNT or graphene loading fraction, Young's modulus of the polymer can increase proportionally; by contrast, the elongation at break can be substantially decreased. A slight decrease in tensile strength is also shown. Both elastomer and thermoplastic polymer exhibit a similar behavior of ductile-to-brittle transition at high loading fractions of CNTs or graphene, whereas thermoset nanocomposites become increasingly brittle over the pristine polymer. The shift of the stress-strain curves with CNT or graphene reinforcements clearly shows that there is a compromise between Young's modulus and toughness, which can be quantified by characterizing the area of the stress-strain curve. (**B**) Loss modulus and electrical conductivity with respect to interfacial volume fraction in CNT or graphene nanocomposites. With increasing interfacial volume and strength between CNT or graphene fillers and polymer matrix, the loss modulus, which represents viscoelastic damping properties, can linearly increase as a result of filler-matrix slippage until the dispersion quality can be maintained. However, at higher loading fraction of CNTs or graphene, the mobility of polymer chains gradually decreases, eventually leading to no further increase in loss modulus. As for electrical conductivity, as the interfacial volume fraction of CNTs or graphene increases, the conductivity of the nanocomposites changes from insulating to percolative conductive regimes, displaying an S-shape curve. Once the random conductive network is formed, the conductivity of the composites rises sharply, which is characterized as the percolation threshold. Beyond that, the composites become conductive. Typical low percolation thresholds of CNTs or graphene in polymer composites are highly advantageous for many electrostatic and electromagnetic applications.

temperature. As a consequence of the stronger interaction at the interface, relaxation process and storage moduli of the functionalized graphene epoxy composites decay more slowly over the pristine graphene composites. It has been observed that without compromising modulus and strength, an order-of-magnitude enhancement in damping properties can be achieved in CNT-polycarbonate nanocomposites, mainly resulting from the interfacial slippage between CNTs and the polymer matrix (*20*). In general for CNT or graphene polymer composites, the intertube, interlayer interaction and sliding, and the sliding between polymer chains, are all contributing factors to the strong viscoelastic behavior observed in composites (Fig. 4B).

The large interfacial volume available can be used to design interface-directed dynamic composites. A few recent studies have shown highly intriguing morphology evolution of rubbery polymer of polydimethylsiloxane (PDMS) to become more ordered around CNTs under repeated loading, resulting in improved bulk composite stiffness for locally crystallized polymer composites (*21*). Discernable stiffening was observed for the composite during cyclic compressive stressing, a phenomenon not observed for the neat polymer. Such self-stiffening behavior was also observed in graphene-based PDMS composites. Because of its relatively higher chain mobility compared to chemical cross-links, the physically cross-linked network allows polymer chains to undergo realignment and reorientation along graphene surfaces under cyclic compression loading, resulting in increased interface stiffness (*22*). The engineering of unique interface-related properties should allow us to build materials that are stimuli-responsive, dynamically reconfigurable, adaptable, and self-strengthening under loading. CNT or graphene particles seem to be ideal fillers for creating such dynamic interfaces in polymers, which can be further tuned through chemical modification of the filler surfaces. In this regard, quantitative interface evaluation between CNT or graphene and matrices enabled by nanomechanical measurement (see Box 2 and Fig. 5) is very important for more effective interface engineering for improving overall composite performances.

In addition to their mechanical properties, nanotubes and graphene have several other functionalities that make them attractive for use in composites. Important examples are their thermal stability and high electrical and thermal conductivities. Nanotubes are used in dispersions in thermoplastic polymers to improve their electrostatic discharge or electromagnetic interference shielding properties (*23*). Graphene exhibits an electrical conductivity comparable to that of CNTs (10^5 to 10^6 S cm^{-1}). However, the percolation threshold for electrical conductivity of graphene in a polymer composite is much higher than for CNTs in composites. The percolation threshold of CNTs in high-density polyethylene matrix is 0.15 wt %, compared to 1 wt % for graphene in the same polymer (*24*). This is because 1D CNTs form a better conductive network than 2D graphene in composite materials (Fig. 4B). In contrast to the electrical conductivity of CNT or graphene composites, the improvement in thermal conductivity appears to be marginal for both. Even though individual CNTs or graphene have thermal conductivities as high as ~3000 W m^{-1} K^{-1}, such particulate composites have a more efficient network, including intertube or interlayer scattering and heat transfer resistance at the filler-matrix interface (*25*). Attempts can still be made to enhance the thermal conductivity of the composites by substantially increasing the weight fraction of graphene or CNTs in a matrix. For the same loading fraction, graphene composites exhibit greater interaction at the interfaces between filler and matrix, which requires a higher activation energy for the glass transition, resulting in better thermal stability of polymer composites.

Compared to the vast number of studies on polymer matrix composites reinforced by CNTs and graphene, ceramic or metal-based composites have received much less attention, possibly because of the fabrication challenges. In addition, mechanical reinforcing effects are more obvious in polymer-based composites. The primary motivation for the development of CNTs or graphene-based ceramic composites is enhancing toughness or resistance to crack growth because ceramics are already stiff and strong. For graphene-reinforced ceramic composites, the primary toughening mechanism is the increased energy dissipation due to pullout of graphene nanosheets (*26*). Other toughening mechanisms observed include crack deflection at the matrix-reinforcement interface and crack bridging. Similar toughening mechanisms were also observed in CNT-reinforced ceramic composites. Additionally, improved thermal and electrical conductance are often attractive properties of ceramic composites. On the other hand, light metals such as Al, Mg, and Cu are commonly used as matrices for CNTs or graphene-reinforced metal composites for mechanical property enhancement (*27*). It is worth noting that the interfacial strength between CNT or graphene fillers and metal matrices is substantially affected by the nature of interactions that include chemisorption (e.g., Ni or Ru) and physisorption (e.g., Pt or Ir) (*28*), and even reactions leading to carbide formation (e.g., Al) (*27*). One important lesson in both ceramic and metal composites is that the processing-induced changes in the matrix, such as grain refinement, are as important in determining the final properties of the composites as the amount of CNTs and graphene reinforcements present. Other functional properties, such as anticorrosion, antioxidation, piezoelectric properties, and biocompatibility, are being explored in both ceramic and metal composites with CNTs or graphene. The large reduction in weight, in addition to property enhancements, is a key reason for adding CNT or graphene dispersants in metal and ceramic matrices.

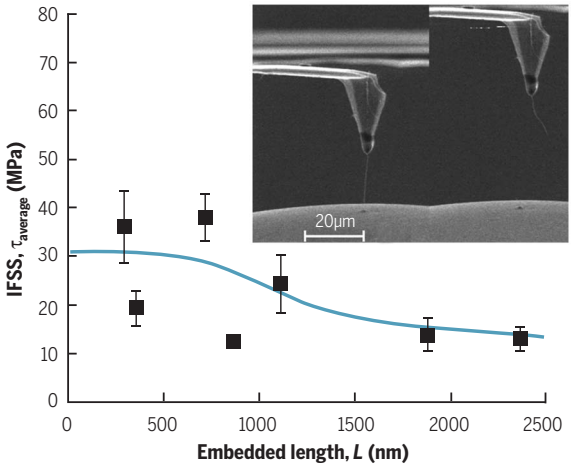


Fig. 5. Nanomechanical characterization of CNT-epoxy interfaces. Shear-lag fits give a value for interfacial shear stress (IFSS) of ~30 ± 7 MPa for pristine carbon nanotubes and epoxy matrix (*25*). (**Inset**) SEM observation of a single nanotube pullout from epoxy using an AFM probe (*43*). It was observed that average IFSS increases as the embedded length decreases, which is consistent with the shear-lag theory and demonstrates the development of interfacial shear stresses at the ends of these fibers. [Reprinted from (*43*) with permission from The Royal Society]

Box. 2. Interfacial strength and nanomechanics in CNT or graphene composites.

Single-fiber pullout tests have been used since the early development of composite materials technology to measure the shear strength of ductile interfaces capable of developing a constant shear stress prior to fiber pullout, and the fracture energy of brittle interfaces that experience crack initiation and propagation prior to fiber pullout. Accordingly, attempts have been made to perform similar experiments on CNT or graphene composites. A successful attempt was made (*40*) to attach a CNT to an atomic force microscopy (AFM) probe that was then dipped into a liquid epoxy, and a pullout test was carried out after the epoxy was cured. More recently, microfabricated devices have been developed to perform the pullout experiments with better control (*41*). The average interfacial shear strength reported by these tests corresponds to the range of ~6 to 200 MPa for CNT-epoxy interfaces (*42*), although such quantitative nanomechanical measurements remain a big challenge for graphene/polymer interfaces.

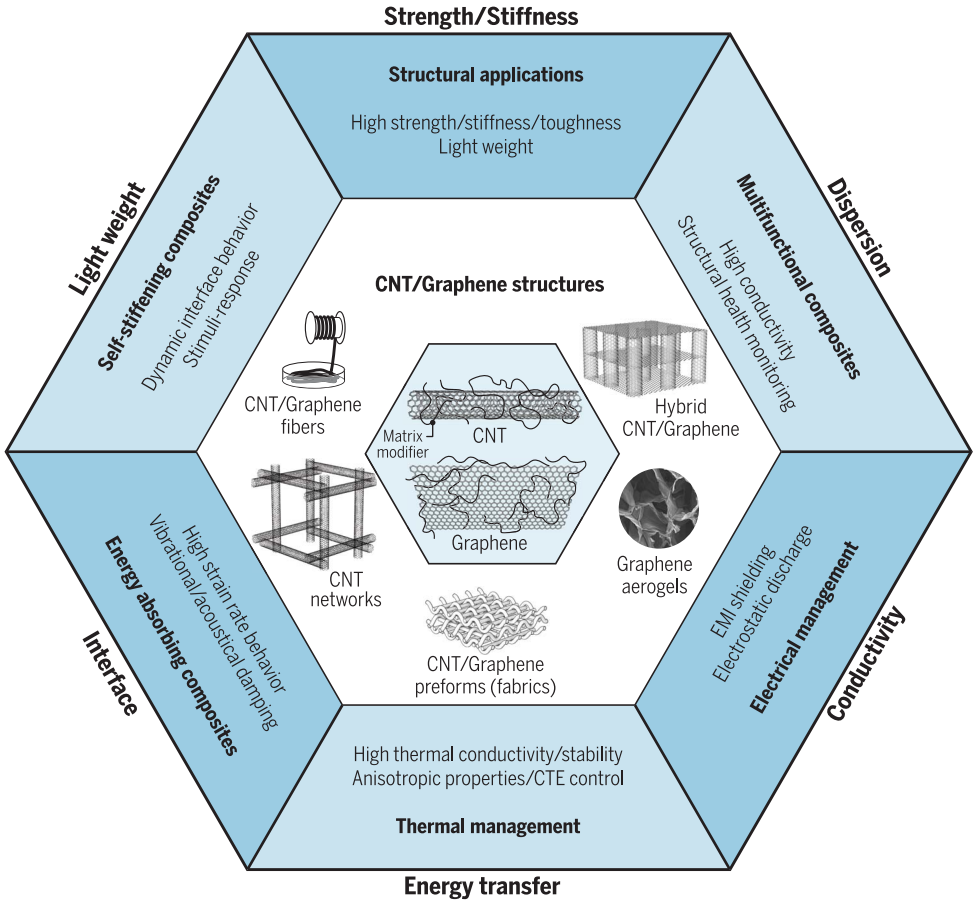


Fig. 6. A schematic summarizing various CNT and graphene additions to composites with anticipated enhanced properties for the prospective applications along with six key parameters in CNT and graphene composites. There has been substantial progress in synthesizing continuous fiber spinning with CNT and graphene, and also various yet controllable CNT and graphene structures. The implementation of these unique structures of CNT and graphene as reinforcements in composite materials could provide the opportunity to translate the exceptional properties of the individual CNT and graphene into a variety of engineering applications. It can be envisioned that, if carefully designed and optimized, CNT and graphene composites could not only replace continuously reinforced carbon fiber composites as state-of-the-art structural composites but also provide new functionalities such as stimuli-responsiveness, energy storage and conversion elements, real-time structural health monitoring, sensing, and vibrational damping in composites. A range of perspective applications including structural applications, multifunctional composites, electrical and thermal management, energy-absorbing composites, and self-stiffening composites with the required properties are summarized along with six key parameters including strength/stiffness, lightweight, interface, energy transfer, conductivity, and dispersion in the composite design and manufacturing.

As discussed above, the Achilles' heel for structural nanocomposites is the difficulty in processing them at meaningful filler loadings. This issue arises from the high interfacial volume between the matrix and the filler, leading to a low rheological percolation threshold. Consequently, the matrix viscosity is found to increase substantially at very low concentrations of CNTs or graphene content, albeit with thixotropic behavior, meaning that the viscosity will drop upon shear (*29*). Furthermore, the dimensionality of the particle has an important role. Indeed, the 1D nature of nanotubes, which makes them ideal for electrical percolation, means that nanotubes become nearly unprocessable in a conventional polymeric dispersion at loadings of a few volume %, whereas the 2D nature of graphene means that the platelets can slide over each other in shear. This allows systems with ~20 wt % GNPs to be processed, giving graphene a competitive advantage over CNTs for higher filler load applications. One successful route for addressing this viscosity increase at moderate concentrations has been to produce liquid crystalline phases, typically nematic, that require near-monodispersed particle size distribution and an excellent degree of dispersion (*30*).

A key approach to achieving high loading fractions is to preform the nanoparticles into an organized architecture (e.g., fibers) before introducing the matrix so that traditional composite production techniques can be subsequently used. In a direct analogy to polyaramid coagulation spinning, graphene, graphene oxide, and nanotubes have been spun from the liquid crystalline phases, by using either superacid, organic solvent, or aqueous/surfactant dispersions (*31*). Alternatively, polymer composite fibers can be produced where the polymer [e.g., poly(vinyl alcohol) or polyacrylonitrile] is used as a binder to improve the interparticle stress transfer, with the binder optionally being pyrolyzed (*32*). Finally, specifically for CNTs, fibers can be drawn from their growth substrate or zone (*33*), which bypasses issues with the liquid-phase rheology, allowing extremely long and continuous CNTs to be used. The 1D nature of CNTs means that they are more appropriate for fibers than graphene; in particular, very–high–aspect ratio CNTs can be used in the fibers, leading to high intertube stress transfer. To date, the mechanical properties of CNT fibers have been considerably better than for graphene fibers. Furthermore, graphene and nanotube fibers could have exceptional multifunctionality, including in applications such as energy harvesting, energy storage, and sensing, that will be essential for future demands of smart fabrics and structures.

Another route to using fiber layup production techniques is hierarchical composites in which CNTs or graphene are grafted onto macroscale carbon or glass fiber by either direct growth or deposition (*34*). These "hairy fibers" show increased out-of-plane improvement in interlaminar shear stress and electrical and thermal conductivity. Alternatively, hybrid systems can be produced where two or more from graphene, nanotubes, and fibers are used within a system through mixing, interacting, and avoiding some of the viscosity and aggregation issues discussed above. For example, nanotubes could combine with graphene to give the low electrical percolation of the former with the high electrical conductivity of the latter. Similarly, additions of graphene or nanotubes to fiber-polymer composites can change the glass transition temperature, and give fire retardancy properties, barrier properties, structural damage sensing, and increased thermal and electrical conductivity. A recent study (*35*) demonstrated that the addition of 2 wt % graphene to the matrix of a carbon fiber epoxy composite can lead to an increase in axial stiffness of ~10 GPa (~9%) as the result of the alignment and constraint of the graphene flakes between the fibers. This technology has been reported (*36*) as being used in the case of the ultralightweight "graphene watch" produced by Richard Mille,

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

the Swiss luxury watchmaker, in conjunction with McLaren Formula 1 motor racing.

Future of CNT or graphene composites

After more than two decades of efforts to create nanotube and graphene composites in various concentrations and with various matrix combinations, the question still stands: How can the spectacular properties of CNT or graphene structures be accessed in their composite structures? The answers have come mainly from trial-and-error–based attempts to control dispersion, filler fraction, and, to a limited extent, interface control. The results thus far do not point to competitive advantages of CNT or graphene particulate additives to composites as real structural reinforcements leading to high strength or increased modulus when compared to carbon fibers. Such an outcome would require continuous fibers and better interfacial engineering, which could still be possible with spun fibers from nanotubes and chemically modified extended graphene structures. Light weight of CNT or graphene fibers could be the best advantage provided by these structures in composites. In addition, in large markets such as the automotive industry, where there is a need for cheap reinforcements that bridge the properties of glass and carbon fibers and to lower production costs, these nanocarbon structures may provide a solution. For all composites, including high-performance ones, the matrix- or fiber-modifying capabilities of these fillers might prove an important advantage, through increasing thermal stability, electrical and thermal conductivities, viscoelastic properties, toughness, and creation of dynamic interfaces toward building multifunctional composites. Indeed, graphene or nanotubes are already used commercially on a large scale as a substitute for traditional conductive additives (e.g., carbon black) for applications such as static dissipative polymers. Figure 6 shows a summary of the important parameters that might determine CNT or graphene additions to composite materials and possible properties that could be enhanced for a range of more advanced applications. The progress in fiber spinning with CNTs and continuous filaments, sheets, and 3D scaffolds with CNT or graphene elements could see a resurgence of CNT or graphene composites for structural applications. If this possibility is realized, then CNT or graphene fiber structures could not only eventually replace today's carbon fibers but also provide new functionalities such as energy storage and conversion elements (e.g., CNT or graphene fibers acting as reinforcements and electrodes) (*37*), real-time structure monitoring, sensing, and vibrational and acoustic damping in composites. These attributes would also be ideally suited for thin composite structures (for which micrometer-size carbon fibers may be too thick) or for providing flexibility and resilience in superflexible structures. Unexplored properties of CNT or graphene composites, such as radiation resistance, could also prove vital in determining whether these unique carbon nanostructures prove serious contenders in composites. Finally, beyond the intrinsic scientific and technological issues, the widespread availability of these nanomaterials in large volumes and at relatively low cost will dictate the ultimate success of nanotube and graphene composites.

REFERENCES AND NOTES

1. A. K. Geim, *Science* **324**, 1530–1534 (2009).
2. J. Zang *et al*., *Nat. Mater.* **12**, 321–325 (2013).
3. O. Lourie, D. Cox, H. Wagner, *Phys. Rev. Lett.* **81**, 1638–1641 (1998).
4. T. Ramanathan *et al*., *Nat. Nanotechnol.* **3**, 327–331 (2008).
5. S. H. Song *et al*., *Adv. Mater.* **25**, 732–737 (2013).
6. D. G. Papageorgiou, I. A. Kinloch, R. J. Young, *Prog. Mater. Sci.* **90**, 75–127 (2017).
7. R. J. Young, L. Deng, T. Z. Wafy, I. A. Kinloch, *J. Mater. Sci.* **51**, 344–352 (2016).
8. M. Terrones *et al*., *Adv. Mater.* **23**, 5302–5310 (2011).
9. P. M. Ajayan, L. S. Schadler, C. Giannaris, A. Rubio, *Adv. Mater.* **12**, 750–753 (2000).
10. S. Bhattacharyya *et al*., *Carbon* **46**, 1037–1045 (2008).
11. M. Hernández, M. del Mar Bernal, R. Verdejo, T. A. Ezquerra, M. A. López-Manchado, *Compos. Sci. Technol.* **73**, 40–46 (2012).
12. P. Calvert, *Nature* **357**, 365–366 (1992).
13. G. Marom, H. D. Wagner, *J. Mater. Sci.* **52**, 8357–8361 (2017).
14. F. M. Blighe *et al*., *Adv. Funct. Mater.* **21**, 364–371 (2011).
15. R. J. Young *et al*., *Compos. Sci. Technol.* **154**, 110–116 (2018).
16. L. Malard, M. Pimenta, G. Dresselhaus, M. Dresselhaus, *Phys. Rep.* **473**, 51–87 (2009).
17. C. Lee, X. Wei, J. W. Kysar, J. Hone, *Science* **321**, 385–388 (2008).
18. Z. Wang, P. Ciselli, T. Peijs, *Nanotechnology* **18**, 455709 (2007).
19. I. Vlassiouk *et al*., *ACS Appl. Mater. Interfaces* **7**, 10702–10709 (2015).
20. N. A. Koratkar *et al*., *Appl. Phys. Lett.* **87**, 063102 (2005).
21. B. J. Carey, P. K. Patra, L. Ci, G. G. Silva, P. M. Ajayan, *ACS Nano* **5**, 2715–2722 (2011).
22. L. Cao *et al*., *Small* **12**, 3723–3731 (2016).
23. M. F. De Volder, S. H. Tawfick, R. H. Baughman, A. J. Hart, *Science* **339**, 535–539 (2013).
24. J. Du *et al*., *Carbon* **49**, 1094–1100 (2011).
25. Y. Liu, S. Kumar, *ACS Appl. Mater. Interfaces* **6**, 6069–6087 (2014).
26. W. A. Curtin, B. W. Sheldon, *Mater. Today* **7**, 44–49 (2004).
27. A. Nieto, A. Bisht, D. Lahiri, C. Zhang, A. Agarwal, *Int. Mater. Rev.* **62**, 241–302 (2017).
28. S. M. Kozlov, F. Vines, A. Gorling, *J. Phys. Chem. C* **166**, 7360–7366 (2016).
29. P.-H. Wang, S. Sarkar, P. Gulgunje, N. Verghese, S. Kumar, *Polymer (Guildf.)* **150**, 10–25 (2018).
30. P. K. Rai *et al*., *J. Am. Chem. Soc.* **128**, 591–595 (2006).
31. N. Behabtu *et al*., *Science* **339**, 182–186 (2013).
32. B. Vigolo *et al*., *Science* **290**, 1331–1334 (2000).
33. Y.-L. Li, I. A. Kinloch, A. H. Windle, *Science* **304**, 276–278 (2004).
34. M. J. Arlen *et al*., *Macromolecules* **41**, 8053–8062 (2008).
35. J. W. Chu *et al*., *Compos., Part A Appl. Sci. Manuf.* **113**, 311–317 (2018).
36. B. Robinson, https://www.manchester.ac.uk/discover/news/science-behind-worlds-lightest-graphene-watch-revealed/ (2018).
37. X. Mao, G. C. Rutledge, T. A. Hatton, *Nano Today* **9**, 405–432 (2014).
38. M. S. Dresselhaus, G. Dresselhaus, R. Saito, A. Jorio, *Phys. Rep.* **409**, 47–99 (2005).
39. O. Frank *et al*., *Nat. Commun.* **2**, 255 (2011).
40. A. H. Barber, S. R. Cohen, A. Eitan, L. S. Schadler, H. D. Wagner, *Adv. Mater.* **18**, 83–87 (2006).
41. Y. Ganesan *et al*., *ACS Appl. Mater. Interfaces* **3**, 129–134 (2011).
42. X. Chen *et al*., *Carbon* **82**, 214–228 (2015).
43. W. Wang, P. Ciselli, E. Kuznetsov, T. Peijs, A. H. Barber, *Philos. Trans. A Math. Phys. Eng. Sci.* **366**, 1613–1626 (2008).

ACKNOWLEDGMENTS

We thank S. K. Lee for her help in creating some of the figures. **Funding:** J.L. acknowledges support from the U.S. Department of Energy, Office of Basic Energy Sciences, under grant DE-SC0018193. P.M.A. and J.L. acknowledge the NSF Nanosystems Engineering Research Center for Nanotechnology-Enabled Water Treatment (ERC-1449500) for support. J.S. acknowledges support from the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (2018R1A2B2001565). I.A.K. and R.J.Y. acknowledge funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 785219. **Competing interests:** The authors declare no competing interests.

10.1126/science.aat7439

RESEARCH

Identifying virus vectors

Babayan et al., *p. 577*

IN *SCIENCE* JOURNALS

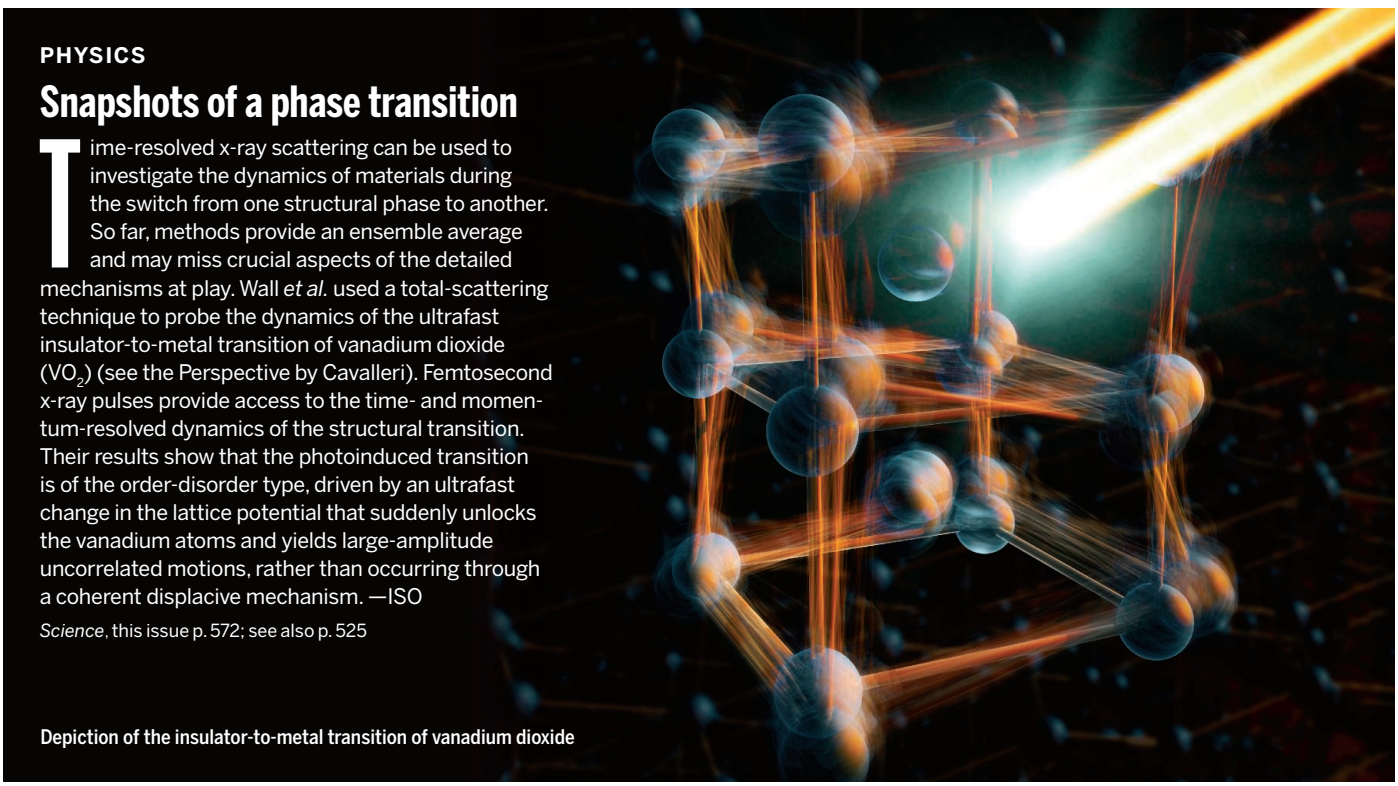
Edited by **Caroline Ash**

PHYSICS

Snapshots of a phase transition

Time-resolved x-ray scattering can be used to investigate the dynamics of materials during the switch from one structural phase to another. So far, methods provide an ensemble average and may miss crucial aspects of the detailed mechanisms at play. Wall *et al.* used a total-scattering technique to probe the dynamics of the ultrafast insulator-to-metal transition of vanadium dioxide (VO_2) (see the Perspective by Cavalleri). Femtosecond x-ray pulses provide access to the time- and momentum-resolved dynamics of the structural transition. Their results show that the photoinduced transition is of the order-disorder type, driven by an ultrafast change in the lattice potential that suddenly unlocks the vanadium atoms and yields large-amplitude uncorrelated motions, rather than occurring through a coherent displacive mechanism. —ISO

Science, this issue p. 572; see also p. 525



Depiction of the insulator-to-metal transition of vanadium dioxide

ARCHAEOLOGY

Stress in Neanderthal children

Although environmental variability, such as extreme seasonal cold, is believed to have influenced human evolution, it is difficult to measure its effects directly because of the coarseness of standard analytical methods. Using oxygen-isotope and trace-element analysis on microsamples of tooth enamel from two Neanderthal children from the archaeological site of Payre, France, Smith *et al*. discovered evidence for extreme wintertime stress in these children, including probable weight loss and exposure to toxic chemicals in the environment, most notably lead. More detailed analyses of one of the children showed that the child was born in the spring and weaned before winter at 2.5 years of age. —MSA

Sci. Adv. 10.1126/sciadv.aau9483 (2018).

STRUCTURAL BIOLOGY

Nucleosomal DNA transcription

In eukaryotes, the basic chromatin unit nucleosome stalls RNA polymerase II (RNAPII) when it transcribes genetic information on DNA. Using cryo–electron microscopy, Kujirai *et al*. explored seven structures of the RNAPII-nucleosome complex, in which RNAPII pauses at four locations on the nucleosome. These serial snapshots of the RNAPII progression reveal the molecular mechanism of how RNAPII peels the nucleosomal DNA off the histone stepwise. —SYM

Science, this issue p. 595

NEUROSCIENCE

Posture in the brain

Our understanding of the neural basis of motor control originates in studies of eye, hand, and arm movements in primates. Mimica *et al*. investigated neuronal representations of body postures in the posterior parietal and frontal motor cortices with three-dimensional tracking of freely moving rodents (see the Perspective by Chen). Both brain regions represented posture rather than movements and self-motion. Decoding the activity of neurons in the two regions accurately predicted an animal's posture. —PRS

Science, this issue p. 584; see also p. 520

MICROBIOTA

Gut microbiota selects fungi

Fungi, such as *Candida albicans*, are found in the mammalian gut, but we know little about what they are doing there. Tso *et al*. put *C. albicans* under evolutionary pressure by serial passage

in mice that were treated with antibiotics and were thus lacking gut bacteria (see the Perspective by d'Enfert). Passage accelerated fungal mutation, especially around the *FLO8* gene, resulting in low-virulence phenotypes unable to form hyphae. Nevertheless, these phenotypes stimulated proinflammatory cytokines and conferred transient cross-protection against several other gut inhabitants. However, if an intact microbiota was present, only the virulent hyphal forms persisted. —CA

Science, this issue p. 589; see also p. 523

ANTI-FLU THERAPY

Durable influenza protection

Vaccines are indispensable for the control and prevention of influenza, but there are several challenges to efficacy. Some individuals respond poorly to vaccination, and virus variation makes targeting optimal antigens difficult. Broadly neutralizing antibodies are one solution, but they have their own pitfalls, including limited cross-reactivity to both influenza A and B strains and the need for repeated injections. Now, Laursen *et al*. have developed multidomain antibodies with breadth and potency. Administered intranasally to mice with an adeno-associated virus vector, the antibodies provided durable and continuous protection from a panoply of influenza strains. —STS

Science, this issue p. 598

PARKINSON'S DISEASE

The benefits of a missing appendix

Misfolded α-synuclein is a pathological hallmark of Parkinson's disease (PD). Killinger *et al*. report that the human appendix contains an abundance of misfolded α-synuclein and that removal of the appendix decreases the risk of developing PD. The appendixes of both individuals with PD and healthy individuals contained abnormally cleaved and aggregated forms of α-synuclein, analogous to those found in postmortem brain tissue from PD patients. Furthermore, α-synuclein derived from the appendix seeds rapid aggregation of recombinant α-synuclein in vitro. In two large-scale epidemiological studies, an appendectomy occurring decades before reduced the risk of developing PD. —OMS

Sci. Transl. Med. **10**, eaar5280 (2018).

NANOMATERIALS

More alloying on silica

Controlling the stoichiometry and achieving a high degree of alloying of metals at ultrasmall scales for catalysis can be difficult. Double complex salts, formed by a cation like $Pd(NH_3)_4^{2+}$ and an anion like $IrCl_6^{2-}$, should be excellent precursors but are poorly soluble and difficult to adsorb directly on metal oxide surfaces. Ding *et al*. show that sequentially adsorbing the cations and anions from organic solvents onto a silica surface, followed by heating in hydrogen, creates well-mixed nanoparticles, most less than 3 nanometers in diameter, for a variety of alloys. These materials were then tested as catalysts for acetylene hydrogenation to ethylene. —PDS

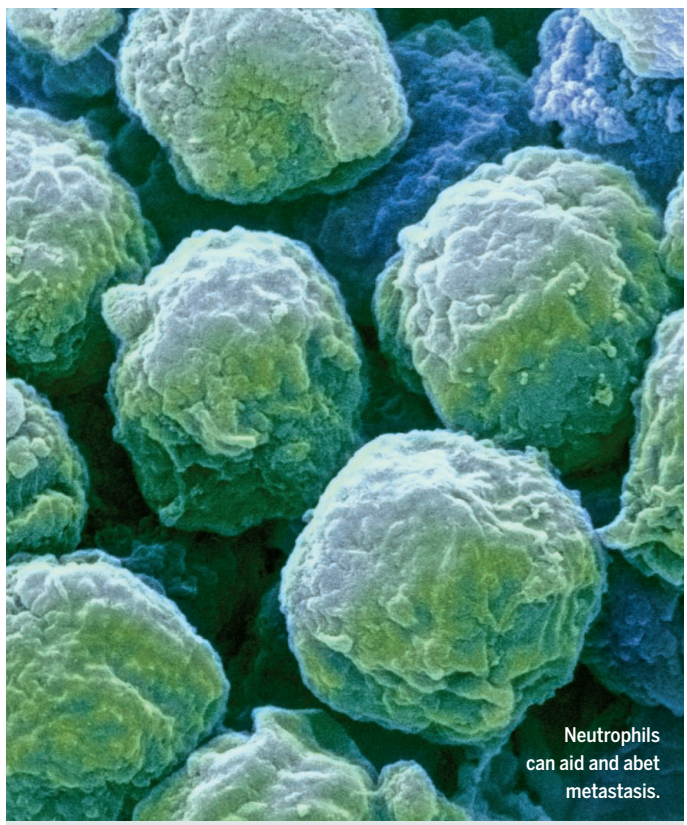
Science, this issue p. 560

ORGANIC CHEMISTRY

Steps to smaller rings

Certain ring-forming reactions in organic chemistry are efficient because the orbital symmetries match up in the reactants and products. Oxyallyl ions tend to react with dienes in this paradigm to form seven-membered rings. Under palladium catalysis, Trost *et al*. redirected this reaction toward more common five-membered tetrahydrofuran rings by appending an ester to the diene. Although that pathway is symmetry-forbidden, the electron-withdrawing ester appears to stabilize a key intermediate along a stepwise route to the smaller ring. —JSY

Science, this issue p. 564



Neutrophils can aid and abet metastasis.

IMAGE: DAVID MCCARTHY/SCIENCE SOURCE

CANCER

An ironclad role in metastasis

In the body's defense against cancer, immune cells are generally viewed as friends, not foes. One exception is neutrophils, which promote the growth of metastatic tumor cells. Liang *et al*. explored the signaling mechanism underlying this effect and identified a surprising culprit. Studying lung metastases in mice, they found that the tumor growth–promoting signal secreted by neutrophils is transferrin, an iron-transporting protein previously thought to be expressed mainly in liver cells. Transferrin synthesis in neutrophils was stimulated by a specific growth factor, called granulocyte-macrophage colony-stimulating factor (GM-CSF), produced by tumor cells in the metastatic microenvironment. Drugs blocking the activity of GM-CSF and its signaling pathway are already available and conceivably could be repurposed as treatments for metastatic disease. —PAK

Proc. Natl. Acad. Sci. U.S.A. **115**, 11060 (2018).

EVOLUTION

Networks of phenotypic variation

Evolution depends on phenotypic variation. Gene regulatory networks are theoretically expected to be likely sources of variation. Schaerli *et al*. explored how the structure of gene regulatory circuits contributes to the range of mutant phenotypes that are produced. Synthetic networks in the bacteria *Escherichia coli* were built to test this idea

ENVIRONMENTAL POLICY

The Clean Water Act actually cleans water

Though the U.S. Clean Water Act has been in effect since 1972, evidence of its effectiveness has been unclear. Keiser and Shapiro compile comprehensive data on water pollution and its drivers, including 50 million pollution readings from 240,000 monitoring sites. They find that $650 billion in federal grants to wastewater treatment plants made rivers significantly cleaner for 25 miles downstream for 30 years ($1.5 million per mile per year). But the economic benefit in terms of home values within 25 miles of improved waters increased by only 25% of the cost of the grants. —BW

Quart. J. Econ. 10.1093/qje/qjy019 (2018).

Wastewater treatment conducted under the mandates of the 1972 Clean Water Act has lead to demonstrably cleaner rivers in the United States.

experimentally. In colonies responding to a gradient of a stimulatory chemical, the networks produced a striped pattern of cells. Ultimately, the structure of the networks was critical in specifying the phenotypes that could be obtained by mutation. —LBR

Mol. Syst. Biol. **14**, e8102 (2018).

EVOLUTIONARY GENOMICS

Mapping genotype to phenotype

Identifying genetic factors that contribute to trait variation among species is difficult. Weiss *et al*. developed a method for unbiased genome-scale mapping of trait differences and used it to work out how a range of yeast species evolved. When crossed, these yeast species produce sterile offspring, preventing the use of traditional genetic-mapping techniques. However, by reciprocal hemizygosity analysis via sequencing (RH-seq), alleles that contributed to differences in thermotolerance were spotted that diverged between two species. Surprisingly, these turned out to map to essential housekeeping genes. Thus, without taking a candidate locus approach, RH-seq allowed identification of genes responsible for traits that differ between species. —LMZ

Nat. Genet. **50**, 1501 (2018).

ANTHROPOLOGY

Osteobiology of Italian miners

There are few historical records of the lives of working-class people. In the late 1400s, alum—hydrated salts of aluminium sulfate—was valued as a primer for wood-panel paintings and as a mordant to stabilize textile dyes. To gain some insight into the diet and lives of the people who mined and processed alum mineral, Baldoni *et al*. analyzed bones from a grave site in the town of Allumiere, one of the first alunite mining sites in Italy. Each phase of the process left telltale musculoskeletal markers, including fractures and osteodegenerative disease. Dietary analyses hint at consumption of beans, cabbage, and sheep and cattle meat, as well as medicinal plants. In this case, the record lies in the very bones of the people. —PJH

PLOS ONE **13**, e0205362 (2018).

CLIMATE CHANGE

Oceans of help

The 2015 Paris Agreement advances the ambitious goal of trying to keep the global average temperature to less than 1.5°C higher than preindustrial levels, but current national pledges to reduce emissions are insufficient to achieve that target. The Ocean Solutions Initiative examined what steps might be taken to mitigate three major climate-related changes in the ocean—warming, acidification, and sea-level rise—as part of that effort. Gattuso *et al*. summarize the suggestions, limitations, and trade-offs discussed in the Initiative, finding that global-scale actions have both the greatest potential impact and the largest limitations and risks, whereas more-local efforts are less risky but also less effective. A combination of global and local efforts has the greatest potential benefit. —HJS

Front. Mar. Sci. **5**, 337 (2018).

METALLIC GLASSES

Packing away fractal structure

Metallic glasses are strong, but relatively elastic, making them interesting for a wide variety of applications. Much debate around the structure of these glasses has recently centered on the scaling between volume and the spacing between atoms. In two papers, Feng *et al*. show that this scaling overall is compact, like other solids. This contrasts with recent claims of more exotic, widespread fractal packing for similar metallic glasses. It also helps us understand the atomic origin of these fascinating materials. —BG

J. Phys. Condens. Matter **30**, 255402 (2018); *Phys. Rev. B* **98**, 024201 (2018).

PHOTO: DANITA DELIMONT/GETTY IMAGES

ALSO IN *SCIENCE* JOURNALS

Edited by **Caroline Ash**

NEURODEGENERATION

PAR promotes α-synuclein toxicity

How pathologic α-synuclein (α-syn) leads to neurodegeneration in Parkinson's disease (PD) remains poorly understood. Kam *et al*. studied the α-syn preformed fibril (α-syn PFF) model of sporadic PD (see the Perspective by Brundin and Wyse). They found that pathologic α-syn–activated poly(adenosine 5′-diphosphate–ribose) (PAR) polymerase–1 (PARP-1) and inhibition of PARP or knockout of PARP-1 protected mice from pathology. The generation of PAR by α-syn PFF–induced PARP-1 activation converted α-syn PFF to a strain that was 25-fold more toxic, termed PAR–α-syn PFF. An increase of PAR in the cerebrospinal fluid and evidence of PARP activation in the substantia nigra of PD patients indicates that PARP activation contributes to the pathogenesis of PD through parthanatos and conversion of α-syn to a more toxic strain. —SMH

Science, this issue p. 557; see also p. 521

DRUG DISCOVERY

Thalidomide-targeted degradation

Thalidomide and its analogs improve the survival of patients with multiple myeloma and other blood cancers. Previous work showed that the drugs bind to the E3 ubiquitin ligase Cereblon, which then targets for degradation two specific zinc finger (ZF) transcription factors with a role in cancer development. Sievers *et al*. found that more ZF proteins than anticipated are destabilized by thalidomide analogs. A proof-of-concept experiment revealed that chemical modifications of thalidomide can lead to selective degradation of specific ZF proteins. The detailed information provided by structural, biochemical, and computational analyses could guide the development of drugs that target ZF transcription factors implicated in human disease. —PAK

Science, this issue p. 558

EPIDEMIOLOGY

Predicting hosts and vectors

During outbreaks of mysterious infections, events can rapidly become dangerous and confusing. A combination of increasing experience with outbreaks and genome-sequencing technology now means the pathogen can often be identified within days. But for some of the most frightening viral pathogens, the originating hosts and possible vectors often remain obscure. Babayan *et al*. took sequence data from more than 500 single-stranded RNA viruses (see the Perspective by Woolhouse) and used machine-learning algorithms to extract evolutionary signals imprinted in the virus sequence that offer information about its original hosts and if an arthropod vector, and what type, plays a part in the virus's natural ecology. —CA

Science, this issue p. 577; see also p. 524

METALLURGY

Stronger copper through twin power

Materials with structural gradients often have unique combinations of properties. Gradient-structured materials are found in nature and can be engineered. Cheng *et al*. made a structural gradient by introducing gradients of crystallographic twins into copper. This strategy creates bundles of dislocations in the crystal interiors, which makes the metal stronger than any of the individual components. This method offers promise for developing high-performance metals. —BG

Science, this issue p. 559

PALEOCLIMATOLOGY

East Asian monsoon mysteries

What exactly does the oxygen isotopic composition of speleothems tell us about the East Asian monsoon? They provide magnificent, detailed records of hydroclimate, but precisely what aspects of hydroclimate they record is unclear. Zhang *et al*. present data from two speleothems from central eastern China for the period from 21,000 to 10,000 years ago and suggest that the cause of the oxygen isotopic variability that they observe is more complex than simple changes in monsoon strength or intensity (see the Perspective by McGee). Alternatively, this variation may reflect the lengths of various phases of the monsoon and the regional heterogeneity of the East Asian hydroclimate. —HJS

Science, this issue p. 580; see also p. 518

VASCULAR BIOLOGY

RSK2 restricts blood vessel diameter

Phosphorylation of myosin proteins enables smooth muscle cells to contract and cause vasoconstriction. This phosphorylation event has generally been attributed to myosin light chain kinase. However, Artamonov *et al*. discovered that the kinase RSK2 was activated by intraluminal pressure or stimuli that promote vasoconstriction in resistance arteries and that it phosphorylated smooth muscle myosin. Mice lacking RSK2 had dilated arteries and lower blood pressure than wild-type littermates. RSK2 is a possible drug target for manipulating blood pressure. —WW

Sci. Signal. **11**, eaar3924 (2018).

QUANTUM OPTICS

Protecting entangled pairs

Photons are readily generated, are fast and can travel vast distances, and are ideal carriers of quantum information. Practical applications, such as quantum computing, will likely be based on an optical-chip platform and require the manipulation of multiphoton states. The inevitable scattering and loss of photons in such a platform would be detrimental for application. Blanco-Redondo *et al*. show how a specially designed optical circuit based on topology can offer protection for propagating biphoton states. The results show that topological design consideration could provide the desired robustness required for quantum optical circuitry. —ISO

Science, this issue p. 568

FIBROSIS

Type 3 injury

Chronic liver injury caused by infection, toxins, or other inflammatory conditions can lead to liver fibrosis. Fabre *et al*. define a role for the type 3 cytokines interleukin-17A (IL-17A) and IL-22 in driving transforming growth factor–β (TGF-β)–dependent fibrosis. IL-17A and IL-22 become increased in intrahepatic lymphocytes from patients with hepatitis. Hepatic stellate cells treated with IL-22 showed enhanced p38 mitogen-activated protein kinase–dependent TGF-β signaling. Liver-resident neutrophils and mast cells were identified as the primary sources of IL-17 in humans, and mouse studies showed that blockade of IL-17A and IL-22 could reduce fibrosis. —CNF

Sci. Immunol. **3**, eaar7754 (2018).

RESEARCH ARTICLE SUMMARY

NEURODEGENERATION

Poly(ADP-ribose) drives pathologic α-synuclein neurodegeneration in Parkinson's disease

Tae-In Kam, Xiaobo Mao, Hyejin Park, Shih-Ching Chou, Senthilkumar S. Karuppagounder, George Essien Umanah, Seung Pil Yun, Saurav Brahmachari, Nikhil Panicker, Rong Chen, Shaida A. Andrabi, Chen Qi, Guy G. Poirier, Olga Pletnikova, Juan C. Troncoso, Lynn M. Bekris, James B. Leverenz, Alexander Pantelyat, Han Seok Ko, Liana S. Rosenthal, Ted M. Dawson*, Valina L. Dawson*

INTRODUCTION: Parkinson's disease (PD) is the second most common neurodegenerative disorder. Intracellular protein aggregates composed primarily of α-synuclein lead to neuronal dysfunction throughout the nervous system, ultimately accumulating in structures called Lewy bodies and neurites. Loss of substantia nigra pars compacta dopamine (DA) neurons and dystrophic striatal projections account for the major motor symptoms of PD, which include a rest tremor, slowness of movement, rigidity, and postural instability. Other neuronal systems are affected by pathologic α-synuclein and contribute to the nonmotor symptoms of PD, which include anxiety, depression, sleep disorders, autonomic dysfunction, constipation, and cognitive impairment.

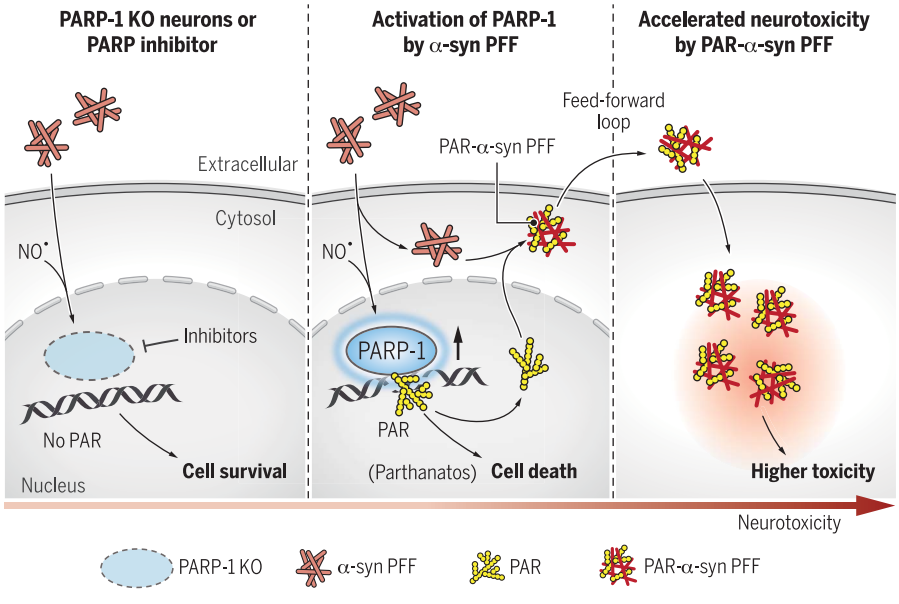
RATIONALE: During the pathogenesis of PD, monomeric α-synuclein assembles into higher-ordered structures that ultimately become pathologic and drive neuronal cell death. Pathologic α-synuclein can spread from cell to cell, contributing to the progressive pathogenesis of PD. What drives the abnormal assembly of pathologic α-synuclein and the cell injury and death mechanisms that are activated by pathologic α-synuclein are not known.

RESULTS: Recombinant α-synuclein preformed fibrils (PFFs), which are similar in structure to those found in PD, were used to model pathologic α-synuclein both in vitro and in vivo. We investigated the cellular cell death pathways that contribute to and drive α-synuclein PFF–mediated neuronal cell death. Pathologic α-synuclein was found to activate nitric oxide synthase (NOS), causing DNA damage and poly(adenosine 5′-diphosphate–ribose) polymerase-1 (PARP-1) activation, leading to cell death via parthanatos. α-Synuclein PFF was found to primarily kill neurons via parthanatos, because necroptosis and autophagy inhibition had no effect on α-synuclein PFF neurotoxicity and there was only modest protection by caspase inhibition. Neuron-to-neuron transmission of pathologic α-synuclein and accompanying pathology and neurotoxicity in primary neuronal cultures were completely attenuated by clinically available PARP inhibitors or by deletion of PARP-1. α-Synuclein PFF–induced loss of DA neurons and biochemical and behavioral deficits in vivo were significantly prevented by PARP inhibition or lack of PARP-1. PAR generated by PARP-1 activation also binds to α-synuclein, accelerating its fibrillization and converting pathologic α-synuclein to a more misfolded compact strain with 25-fold enhanced toxicity. PAR-modified α-synuclein PFF–injected mice showed accelerated disease progression and phenotype compared to α-synuclein PFF–injected mice. Moreover, PAR levels were increased in the cerebrospinal fluid (CSF) in two independent patient cohorts and brains of PD patients, providing evidence that parthanatos may contribute to the pathogenesis of PD.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat8407

CONCLUSION: We identified PARP-1 activation and the generation of PAR as a key mediator of pathologic α-synuclein toxicity and transmission. Activation of parthanatos is the primary driver of pathologic α-synuclein neurodegeneration. Inhibition of PARP and depletion of PARP-1 substantially reduces the pathology induced by the transmission of pathologic α-synuclein. In a feed-forward loop, PAR converted pathologic α-synuclein to a more toxic strain and accelerated neurotoxicity both in vitro and in vivo. Consistent with the notion that PARP-1 activation plays a role in PD pathogenesis, PAR levels were increased in the CSF and brains of PD patients. Thus, strategies aimed at inhibiting PARP-1 activation could hold promise as a disease-modifying therapy to prevent the loss of DA neurons in PD and related α-synucleinopathies. Moreover, assessment of PAR levels in the CSF could serve as a theranostic biomarker for disease-modifying therapies in these disorders. ■



PARP-1 is activated by α-synuclein PFF and PAR mediates cell death. Inhibition of PARP or deletion of PARP-1 reduces α-synuclein PFF–induced cell death. In a feed-forward loop, PAR causes the formation of a more toxic α-synuclein strain, resulting in accelerated pathologic α-synuclein transmission and toxicity.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: tdawson@jhmi.edu (T.M.D.); vdawson@jhmi.edu (V.L.D.)
Cite this article as T.-I. Kam *et al.*, *Science* **362**, eaat8407 (2018). DOI: 10.1126/science.aat8407

RESEARCH ARTICLE

NEURODEGENERATION

Poly(ADP-ribose) drives pathologic α-synuclein neurodegeneration in Parkinson's disease

Tae-In Kam[1,2,3]*, Xiaobo Mao[1,2,3]*, Hyejin Park[1,2,3]*, Shih-Ching Chou[1,4], Senthilkumar S. Karuppagounder[1,2,3], George Essien Umanah[1,2], Seung Pil Yun[1,2,3], Saurav Brahmachari[1,2,3], Nikhil Panicker[1,2,3], Rong Chen[1,2,3], Shaida A. Andrabi[1,2]†, Chen Qi[1,2,5], Guy G. Poirier[6], Olga Pletnikova[7], Juan C. Troncoso[2,7], Lynn M. Bekris[8], James B. Leverenz[9], Alexander Pantelyat[2], Han Seok Ko[1,2,3], Liana S. Rosenthal[2], Ted M. Dawson[1,2,3,4,10]‡, Valina L. Dawson[1,2,3,10,11]‡

The pathologic accumulation and aggregation of α-synuclein (α-syn) underlies Parkinson's disease (PD). The molecular mechanisms by which pathologic α-syn causes neurodegeneration in PD are not known. Here, we found that pathologic α-syn activates poly(adenosine 5′-diphosphate–ribose) (PAR) polymerase-1 (PARP-1), and PAR generation accelerates the formation of pathologic α-syn, resulting in cell death via parthanatos. PARP inhibitors or genetic deletion of PARP-1 prevented pathologic α-syn toxicity. In a feed-forward loop, PAR converted pathologic α-syn to a more toxic strain. PAR levels were increased in the cerebrospinal fluid and brains of patients with PD, suggesting that PARP activation plays a role in PD pathogenesis. Thus, strategies aimed at inhibiting PARP-1 activation could hold promise as a disease-modifying therapy to prevent the loss of dopamine neurons in PD.

Parkinson's disease (PD) is an age-related neurodegenerative disease in which α-synuclein (α-syn) deposits as fibrils in intracytoplasmic inclusions in structures termed Lewy bodies and neurites (*1*). Recombinant α-syn can be aggregated in vitro to form fibrils similar in structure to those found in vivo (*2*), and these α-syn preformed fibrils (α-syn PFFs) can spread in a prion-like manner: both in in vitro neuronal cultures and in vivo when injected into the mouse brain with accompanying phosphorylation of α-syn on serine-129, a marker of pathologic α-syn (*3*) and neurotoxicity (*2*, *4*, *5*). Although it is clear that aggregated α-syn underlies the pathology of PD, what drives abnormal aggregation of α-syn and the cell injury and death mechanisms that are activated by this aggregation are not yet known. Because poly(adenosine 5′-diphosphate–ribose) (PAR) polymerase-1 (PARP-1) and PAR play a major contributing role in cell death relevant to neurologic disorders (*6*, *7*), here, we evaluated a role for PARP-1 and PAR in pathologic α-syn–induced neurodegeneration.

[1]Neuroregeneration and Stem Cell Programs, Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [2]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [3]Adrienne Helis Malvin Medical Research Foundation, New Orleans, LA 70130-2685, USA. [4]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [5]Department of Neurology, Xin Hua Hospital affiliated to Shanghai Jiaotong University School of Medicine, Shanghai 200092, China. [6]Centre de recherche du CHU de Québec-Pavillon CHUL, Faculté de Médecine, Université Laval, Québec G1V 4G2, Canada. [7]Department of Pathology (Neuropathology), Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [8]Lerner Research Institute, Genomic Medicine, Cleveland Clinic, Cleveland, OH 44195, USA. [9]Lou Ruvo Center for Brain Health, Neurological Institute, and Department of Neurology, Cleveland Clinic, Cleveland, OH 44195, USA. [10]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [11]Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
*These authors contributed equally to this work.
†Present address: Departmental of Pharmacology and Toxicology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294-6810, USA.
‡Corresponding author. Email: tdawson@jhmi.edu (T.M.D.); vdawson1@jhmi.edu (V.L.D.)

α-Syn PFF–induced neurotoxicity is PARP-1 dependent

To determine whether α-syn PFF induces the activation of PARP, levels of PAR were measured using a highly sensitive and specific PAR monoclonal antibody after administration of α-syn PFF to primary mouse cortical neurons (Fig. 1). α-Syn PFF (1 μg/ml) induced PARP activation peaks between 3 and 7 days and remained elevated for up to 14 days (Fig. 1A). The elevation of PAR was accompanied by neuron death as assessed by propidium iodide (PI) staining (Fig. 1, B and C). Treatment of cortical neurons with 1 μM of the PARP inhibitors [ABT-888 (veliparib), AG-014699 (rucaparib), or BMN 673 (talazoparib)] prevented the α-syn PFF–mediated PARP activation and cell death (Fig. 1, B to D). Consistent with known half-maximal inhibitory concentration values for inhibition of PARP-1 (5.2 nM ABT-888, 1.4 nM AG-014699, and 1.2 nM BMN 673) (*8*), 10 nM ABT-888, 1 nM AG-014699, or 1 nM BMN 673 partially prevented PARP-1 autoribosylation activity in vitro (fig. S1, A and B). Complete inhibition was observed at higher concentration of these inhibitors (fig. S1, A and B). These PARP inhibitors prevented α-syn PFF–induced cell death and PARP activation at concentrations as low as 10 nM (fig. S1, C to E). They also reduced α-syn PFF–mediated phosphorylation of α-syn at serine-129 (p–α-syn) (fig. S1, F and G) and α-syn aggregation (fig. S1, H and I), both of which are associated with pathology in α-synucleinopathies (*4*). Because PARP-1 plays a major role in parthanatos (*9*, *10*), we deleted PARP-1 from cortical neurons using CRISPR-Cas9 via adeno-associated virus (AAV) transduction carrying a guide RNA against PARP-1 (Fig. 1, E and F, and fig. S2A) (*11*) or used cortical cultures from PARP-1 knockouts (KOs) (Fig. 1, G and H, and fig. S2, B to F). Deletion or knockout of PARP-1 prevented α-syn PFF–mediated PARP activation and cell death (Fig. 1, E to H, and fig. S2, A and B). Knockout of PARP-1 also reduced p–α-syn immunostaining and α-syn aggregation (fig. S2, C to F). Treatment of cortical neurons with the broad-spectrum caspase inhibitor carbobenzoxy-valyl-alanyl-aspartyl-[*O*-methyl]-fluoromethylketone (Z-VAD-FMK) (Z-VAD) partially reduced α-syn PFF toxicity. The necroptosis inhibitor necrostatin-1 (Nec-1) and the autophagy inhibitor 3-methyladenine (3-MA) had no effect, whereas the PARP inhibitor ABT-888 prevented α-syn PFF toxicity (fig. S2, G to M). Because PARP inhibition and knockout of PARP-1 reduced the accumulation of pathologic α-syn as indicated by a reduction of p–α-syn immunostaining, we assessed cell-to-cell transmission of α-syn (*5*). Knockout of PARP-1 or PARP inhibition did not show a difference in the levels of α-syn–biotin PFF in endosomal-enriched fractions (fig. S3, A to D) (*5*), indicating that PARP-1 did not affect the uptake of α-syn PFF. However, knockout of PARP-1 reduced the cell-to-cell transmission of pathologic α-syn by inhibiting propagation of α-syn PFF into recipient cells (fig. S3, E to G).

α-Syn PFF activates PARP-1 via nitric oxide–induced DNA damage

To determine how α-syn PFF activates PARP-1, we measured levels of damaged DNA and activation of nitric oxide synthase (NOS) (Fig. 2 and fig. S4) (*12*–*14*). α-Syn PFF treatment increased NO levels in primary cultured neurons, whereas pretreatment with the NOS inhibitor, Nω-nitro-L-arginine methyl ester hydrochloride (L-NAME), prevented α-syn PFF–induced NO generation (Fig. 2A). NO levels also increased in α-syn PFF–injected brain (Fig. 2B). In both α-syn PFF–treated primary cultured neurons and α-syn PFF–injected brain, expression of γH2A.X, a marker of DNA strand breaks (*15*), was increased (Fig. 2, C to G, and fig. S4A).

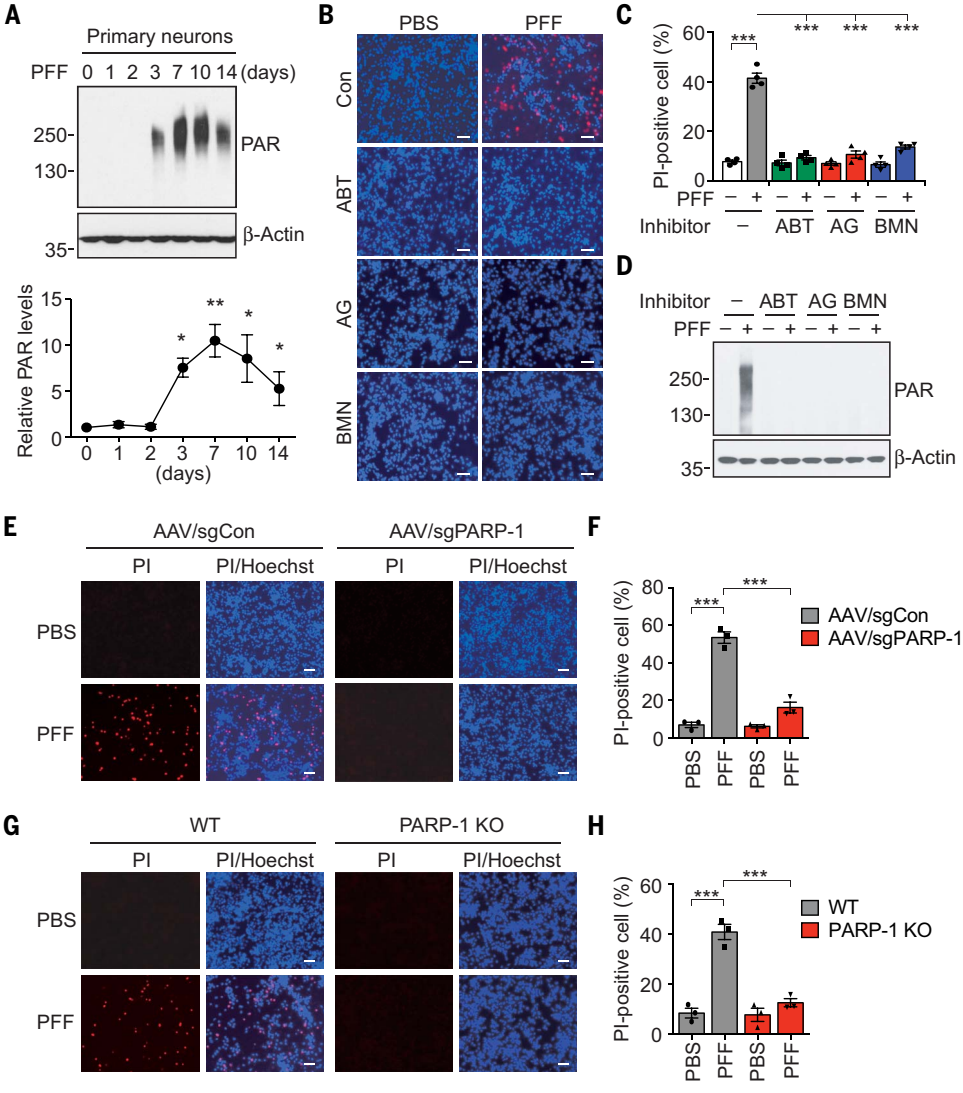


Fig. 1. α-Syn PFF induces parthanatos in neurons. (**A**) Activation of PARP-1 in α-syn PFF–treated primary cortical neurons. The representative Western blot analysis (top) and quantification (bottom) of the levels of PAR accumulation. Numbers to the left of the blot indicate molecular mass in kDa. Bars represent means ± SEM. One-way analysis of variance (ANOVA) followed by Tukey's post hoc test (n = 3 to 4). (**B**) Representative images of Hoechst and PI staining from primary cortical neurons pre-incubated with ABT-888 (1 μM), AG-014699 (1 μM), or BMN 673 (1 μM) for 1 hour and further incubated with α-syn PFF (5 μg/ml) for 14 days. Scale bars, 20 μm. (**C**) Quantification of cell death. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 4). (**D**) Inhibition of α-syn PFF–induced PAR accumulation was determined by Western blot analysis. (**E**) Representative images of Hoechst and PI staining from primary cortical neurons transduced with AAV containing single-guide RNA control (AAV-sgCon) or AAV–sgPARP-1 and further incubated with α-syn PFF for 14 days. Scale bar, 20 μm. (**F**) Quantification of cell death. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 3). (**G**) Representative images of Hoechst and PI staining from WT or PARP-1 KO primary cortical neurons and further incubated with α-syn PFF for 14 days. Scale bars, 20 μm. (**H**) Quantification of cell death. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 3). $^*P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$.

α-Syn PFF resulted in substantial numbers of neurons with DNA damage, whereas no such damage was detected by treatment with L-NAME (Fig. 2, C to E, and fig. S4, B to D). Consistent with suppression of PARP-1 activation (Fig. 2C), α-syn PFF–induced cell death was prevented by L-NAME (fig. S4, E to G). Thus, α-syn PFF activates NOS, leading to DNA damage and PARP-1 activation. Moreover, α-syn PFF primarily kills neurons through parthanatos, and PARP-1 contributes to generation of pathologic α-syn.

PARP-1 activation mediates α-syn PFF–induced loss of DA neurons

Because synthetic α-syn PFF killed primary cortical neurons via parthanatos, we sought to determine whether parthanatos plays a role in the loss of dopamine (DA) neurons after the intrastriatal injection of α-syn PFF (Fig. 3) (*4*, *5*). A single intrastriatal injection of α-syn PFF (5 μg) induced PARP activation and increased PAR levels (Fig. 3A). Intrastriatal injection of α-syn PFF into PARP-1 KO mice failed to increase PAR levels (Fig. 3A). As previously described (*4*), a single intrastriatal injection of α-syn PFF leads to an approximate 50% loss of DA neurons 6 months after the injection in wild-type (WT) mice (Fig. 3, B and C). In contrast, a single intrastriatal injection of α-syn PFF into PARP-1 KO mice failed to induce DA cell loss (Fig. 3, B and C). WT mice were also fed a diet containing the PARP inhibitor ABT-888 (125 mg/kg) and compared with mice given a control diet (Fig. 3, B and C). Mice treated with ABT-888 exhibited significantly less loss of DA neurons after an intrastriatal injection of α-syn PFF compared to mice on the control diet (Fig. 3, B and C). ABT-888 also reduced the formation of α-syn PFF–induced increase in PAR levels (fig. S5A). Tyrosine hydroxylase (TH) and DA transporter (DAT) levels were also reduced in WT mice in response to α-syn PFF, whereas the reduction in TH and DAT levels was prevented in PARP-1 KO or ABT-888–treated WT mice (fig. S5, A to C). The loss of DA neurons was accompanied by a reduction in striatal DA and its metabolites in WT mice, but not PARP-1 KO or ABT-888–treated mice (Fig. 3D and fig. S5, F to H). As previously described (*4*, *5*), injection of intrastriatal α-syn PFF leads to α-syn pathology in DA neurons of WT mice (figs. S5, A, D, and E, and S6, A to D). α-Syn pathology was markedly reduced in PARP-1 KO mice and ABT-888–treated WT mice consistent with the absence and reduction of neurodegeneration, respectively. Intrastriatal injection of α-syn PFF caused deficits on the pole test (*16*), a sensitive behavioral measurement of dopaminergic function, in WT mice, whereas there were no deficits in PARP-1 KO and ABT-888–treated WT mice (Fig. 3E and fig. S6E). Both forelimb plus hindlimb and forelimb grip strength were also reduced in WT mice after α-syn PFF injection, but not in PARP-1 KO or ABT-888–treated WT mice (Fig. 3F and fig. S6F). Thus, the striatal α-syn PFF–induced loss of DA neurons is dependent on PARP-1.

PAR accelerates α-syn fibrillization

Because PAR causes liquid demixing of intrinsically disordered proteins leading to their aggregation (*17*), experiments were performed to determine whether PAR could seed and accelerate

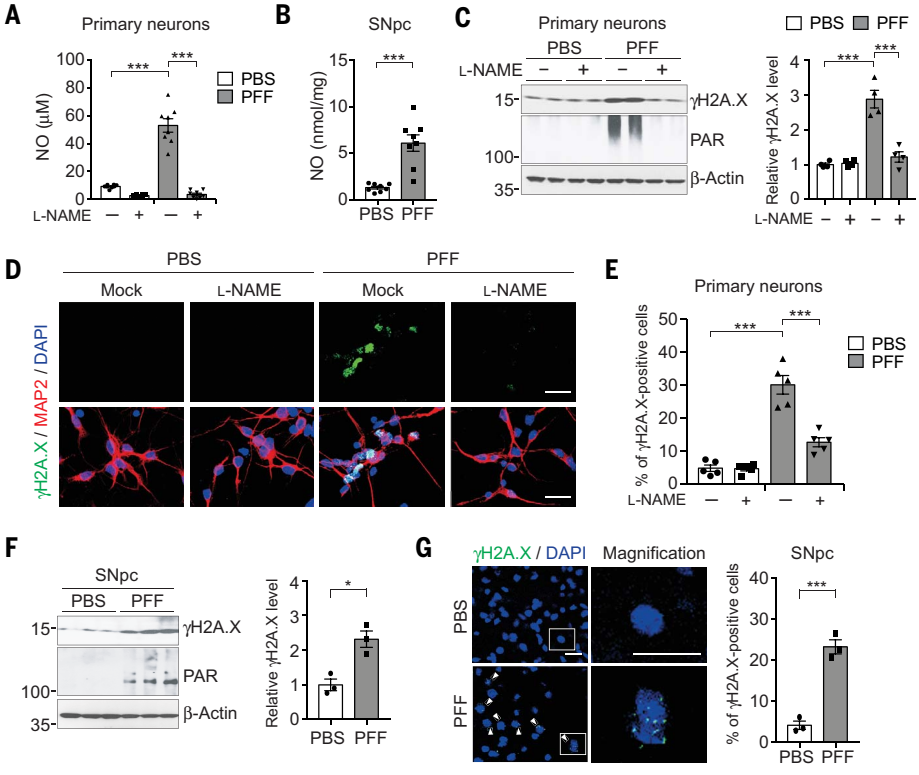


Fig. 2. Increased NO levels and damaged DNA by α-syn PFF. (**A**) NO (nitrite + nitrate) levels in primary cortical neurons preincubated with L-NAME for 1 hour and further incubated with α-syn PFF for 3 days. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 8). (**B**) NO (nitrite + nitrate) levels in SNpc of α-syn PFF–injected mice. Bars represent means ± SEM. Unpaired Student's t test (n = 8). (**C**) Representative immunoblots (left) and quantification (right) of γH2A.X and PAR levels in primary cortical neurons preincubated with L-NAME for 1 hour and further incubated with α-syn PFF. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 4). (**D**) Representative images of γH2A.X (green) and MAP2 (red) in primary cortical neurons preincubated with L-NAME for 1 hour and further incubated with α-syn PFF. Scale bars, 20 μm. (**E**) Quantification of γH2A.X-positive neurons. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 5). (**F**) Representative immunoblots (left) and quantification (right) of γH2A.X levels in the SNpc of α-syn PFF–injected mice. Bars are means ± SEM. Unpaired Student's t test (n = 3). (**G**) Representative images (left) and quantification (right) of γH2A.X-positive cells (green) in the SNpc of α-syn PFF–injected mice. Scale bars, 20 μm. Bars are means ± SEM. Unpaired Student's t test (n = 3). $^*P < 0.05$, $^{***}P < 0.0005$.

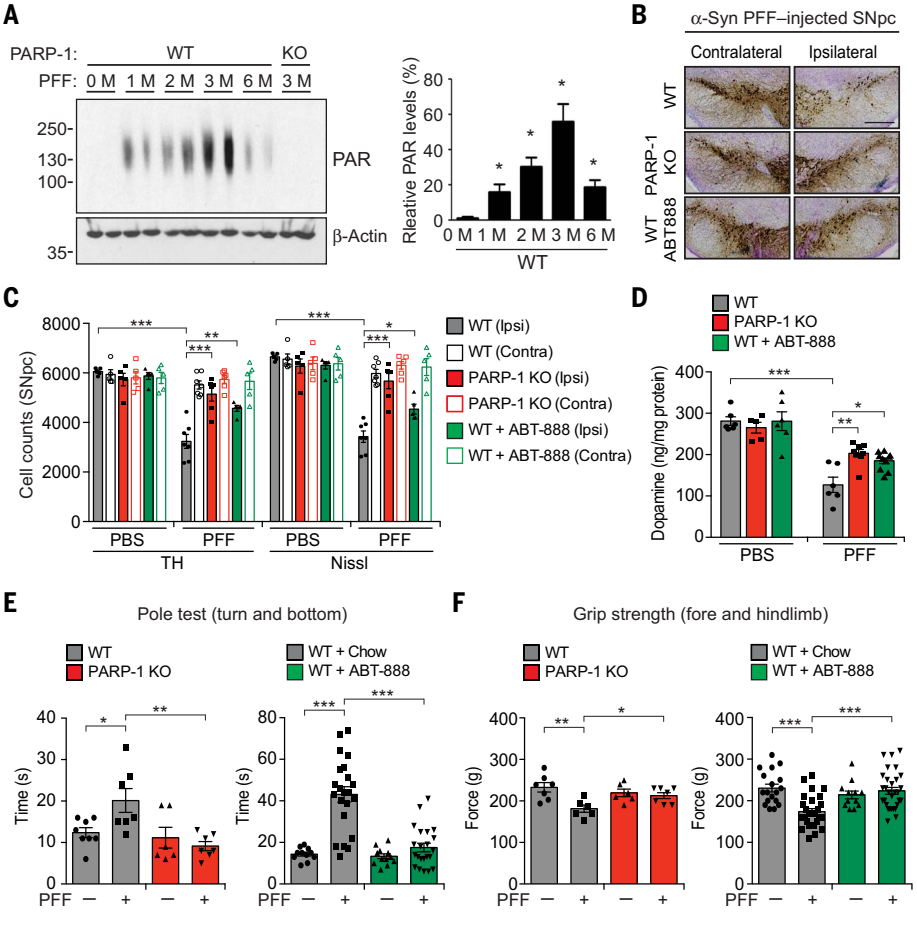


Fig. 3. α-Syn PFF–induced pathology is reduced by deletion of PAPR-1 or a PARP inhibitor, ABT-888, in vivo. (**A**) Representative immunoblots (left) and quantification (right) of the levels of PAR accumulation in the striatum of α-syn PFF–injected mice. Bars represent means ± SEM. One-way ANOVA followed by Tukey's post hoc test (n = 4). (**B**) Representative TH and Nissl staining of SNpc DA neurons of α-syn PFF–injected WT, PARP-1 KO, and WT mice fed with ABT-888 at 6 months after intrastriatal α-syn PFF or phosphate-buffered saline (PBS) injection. Scale bar, 400 μm. (**C**) Stereological counts. Data are means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 5 to 7 mice per group). (**D**) DA concentrations in the striatum of WT, PARP-1 KO, and WT mice fed with ABT-888 at 6 months after intrastriatal α-syn PFF or PBS injection measured by high-performance liquid chromatography (HPLC). Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 5 to 10 mice per group). (**E** and **F**) One hundred eighty days after α-syn PFF injection, the pole test (E) and grip strength test (F) were performed in WT, PARP-1 KO, or WT mice fed with ABT-888. Data are the means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 6 to 8 mice for PARP-1 KO, n = 11 to 30 mice for ABT-888). $^*P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$.

α-syn aggregation. Recombinant α-syn was incubated at 37°C and agitated in the presence and absence of 5 nM PAR, concentrations that are observed in brain tissue (*9*). High–molecular weight forms of α-syn in the absence of PAR were observed as early as 4 hours of incubation, and α-syn continued to fibrillize with time (Fig. 4A). Different molecular weight forms of α-syn were observed at 72 hours. In the presence of PAR, the fibrillization of α-syn was markedly accelerated with different molecular weight forms of α-syn being observed as early as 24 hours of incubation (Fig. 4A). PAR accelerated the fibrillization of α-syn as indicated by thioflavin T fluorescence, whereas PAR alone had no effect (Fig. 4B). PAR caused the fibrillization of α-syn at lower temperatures than required for α-syn fibrillization in the absence of PAR (fig. S7A). At 12 hours, α-syn fibrils were observed in the absence of PAR and were fully formed and extensive after 72 hours of agitation and incubation at 37°C (Fig. 4C). In contrast, in the presence of PAR, α-syn was extensively fibrillated at 12 hours and was more extensively fibrillated at 36 and 72 hours (Fig. 4C). We monitored the concentration dependence of fibrillization at 36 hours. PAR (1 nM) enhanced α-syn fibrillization, with peak aggregation occurring at 5 nM PAR, whereas 20 nM PAR did not appreciably increase α-syn fibrillization (fig. S7, B and C). α-Syn was not PARylated by PARP-1 activity, but it bound to PAR through its N-terminal domain (fig. S8, A to C). In the α-syn PFF–injected mouse brain, about 20% of α-syn was PAR bound (fig. S8, D and E). Because PAR is a highly negatively charged molecule, we tested the effects of another highly charged polymer, PolyA. α-Syn failed to interact with polyadenylic acid (PolyA), and it had no effect on α-syn fibrillization (fig. S7, D to F). ADP ribose monomer also failed to interact with α-syn and had no effect on α-syn fibrillization (fig. S7, D to F). To determine whether endogenous PAR formation accelerates α-syn fibrillization, primary mouse cortical neurons overexpressing WT human α-syn after AAV–α-syn transduction were treated with a toxic dose of *N*-methyl-D-aspartate (NMDA). In WT cultures, NMDA treatment

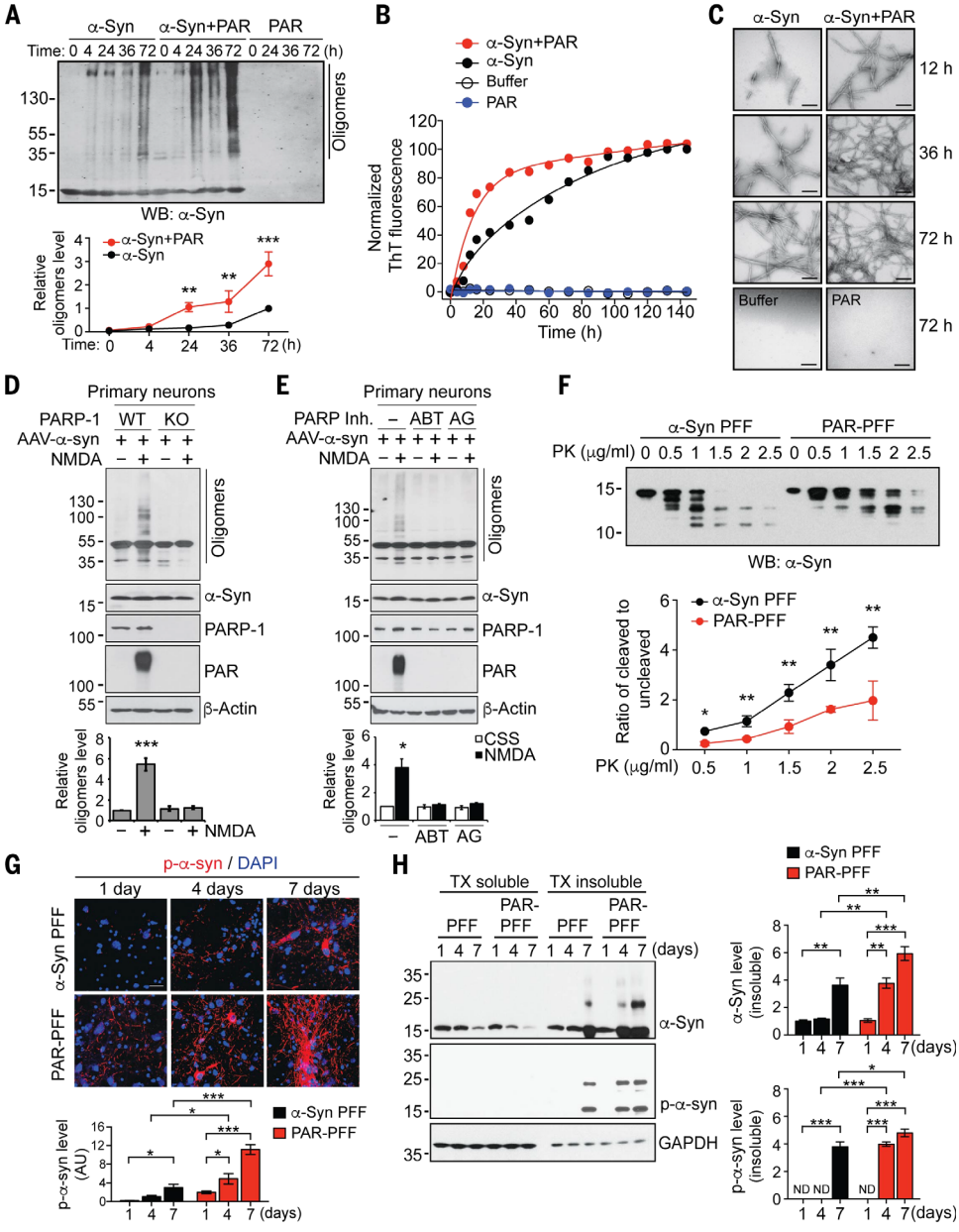


Fig. 4. PAR accelerates α-syn fibrillization in vitro. (**A**) Acceleration of α-syn fibrillization by PAR. Monomeric α-syn either with or without 5 nM purified PAR was incubated at 37°C for indicated times. Fibrillization of α-syn was detected by immunoblotting using an α-syn antibody. Data are means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 3). (**B**) The rate of formation of α-syn fibrils either with or without PAR was monitored by thioflavin T (ThT) fluorescence (n = 3). (**C**) Representative transmission electron microscopy (TEM) images for α-syn fibrils. Scale bars, 200 nm. (**D** and **E**) Suppression of NMDA-induced α-syn fibrillization in (D) PARP-1 KO neurons and by (E) PARP inhibitors. Primary cortical neurons from WT or PARP-1 KO embryos were transduced with AAV–α-syn and then further incubated with 500 μM NMDA for 5 min. ABT-888 (10 μM) or AG-014699 (1 μM) for 1 hour was pretreated for 1 hour. α-Syn fibrillization was detected by Western blot (WB) analysis 6 hours after NMDA treatment. (**F**) α-Syn PFF or PAR–α-syn PFF was incubated with increasing concentrations of PK (0 to 2.5 μg/ml) and immunoblotted with α-syn antibody (top). Quantification represents the ratio of cleaved to uncleaved α-syn (bottom). Data are means ± SEM. (bottom). Unpaired two-tailed Student's t test in each concentration of PK (n = 3). (**G**) Representative immunostaining and quantification of p–α-syn (red) in primary cortical neurons treated with α-syn PFF or PAR–α-syn PFF for 1, 4, and 7 days. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 3). Scale bar, 20 μm. (**H**) Primary cortical neurons treated with α-syn PFF or PAR–α-syn PFF were sequentially extracted with 1% Triton X-100 (TX-100) (TX soluble) and 2% SDS (TX insoluble). Lysates were subjected to immunoblotting using α-syn, p–α-syn, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) antibodies. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 3). ND, not detected. *P < 0.05, **P < 0.005, ***P < 0.0005. AU, arbitrary units; DAPI, 4′,6-diamidino-2-phenylindole.

activated PARP and led to a concomitant aggregation of α-syn, whereas α-syn did not aggregate in PARP-1 KO cultures treated with NMDA (Fig. 4D). Exogenous administration of PAR via BioPORTER (*9*) increased the aggregation of α-syn in both WT and PARP-1 KO cultures transduced with AAV–α-syn (fig. S9A). Thus, PAR, not PARP-1, can directly increase α-syn aggregation. Two different PARP inhibitors, ABT-888 and AG-014699, prevented α-syn aggregation and PARP activation in response to NMDA administration (Fig. 4E). In SH-SY5Y neuroblastoma cells, the potent PARP activator *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG) enhanced the aggregation of overexpressed WT or A53T α-syn, whereas MNNG had no effect in PARP-1 KO SH-SY5Y cells (fig. S9, B and C). Exogenous administration of PAR increased the aggregation of α-syn in both SH-SY5Y WT and SH-SY5Y PARP-1 KO cultures (fig. S9D). Two different PARP inhibitors, ABT-888 and AG-014699, prevented α-syn aggregation and PARP activation in response to MNNG administration in SH-SY5Y cells (fig. S9E). Thus, PAR seeds and accelerates α-syn aggregation.

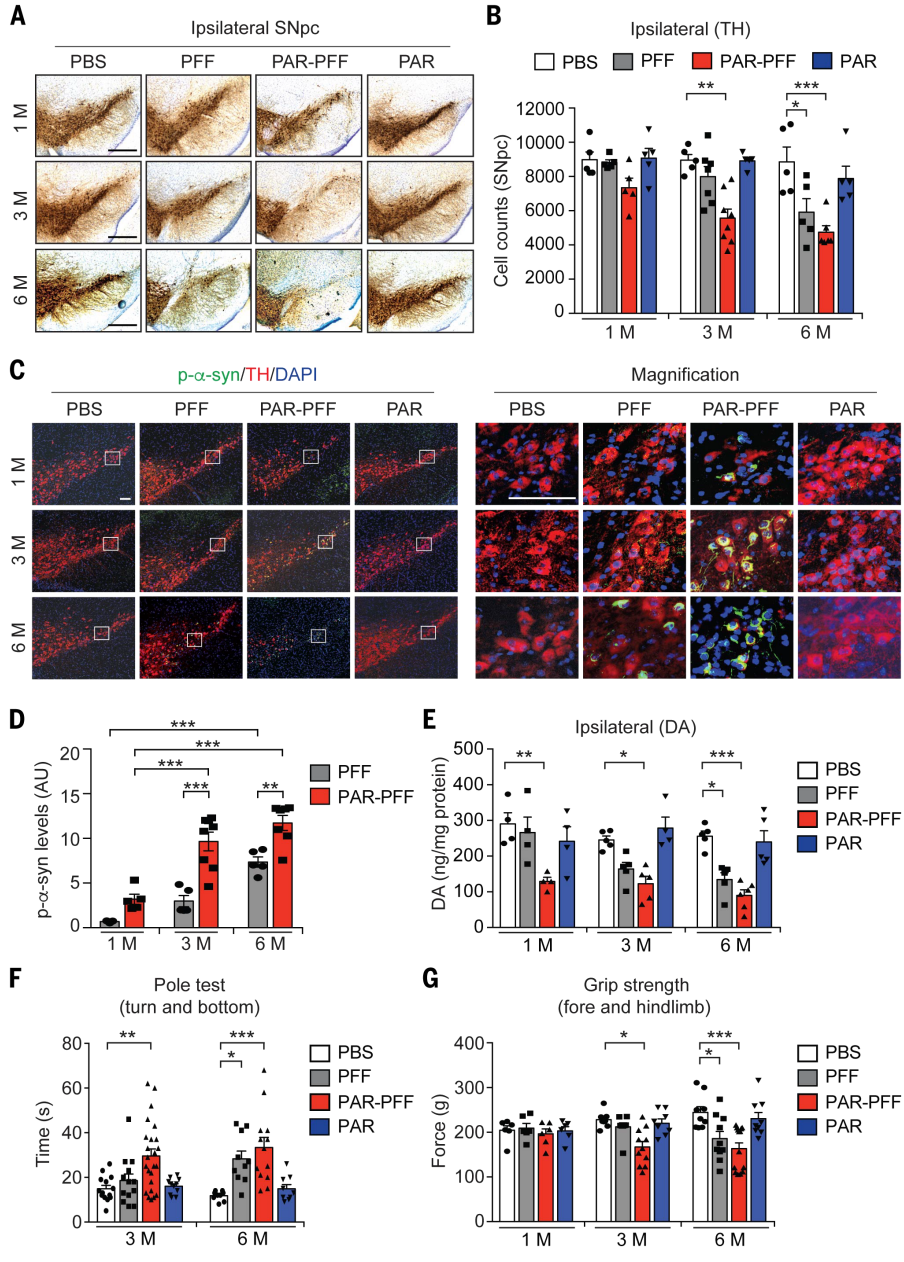


Fig. 5. PAR-α-syn PFF strains are more neurotoxic in vivo. (**A**) Representative TH and Nissl staining of SNpc DA neurons of WT mice at 1, 3, and 6 months after instrastriatal PBS, α-syn PFF, PAR–α-syn PFF, or PAR injection. Scale bars, 400 μm. (**B**) Stereological counts. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 5 to 8 mice per group). (**C**) Representative p–α-syn immunostaining in the SNpc of WT mice at 1, 3, and 6 months after instrastriatal PBS, α-syn PFF, PAR–α-syn PFF, or PAR injection. Scale bar, 100 μm. (**D**) Quantification of p–α-syn levels. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 5 to 8 mice per group). (**E**) DA concentrations in the striatum of PBS-, α-syn PFF–, PAR–α-syn PFF–, or PAR-injected mice at 1, 3, and 6 months measured by HPLC. Bars represent means ± SEM. Two-way ANOVA followed by Tukey's post hoc test (n = 4 to 6 mice per group). (**F** and **G**) Behavioral abnormalities of PBS-, α-syn PFF–, PAR–α-syn PFF–, or PAR-injected mice at 1, 3, and 6 months measured by the pole test (F, n = 9 to 24 mice per group) and the grip strength test (G, n = 6 to 14 mice per group). Data are the means ± SEM. Two-way ANOVA followed by Tukey's post hoc test. *P < 0.05, **P < 0.005, ***P < 0.0005.

PAR promotes formation of more toxic α-syn PFF strains

To determine whether PAR changes the biophysical properties of α-syn PFF, a series of biochemical analysis were conducted using α-syn PFF and α-syn PFF in the presence of PAR (PAR–α-syn PFF). First, proteinase K (PK) digestion of α-syn PFF was performed and monitored by α-syn immunoblots. α-Syn PFF and PAR–α-syn PFF showed very distinct banding patterns after PK digestion, with PAR–α-syn PFF's being more resistant to increasing concentrations of PK (Fig. 4F). PAR–α-syn PFF predominantly showed an undigested band of α-syn (first band), with comparable digested bands only at higher concentrations of PK, whereas α-syn PFF degraded into smaller fragments (second to fifth band) at lower concentrations of PK (0.5 and 1 μg/ml), and these bands became predominant at higher concentrations of PK (1.5 to 2.5 μg/ml) (Fig. 4F). Epitope-specific antibodies to α-syn revealed that PAR rendered the majority of the α-syn regions resistant to PK digestion (fig. S10A). The resistance to PK digestion of PAR–α-syn PFF suggests that PAR induces the formation of a distinct α-syn PFF strain with a more misfolded and compact structure than α-syn PFF. We then compared the α-syn PFF– and PAR–α-syn PFF–induced neuronal cell death in cultured neurons. After 14 days of treatment, cell death was enhanced in cultures treated with PAR–α-syn PFF as compared to that with α-syn PFF (fig. S10B). PAR itself did not cause significant cell death even at a higher concentration (20 nM) (fig. S10B). To further confirm the potencies of neuropathology, p–α-syn immunoreactivity was monitored after treatment with varying concentrations of α-syn PFF or PAR–α-syn PFF. P–α-syn immunoreactivity was observed at 20 ng of PAR–α-syn PFF at an equivalent level to 500 ng of α-syn PFF, suggesting that the PAR modification of α-syn PFF increased toxicity by 25-fold (fig. S10, C and D). PAR–α-syn PFF significantly increased p–α-syn immunoreactivity at higher concentrations (fig. S10, C and D). P–α-syn immunoreactivity at different time points was also monitored in cultured neurons exposed to α-syn PFF or PAR–α-syn PFF. In the absence of PAR, α-syn PFF treatment leads to barely detectable p–α-syn immunoreactivity 1 day after treatment, whereas PAR–α-syn PFF

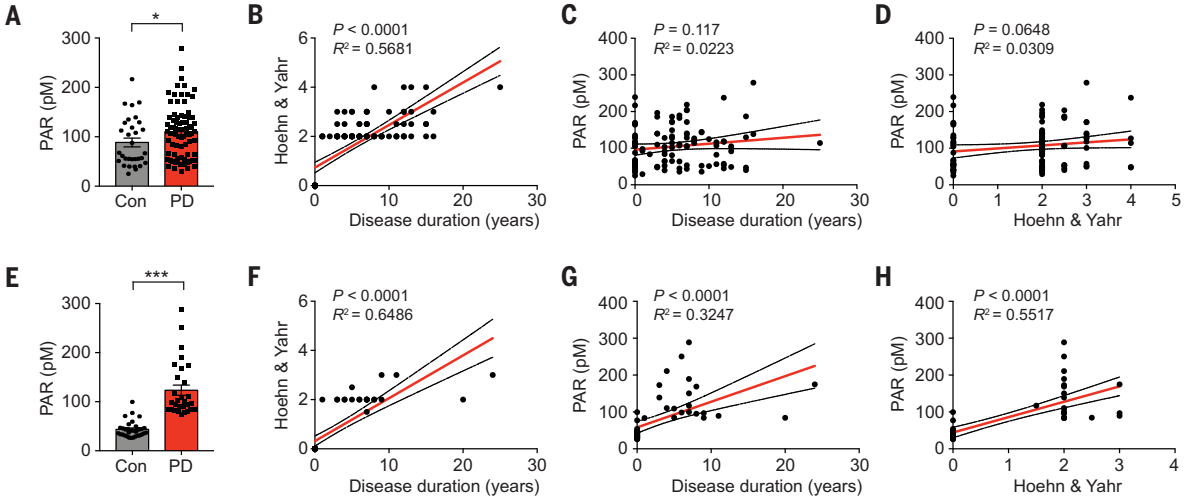


Fig. 6. Increase of PAR levels in CSF of patients with PD. (**A** and **E**) Increase of PAR in CSF of patients with PD. The levels of PAR in CSF of healthy controls (Con) (A, n = 31; E, n = 33) and patients with PD (A, n = 80; E, n = 21) were determined by PAR ELISA. Bars represent means ± SEM. Student's t test with Welch's correction (A) and Mann-Whitney U test (E). *P < 0.05, ***P < 0.0005. (**B** and **F**) Correlation analysis between disease duration and Hoehn & Yahr. (**C** and **G**) Correlation analysis between disease duration and PAR levels. (**D** and **H**) Correlation analysis between Hoehn & Yahr and PAR levels in the two independent patient cohorts used in (A) and (E). R^2, correlation coefficient.

treatment leads to detectable levels of p–α-syn immunoreactivity as early as 1 day after treatment and markedly enhances immunoreactivity at 7 days (Fig. 4G). Aggregated and phosphorylated α-syn were detectable at 4 days, whereas in the absence of PAR, these species of α-syn were only detectable after 7 days of treatment (Fig. 4H). In the presence of PAR, there was an increase in the aggregated form of α-syn (Fig. 4H). We then compared the cell-to-cell transmission of α-syn PFF and PAR–α-syn PFF in neuronal cultures transduced with AAV–α-syn (fig. S11). The exosomal secretion (*18*) of PAR-induced α-syn oligomers by NMDA treatment was significantly increased in WT neurons, but not in PARP-1 KO neurons or in WT neurons with treatment of a PARP inhibitor (fig. S11, A and B). PAR–α-syn PFF was enriched in the endosome-enriched fraction, and propagation of pathologic p–α-syn was enhanced in recipient cells as compared to those of α-syn PFF (fig. S11, C to F). Thus, PAR converts pathologic α-syn to a more misfolded compact toxic strain.

PAR–α-syn PFF strain is more neurotoxic in vivo

To determine whether the PAR–α-syn PFF strain exhibits enhanced neurotoxicity in vivo, a single intrastriatal injection of PAR–α-syn PFF (5 μg) was compared to that of α-syn PFF (5 μg). We observed a trend toward the loss of DA neurons ipsilateral to the injection side of SNpc after 1 month and a significant loss of DA neurons 3 months after PAR–α-syn PFF injection, whereas α-syn PFF injection had no effect at these time points (Fig. 5, A and B, and fig. S12A). Six months after PAR–α-syn PFF or α-syn PFF injection, there was no significant difference in the loss of DA neurons (Fig. 5, A and B, and fig. S12A). There was no significant loss of DA neuron contralateral to the injection side at any time point (fig. S12, B and C). PAR injection by itself had no effect on DA neuron number (Fig. 5, A and B, and fig. S12A). PAR–α-syn PFF also accelerated the loss of striatal DA and its metabolites, with significant reductions in DA and its metabolites 1 month after the PAR–α-syn PFF injection in contrast to α-syn PFF (Fig. 5E and fig. S13). Three and 6 months after PAR–α-syn PFF or α-syn PFF injection, there was no significant difference in the loss of DA and its metabolites (Fig. 5E and fig. S13). TH and DAT levels were also reduced after PAR–α-syn PFF compared to α-syn PFF 3 months after the injection, whereas there was no difference in the degree of loss at 6 months (fig. S14, A and B). α-Syn pathology as assessed by immunostaining for p–α-syn in DA neurons was increased by PAR–α-syn PFF compared to α-syn PFF at 3 and 6 months after injection (Fig. 5, C and D). PAR–α-syn PFF caused a deficit on the pole test at 3 months consistent with loss of DA neurons and DA deficits at 3 months, whereas there were no significant deficits in α-syn PFF–injected or PAR-injected mice (Fig. 5F and fig. S14C). Both forelimb plus hindlimb and forelimb grip strength were also reduced in PAR–α-syn PFF–injected mice but not in α-syn PFF–injected or PAR-injected mice at 3 months (Fig. 5G and fig. S14D). At 6 months, there was no significant difference in the behavioral deficits induced by PAR–α-syn PFF or α-syn PFF (Fig. 5, F and G, and fig. S14, C and D). Thus, PAR–α-syn PFF is substantially more neurotoxic than α-syn PFF in vivo.

Increased levels of PAR in the CSF of patients with PD

To determine whether PAR plays a role in patients with PD, PAR levels were monitored in the cerebrospinal fluid (CSF) of patients with PD versus controls (tables S1 and S2) using a sensitive enzyme-linked immunosorbent assay (ELISA) for PAR (fig. S15A). PAR levels were elevated in patients with PD compared to controls in two independent patient cohorts (Fig. 6, A and E). One of the cohorts showed a positive correlation between PAR levels and either disease duration or progression (Fig. 6, B to D and F to H). As previously reported, PAR levels were increased in the substantia nigra of patients with PD compared to controls (fig. S15, B and C, and table S3) (*19*).

Discussion

Our results indicate that α-syn PFF kills neurons both in vitro and in vivo via activation of PARP-1 in a cell death process designated parthanatos (*20*). Knockout of PARP-1 and inhibition of PARP prevents the neurodegeneration and behavioral deficits initiated by an intrastriatal α-syn PFF injection. Activation of parthanatos seems to be the primary driver of α-syn PFF neurodegeneration because necroptosis and autophagy inhibitors had no effect on α-syn PFF neurotoxicity, and there is only modest protection by caspase inhibition. It is known that α-syn PFFs induce inflammatory mediator activation (*21*, *22*), which likely contributes, in part, to cell death and accounts for the modest neuroprotection by the broad-spectrum caspase inhibitor Z-VAD. It will be important in future studies to explore the role of caspase activation, neuroinflammation, and neurodegeneration induced by α-syn PFF.

For the studies reported here, we used α-syn PFF (5 μg/ml) or PAR–α-syn PFF (5 μg/ml) for the primary neuronal culture experiments, which represents a concentration of about 340 nM in monomeric equivalents. The actual concentration of α-syn PFFs would be lower. Because α-syn is thought to represent about 1% of soluble brain protein (*23*), its monomeric concentration is estimated to be in the low micromolar range in the

brain. The concentration of higher-order species of α-syn, which is not known, would be lower in the brain. For both the primary neuronal cultures and the intrastriatal injections, 5 μg of α-syn PFF or PAR–α-syn PFF was used for each experiment. Because the concentrations used for these and related studies are within the range of the concentration of monomeric α-syn in the brain, the concentration of α-syn PFF or PAR–α-syn PFF is likely to be within the range of the concentration of higher-order species of α-syn in the brain. Providing some specificity to these experiments is the observation that the concentration of α-syn PFF used here and in other studies has no toxicity in the absence of endogenous α-syn (*2*, *4*).

Recent studies have identified conformational variants of α-syn strains that exhibit distinct neurotoxicity, seeding abilities, and propagation, which contribute to different properties of α-synucleinopathies (*24*, *25*). Given that α-syn PFF induces PARP activation and PAR accumulation, PAR then accelerates α-syn fibrillization and changes the biochemical properties of α-syn PFF, converting it to a more toxic strain. Consistent with this notion, PAR–α-syn PFF shows an approximate 25-fold increase in α-syn aggregation and neurotoxicity compared to the parental α-syn PFF. Moreover, PAR–α-syn PFF–injected mice show an accelerated disease progression and phenotype compared to α-syn PFF–injected mice.

In addition to PAR levels being increased in cultured neurons and mouse brain, PAR levels in PD are elevated in the substantia nigra and in the CSF. The elevation of PAR in the CSF and brains of patients with PD and evidence of PARP activation in the substantia nigra of patients with PD suggest that PARP activation contributes to the pathogenesis of PD through parthanatos and conversion of α-syn to a more toxic strain. In future studies, it will be important to determine whether the increase of PAR in CSF from human PD correlates with disease severity or progression. Moreover, it will be important to determine whether it can serve as a theranostic biomarker for disease-modifying therapies. Because PARP inhibitors are currently being used clinically as synergizing agents in the treatment of cancer (*26*), they could be considered for disease modification in PD (*6*).

Materials and methods

Animals

C57BL/6 WT and PARP-1 KO mice were obtained from the Jackson Laboratories (Bar Harbor, ME). The littermates of WT and PARP-1 KO mice were used in experiments. All housing, breeding, and procedures were performed according to the NIH Guide for the Care and Use of Experimental Animals and approved by Johns Hopkins University Animal Care and Use Committee.

Preparation of α-syn PFF and PAR-α-syn PFF

Recombinant mouse α-syn proteins were purified as previously described (*5*). After purification, bacterial endotoxins were removed using Toxineraser endotoxin removal kit (GeneScript). α-syn PFF were prepared in PBS by constantly agitating α-syn with a thermomixer (1,000 rpm at 37°C) (Eppendorf, Hamburg, Germany). After 7 days of incubation, the α-syn aggregates were diluted to 0.1 mg/ml with PBS and sonicated for 30 s (0.5 sec pulse on/off) at 10% amplitude (Branson Digital Sonifier, Danbury, CT). Synthesis and purification of PAR polymer were performed as described (*27*). PAR-α-syn PFF was prepared by adding 5 nM or indicated dose of PAR in α-syn fibrillization reaction.

Stereotaxic injection of α-syn PFF

Two to 3-month-old WT and PARP-1 KO mice were deeply anesthetized with a mixture of ketamine (100 mg/kg) and xylazine (10 mg/kg). PBS, α-syn PFF (5 μg), PAR-α-syn PFF (5 μg) or PAR was unilaterally injected into striatum (2 μl per hemisphere at 0.4 μl/min) with the following coordinates: anteroposterior (AP) = +0.2 mm, mediolateral (ML) = + 2.0 mm, dorsoventral (DV) = +2.8 mm from bregma. After the injection, the needle was maintained for an additional 5 min for a complete absorption of the solution. After surgery, animals were monitored and post-surgical care was provided. Behavioral tests were performed 1, 3 and 6 months after injection and mice were euthanized for biochemical and histological analysis. For biochemical studies, tissues were immediately dissected and frozen at -80°C. For histological studies, mice were perfused with PBS and 4% PFA and brains were removed, followed by fixation in 4% PFA overnight and transfer to 30% sucrose for cryoprotection.

Thioflavin T (ThT) binding assay

α-syn fibrillization with or without PAR was monitored with ThT fluorescence. Aliquots of 5 ul from the incubation mixture were taken at various time points, diluted to 100 μl with 25 μM ThT in PBS, and incubate for 10 min at room temperature. The fluorescence was recorded at 450 nm excitation and 510 nm emission using SpectraMax plate reader (Molecular Devices, Sunnyvale, CA). The experiments were performed in triplicate.

Transmission electron microscopy (TEM) measurements

α-syn PFF or PAR-α-syn PFF were adsorbed to glow discharged 400 meshed carbon coated copper grids (Electron Microscopy Sciences, Hatfield, PA) for 2 min, quickly washed twice with Tris-HCl (50 mM, pH 7.4), and floated upon two drops of 0.75% uranyl formate for 30 s each. The grids were allowed to dry before imaging on a Phillips CM 120 TEM operating at 80 kV. The images were captured and digitized with an ER-80 CCD (8 megapixel) by advanced microscopy techniques.

Intracellular delivery of PAR

Purified PAR was intracellularly delivered using BioPORTER (Genelnatis, San Diego, CA) according to the manufacturer's instructions (*9*). PAR polymer was diluted to desired concentration with PBS. The diluted solution was added to the dried BioPORTER reagent and mixed gently, followed by incubation at room temperature for 5 min. The BioPORTER-PAR complex was added to cell culture after a wash in serum-free media and incubated for 3–4 h at 37°C. Cultures were subsequently used for experiments.

Tissue lysate preparation and Western blot analysis

Human post mortem brain (Table S3) or mouse brain tissue were homogenized and prepared in lysis buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% Triton x-100, 0.5% SDS, 0.5% sodium-deoxycholate, phosphatase inhibitor mixture I and II (Sigma-Aldrich, St. Louis, MO), and complete protease inhibitor mixture (Roche, Indianapolis, IN)], using a Diax 900 homogenizer (Sigma-Aldrich). After homogenization, samples were rotated at 4°C for 30 min for complete lysis, the homogenate was centrifuged at 15,000 x g for 20 min and the supernatants were used for further analysis. Protein levels were quantified using the BCA assay (Pierce, Rockford, IL), samples were separated using SDS-polyacrylamide gels and transferred onto nitrocellulose membranes. The membranes were blocked with 5% non-fat milk in TBS-T (Tris-buffered saline with 0.1% Tween-20) for 1 h, probed using primary antibodies (Table S4) and incubated with appropriate HRP-conjugated secondary antibodies (Cell signaling, Danvers, MA). The bands were visualized by ECL substrate.

Immunoprecipitation (IP), overlay, and in vitro ribosylation assay of PAR

The deletion mutants of α-syn-GFP were generated using site-directed mutagenesis (Agilent, Santa Clara, CA) according to the manufacturer's instructions. SH-SY5Y cells were transfected with either α-syn-GFP or its deletion mutants by Polyfect (Qiagen, Hilden, Germany) for 36 h and then further incubated with 50 μM MNNG for 15 min. The cells were washed 2 times with PBS and harvested with lysis buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% Triton x-100, 0.5% SDS, 0.5% sodium-deoxycholate, phosphatase inhibitor mixture I and II (Sigma-Aldrich), and complete protease inhibitor mixture (Roche)]. The supernatants after centrifugation at 15,000 x g for 20 min were used for measuring the protein concentration by BCA assay. The same amount of proteins were incubated with anti-PAR antibody (Table S4) overnight at 4°C, followed by incubation with PureProteome kappa Ig binder magnetic beads (Millipore, Burlington, MA) for 3 h at 4°C. The IP complexes were washed 5 times with IP buffer and then denatured by boiling for 5 min after adding 2x Laemlli Buffer plus β-mercaptoethanol. For the PAR overlay assay, the equal amount of BSA or purified α-syn protein was spotted onto a nitrocellulose membrane. The membranes were washed once with TBS-T buffer and air-dried, followed by incubation with biotin-labeled PAR polymer for 1 h at room temperature in TBS-T buffer.

After washing 3 times in TBS-T buffer, the membranes were blocked with 5% non-fat milk for 1 h, probed using anti-PAR antibody. The bands were visualized by ECL substrate. For PARP-1 in vitro ribosylation assay, 1 μg of recombinant PARP-1, activated DNA and NAD^+ (Trevigen, Gaithersburg, MD) in the presence or absence of indicated concentration of PARP inhibitors were incubated in the PARP assay buffer (Trevigen) for 30 min at room temperature. For the substrate ribosylation assay of PARP-1, we used 2 or 10 μg of purified α-syn and histone 2B (Abcam) protein under the same condition.

Endosome and exosome enrichment

Endosomes were enriched to detect internalized α-syn-biotin PFF as previously described (*5*). Primary cultured neurons from WT or PARP-1 KO were incubated with α-syn-biotin PFF for 1.5 h, followed by adding trypsin to remove the membrane-bound α-syn-biotin PFF. After washing with PBS, neurons were harvested and lysed by aspirating a syringe 20 times in lysis buffer [250 mM sucrose, 50 mM Tris-HCl (pH 7.4), 5 mM MgCl2, 1mM EDTA, 1mM EGTA] with a complete protease inhibitor mixture (Roche). The last pellet containing the endosomes were used for immunoblot analysis after sequential centrifugation at 1000 x g for 10 min, 16,000 x g for 20 min, and 100,000 x g for 60 min at 4°C). Exosomes were enriched to detect secreted α-syn as previously described (*18*). Culture supernatants of primary cortical neurons transduced with AAV-α-syn followed by incubation with 500 μM NMDA for 5 min were collected and spun at 300 x g for 10 min to remove cells. The supernatants were then sequentially centrifuged at 2000 × g for 10 min, 10,000 × g for 30 min, and 100,000 × g for 90 min at 4°C). The last pellet containing exosomes was washed once with PBS and centrifuged again at 100,000 × g for 90 min. The remaining pellet was resuspended with lysis buffer.

Cell culture, transfection, primary neuronal culture and treatment

SH-SY5Y cells (ATCC, Manassas, VA) were cultured in DMEM containing 10% fetal bovine serum and penicillin/streptomycin at 37°C under 5% CO_2. The cells were transfected using PolyFect reagent (Qiagen) according to the manufacturer's instructions. Primary cortical neurons from WT or PARP-1 KO embryos were prepared as described previously (*28*). Briefly, the primary cortical neurons were cultured at embryonic day 16 in neurobasal media supplemented with B-27, 0.5 mM L-glutamine, penicillin and streptomycin (Invitrogen, Carlsbad, CA). ABT-888 (1 or 10 μM), AG-014699 (1 μM), BMN 673 (1 or 10 μM), Z-VAD (10 μM), Nec-1 (10 μM) or 3-MA (500 μM) were applied to neurons 1 h before α-syn PFF treatment. The neuron culture media was replaced with fresh medium alone or including cell death inhibitors every 3-4 days. α-syn PFF was added at 7 days in vitro (DIV) and further incubated for indicated times followed by the cell death assay or biochemical experiments. Primary neurons were infected with AAV9-control sgRNA or AAV9-PARP-1 sgRNA (ViGene Biosciences, Rockville, MD), and AAV-α-syn WT or AAV-α-syn A53T at DIV 4-5.

Cell death and viability assessment

Primary cultured cortical neurons were treated with 5 μg/ml of α-syn PFF or PAR-α-syn PFF for 14 days. Percent of cell death was determined by staining with 7 μM Hoechst 33342 and 2 μM propidium iodide (PI) (Invitrogen). Images were taken and counted by a Zeiss microscope equipped with automated computer assisted software (Axiovision 4.6, Carl Zeiss, Dublin, CA). After adding Alamar Blue (Invitrogen), cell viability was determined by fluorescence at an excitation wavelength 570 nm and an emission wavelength 585 nm (*29*).

Microfluidic chambers

Triple compartment microfluidic devices (TCND1000) were obtained from Xona Microfluidic, LLC (Temecula, CA). Glass coverslips were prepared and coated as described, before being affixed to the microfluidic device. Approximately 100,000 WT or PARP-1 KO neurons were plated per chamber individually. At 7 DIV, 5 μg/ml of α-syn PFF was added into chamber 1. To control for direction of flow, a 50 μl difference in media volume was maintained between chamber 1 and chamber 2 and chamber 2 and chamber 3 according to the manufacturers' instructions. Neurons were fixed on day 14 after α-syn PFF treatment using 4% paraformaldehyde in PBS. The chambers were then processed for immunofluorescence staining with a p-α-syn antibody (Table S4).

Behavioral tests

The behavioral deficits in α-syn PFF injected WT or PARP-1 KO mice, α-syn PFF injected mice fed ABT-888, and α-syn PFF or PAR-α-syn PFF injected mice were assessed by the pole test and the grip strength test 1 week prior to sacrifice of the different cohorts. All the experiments were performed by investigators who were blind to genotypes or treatment condition and randomly allocated to groups.

Pole test

A metal rod (75 cm long with a 9 mm diameter) wrapped with bandage gauze was used as the pole. Before the actual test, the mice were trained for two consecutive days and each training session consisted of three test trials. Mice were placed 7.5 cm from the top of the pole. The time to turn and total time to reach the base of the pole were recorded. The end of test was defined by placement of all 4 paws on the base. The maximum cutoff time to stop the test and recording was 60 s. After each trial, the maze was cleaned with 70% ethanol. In Fig. 3E, we used 4 males and 4 females for WT mice injected with PBS, 4 males and 3 females for WT mice injected with PFF, 3 males and 3 females for PARP-1 KO mice injected with PBS, and 5 males and 2 females for PARP-1 KO mice injected with PFF (Fig. 3E, left). We used 6 males and 5 females for PBS injected mice fed with chow, 16 males and 9 females for PFF injected mice fed with chow, 6 males and 6 females for PBS injected mice fed with ABT-888, and 15 males and 10 females for PFF injected mice fed with ABT-888 (Fig. 3E, right). In Fig. 5F, we used 7 males and 7 females for PBS, 8 males and 5 females for PFF, 13 males and 11 females for PAR-PFF, 6 males and 5 females for PAR-injected mice at 3 months, and 4 males and 5 females for PBS, 5 males and 5 females for PFF, 7 males and 6 females for PAR-PFF, 5 males and 4 females for PAR-injected mice at 6 months.

Grip strength test

Neuromuscular function was measured by determining the maximal peak force developed by the mice using a grip-strength meter (Bioseb, USA). Mice were placed onto a metal grid to grasp with either fore or both limbs that are recorded as 'fore limb' and 'fore and hindlimb', respectively. The tail was gently pulled and the force applied to the grid before the mice lose their grip was recorded as the peak tension displayed in grams (g). In Fig. 3F, we used 4 males and 3 females for WT mice injected with PBS, 3 males and 3 females for WT mice injected with PFF, 3 males and 3 females for PARP-1 KO mice injected with PBS, and 5 males and 2 females for PARP-1 KO mice injected with PFF (Fig. 3F, left). We used 10 males and 8 females for PBS injected mice fed with chow, 18 males and 9 females for PFF injected mice fed with chow, 6 males and 6 females for PBS injected mice fed with ABT-888, and 19 males and 11 females for PFF injected mice fed with ABT-888 (Fig. 3F, right). In Fig. 5G, we used 3 males and 3 females for PBS, 3 males and 3 females for PFF, 3 males and 3 females for PAR-PFF, 3 males and 3 females for PAR-injected mice at 1 months, 4 males and 4 females for PBS, 4 males and 4 females for PFF, 6 males and 5 females for PAR-PFF, 4 males and 4 females for PAR-injected mice at 3 months, and 5 males and 4 females for PBS, 5 males and 5 females for PFF, 8 males and 6 females for PAR-PFF, 5 males and 4 females for PAR-injected mice at 6 months.

Dopamine and derivatives measurement using HPLC

Biogenic amine concentrations were measured by high-performance liquid chromatography with electrochemical detection (HPLC-ECD). The striatum was rapidly removed from the brain, followed by weighing, then sonication in ice cold 0.01 mM of perchloric acid containing 0.01% EDTA. The 60 ng of 3,4-dihydroxybenzylamine (DHBA) was used as an internal standard. After centrifugation at 15,000 g for 30 min at 4°C, the supernatant was cleaned using a 0.2 μm filter and 20 μl of supernatant was analyzed in the HPLC column (3 mm × 150 mm, C-18 reverse phase column, AcclaimTM Polar Advantage II, Thermo Scientific) by a dual channel coulochem III electrochemical detector (Model 5300, ESA, Inc. Chelmsford, MA). The protein concentrations of tissue homogenates were measured using the BCA protein assay kit (Pierce).

Data were normalized to protein concentrations and expressed in ng/mg protein.

Immunohistochemistry and immunofluorescence

Mice were perfused with PBS and 4% PFA and brains were removed, followed by fixation in 4% PFA overnight and transfer to 30% sucrose for cryoprotection. Immunohistochemistry (IHC) and immunofluorescence (IF) was performed on 40 μm thick serial brain sections. Primary antibodies and working dilutions are detailed in Table S4. For histological studies, free-floating sections were blocked with 10% goat serum in PBS with 0.2% Triton X-100 and incubated with TH or p-α-syn antibodies followed by incubation with biotin-conjugated anti-rabbit or mouse antibody, respectively. After three washes, ABC reagent (Vector laboratories, Burlingame, CA) was added and the sections were developed using SigmaFast DAB peroxidase substrate (Sigma-Aldrich). Sections were counterstained with Nissl (0.09% thionin). For the quantification, both TH- and Nissl-positive DA neurons from the SNpc region were counted by an investigator who was blind to genotypes or treatment condition with randomly allocated groups through optical fractionators, the unbiased method for cell counting, using a computer-assisted image analysis system consisting of an Axiophot photomicroscope (Carl Zeiss) equipped with a computer controlled motorized stage (Ludl Electronics, Hawthorne, NY), a Hitachi HV C20 camera, and Stereo Investigator software (MicroBright-Field, Williston, VT). The total number of TH-stained neurons and Nissl counts were analyzed as previously described (*16*). For immunofluorescent studies, double-labeled sections with TH and p-α-syn antibodies were incubated with a mixture of Alexa-fluor 488- and 594-conjugated secondary antibodies (Invitrogen). The fluorescent images were acquired by confocal scanning microscopy (LSM710, Carl Zeiss). All the images were processed by the Zen software (Carl Zeiss). The selected area in the signal intensity range of the threshold was measured using ImageJ software.

Nitric oxide (NO) measurement

NO levels were measured using a NO assay kit (Abcam, Cambridge, MA) according to the manufacturer's instructions. Briefly, the pellets of primary neurons or SNpc tissues were washed with ice-cold PBS, resuspended, and homogenized in ice-cold assay buffer. The supernatant after centrifugation at 16,000 x g for 5 min at 4°C were deproteinized and neutralized (pH 6.5-8) by adding 1M perchloric acid and 2M Potassium Hydroxide. After centrifugation at 13,000 x g for 15 min at 4°C, the supernatant was reacted with nitrate reductase and cofactors at room temperature for 1 h to convert nitrate to nitrite. The amount of the azo compound converted from nitrite by Griess reagents were measured at OD540 nm.

Comet assay

The comet assay was performed according to the manufacturer's instructions (Trevigen). Briefly, primary cortical neurons pre-treated with L-NAME (Sigma-Aldrich) followed by further incubation with α-syn PFF were washed with ice-cold PBS (Ca^{2+} and Mg^{2+} free), harvested, and resuspended in PBS at 1 x 10^5 cells/ml. After combining with 1% low melting point agarose in PBS at 42°C, 50 μl of the cell-agarose mixture was immediately placed on the Comet slide at 4°C in the dark for 30 min. Slides were lysed in lysis buffer and immersed with alkaline unwinding solution (200 mM NaOH, pH >13, 1 mM EDTA) for 1 hour at room temperature. After electrophoresis at 21 V for 30 min at 4°C, slides were rinsed twice with dH_2O), fixed with 70% ethanol for 5 min, and then stained with SYBR green for 5 min at 4°C. Images were captured using a Zeiss epifluorescent microscope (Axiovert 200M) and analyzed with the tail positive cells (% of total cells) and tail length (the length from the edge of the nucleus to the end of the comet tail).

Proteinase K (PK) digestion of α-syn PFF

PK digestion was performed as previously described (*30*). Ten micrograms of α-syn PFF or PAR-α-syn PFF were mixed with 0.5 to 2.5 μg/ml of PK in PBS and incubated at 37°C for 30 min. The reaction was stopped by adding 1 mM PMSF, boiled with SDS-sample buffer for 5 min. The bands of the PK digestion products were detected by immunoblotting using epitope-specific α-syn antibodies (Table S4).

Human CSF samples and PAR ELISA

Participants at the Johns Hopkins University site of the NINDS Parkinson's Disease Biomarker Program (PDBP) underwent extensive clinical and cognitive testing and a lumbar puncture annually. The CSF was centrifuged, aliquoted, and stored at -80°C within one hour of acquisition. CSF samples were also obtained from the Cleveland Clinic Lou Ruvo Center Brain Health Biobank (CBH-biobank) under similar protocols. Two different clones (#19 and #25) of monoclonal anti-PAR antibody were used for PAR ELISA. Anti-PAR antibody (capture antibody, clone #19) (5 μg/ml) was coated on 96-well microtiter plate (NUNC, Cat #46051), various concentration of purified PAR (0-200 nM, positive control) and CSF samples from either normal or PD patients were added to each well and incubated for 1 h at room temperature (RT). After washing the plates five times with PBST (0.05% Tween20 in PBS buffer), the biotinylated PAR antibody (detection antibody, clone #25) was incubated for 1 h at RT. The color change was detected via HRP-conjugated streptavidin antibody (Thermo Scientific). Our assay can detect the PAR as low as 3 pM and is saturated at 50 nM.

Statistical analysis

All data are represented as mean ± s.e.m. with at least 3 independent experiments. Statistical analysis was performed using GraphPad Prism 7. Differences between 2 means and among multiple means were assessed by unpaired two-tailed student *t* test and ANOVA followed by Tukey's post hoc test, respectively. The distribution of data from human CSF samples were assessed with D'Agostino & Pearson omnibus normality test and non-normally distributed data were analyzed with nonparametric test (Mann-Whitney test). Assessments with $P < 0.05$ were considered significant.

REFERENCES AND NOTES

1. M. Baba *et al*., Aggregation of α-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am. J. Pathol.* **152**, 879–884 (1998). pmid: 9546347
2. L. A. Volpicelli-Daley *et al*., Exogenous α-synuclein fibrils induce Lewy body pathology leading to synaptic dysfunction and neuron death. *Neuron* **72**, 57–71 (2011). doi: 10.1016/j.neuron.2011.08.033; pmid: 21982369
3. H. Fujiwara *et al*., α-Synuclein is phosphorylated in synucleinopathy lesions. *Nat. Cell Biol.* **4**, 160–164 (2002). doi: 10.1038/ncb748; pmid: 11813001
4. K. C. Luk *et al*., Pathological α-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **338**, 949–953 (2012). doi: 10.1126/science.1227157; pmid: 23161999
5. X. Mao *et al*., Pathological α-synuclein transmission initiated by binding lymphocyte-activation gene 3. *Science* **353**, aah3374 (2016). doi: 10.1126/science.aah3374; pmid: 27708076
6. N. A. Berger *et al*., Opportunities for the repurposing of PARP inhibitors for the therapy of non-oncological diseases. *Br. J. Pharmacol.* **175**, 192–222 (2018). doi: 10.1111/bph.13748; pmid: 28213892
7. T. M. Dawson, V. L. Dawson, Mitochondrial mechanisms of neuronal cell death: Potential therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **57**, 437–454 (2017). doi: 10.1146/annurev-pharmtox-010716-105001 pmid: 28061689
8. M. Sisay, D. Edessa, PARP inhibitors as potential therapeutic agents for various cancers: Focus on niraparib and its first global approval for maintenance therapy of gynecologic cancers. *Gynecol. Oncol. Res. Pract.* **4**, 18 (2017). doi: 10.1186/s40661-017-0055-8; pmid: 29214031
9. S. A. Andrabi *et al*., Poly(ADP-ribose) (PAR) polymer is a death signal. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18308–18313 (2006). doi: 10.1073/pnas.0606526103; pmid: 17116882
10. S.-W. Yu *et al*., Mediation of poly(ADP-ribose) polymerase-1-dependent cell death by apoptosis-inducing factor. *Science* **297**, 259–263 (2002). doi: 10.1126/science.1072221; pmid: 12114629
11. L. Swiech *et al*., In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* **33**, 102–106 (2015). doi: 10.1038/nbt.3055; pmid: 25326897
12. A. A. Fatokun, V. L. Dawson, T. M. Dawson, Parthanatos: Mitochondrial-linked mechanisms and therapeutic opportunities. *Br. J. Pharmacol.* **171**, 2000–2016 (2014). doi: 10.1111/bph.12416; pmid: 24684389
13. J. Zhang, V. L. Dawson, T. M. Dawson, S. H. Snyder, Nitric oxide activation of poly(ADP-ribose) synthetase in neurotoxicity. *Science* **263**, 687–689 (1994). doi: 10.1126/science.8080500; pmid: 8080500
14. M. Y. Kim, T. Zhang, W. L. Kraus, Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD^+ into a nuclear signal. *Genes Dev.* **19**, 1951–1967 (2005). doi: 10.1101/gad.1331805; pmid: 16140981
15. W. M. Bonner *et al*., γH2AX and cancer. *Nat. Rev. Cancer* **8**, 957–967 (2008). doi: 10.1038/nrc2523; pmid: 19005492
16. S. S. Karuppagounder *et al*., The c-Abl inhibitor, nilotinib, protects dopaminergic neurons in a preclinical animal model of Parkinson's disease. *Sci. Rep.* **4**, 4874 (2014). doi: 10.1038/srep04874; pmid: 24786396
17. M. Altmeyer *et al*., Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088 (2015). doi: 10.1038/ncomms9088; pmid: 26286827
18. E. Emmanouilidou *et al*., Cell-produced α-synuclein is secreted in a calcium-dependent manner by exosomes and impacts neuronal survival. *J. Neurosci.* **30**, 6838–6851 (2010). doi: 10.1523/JNEUROSCI.5699-09.2010; pmid: 20484626
19. Y. Lee *et al*., Parthanatos mediates AIMP2-activated age-dependent dopaminergic neuronal loss. *Nat. Neurosci.* **16**, 1392–1400 (2013). doi: 10.1038/nn.3500; pmid: 23974709
20. L. Galluzzi *et al*., Molecular mechanisms of cell death: Recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ.* **25**, 486–541 (2018). doi: 10.1038/s41418-017-0012-4; pmid: 29362479
21. G. Codolo *et al*., Triggering of inflammasome by aggregated α-synuclein, an inflammatory response in synucleinopathies.

PLOS ONE **8**, e55375 (2013). doi: 10.1371/journal.pone.0055375; pmid: 23383169
22. W. Wang *et al*., Caspase-1 causes truncation and aggregation of the Parkinson's disease-associated protein α-synuclein. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9587–9592 (2016). doi: 10.1073/pnas.1610099113; pmid: 27482083
23. A. Iwai *et al*., The precursor protein of non-Aβ component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system. *Neuron* **14**, 467–475 (1995). doi: 10.1016/0896-6273(95)90302-X; pmid: 7857654
24. P. Brundin, R. Melki, Prying into the prion hypothesis for Parkinson's disease. *J. Neurosci.* **37**, 9808–9818 (2017). doi: 10.1523/JNEUROSCI.1788-16.2017; pmid: 29021298
25. C. Peng, R. J. Gathagan, V. M.-Y. Lee, Distinct α-synuclein strains and implications for heterogeneity among α-synucleinopathies. *Neurobiol. Dis.* **109** (Pt. B), 209–218 (2018). doi: 10.1016/j.nbd.2017.07.018; pmid: 28751258
26. C. J. Lord, A. Ashworth, PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017). doi: 10.1126/science.aam7344; pmid: 28302823
27. E. B. Affar *et al*., Immunological determination and size characterization of poly(ADP-ribose) synthesized in vitro and in vivo. *Biochim. Biophys. Acta* **1428**, 137–146 (1999). doi: 10.1016/S0304-4165(99)00054-9; pmid: 10434031
28. T.-I. Kam *et al*., FcγRIIb-SHIP2 axis links Aβ to tau pathology by disrupting phosphoinositide metabolism in Alzheimer's disease model. *eLife* **5**, e18691 (2016). doi: 10.7554/eLife.18691; pmid: 27834631
29. Y. Wang *et al*., A nuclease that mediates cell death induced by DNA damage and poly(ADP-ribose) polymerase-1. *Science* **354**, aad6872 (2016). doi: 10.1126/science.aad6872; pmid: 27846469
30. J. L. Guo *et al*., Distinct α-synuclein strains differentially promote tau inclusions in neurons. *Cell* **154**, 103–117 (2013). doi: 10.1016/j.cell.2013.05.057; pmid: 23827677

ACKNOWLEDGMENTS

We thank I.-H. Wu for graphic art assistance, H. Gu for collecting data and writing assistance, N. Yoritomo and M. Gudavalli for assistance in patient recruitment and biospecimen collection, and the patients and families that volunteer and participate in research. **Funding:** This work was supported by grants from the NIH/NINDS (P50NS38377, R37NS067525, NS082205, U01NS082133, and U01NS097049) and the JPB Foundation and by U01 NS100610 and the Jane and Lee Seidman Fund to J.B.L. We acknowledge the joint participation by the Adrienne Helis Malvin Medical Research Foundation through its direct engagement in the continuous active conduct of medical research in conjunction with the Johns Hopkins Hospital and the Johns Hopkins University School of Medicine and the Foundation's Parkinson's Disease Programs M-2016. T.M.D. is the Leonard and Madlyn Abramson Professor in Neurodegenerative Diseases. **Author contributions:** Conceptualization: T.-I.K., V.L.D., and T.M.D. Methodology: T.-I.K., X.M., H.P., V.L.D., and T.M.D. Validation: T.-I.K., X.M., H.P., S.-C.C., and S.S.K. Formal analysis: T.-I.K., X.M., H.P., S.-C.C., and S.S.K. Investigation: T.-I.K., X.M., H.P., S.-C.C., S.S.K., G.E.U., S.B., S.P.Y., N.P., and C.Q. Resources: R.C., S.A.A., G.G.P., O.P., J.C.T., L.M.B., J.B.L., A.P., H.S.K., L.S.R., V.L.D., and T.M.D. Writing—original draft: T.-I.K., X.M., H.P., V.L.D., and T.M.D. Writing—review and editing: T.-I.K., H.P., V.L.D., and T.M.D. Visualization: T.-I.K., X.M., and H.P. Supervision: V.L.D. and T.M.D. Funding acquisition: J.B.L., V.L.D., and T.M.D. **Competing interests:** J.B.L. is a consultant for Acadia Pharmaceuticals, Avid Radiopharmaceuticals, Axovant, Bracco Radiopharmaceuticals, Eisai, GE Healthcare, and Takeda. T.M.D. is a member of the Board of Directors of the Bachmann Strauss Dystonia and Parkinson's Disease Foundation, is a member of the Executive Scientific Advisory Board of the Michael J. Fox Foundation for Parkinson's Research, is a consultant and advisor to Sun Pharma Advanced Research Company Ltd., and serves on the advisory council of Aligning Science Across Parkinson's. These arrangements have been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies. T.-I.K., L.R., S.A.A.,V.L.D., and T.M.D have filed a U.S. patent entitled "Detection of PAR in the CSF of Patients with Parkinson's Disease." **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/eaat8407/suppl/DC1
Figs. S1 to S15
Tables S1 to S4

10 April 2018; resubmitted 13 August 2018
Accepted 26 September 2018
10.1126/science.aat8407

RESEARCH ARTICLE SUMMARY

DRUG DISCOVERY

Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN

Quinlan L. Sievers*, Georg Petzold*, Richard D. Bunker, Aline Renneville, Mikołaj Słabicki, Brian J. Liddicoat, Wassim Abdulrahman, Tarjei Mikkelsen, Benjamin L. Ebert†, Nicolas H. Thomä†

INTRODUCTION: Thalidomide, lenalidomide, and pomalidomide are clinically approved therapies for the treatment of multiple myeloma and other hematologic malignancies. These drugs induce rapid ubiquitination and proteasomal degradation of two transcription factors, Ikaros (IKZF1) and Aiolos (IKZF3), by recruiting them to the $CRL4^{CRBN}$ E3 ubiquitin ligase through a Cys2-His2 (C2H2) zinc finger (ZF) domain that is present in both proteins and required for their destruction.

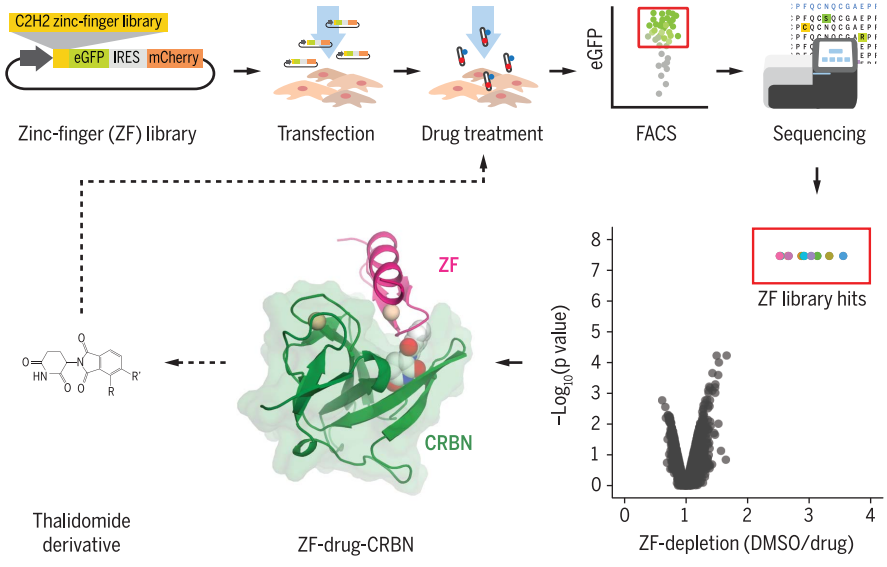
RATIONALE: Transcription factors have been challenging drug targets because they lack discrete catalytic domains amenable to small-molecule inhibition. Thalidomide analog–induced degradation of IKZF1 and IKZF3 through a C2H2 ZF domain raised the possibility that the >800 C2H2 ZF–containing proteins encoded by the human genome, many of which are putative transcription factors, could be similarly destabilized. We therefore set out to (i) define the human ZF "degrome" in the context of thalidomide, lenalidomide, and pomalidomide; (ii) characterize the ZF-drug-CRBN interaction structurally and functionally; and (iii) determine whether different thalidomide analogs degrade distinct ZFs.

RESULTS: Using a reporter of substrate degradation, we screened 6572 C2H2 ZFs for degradation in the presence of thalidomide, lenalidomide, and pomalidomide, identifying 11 ZF degrons, motifs that are capable of mediating drug-dependent degradation, of which six were found to mediate degradation of their full-length protein. Surprisingly, the 11 ZF degrons lack an identifiable consensus sequence. Saturation mutagenesis of the IKZF1/3 ZF degron and crystal structures of two ZF degrons bound to pomalidomide-engaged CRBN demonstrate that the drug-CRBN interface accommodates ZF degrons with diverse amino acid sequences. Computational docking in combination with in vitro binding assays revealed that a large number of ZFs are capable of weakly binding the drug-CRBN interface, indicating that this interface may be more permissive than suggested by the 11 ZF degrons identified in the degradation screen. To test this hypothesis, we screened the ZF library against two thalidomide analogs with chemical alterations at the ZF-drug-CRBN interface. The two thalidomide analogs induced degradation of distinct sets of C2H2 ZF degrons, including ZFs that bind the CRBN-pomalidomide complex weakly in vitro, but were not degraded by pomalidomide in cells.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat0572

CONCLUSION: We found that thalidomide analogs mediate $CRL4^{CRBN}$-dependent degradation of a larger number of C2H2 ZF proteins than previously anticipated. ZFs compatible with the drug-CRBN interface show little sequence conservation apart from residues that stabilize the ternary ZF fold. In addition to the complex ZF-CRBN side chain interactions, direct contacts between thalidomide analogs and varying ZF residues provide another layer of specificity. Thalidomide analogs with altered chemical scaffolds thus allow selective degradation of distinct ZF targets. Our results provide a structural and functional basis for the chemical modulation of $CRL4^{CRBN}$ to degrade C2H2 ZF transcription factors. Degradation of C2H2 ZF–containing proteins through derivatized thalidomide analogs may be a general paradigm for therapeutically targeting C2H2 ZF transcription factors, a class of proteins previously perceived to be "undruggable." ■



Thalidomide analogs mediate $CRL4^{CRBN}$-dependent ubiquitination and degradation of C2H2 ZFs. We created a cellular library in which each cell expresses one of 6572 individual C2H2 ZF domains from the human proteome fused to enhanced green fluorescent protein (eGFP). Cells expressing a C2H2 ZF reporter that is susceptible to thalidomide analog–induced degradation lose their eGFP signal, allowing their identification through a combination of fluorescence-activated cell sorting (FACS) and high-throughput sequencing. Structural and functional studies revealed how the drug-CRBN complex accommodates ZFs with diverse amino acid sequences. On the basis of these results, we tested thalidomide analogs with chemical modifications at the drug-ZF interface and found that these derivatives can degrade different sets of C2H2 ZFs. Our results suggest that chemical modulation of $CRL4^{CRBN}$ may be a more generalizable paradigm to inactivate C2H2 ZF–containing proteins, the largest group of putative transcription factors in the human genome. IRES, internal ribosomal entry site; DMSO, dimethyl sulfoxide.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: benajmin_ebert@dfci.harvard.edu (B.L.E.); nicolas.thoma@fmi.ch (N.H.T.)
Cite this article as Q. L. Sievers *et al.*, *Science* **362**, eaat0572 (2018). DOI: 10.1126/science.aat0572

RESEARCH ARTICLE

DRUG DISCOVERY

Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN

Quinlan L. Sievers[1,2]*, Georg Petzold[3]*, Richard D. Bunker[3], Aline Renneville[1,2], Mikołaj Słabicki[1,2,4], Brian J. Liddicoat[1,2], Wassim Abdulrahman[3]†, Tarjei Mikkelsen[1]‡, Benjamin L. Ebert[1,2,5]§, Nicolas H. Thomä[3]§

The small molecules thalidomide, lenalidomide, and pomalidomide induce the ubiquitination and proteasomal degradation of the transcription factors Ikaros (IKZF1) and Aiolos (IKZF3) by recruiting a Cys_2-His_2 (C2H2) zinc finger domain to Cereblon (CRBN), the substrate receptor of the $CRL4^{CRBN}$ E3 ubiquitin ligase. We screened the human C2H2 zinc finger proteome for degradation in the presence of thalidomide analogs, identifying 11 zinc finger degrons. Structural and functional characterization of the C2H2 zinc finger degrons demonstrates how diverse zinc finger domains bind the permissive drug-CRBN interface. Computational zinc finger docking and biochemical analysis predict that more than 150 zinc fingers bind the drug-CRBN complex in vitro, and we show that selective zinc finger degradation can be achieved through compound modifications. Our results provide a rationale for therapeutically targeting transcription factors that were previously considered undruggable.

Thalidomide and its derivatives, lenalidomide and pomalidomide, are effective therapies for the hematologic malignancies multiple myeloma, del(5q) myelodysplastic syndrome, and mantle cell lymphoma (*1*). Thalidomide analogs bind Cereblon (CRBN), the substrate receptor of the CUL4-RBX1-DDB1-CRBN ($CRL4^{CRBN}$) E3 ubiquitin ligase, and alter its substrate selectivity to recruit, ubiquitinate, and degrade seemingly unrelated proteins, including Ikaros (IKZF1), Aiolos (IKZF3), casein kinase 1 α (CK1α), and GSPT1 (*2*–*7*). Degradation of these targets in part explains the therapeutic effects of thalidomide analogs. IKZF1 and IKZF3 are lymphocyte lineage transcription factors (*8*, *9*) that are essential for the survival of the malignant plasma cells in multiple myeloma. CK1α is essential for the survival of hematopoietic stem cells, and heterozygous deletion of its gene in del(5q) myelodysplastic syndrome provides a therapeutic window for eliminating the malignant stem cell clone (*10*).

IKZF1 and IKZF3 belong to the family of Cys_2-His_2 (C2H2) zinc finger (ZF) proteins (*11*). These ZF proteins share a conserved C2H2 ZF fold composed of a β-hairpin and an α-helix held together by pairs of zinc-coordinating cysteine and histidine residues. Because amino acids in the α-helical portion of some ZF domains recognize DNA base pairs in a sequence-specific manner (*12*), the approximately 800 C2H2 ZF proteins are predicted to comprise the largest group of transcription factors in the human genome (*13*). Transcription factors have remained challenging drug targets because of the absence of druggable active sites (*14*). Thalidomide analogs, however, induce degradation of IKZF1 and IKZF3, raising the possibility that other C2H2 ZF–containing transcription factors are similarly destabilized.

E3 ubiquitin ligases recognize their substrates through degrons, short stretches of primary sequence that are necessary and sufficient for the interaction with substrate receptors of ubiquitin ligases (*15*). Previous work has implicated the second C2H2 ZF domain in IKZF1 and IKZF3 (*5*, *7*) and ZF4 of ZFP91 as the drug-inducible degrons (*5*, *7*, *16*). Unexpectedly, the known thalidomide analog targets—IKZF1/IKZF3, ZFP91, CK1a, and GSPT1—do not share obvious primary sequence similarity, with the exception of a glycine residue located in a β-hairpin, which diverges from the canonical destruction motif paradigm. Although the majority of C2H2 ZFs are structurally similar to the ZF degrons of IKZF1/IKZF3 and ZFP91, with 4661 out of 6572 carrying a glycine residue at an equivalent position, proteome-wide mass spectrometry demonstrated selective degradation of endogenous IKZF1/IKZF3 and ZFP91 in multiple cell lines (*3*, *4*, *16*). To identify determinants of drug-induced substrate specificity, we set out to characterize the human ZF "degrome" amenable to degradation in the presence of $CRL4^{CRBN}$ and thalidomide analogs and examined whether compound modifications change ZF selectivity.

A C2H2 ZF library screen identifies previously unknown degrons and protein targets

To characterize the minimal C2H2 ZF degron of IKZF1/IKZF3, we first generated different IKZF1 deletion constructs and measured their affinity for the drug-CRBN complex in vitro using time-resolved fluorescence resonance energy transfer (TR-FRET) (Fig. 1A and fig. S1, A and B). IKZF1 ZF2 (amino acid residues 141 to 174) was the shortest construct that showed binding to CRBN-pomalidomide, with an inhibition constant (K_i) of 2314 ±81 nM. Higher affinity binding to CRBN-pomalidomide in vitro was observed with an IKZF1 construct spanning ZF2 and ZF3 [amino acid residues 141 to 243 (Δ197 to 238); K_i 165 ± 37 nM]. Replacing ZF3 with ZF1 in the context of the ZF2-ZF3 construct (ZF2-ZF1) decreased the binding affinity by a factor of 6 (K_i 1027 ± 302 nM). IKZF1 binding to CRBN-pomalidomide is thus driven by ZF2 in vitro, with minor but specific contributions from the C-terminal ZF3.

To identify the minimal construct required for IKZF1/IKZF3 degradation in cells, we created a lentiviral degradation reporter vector that enabled us to compare the fluorescence of degron-tagged enhanced green fluorescent protein (eGFP) to untagged mCherry using flow cytometry (Fig. 1B) (*17*). We transduced WT and $CRBN^{-/-}$ human embryonic kidney cells (HEK293T) with the degradation reporter that contains IKZF3 deletion constructs and treated the cells with thalidomide, lenalidomide, or pomalidomide. Deletion of ZF2 (amino acid residues 146 to 168) in full-length IKZF3 rendered the reporter resistant to drug treatment, whereas deletion of the other ZFs had little or no effect (fig. S1C). Accordingly, IKZF3 ZF2 (amino acid residues 146 to 168), which is identical to IKZF1 ZF2 (amino acid residues 145 to 167) (fig. S1D), was sufficient to confer CRBN-dependent degradation of the reporter (Fig. 1C). Together, these results establish IKZF1/IKZF3 ZF2 as the minimal unit required for thalidomide analog-induced CRBN binding in vitro and for $CRL4^{CRBN}$-dependent degradation in cells.

Having demonstrated that a single ZF is sufficient to induce degradation of the eGFP/mCherry reporter, we sought to screen the entire human C2H2 ZF proteome for degradation in the presence of thalidomide analogs. To analyze the

[1]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. [2]Brigham and Women's Hospital, Division of Hematology, Harvard Medical School, Boston, MA 02115, USA. [3]Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland. [4]Division of Translational Oncology, National Center for Tumor Diseases, Heidelberg and German Cancer Research Center, 69120 Heidelberg, Germany. [5]Dana Farber Cancer Institute, Department of Medical Oncology, Boston, MA 02215, USA.
*These authors contributed equally to this work. †Present address: LeadXpro AG Park InnovAARE, 5232 Villigen, Switzerland. ‡Present address: 10x Genomics, 7068 Koll Center Parkway 401, Pleasanton, CA 94566, USA.
§Corresponding author. Email: benajmin_ebert@dfci.harvard.edu (B.L.E.); nicolas.thoma@fmi.ch (N.H.T.)

human ZF degrome independent of cell-type specific expression levels (fig. S1, E and F) and accessibility of ZFs in the context of full-length proteins engaged in macromolecular assemblies, we synthesized cDNAs for 6572 distinct C2H2 ZFs from the human proteome that match the PROSITE (*18*) ZF motif *x*(2)-C-*x*(2,4)-C-*x*(3)-[LIVMFYWC]-*x*(7)-H-*x*(3,5)-H and inserted the cDNAs into the degradation reporter vector. HEK293T cells were transduced with this C2H2 ZF library and then treated with dimethyl sulfoxide (DMSO), thalidomide, lenalidomide, or pomalidomide. Fluorescence-activated cell sorting (FACS) was used to isolate eGFP$^+$/mCherry$^+$ cells, and the relative number of read counts of each ZF was quantified by means of next-generation sequencing (Fig. 1D). A ZF was considered degraded if read counts were significantly underrepresented in drug-treated eGFP$^+$/mCherry$^+$ cells relative to DMSO-treated control cells.

Of the 6572 ZFs, 5611 had sufficient representation in the sequencing data to be assayed (>200 read counts). At a false discovery rate (FDR) of <0.01, pomalidomide depleted 11 C2H2 ZFs, each from different proteins, including IKZF1/IKZF3 ZF2 (Fig. 1E). Lenalidomide and thalidomide targeted a variable subset of these 11 ZFs. When cloned into the degradation reporter vector and tested in isolation, the 11 ZFs exhibited degradation in the presence of all three compounds (Fig. 2A and fig. S2A).

We next used the degradation reporter to determine whether the 11 ZFs destabilized their respective full-length protein and found that 6 of the 11 full-length proteins were degraded in the presence of the drug (IKZF1/IKZF3, ZNF692,

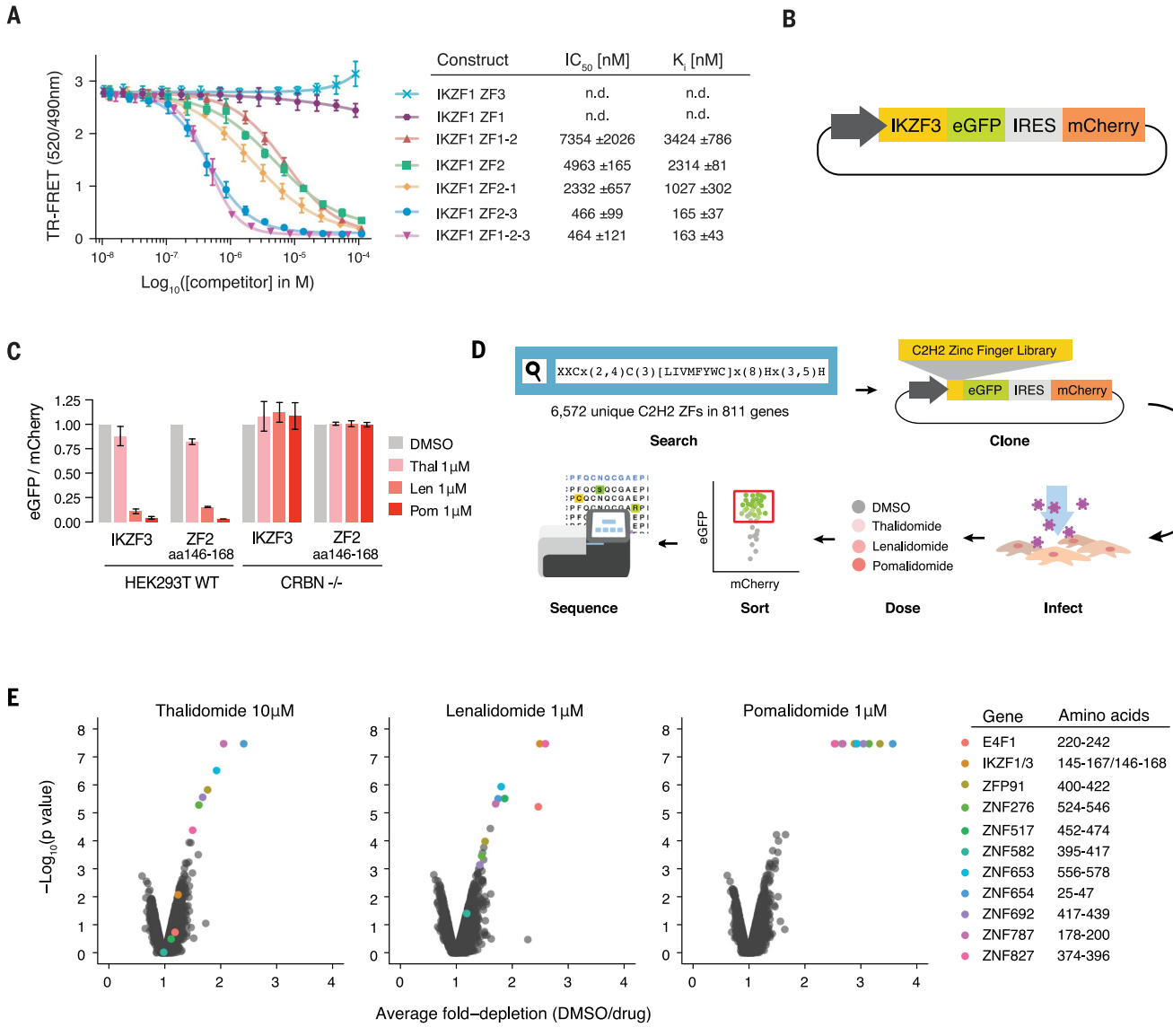


Fig. 1. C2H2 ZF library screen identifies 11 C2H2 ZFs degraded in the presence of thalidomide, lenalidomide, or pomalidomide. (**A**) TR-FRET counter titration. Unlabeled IKZF1 constructs (0.01 to 100 μM) titrated to preassembled $_{Alexa488}$DDB1ΔB-CRBN-pomalidomide-$_{biotin}$IKZF1$^{ZF1-2-3}$ (200 nM $_{Alexa488}$DDB1ΔB-CRBN, 100nM $_{biotin}$IKZF1$^{ZF1-2-3}$, 5μM pomalidomide). Data are mean ± standard deviation (SD). (**B**) Schematic of the protein degradation reporter vector. IRES, internal ribosome entry site. (**C**) HEK293T WT and CRBN$^{-/-}$ cells expressing IKZF3 constructs in the degradation reporter were treated for 20 hours with DMSO or drug then analyzed by flow cytometry to quantify the DMSO-normalized ratio of eGFP/mCherry fluorescence [experimental replicates = 3, technical replicates = 3, bar heights indicate mean of experimental replicates, error bars indicate 95% confidence interval (CI)]. (**D**) Schematic of the human C2H2 ZF library screen. (**E**) Average fold-depletion of sequencing read counts (DMSO/drug) and corresponding *P* values (empirical rank-sum test-statistic) for the 5611 C2H2 ZFs with raw read count >200 in all three control replicates (experimental replicates = 3, labeled data points possess FDR < 0.01 in at least one of the three drugs).

ZFP91, ZNF276, ZNF653, and ZNF827) (Fig. 2B and fig. S2B). All six ZFs that mediated degradation of their full-length protein carry an additional ZF C-terminal to the one identified in the library screen, whereas five of six ZFs that did not mediate degradation of their corresponding full-length protein do not posses a proximal C-terminal ZF. These results are consistent with our findings that although IKZF1/IKZF3 ZF2 constitutes the minimal ZF degron, an IKZF1 construct spanning ZF2 and ZF3 conferred higher affinity binding to CRBN-pomalidomide in vitro than ZF2 alone (Fig. 1A). We confirmed degradation of endogenous ZNF692, ZFP91, ZNF276, ZNF653, and ZNF827 in the presence of pomalidomide using Western blotting (Fig. 2, C to E). Additionally, we demonstrated binding of hemagglutinin (HA)–tagged, full-length ZFP91 and ZNF692 to endogenous CRBN in the presence of all three compounds by means of immunoprecipitation (fig. S2C).

A screen of 6572 C2H2 ZFs identified 11 C2H2 ZFs degraded in the presence of thalidomide, lenalidomide, or pomalidomide, six of which mediate degradation of their respective full-length protein. Four of these full-length proteins were previously unknown thalidomide analog targets (fig. S2D).

Identification of amino acid loci critical to degradation

Sequence alignment of the C2H2 ZF hits (Fig. 3A) highlighted shared residues that are part of the PROSITE C2H2 ZF search motif (IKZF3 C148, C151, H164, H168, and F155), residues that stabilize the ternary fold of the ZF domain (IKZF3 L161), and residues that frequently appear within C2H2 ZF domains (IKZF3 G152). The PROSITE ZF search motif covered sequences in

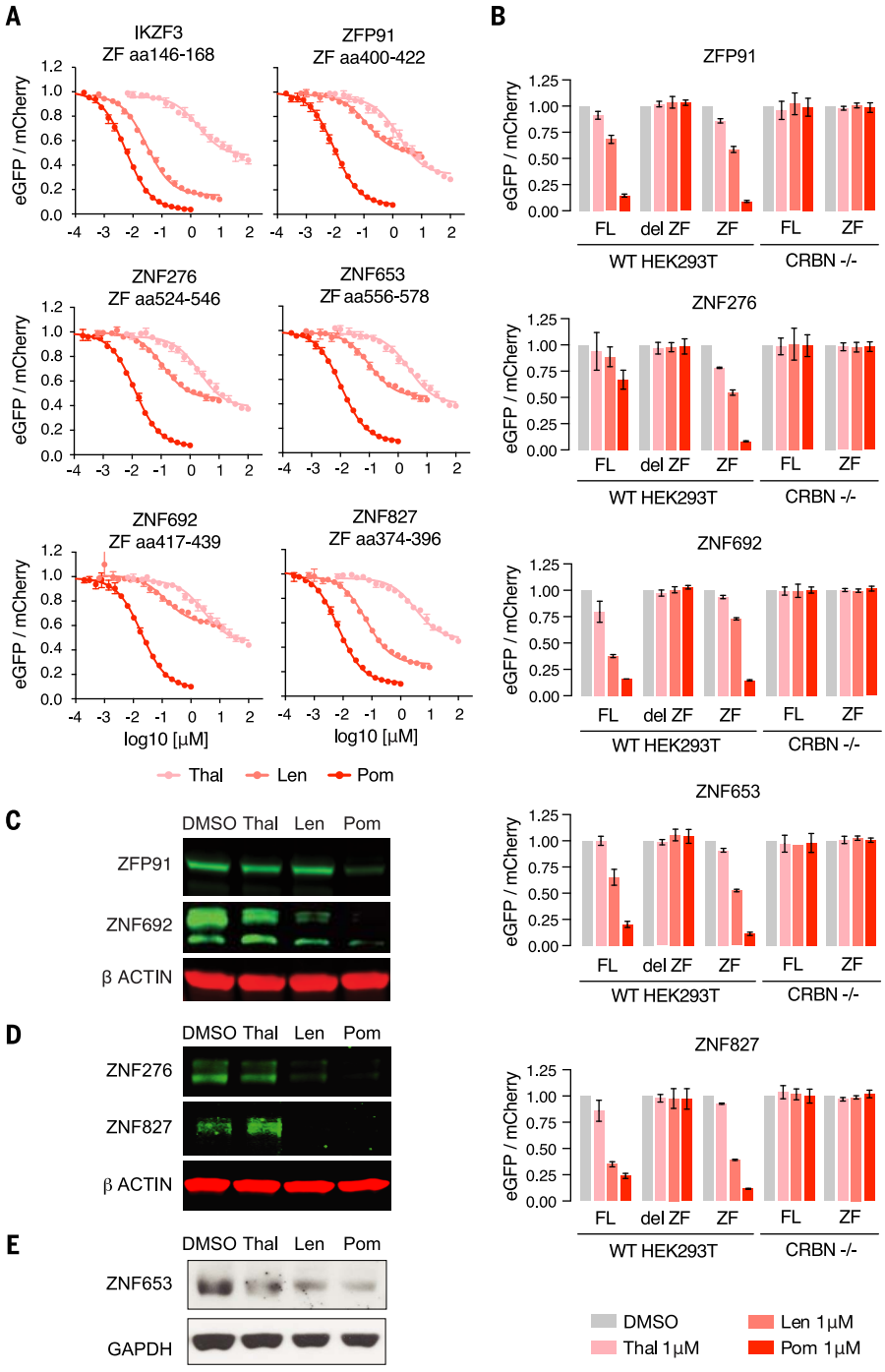


Fig. 2. Six proteins are degraded by thalidomide, lenalidomide, or pomalidomide conditional on the presence of a C2H2 ZF degron. (**A**) HEK239T WT cells expressing the 11 ZFs identified in the screen in the degradation reporter were treated for 20 hours with drug then analyzed by means of flow cytometry to measure the DMSO-normalized ratio of eGFP/mCherry fluorescence (experimental replicates = 3, technical replicates =3, dots represent average of experimental replicates, error bars indicate range). ZFs not shown are in fig. S2A. (**B**) HEK293T WT or CRBN$^{-/-}$ cells expressing full-length target proteins (FL), FL with ZF deleted (del ZF), or ZF alone were treated for 20 hours with DMSO or 1 μM drug then analyzed by using flow cytometry to measure the DMSO-normalized ratio of eGFP/mCherry fluorescence (experimental replicates = 3, technical replicates = 3, bar heights indicate mean of experimental replicates, error bars indicate 95% confidence interval). Proteins not shown are in fig. S2B. (**C**) KG1 (human acute myeloid leukemia), (**D**) WM266-4 (human melanoma), and (**E**) MOLM-16 cells (human acute myeloid leukemia) were treated with DMSO or 1 μM drug for 20 or 24 hours, after which protein lysates were harvested, run on a polyacrylamide gel, and immunoblotted for the specified targets (images representative of three experimental replicates).

which the β-hairpin zinc-coordinating cysteines are separated by two, three, or four residues [C-x(2,4)-C], but only C-x(2)-C ZFs were found destabilized in the screen (Fig. 3A). Insertion of a glycine residue between the β-hairpin residues N148 and Q149 of the IKZF1 ZF2-ZF3 construct (IKZF3 N149 and Q150) (fig. S1D) compromised binding in vitro (fig. S3A). This suggests that the ZF degrome is restricted to an inter-cysteine spacing of two residues. Besides the inter-cysteine spacing, residues that stabilize the ternary ZF fold, or residues that frequently appear in C2H2 ZF domains, no discernable sequence consensus could be detected among the 11 ZFs (Fig. 3A). The identified ZF degrons thus revealed structural features common to C2H2 ZF domains that alone cannot explain selective degradation of only 11 ZFs, whereas 4661 structurally similar ZFs were present in the library.

To test whether the divergent, nonstructural residues contribute to a functional ZF degron, we synthesized a mutagenesis library of IKZF3 spanning amino acid residues 130 to 189 so that at each of the 60 loci, all 19 possible amino acid substitutions were represented. The pooled library was inserted into the degradation reporter vector and transduced into HEK293T cells. We treated the cells with DMSO, thalidomide, lenalidomide, or pomalidomide and used FACS to isolate drug-treated eGFP$^-$/mCherry$^+$ cells or DMSO-treated eGFP$^+$/mCherry$^+$ control cells. Next-generation sequencing was used to quantify the relative number of read counts for each substitution (fig. S3B).

The screen highlighted nine loci in IKZF3 ZF2 whose amino acid identities were critical for degradation (Fig. 3, B and C, and fig. S3, C and D). These loci again included residues that maintain the tertiary ZF fold (IKZF3 C148, C151, F155, L161, H164, and H168) and the β-hairpin glycine residue (IKZF3 G152). In addition, the screen highlighted two nonstructural residues (IKZF3 Q147 and A153) that varied among the other ZF degrons (Fig. 3A) but were important for degradation of the IKZF3 ZF2 reporter (Fig. 3, B and C). Mutation of these two residues impaired degradation in validation experiments, whereas N149, a residue that was not highlighted in the mutagenesis screen, tolerated mutation (Fig. 3D). These data show that in addition to the residues maintaining the ZF fold, nonstructural amino acids (IKZF3 Q147, G152, and A153) contribute to degron specificity.

Structures of CRBN bound to pomalidomide and two different ZF degrons

To examine how the varying, nonstructural residues within ZF degrons contribute to selective CRBN binding, we determined crystal structures of DDB1$^{\Delta BPB}$-CRBN bound to pomalidomide and IKZF1 ZF2 (amino acid residues 141 to 174) or ZNF692 ZF4 (amino acid residues 416 to 442) (Fig. 4, A and B). Unlike previous CRBN structures, CRBN bound IKZF1 ZF2 in an "open" conformation, with its N- and C-terminal domains

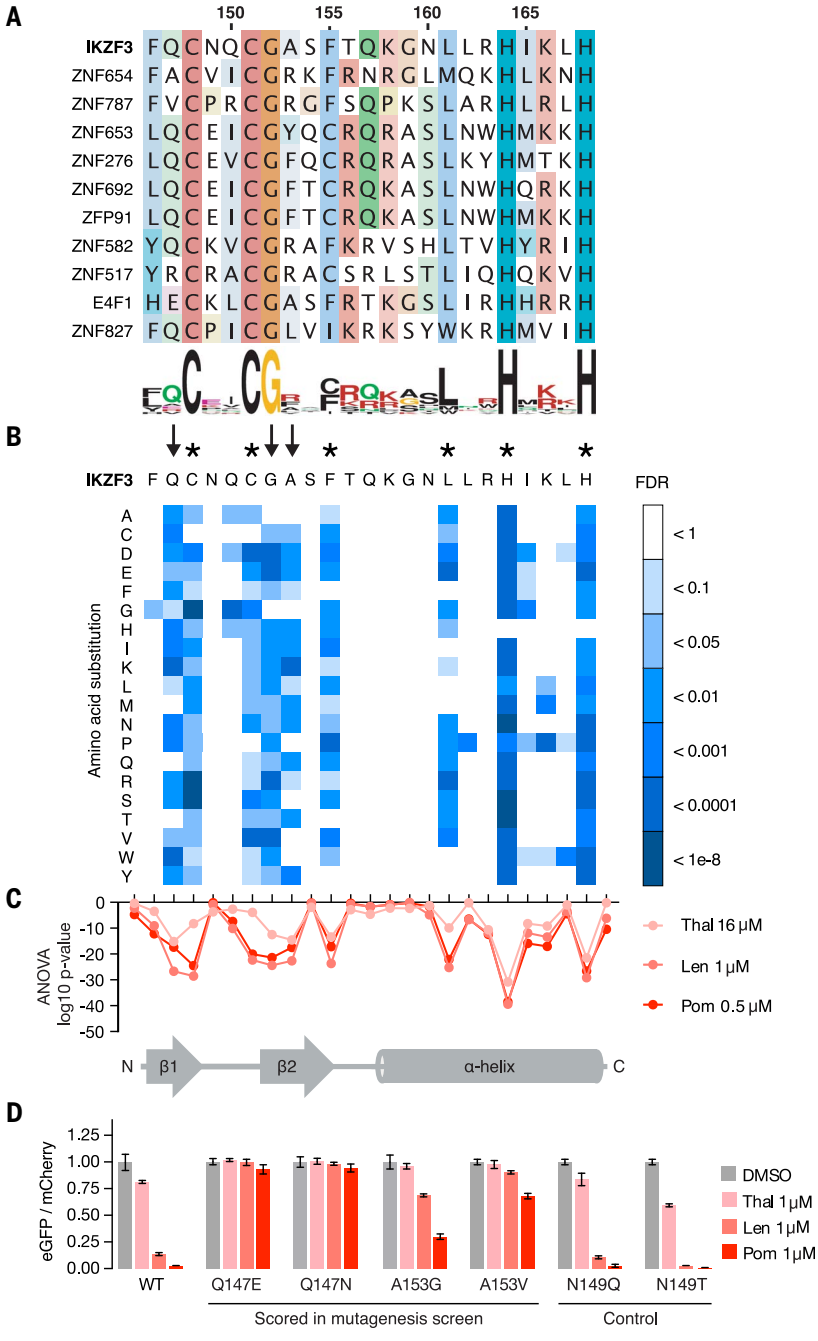


Fig. 3. Identification of amino acids required for drug-induced degradation of a C2H2 ZF degron. (**A**) Sequence alignment of the 11 C2H2 ZFs with FDR < 0.01 in at least one drug condition (amino acids colored by property). Numbers correspond to amino acid residues in full-length IKZF3. (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr). (**B**) Saturation mutagenesis screen of IKZF3 amino acid residues 130 to 189 in presence of lenalidomide displayed as heat map of the FDR for mutant amino acids (unpaired, one-sided Welch t test, FDR correction performed within each column, technical replicates = 3). Asterisks indicate amino acids required for the ZF fold, and arrows indicate nonstructural IKZF3 residues required for degradation. Complete results for all 60 amino acids are shown in fig. S3C. (**C**) Analysis of variance (ANOVA) P values for difference in frequency of mutant amino acids at each position in IKZF3 amino acid residues 130 to 189 (DMSO versus drug). Complete results for all 60 amino acid residues are shown in fig. S3D. (**D**) HEK293T cells expressing IKZF3 ZF2 constructs in the degradation reporter were treated for 20 hours with DMSO or 1 µM drug, after which flow cytometry was used to measure the DMSO-normalized ratio of eGFP/mCherry fluorescence (experimental replicates = 1, technical replicates = 3, bar height is the average of technical replicates, error bars denote 95% CI).

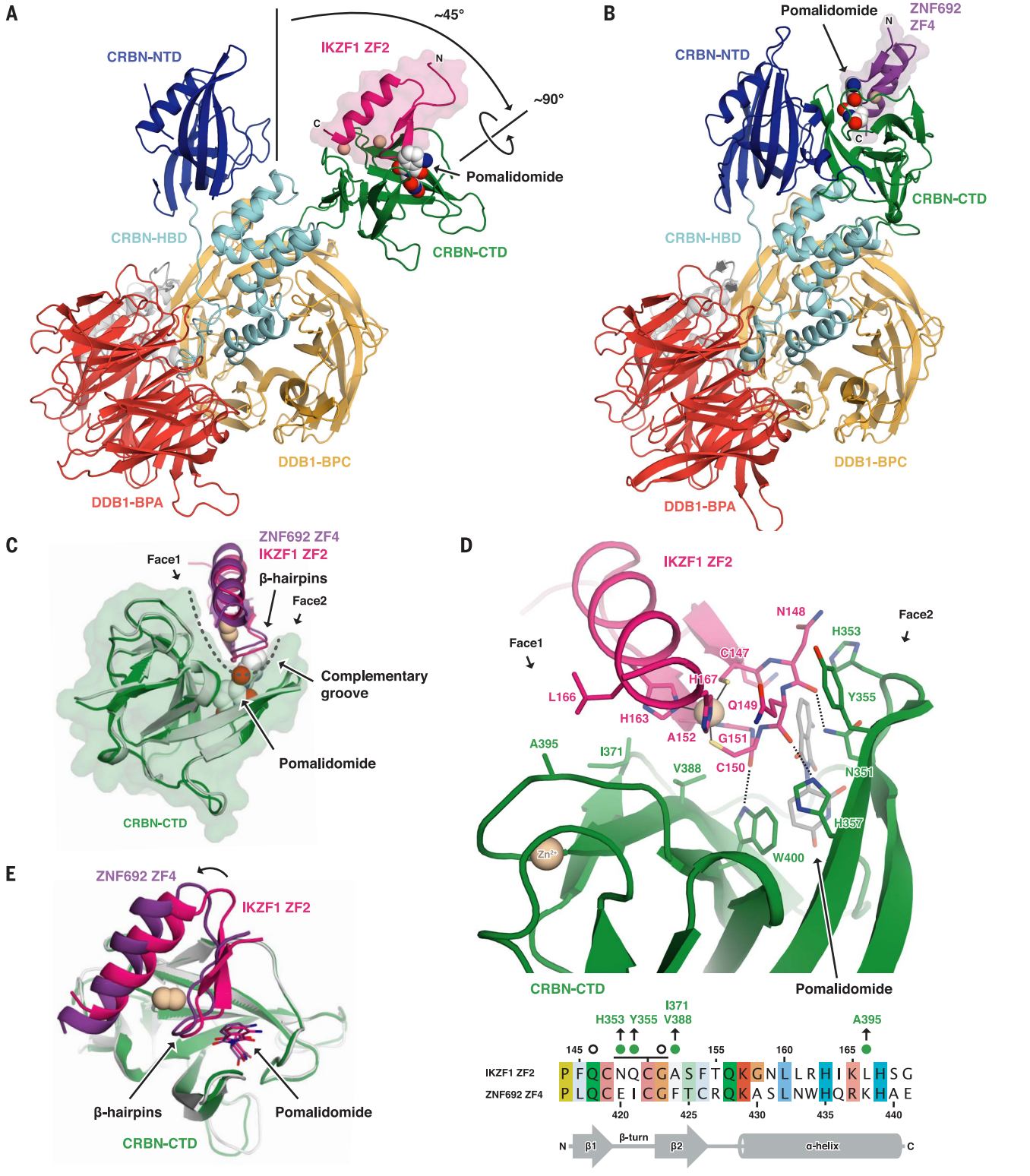


Fig. 4. The CRBN-pomalidomide interface accommodates C2H2 ZF degrons with diverse amino acid sequences. (**A**) Crystal structure of $DDB1^{\Delta B}$-$CRBN^{\Delta N40}$ bound to pomalidomide and $IKZF1^{ZF2}$. Zinc ions are shown as beige spheres. (**B**) Crystal structure of $DDB1^{\Delta B}$-$CRBN^{\Delta N40}$ bound to pomalidomide and $ZNF692^{ZF4}$. CTD, carboxy-terminal domain; HBD, helical bundle domain; NTD, amino-terminal domain; BPA and BPC, β-propeller A and C, respectively. (**C**) Superposition of the CRBN-CTDs bound to $IKZF1^{ZF2}$ and $ZNF692^{ZF4}$. A dashed line indicates the complementary groove. (**D**) Side chain interactions between $IKZF1^{ZF2}$, CRBN, and pomalidomide (dashed lines indicate hydrogen bonds) and sequence alignment of $IKZF1^{ZF2}$ and $ZNF692^{ZF4}$ (amino acids colored by property). Green dots or open circles indicate side chain interactions with CRBN or pomalidomide, respectively. A black line indicates ZF residues involved in backbone interactions with CRBN and/or pomalidomide. (**E**) Superposition of both CRBN-CTDs shows differences in IKZF1/ZNF692 ZF orientation.

separated and stabilized by crystal contacts (Fig. 4A and fig. S4A). As supported by the ZNF692 structure, release of the C-terminal domain is not required for ZF binding (Fig. 4B).

Attempts to crystalize the IKZF1 ZF2-ZF3 construct resulted in poorly diffracting crystals. To dissect the contribution of the C-terminal ZF to binding affinity, we used an unbiased docking approach that placed ZF3 on the neighboring N-terminal domain of "closed" CRBN (fig. S4B). This model positions the C terminus of ZF2 adjacent to the N terminus of ZF3 consistent with the five-amino-acid linker between the two ZFs. The model highlights three consecutive arginine residues of ZF3 (IKZF1 R183 to R185) at the CRBN-ZF2 interface that are not present in the ZF2-ZF1 construct (Fig. 1A and fig. S4C). Mutation of R185 to alanine, or replacement of all three arginines with the corresponding amino acids in ZF1 (IGP), reduced the affinity of the ZF2-ZF3 construct similar to that of ZF2-ZF1 (Fig. 1A) or ZF2 alone as predicted by our model (fig. S4D). These results show that IKZF1 ZF2 is the primary determinant for drug binding and identify features of ZF3 that contribute to high-affinity binding to the drug-CRBN complex (Fig. 1A and fig. S4B).

The crystal structures of CRBN-pomalidomide bound to IKZF1 ZF2 and ZNF692 ZF4 reveal the detailed side chain interactions between CRBN, pomalidomide, and the ZFs. Both IKZF1 ZF2 and ZNF692 ZF4 bind a complementary groove on the CRBN C-terminal domain (Fig. 4C), interact with the compound through their β-hairpin loops, and match the overall backbone conformation of previous models of the ZF-drug-CRBN complex (fig. S4E). In each structure, the β-hairpin glycine of the ZF packs against the phthalimide ring of pomalidomide similar to the β-hairpins of CK1α (*6*) and GSPT1 (*5*), affirming the structural importance of a glycine residue at this position (Fig. 4D and fig. S4F). Because the CRBN groove is narrow (Fig. 4C), ZF-CRBN interactions at both sides of the groove together determine binding. IKZF1 ZF2 N148 and Q149 (IKZF3 N149 and Q150) (fig. S1D) intercalate through their amphipathic side chains with CRBN residues H353 and Y355 (Fig. 4D), explaining CRBN compatibility with diverse amino acids at these ZF positions. IKZF1 residues A152 and L166 (IKZF3 A153 and L167) face CRBN residues V388, I371, and A395, located on the opposite side of the groove (Fig. 4, C and D). Although ZNF692 ZF4 and IKZF1 ZF2 show similar binding modes (Fig. 4C), the position of ZNF692 ZF4 is offset with respect to IKZF1 ZF2 (Fig. 4E and fig. S4, G and H), without significant differences in binding affinity (fig. S4I). Thus, the two structures demonstrate how CRBN accommodates ZF degrons with varying side chain properties at the ZF-CRBN interface (IKZF3 N148, E, N, K, R, V, P; Q149, I, V, R, L, A; A153, Y, F, L, R; and L167, V, I, K, N) (Fig. 3A).

IKZF1 Q146 forms the only side-chain interaction with the compound (Fig. 5A), and mutation of the equivalent residue in IKZF3 (Q147) stabilized the reporter in the saturation mutagenesis experiment (Fig. 3, B and D). The IKZF1 Q146 side chain (IKZF3 Q147) packs against the phthalimide group of pomalidomide and forms a water-mediated hydrogen bond with the C4 amino group of the compound (Fig. 5, A and B, and fig. S5A). Thalidomide lacks the C4 amino group (Fig. 5B), and mutation of IKZF1 Q146 (IKZF3 Q147) to isoleucine, which removes the ability to form a hydrogen bond with the drug, equalizes the binding affinity of IKZF1 ZF2-ZF3 to CRBN across the three compounds accordingly (Fig. 5C and fig. S5, B and C). Equivalent mutations in IKZF3 ZF2 (Q147I/A/H) stabilized the degradation reporter in cells (Fig. 5D and fig. S5D), explaining preferential binding of IKZF1/IKZF3 to CRBN engaged with pomalidomide or lenalidomide over thalidomide in vitro (*6*) and the contribution of this residue to compound selectivity in vivo (fig. S5E). Seven of the 11 ZF degrons carry an equivalent glutamine residue at this position (Fig. 3A).

ZF degrons show epistatic properties during CRL4^{CRBN} engagement

The crystal structures demonstrate that ZFs with diverse amino acid sequences are compatible with the drug-CRBN interface. To test whether the variable residues are interchangeable between ZF degrons, we swapped the β-hairpin and α-helix of IKZF3 ZF2 with that of ZFP91 ZF4. Replacing the IKZF3 ZF2 β-hairpin with the ZFP91 ZF4 β-hairpin resulted in a construct that was degraded more efficiently than either of the parent molecules; however, replacing the IKZF3 α-helix with the ZFP91 α-helix resulted in a ZF domain that was resistant to degradation (fig. S5F). Furthermore, introducing a single-residue mutation at the drug-ZF interface of IKZF3 (Fig. 5A and fig. S1D), Q147 to a glutamic acid residue, stabilized the IKZF3 reporter (IKZF3 ZF2 Q147E) (Fig. 3D), whereas an equivalent glutamic acid allows robust reporter degradation in the sequence context of the ZF degron

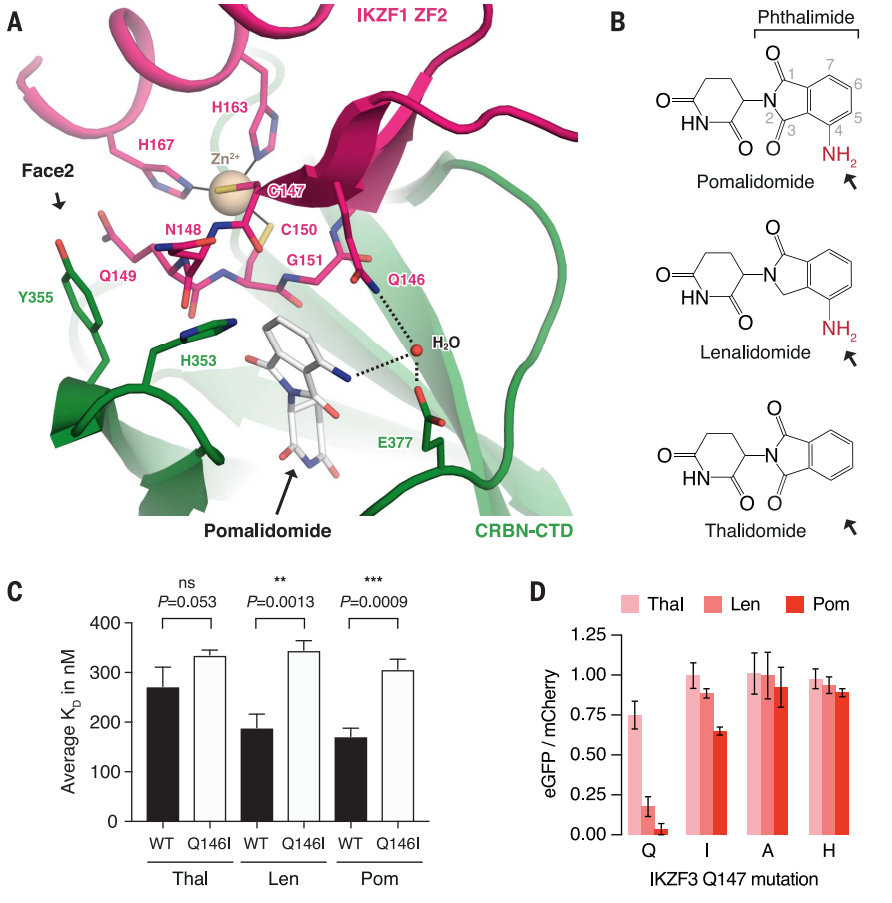


Fig. 5. IKZF3 Q147 interacts with the amino group on pomalidomide. (**A**) Side chain interactions between IKZF1^{ZF2}, CRBN, and pomalidomide. Dashed lines indicate hydrogen bonds. (**B**) Chemical structures of pomalidomide, lenalidomide, and thalidomide, with differences highlighted in red. (**C**) Inhibition constants (K_i) of wild-type (WT) and Q146I mutant IKZF1$^{ZF2-ZF3}$ in the presence of thalidomide, lenalidomide, and pomalidomide in TR-FRET counter titration experiments. Data are mean ± SD. (**D**) HEK293T cells expressing IKZF3 ZF2 Q147 mutants in the degradation reporter were treated for 20 hours with DMSO or drug and then analyzed by means of flow cytometry to measure the DMSO-normalized ratio of eGFP/mCherry fluorescence (experimental replicates = 3, technical replicates = 3, bar height is average of experimental replicates, error bars indicate 95% CI). The complete set of Q147 amino acid mutants is shown in fig. S5D.

from E4F1 (Fig. 3A and fig. S2A). Despite the low degree of sequence similarity across the 11 ZF degrons (Fig. 3A), only distinct amino acid combinations at the drug-CRBN interface result in degradation. ZF degrons are thus defined by the sequence context of their amino acid side chains that contact the drug-CRBN interface (Fig. 4D), which suggests an epistatic relationship of these variable ZF residues during $CRL4^{CRBN}$ engagement and explains the absence of a primary sequence consensus.

In silico identification of ZFs that interact with the drug-CRBN interface

Only 11 of the 4661 structurally similar ZFs were degraded in the library screen, yet the ZF degron appears to be complex. To create a semiquantitative model of the observed degradation pattern with respect to our crystal structures, we used computational docking to account for multiple, potentially epistatic ZF interactions at the drug-CRBN interface. The Rosetta software package (*19*) was used to computationally dock all human C2H2 ZFs into the complementary binding groove of CRBN. The Rosetta interface energy scores and the structural similarity to our ZF-pomalidomide-CRBN crystal structures were calculated to assess the docking trials (Fig. 6A and fig. S6, A and B). Using the ZNF692 template structure, we accurately predicted all 11 ZFs identified in the library screen (Fig. 6A), found 40 ZFs with lower interface energy scores than the ZNF692 ZF4 template, and identified 108 ZFs that score better than the lowest ranking ZF found degraded in the library screen. Overall, different docking trials revealed an overlapping set of ~50 to 150 ZFs with interface scores similar to or better than those of ZFs found degraded in the library screen (Fig. 6A and fig. S6, A and B), suggesting that an unexpectedly large number of ZFs are capable of binding the drug-CRBN interface.

On the basis of biological relevance, 21 of these ZFs were selected and tested for CRBN engagement in TR-FRET experiments, either as individual ZFs (fig. S6C) or fused to their respective C-terminal ZF (equivalent to the IKZF1 ZF2-ZF3 construct) (Fig. 6B). Sixteen ZFs bound CRBN-pomalidomide in vitro [BCL6, BCL6B, EGR1, EGR4, GZF1, HIC1, HIC2, SALL1, SALL3, SALL4, OSR1, OSR2, WIZ (ZF6 and ZF7), ZBTB7A, and ZBTB7B), with similar (WIZ ZF7)] or lower binding affinities (~5 to 20 fold) than the respective IKZF1 ZF2-ZF3 or ZF2 constructs (Fig. 6B and fig. S6, C, D, and E). Analogously, IKZF1 ZF1, a ZF that was not predicted to bind CRBN in silico (Fig. 6A), did not bind CRBN in vitro (fig. S6C). Thus, of the 33 ZFs individually tested in this study, 28 are recruited to pomalidomide-engaged CRBN (fig. S6F). Our in silico ZF docking approach is thus capable of predicting ZFs that interact with CRBN-pomalidomide with greater than 80% accuracy.

Those ZFs that bound CRBN-pomalidomide in the biochemical assay were subsequently tested for degradation in cells by using the degradation reporter. Of the 16 ZFs, WIZ ZF7 showed partial degradation in the presence of pomalidomide, the best in vitro binder in the set (fig. S6E and fig. S7, A and B). Examination of published proteome-wide mass spectrometry data confirmed degradation of endogenous, full-length WIZ in WSU-DLCL2 (human diffuse large B cell lymphoma) and TMD8 (human diffuse large B cell lymphoma) cells in the presence of lenalidomide and the thalidomide analog CC-122 (fig. S7C) (*20*). Because the majority of these ZFs are recruited to pomalidomide-engaged CRBN in vitro, but not destabilized by the same compound in

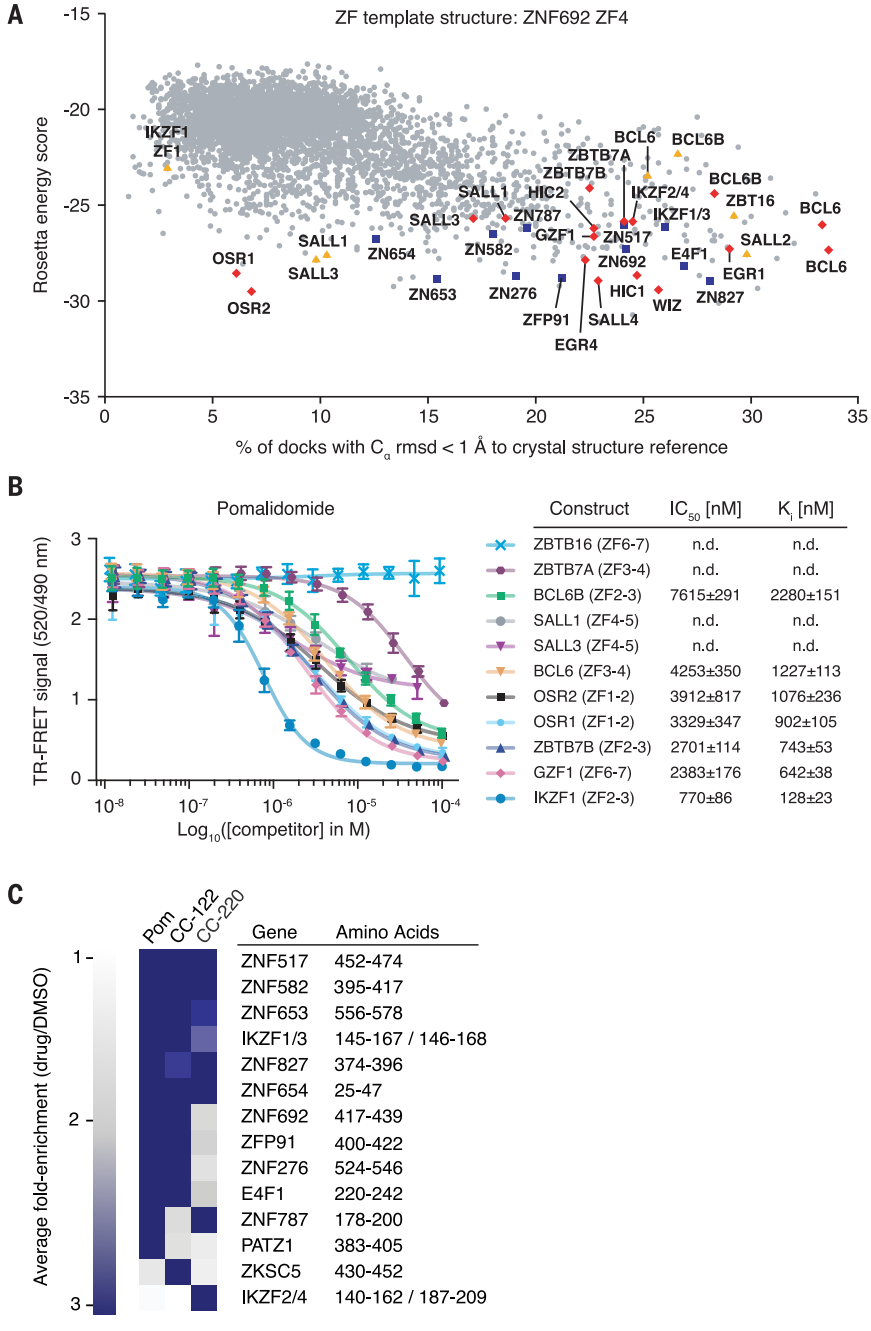


Fig. 6. Thalidomide analogs with chemical modifications at the ZF interface target different sets of C2H2 ZF degrons. (**A**) ZF library docking results by use of RosettaDock, with the CRBN C-terminal domain bound to pomalidomide and $ZNF692^{ZF4}$ as a crystal structure reference. Blue squares indicate degraded ZFs from the library screen, red diamonds highlight ZFs that bind pomalidomide-CRBN in vitro, and yellow triangles mark ZFs for which CRBN-binding could not be confirmed in vitro or in vivo. (**B**) TR-FRET counter titration. Unlabeled dual ZF candidate constructs (0.01 to 100 μM) titrated to preassembled $_{Alexa488}$DDB1ΔB-CRBN-pomalidomide-$_{biotin}$IKZF1$^{ZF1\text{-}2\text{-}3}$ (500 nM $_{Alexa488}$DDB1ΔB-CRBN, 50 nM $_{biotin}$IKZF1$^{ZF1\text{-}2\text{-}3}$, 5 μM pomalidomide). Data are mean ± SD. (**C**) Average fold-enrichment of sequencing read counts (drug/DMSO) for the 11 previously identified ZF degrons and 3 previously unidentified ZF hits.

cells, our results suggest that small changes in ZF-binding affinity disproportionally influence ZF degradation in cells.

Different thalidomide analogs target distinct sets of ZFs for degradation

We next examined whether thalidomide derivatives with chemical alterations at the drug-ZF interface induce degradation of the computationally predicted and biochemically validated ZF hits that are not destabilized by pomalidomide in cells. We therefore treated cells that express these ZFs in the eGFP/mCherry degradation reporter with the previously reported thalidomide derivatives CC-122 (*20*), CC-220 (*21*), and CC-885 (*5*) (fig. S7, A and B). The modified thalidomide analogs induced mild but significant degradation for some of these ZFs, including two ZFs destabilized by CC-122 (WIZ ZF6 and BCL6 ZF3), two destabilized by CC-220 (BCL6B ZF2 and HIC2 ZF5), and eight destabilized by CC-885 (BCL6 ZF3, BCL6B ZF2, OSR1 ZF1, ZBTB7B ZF2, SALL3 ZF4, SALL4 ZF2, ERG1 ZF3, and SALL1 ZF4). Although the effects are small, our findings suggest that different thalidomide analogs target different ZFs for degradation, including ZFs identified by computational docking, rendering them suitable candidates for drug development.

To comprehensively determine whether the chemically distinct thalidomide analogs induce degradation of different sets of ZFs, we repeated the C2H2 ZF library screen in the presence of pomalidomide, CC-122, and CC-220 (Fig. 6C and fig. S8, A and B). The 11 previously identified ZF hits again scored in the screen with pomalidomide (fig. S8B). Consistent with our earlier finding that CC-122 and CC-220 were capable of degrading ZFs not affected by pomalidomide, all three thalidomide analogs exhibited distinct ZF degradation patterns across the ZF library (Fig. 6C). IKZF2/4 ZF2 was selectively degraded by CC-220 (Fig. 6C). IKZF2/4 ZF2 differs from IKZF/13 ZF2 by a single amino acid substitution at the drug-ZF interface, IKZF3 Q147H (Fig. 5A and fig. S6F). This amino acid change decreases binding to CRBN-pomalidomide in vitro by a factor of 2 to 3 (fig. S5B) and stabilizes the respective ZF reporter in cells (Fig. 5D). In validation experiments, the IKZF2/4 ZF2 reporter was degraded by more than 40% in the presence of CC-220 (fig. S8, C and D). These results demonstrate that thalidomide analogs with chemical alterations at the drug-ZF interface promote degradation of distinct sets of ZFs and are capable of converting ZFs that bind CRBN-pomalidomide weakly in vitro into degraded ZFs in cells.

Discussion

We used a combination of functional and structural approaches to define the molecular basis of C2H2 ZF recruitment to the drug-engaged $CRL4^{CRBN}$ ubiquitin ligase. Thalidomide, lenalidomide, and pomalidomide mediate $CRL4^{CRBN}$-dependent degradation of a larger number of proteins than previously appreciated through a C2H2 ZF degron (*3*, *4*, *16*). We identified 15 individual ZFs and seven full-length ZF-containing proteins that are degraded by thalidomide derivatives in functional or computational screens. Crystal structures, in vitro binding, and cellular degradation assays illustrate that 28 ZFs (including IKZF2/4 ZF2) with diverse amino acid sequences bind the same drug-CRBN interface (fig. S6E).

The majority of E3 ligase degrons are characterized by conserved primary sequence motifs that lie in unstructured regions of the protein substrate (*15*). By contrast, the more than 28 C2H2 ZFs accommodated by the drug-CRBN interface show surprisingly little sequence conservation, apart from residues that stabilize the ternary ZF fold. ZF degrons bind a complementary groove on the CRBN surface that fits the overall shape of the C2H2 ZF domain, bringing different amino acids of different ZF degrons in contact with the same drug-CRBN interface. Substitution of a ZF residue at one site of the CRBN interface can be compensated by amino acid changes at another site of the ZF, explaining the observed epistatic properties of the ZF degrons. Recruitment and degradation are thus influenced by the complementarity of the substrate to the drug-CRBN binding groove and the sequence context of the contacting ZF residues, making the overall "shape" of the ZF the important binding determinant rather than its primary amino acid sequence. In addition to the complex ZF-CRBN side chain interaction, direct contacts between thalidomide analogs and the ZF provide additional layers of specificity (fig. S5D). Unbiased functional and computational approaches are thus required to identify the full complement of ZF degrons capable of drug-dependent binding and degradation.

The discovery that multiple C2H2 ZFs are drug targets raises the possibility that a large number of ZFs could be subject to thalidomide analog–based degradation. Supporting this notion, 28 of the ~50 to 150 additional C2H2 ZFs nominated by means of in silico docking were shown to bind the drug-CRBN complex, demonstrating that the CRBN surface is more permissive than suggested by the number of targets identified in the ZF library screens. Small differences in affinity between the ZF and the drug-CRBN complex in vitro translate into large differences in degradation in cells, which are likely due to competition for CRBN occupancy by multiple ZFs as well as other substrates that use the same binding surface (Fig. 6, C and D, and figs. S5 and S7). If ZF degradation depends on the ZF occupancy of $CRL4^{CRBN}$, it is possible that high local concentration of the ZF or $CRL4^{CRBN}$, or low concentrations of competing ZFs, could compensate for low binding affinity, leading to degradation under such conditions. Given the $CRL4^{CRBN}$ protein architecture and its ~100-Å ubiquitination radius (*6*, *7*, *22*), we expect degradation of full-length ZF proteins to be less dependent on lysine accessibility and instead primarily determined by the protein synthesis rate and the affinity and binding kinetics of the internal ZF degron to CRBN, which can be modulated by the drug.

We observed that thalidomide analogs with chemical modifications at the drug-ZF interface are capable of converting ZFs with weak affinity for the CRBN-pomalidomide complex into degraded targets. IKZF2/4 ZF2 is not degraded by pomalidomide, lenalidomide, or CC-122 but is efficiently degraded by CC-220. The crystal structures suggest that ZF side chains—particularly those proximal to the drug at IKZF1 position 146, 148, and 153 (IKZF3 147, 149 and 154)—interact with thalidomide analogs and that chemical alterations can modulate ZF specificity (fig. S5, D and E). ZFs that bind weakly but are not degraded in response to one compound could therefore serve as starting points for the development of new thalidomide analogs that selectively and efficiently degrade such ZFs. Several of the ZFs identified as degraded in functional experiments, or in vitro binders that could be targeted with further drug development, have been implicated in human disease. For example, BCL6 is an oncoprotein in lymphomas, ZFP91 is implicated in NF-κB signaling (*23*), and ZNF827 is reported to be an essential scaffolding protein for alternative lengthening of telomeres (*24*). HIC1, HIC2, GZF1, OSR1, OSR2, and SALL4 have all been implicated in development and could contribute to developmental abnormalities in fetuses exposed to thalidomide (*25*–*30*).

The approximately 800 C2H2 ZF-containing proteins are the largest class of putative transcription factors in the human proteome (*13*). Our results suggest that degradation of C2H2 ZF–containing proteins through different thalidomide derivatives may be a generalizable paradigm for targeting "undruggable" transcription factors and provides a structural and functional starting point from which to explore the extent to which CRBN-binding small molecules may be used to target ZF proteins for therapeutic intervention.

Materials and methods summary

Binding assays

In vitro CRBN binding was measured using a TR-FRET assay, and cellular CRBN binding was assayed via CRBN immunoprecipitation.

Protein degradation assays

HEK293T cells expressing candidate proteins in an eGFP/mCherry protein degradation reporter vector were treated with DMSO or drug and the eGFP:mCherry ratio was quantified by flow cytometry. Western blotting was used to assay degradation of endogenous protein targets.

C2H2 ZF library screen

HEK293T cells expressing a library of 6572 C2H2 zinc fingers in the protein degradation reporter vector were treated with DMSO or drug, $eGFP^+$ and $eGFP^-$ cell populations were isolated by FACS, and the relative frequency of individual ZFs was quantified with next-generation sequencing.

Saturation mutagenesis

HEK293T cells expressing an IKZF3 saturation mutagenesis library spanning amino acid

residues 130 to 189 were exposed to DMSO or drug, eGFP$^+$ and eGFP$^-$ cell populations were isolated using FACS, and the relative frequency of each mutation was quantified with next-generation sequencing.

X-ray structure determination

IKZF1 ZF2 and ZNF692 ZF4 were crystalized in the presence of DDB1$^{\Delta BPB}$-CRBN$^{\Delta N40}$ and pomalidomide. The structures were determined by molecular replacement using PHASER, with iterative cycles of model building in COOT followed by refinement in phenix.refine or autoBUSTER.

Docking simulations

ZF docking was carried out with the *RosettaDock* pipeline between the CRBN^{CTD} (residues 320 to 425) bound to pomalidomide and 4645 ZF models matching the pattern [x(6)- C-x(2)-C-G-x(2)-[LIVMFYWC]-x(8)-H-x(3)-H-x(2)]. The docking runs were evaluated with the *beta_nov16* scoring function.

REFERENCES AND NOTES

1. E. C. Fink, B. L. Ebert, The novel mechanism of lenalidomide activity. *Blood* **126**, 2366–2369 (2015). doi: 10.1182/blood-2015-07-567958; pmid: 26438514
2. G. Lu *et al*., The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014). doi: 10.1126/science.1244917; pmid: 24292623
3. J. Krönke *et al*., Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014). doi: 10.1126/science.1244851; pmid: 24292625
4. J. Krönke *et al*., Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS. *Nature* **523**, 183–188 (2015). doi: 10.1038/nature14610; pmid: 26131937
5. M. E. Matyskiela *et al*., A novel cereblon modulator recruits GSPT1 to the CRL4(CRBN) ubiquitin ligase. *Nature* **535**, 252–257 (2016). doi: 10.1038/nature18611; pmid: 27338790
6. G. Petzold, E. S. Fischer, N. H. Thomä, Structural basis of lenalidomide-induced CK1α degradation by the CRL4(CRBN) ubiquitin ligase. *Nature* **532**, 127–130 (2016). doi: 10.1038/nature16979; pmid: 26909574
7. E. S. Fischer *et al*., Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49–53 (2014). doi: 10.1038/nature13527; pmid: 25043012
8. K. Georgopoulos *et al*., The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143–156 (1994). doi: 10.1016/0092-8674(94)90407-3; pmid: 7923373
9. J. H. Wang *et al*., Aiolos regulates B cell activation and maturation to effector state. *Immunity* **9**, 543–553 (1998). doi: 10.1016/S1074-7613(00)80637-8; pmid: 9806640
10. R. K. Schneider *et al*., Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer Cell* **26**, 509–520 (2014). doi: 10.1016/j.ccr.2014.08.001; pmid: 25242043
11. B. S. Cobb *et al*., Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes Dev.* **14**, 2146–2160 (2000). doi: 10.1101/gad.816400; pmid: 10970879
12. N. P. Pavletich, C. O. Pabo, Crystal structure of a five-finger GLI-DNA complex: New perspectives on zinc fingers. *Science* **261**, 1701–1707 (1993). doi: 10.1126/science.8378770; pmid: 8378770
13. H. S. Najafabadi *et al*., C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* **33**, 555–562 (2015). doi: 10.1038/nbt.3128; pmid: 25690854
14. C. V. Dang, E. P. Reddy, K. M. Shokat, L. Soucek, Drugging the 'undruggable' cancer targets. *Nat. Rev. Cancer* **17**, 502–508 (2017). pmid: 28643779
15. M. Guharoy, P. Bhowmick, M. Sallam, P. Tompa, Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system. *Nat. Commun.* **7**, 10239 (2016). doi: 10.1038/ncomms10239; pmid: 26732515
16. J. An *et al*., pSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4^{CRBN} ubiquitin ligase. *Nat. Commun.* **8**, 15398 (2017). doi: 10.1038/ncomms15398; pmid: 28530236
17. H.-C. S. Yen, Q. Xu, D. M. Chou, Z. Zhao, S. J. Elledge, Global protein stability profiling in mammalian cells. *Science* **322**, 918–923 (2008). doi: 10.1126/science.1160489; pmid: 18988847
18. C. J. A. Sigrist *et al*., New and continuing developments at PROSITE. *Nucleic Acids Res.* **41** (D1), D344–D347 (2013). doi: 10.1093/nar/gks1067; pmid: 23161676
19. S. Chaudhury *et al*., Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLOS ONE* **6**, e22477 (2011). doi: 10.1371/journal.pone.0022477; pmid: 21829626
20. P. R. Hagner *et al*., CC-122, a pleiotropic pathway modifier, mimics an interferon response and has antitumor activity in DLBCL. *Blood* **126**, 779–789 (2015). doi: 10.1182/blood-2015-02-628669; pmid: 26002965
21. M. E. Matyskiela *et al*., A cereblon modulator (CC-220) with improved degradation of ikaros and aiolos. *J. Med. Chem.* **61**, 535–542 (2018). doi: 10.1021/acs.jmedchem.6b01921; pmid: 28425720
22. S. Angers *et al*., Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006). pmid: 16964240
23. X. Jin *et al*., An atypical E3 ligase zinc finger protein 91 stabilizes and activates NF-kappaB-inducing kinase via Lys63-linked ubiquitination. *J. Biol. Chem.* **285**, 30539–30547 (2010). doi: 10.1074/jbc.M110.129551; pmid: 20682767
24. D. Conomos, R. R. Reddel, H. A. Pickett, NuRD-ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. *Nat. Struct. Mol. Biol.* **21**, 760–770 (2014). doi: 10.1038/nsmb.2877; pmid: 25150861
25. T. Valenta, J. Lukas, L. Doubravska, B. Fafilek, V. Korinek, HIC1 attenuates Wnt signaling by recruitment of TCF-4 and beta-catenin to the nuclear bodies. *EMBO J.* **25**, 2326–2337 (2006). doi: 10.1038/sj.emboj.7601147; pmid: 16724116
26. I. M. Dykes *et al*., HIC2 is a novel dosage-dependent regulator of cardiac development located within the distal 22q11 deletion syndrome region. *Circ. Res.* **115**, 23–31 (2014). doi: 10.1161/CIRCRESAHA.115.303300; pmid: 24748541
27. P. Vallecillo-García *et al*., Odd skipped-related 1 identifies a population of embryonic fibro-adipogenic progenitors regulating myogenesis during limb development. *Nat. Commun.* **8**, 1218 (2017). doi: 10.1038/s41467-017-01120-3; pmid: 29084951
28. Y. Lan, P. D. Kingsley, E. S. Cho, R. Jiang, Osr2, a new mouse gene related to Drosophila odd-skipped, exhibits dynamic expression patterns during craniofacial, limb, and kidney development. *Mech. Dev.* **107**, 175–179 (2001). doi: 10.1016/S0925-4773(01)00457-9; pmid: 11520675
29. T. Morinaga *et al*., GDNF-inducible zinc finger protein 1 is a sequence-specific transcriptional repressor that binds to the HOXA10 gene regulatory region. *Nucleic Acids Res.* **33**, 4191–4201 (2005). doi: 10.1093/nar/gki734; pmid: 16049025
30. J. Kohlhase *et al*., Mutations at the SALL4 locus on chromosome 20 result in a range of clinically overlapping phenotypes, including Okihiro syndrome, Holt-Oram syndrome, acro-renal-ocular syndrome, and patients previously reported to represent thalidomide embryopathy. *J. Med. Genet.* **40**, 473–478 (2003). doi: 10.1136/jmg.40.7.473; pmid: 12843316

ACKNOWLEDGMENTS

We thank D. Neuberg and R. Redd for statistical analysis of the saturation mutagenesis screen, C. Zhu for helpful discussions on how to clone the saturation mutagenesis library, D. Haldar for biological characterization of the novel targets, J. Chen for assistance with analysis of the C2H2 ZF library screen, P. Rogers for help with FACS, and S. S. Roy Burman for advice with Rosetta docking. **Funding:** B.L.E. received funding from the NIH (R01HL082945 and P01CA108631), the Howard Hughes Medical Institute, the Edward P. Evans Foundation, and the Leukemia and Lymphoma Society. N.H.T. received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 666068). Q.L.S. was supported by award T32GM007753 from the National Institute of General Medical Sciences. G.P. was supported by the Human Frontier Science Program (HFSP Long-Term Fellowship LT000210/2014) and the European Molecular Biology Organization (EMBO Advanced Fellowship aALTF 761-2016). M.S. received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 702642. B.J.L. was supported by the National Health and Medical Research Council (Early Career Fellowships Grant APP1124979). **Author contributions:** N.H.T. and B.L.E. conceived of the concept for the identification of previously unidentified C2H2-ZF–containing targets of thalidomide analogs and advised those involved in the project. Q.L.S. generated the degradation reporter vector; identified the minimal degron required for in vivo degradation; conducted the C2H2 ZF library screen with thalidomide, lenalidomide, and pomalidomide; validated the ZF degrons; and executed the saturation mutagenesis of IKZF3 with advice and experimental and analytic assistance from T.M. G.P. purified proteins, performed TR-FRET experiments, and crystalized the IKZF1 and ZNF692 ZF degrons bound to CRBN and pomalidomide, with help from W.A. G.P. and R.D.B. carried out the structural analysis. R.D.B. performed Rosetta docking and interpreted results with input from G.P. A.R. and B.J.L. validated new targets and A.R. tested new thalidomide analogs against IKFZ3 Q147 mutants. M.S. conducted and analyzed the C2H2 ZF library screen with pomalidomide, CC-122, and CC-220. N.H.T., B.L.E., G.P., R.D.B., and Q.L.S. wrote the manuscript. **Competing interests:** B.L.E., Q.L.S., and T.S.M. are inventors on a patent application (U.S. 15/759,168, EP 168451557.7), submitted by the Broad Institute, Harvard University, and Brigham and Women's Hospital that covers methods of identifying drug-modulated polypeptide targets for degradation. B.L.E. has received research funding from Celgene. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. **Data and materials availability:** All data are publicly available. The insect cell expression plasmids are available from the Friedrich Miescher Institute for Biomedical Research under a materials transfer agreement. Structural coordinates have been deposited in the Protein Data Bank under the accession nos. 6HOF (bound to IKZF1^{ZF2}) and 6HOG (bound to ZNF692^{ZF4}).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/eaat0572/suppl/DC1
Figs. S1 to S8
Table S1
References (*31*–*43*)
Data Files S1 to S7

19 January 2018; resubmitted 28 May 2018
Accepted 12 September 2018
10.1126/science.aat0572

RESEARCH ARTICLE SUMMARY

METALLURGY

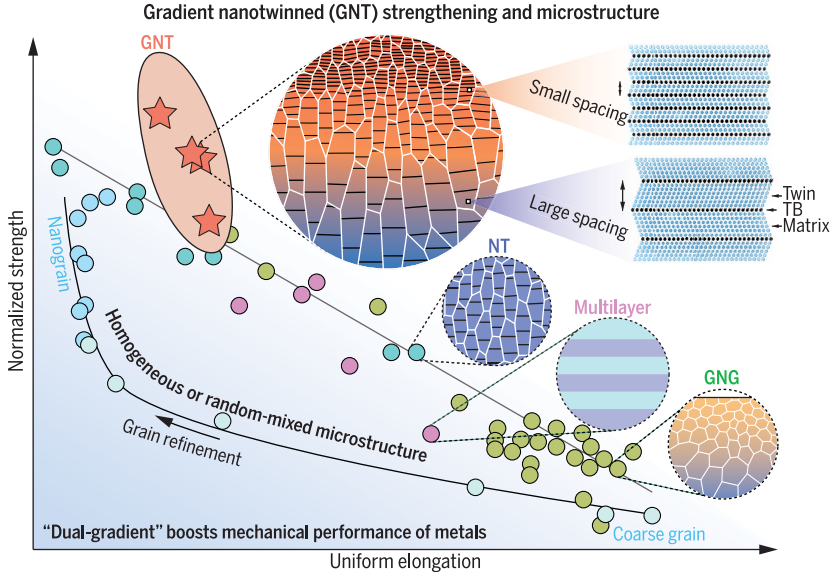
Extra strengthening and work hardening in gradient nanotwinned metals

Zhao Cheng*, Haofei Zhou*, Qiuhong Lu, Huajian Gao†, Lei Lu†

INTRODUCTION: Gradient structures ubiquitously exist in natural materials such as bone, shells, and trees. Microstructural gradients are increasingly being introduced into a wide range of engineering materials, providing them with enhanced strength, hardness, work hardening, ductility, and fatigue resistance through deformation mechanisms that are distinct from those operating in gradient-free counterparts. However, understanding structural gradient–related mechanical behaviors in all gradient structures, including those in engineering materials, has been challenging.

Although control of the structural gradient is essential to achieving optimized mechanical behaviors, existing manufacturing approaches are limiting for bulk gradient materials. For instance, surface tooling and mechanical treatments generate either limited volume fractions of gradient layers localized only near the surface or a negligible degree of structural gradient along the gradient direction. All of these limit our ability to tailor the mechanical properties and understand the deformation mechanisms of gradient-structured metals.

RATIONALE: We used a direct-current electrodeposition method to synthesize bulk gradient nanotwinned pure Cu samples with controllable patterning of homogeneous nanotwinned components. The individual components are composed of high-density, preferentially oriented nanometer-scale twin boundaries embedded within micrometer-scale columnar-shaped grains. We observed gradient-induced enhancements in both tensile strength and work-hardening rate through a wide range of structural gradients (in both twin boundary spacing and grain size) that span across the entire thickness of the sample. We combined scanning electron microscopy, a two-beam diffraction technique in transmission electron microscopy, and massively parallel atomistic simulations to identify the underlying strengthening mechanism in gradient nanotwinned Cu. We also prepared homogeneous nanotwinned samples as controls to demonstrate the importance of the structural gradient.



Highly tunable structural gradient for extra strengthening and ductility in metals. A gradient nanotwinned microstructure with a spatial gradient in both twin boundary (TB) spacing and grain size offers an unusual combination of strength, uniform elongation, and work hardening, which is superior to its strongest component and to existing heterogeneous strengthening approaches through gradient nanograined (GNG), homogeneous nanotwinned (NT), and multilayered microstructures.

RESULTS: Our findings indicate that simultaneous enhancement in strength and work hardening can be achieved by solely increasing the structural gradient in pure Cu. The maximum structural gradient led to an improved work-hardening rate and tensile strength that can exceed even the strongest component of the gradient microstructure—an unusual phenomenon that appears to be absent from the existing literature on metals and alloys with gradient nanograined or any other heterogeneous microstructures. We attribute the extra strengthening and work hardening of gradient nanotwinned structures to the unique patterning of geometrically necessary dislocations in the form of bundles of concentrated dislocations uniformly distributed in grain interiors. Such dislocation patterns in grain interiors are fundamentally different from randomly distributed, statistically stored dislocations in homogeneous structures. The bundles of concentrated dislocations with ultrahigh density of dislocations act as strong obstacles to slip, help to delocalize plastic deformation inside the grains, and accelerate the work-hardening process.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau1925

CONCLUSION: In our simple bottom-up approach to create gradient nanotwinned structure in pure Cu, a large structural gradient allows for an exceptional combination of high strength and work hardening that can exceed even the strongest component of the gradient structure. Both experimental and computational evidence suggest the importance of covering the whole structure with tunable structural gradients for the development of high densities of dislocations in grain interiors. The gradient nanotwinned strengthening concept proposed in this work provides insights into combining structural gradients at different length scales in order to push forward the strength limit of materials and may be essential to creating the next generation of metals with both high strength and high ductility. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: huajian_gao@brown.edu (H.G.); llu@imr.ac.cn (L.L.)
Cite this article as Z. Cheng *et al.*, *Science* **362**, eaau1925 (2018). DOI: 10.1126/science.aau1925

RESEARCH ARTICLE

METALLURGY

Extra strengthening and work hardening in gradient nanotwinned metals

Zhao Cheng[1,2]*, Haofei Zhou[3]*, Qiuhong Lu[1], Huajian Gao[3]†, Lei Lu[1]†

Gradient structures exist ubiquitously in nature and are increasingly being introduced in engineering. However, understanding structural gradient–related mechanical behaviors in all gradient structures, including those in engineering materials, has been challenging. We explored the mechanical performance of a gradient nanotwinned structure with highly tunable structural gradients in pure copper. A large structural gradient allows for superior work hardening and strength that can exceed those of the strongest component of the gradient structure. We found through systematic experiments and atomistic simulations that this unusual behavior is afforded by a unique patterning of ultrahigh densities of dislocations in the grain interiors. These observations not only shed light on gradient structures, but may also indicate a promising route for improving the mechanical properties of materials through gradient design.

Making metals stronger and tougher by tailoring their microstructures has been an enduring pursuit in materials science and engineering (*1*–*3*). Gradient metals have emerged as a class of hierarchical materials with tunable distribution of microstructures (*4*–*8*). As a result of the substantial strain gradients (*9*–*13*) intrinsic to these materials, gradient metals could be designed to exhibit an unusual combination of high strength and hardness, considerable ductility, and enhanced strain hardening and fatigue resistance (*4*–*6*, *14*). However, it remains a challenge to understand the underlying mechanisms that control the strengthening of gradient materials because of the often complex dislocation structures in plastic deformation (*6*, *15*, *16*).

Meanwhile, the design of nanotwinned structures has been recognized as a universal strategy in the pursuit of superior mechanical and physical properties (*1*, *5*, *17*–*20*). Various processing techniques have been adopted to synthesize highly oriented nanotwinned structures (*20*) in a range of pure metals and alloys (*21*). Such materials usually exhibit plastic anisotropy, as their deformation mechanisms are strongly dependent on the orientation of loading axis with respect to the twin boundaries (TBs) (*18*, *22*, *23*). When the loading direction is oriented in parallel to the twin planes, single slip by threading dislocations (*18*) is activated—a mechanism similar to one previously observed in thin films and nanolaminated metallic composites (*24*, *25*). Most recently, a history-independent and stable cyclic response in highly oriented nanotwinned copper was reported (*19*). Such unique cyclic behavior is governed by a type of correlated "necklace" dislocations formed by the linking of threading dislocations in adjacent twins, leading to a superstable twin structure and negligible cyclic damage, which is fundamentally distinct from the inevitably irreversible damage accumulation in single-crystal, coarse-grained, ultrafine-grained, and nanograined metals (*19*).

We performed experiments and atomistic simulations on bulk copper samples with a controllable spatial gradient in the highly oriented nanotwinned structure. We found an interesting strengthening effect and obtained simultaneous enhancement in strength and work hardening by increasing the structure gradient. By taking advantage of the simple deformation patterns of highly oriented nanotwins (*18*, *26*), we show that this unusual strengthening behavior is governed by bundles of concentrated dislocations in the grain interiors that help to accommodate large strain gradients and suppress strain localization. The gradient nanotwinned (GNT) structures provide a promising route to the design of new generations of high-performance metals.

Gradient nanotwinned structures

We used direct-current electrodeposition to prepare GNT Cu samples with a controllable patterning of homogeneous nanotwinned components. A typical GNT sample (GNT-1, Fig. 1A) has a patterning of homogeneous nanotwinned components we labeled ⒶⒷⒸⒹ, from bottom to top. We found no sharp interfaces between the homogeneous components from cross-sectional scanning electron microscopy (SEM) observations (Fig. 1B). Magnified SEM (Fig. 1, C and E) and transmission electron microscopy (TEM) (Fig. 1, D and F) observations of components Ⓐ and Ⓓ reveal that the individual components are composed of preferentially oriented nanometer-scale TBs embedded within the micrometer-scale columnar-shaped grains, with a mean aspect ratio of ~3. The selected-area electron diffraction (SAED) pattern and high-resolution TEM image (Fig. 1D, insets) show that the twin planes are coherent Σ3 boundaries. From component Ⓐ to component Ⓓ, the average twin thickness increases from 29 nm to 72 nm and the average grain size (defined as the size measured along the short axis) increases from 2.5 μm to 15.8 μm (Fig. 1G and fig. S1). To quantify the structural gradient of the GNT microstructure, we define a parameter, s, as the increment in hardness per unit thickness along the gradient direction. The cross-sectional hardness decreases gradually from 1.5 GPa in component Ⓐ to 0.8 GPa in component Ⓓ, resulting in a structural gradient of s = 1.75 GPa/mm in sample GNT-1 (Fig. 1H). By changing the stacking sequence of the homogeneous nanotwinned components while keeping the total sample thickness fixed at 400 μm, we designed three other GNT structures (GNT-2, GNT-3, and GNT-4) that possess various spatial distributions of the homogeneous components: ⒶⒷⒸⒹⒹ-ⒸⒷⒶ, 2×ⒶⒷⒸⒹⒹⒸⒷⒶ, and 4×ⒶⒷⒸⒹ-ⒹⒸⒷⒶ, respectively (fig. S2). Each homogeneous nanotwinned component has the same volume fraction of 25% and thickness of 100 μm in these GNT samples. The structural gradients of the GNT samples range from 1.75 to 11.6 GPa/mm, occupying the entire thickness of the sample (fig. S2D and tables S1 and S2) (*4*–*6*, *27*).

Strengthening and work-hardening ability

We compared engineering tensile stress-strain curves of the GNT samples to the tensile behaviors of the homogeneous nanotwinned components in free-standing conditions (Fig. 2A). The yield strength (σ_y), stress at 1% strain ($\sigma_{1\%}$), and ultimate tensile strength (σ_{uts}) of GNT-1, with the smallest s = 1.75 GPa/mm, are close to the average of the corresponding values of the four individual components indicated by the open symbols in Fig. 2B. As the structural gradient increases, both σ_y and σ_{uts} of the GNT structure gradually increase at the expense of a slight decrease in the uniform elongation (Table 1). In particular, GNT-4 exhibits a σ_y value of 481 ± 15 MPa that even exceeds the corresponding value of the strongest component Ⓐ, while its uniform elongation is 7%, larger than the corresponding values of components Ⓐ and Ⓑ. Meanwhile, the work-hardening rates, Θ, of the GNT samples are much higher than the average of the four homogeneous components, especially at strains less than 2% (Fig. 2C). We observed an increase with structural gradient for the work-hardening rate at ε = 1% (Fig. 2C, inset, and fig. S3). Surface morphologies indicate no surface cracking in GNT-4 before fracture

[1]Shengyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China. [2]School of Materials Science and Engineering, University of Science and Technology of China, Shenyang 110016, China. [3]School of Engineering, Brown University, Providence, RI 02912, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: huajian_gao@brown.edu (H.G.); llu@imr.ac.cn (L.L.)

failure, which further confirms that component Ⓐ sustains a large uniform elongation under a stress level much higher than its tensile strength measured in free-standing conditions. The above observations strongly suggest that a substantial strengthening effect can be achieved solely by increasing the structural gradient.

Post-mortem microstructures of GNT samples

Taking GNT-4 as an example, we characterized the post-mortem microstructures with SEM and TEM observations. Our observations (Fig. 3, A and E) show a class of unexpected parallel aligned dislocation bundles in the interior of the columnar grains at ε = 1% in both components Ⓐ and Ⓓ. Each dislocation bundle spans across multiple nanotwins and extends through almost the entire length of the grain perpendicular to TBs, marked by red arrows (Fig. 3, A and E), which coincides with the direction of the structural gradient in the GNT structure. Such dislocation patterns were observed in all nanotwinned components and have an average width that varies from 0.5 to 1.5 μm, as shown in the cross-sectional SEM images obtained by backscattered electrons (fig. S4).

We then adopted a two-beam diffraction technique (*26*) in TEM to determine the detailed configurations of the bundles of concentrated dislocations (BCDs), including the type, density, spatial arrangement, and Burgers vector of dislocations therein. We examined a typical BCD in component Ⓓ (Fig. 3A, white square) in TEM images (Fig. 3, B and C). We observed dislocation segments (identified by green arrows in Fig. 3B) with Burgers vectors of $\mathbf{b}_I$ = **DA**, **DB**, and **DC** (see the inset Thompson tetrahedron of Fig. 3A) inclined to the TBs by using a diffraction vector of $\mathbf{g}_M = \mathbf{g}_T$ = 111. We refer to these as mode I dislocations, following the definition in (*26*). Mode I dislocations are known to control the plastic deformation in highly oriented nanotwins when the loading direction is perpendicular to the TBs, leading to the classical Hall-Petch strengthening with reduced twin thickness (*18*). TEM observations confirmed that mode I dislocations are rarely observed in free-standing, homogeneous, highly oriented nanotwins when the loading direction is parallel to TBs (fig. S5).

Meanwhile, we observed a large amount of short dislocation segments (indicated by orange arrows in Fig. 3, C and D) located in the moiré fringes of the TBs with Burgers vectors of $\mathbf{b}_{II}$ =**AB**, **BC**, **CA** (Fig. 3A, inset) that are parallel to the twin planes by using a diffraction vector of $\mathbf{g}_M$ = 200. We call these mode II dislocations (*26*), which are often found to be prevalent in thin films (*24*), multilayered composites (*25*), and highly oriented nanotwinned structures (*18*, *26*).

In both components Ⓐ and Ⓓ, our observations show similar dislocation structures consisting of a majority of mode II dislocations (Fig. 3, C, D, G, and H) measured in and out of the BCDs (marked by the white square in Fig. 3, A and E), and much fewer mode I dislocations that span across multiple twin lamellae (Fig. 3, B and F). For instance, the BCDs we observed in components Ⓐ and Ⓓ contain mode II dislocations with a density of 2.1×10^{14} to 3.7×10^{14} m^{-2}, about

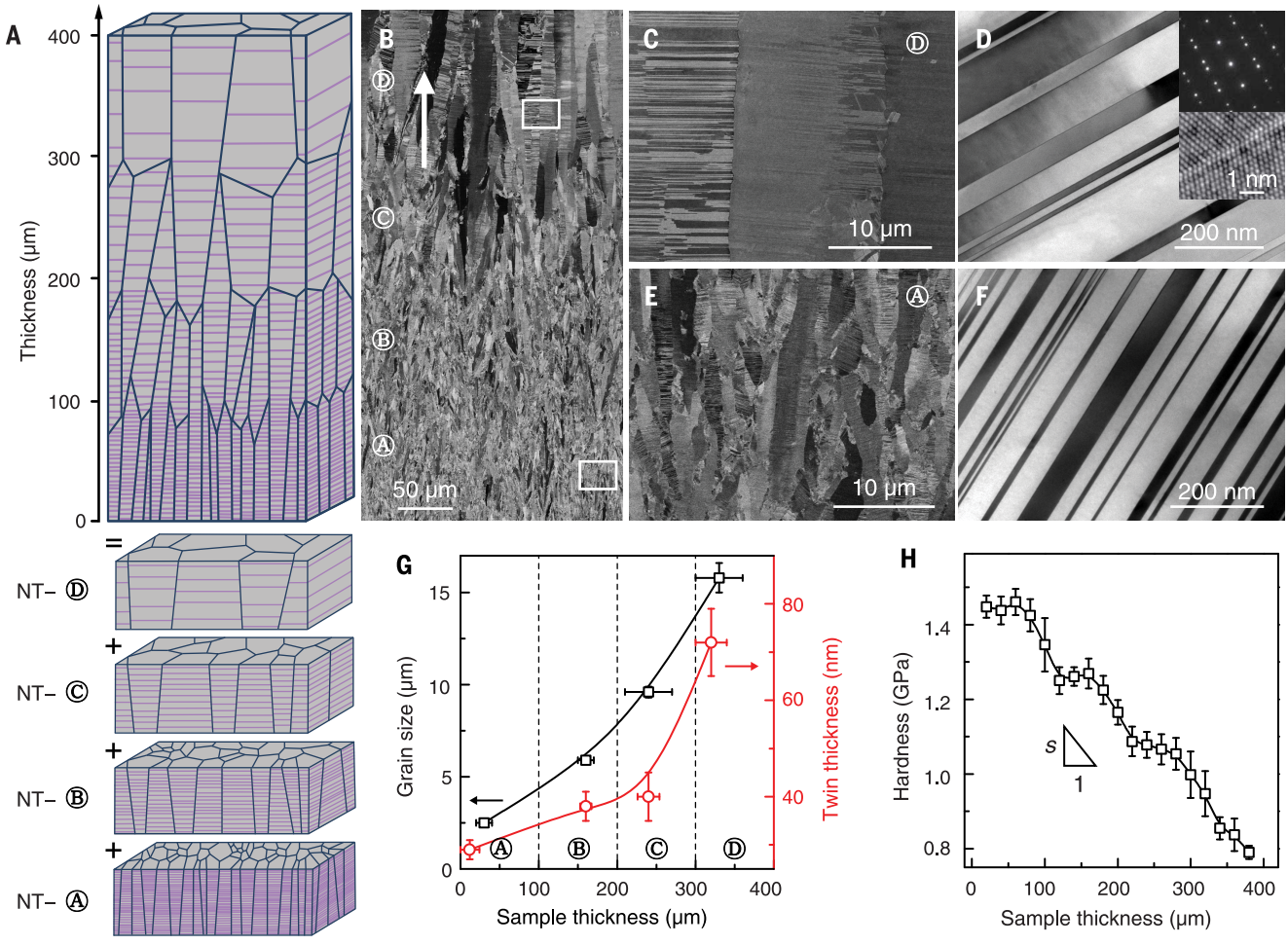


Fig. 1. Microstructure and structural gradient of a typical GNT specimen, GNT-1. (**A** and **B**) Schematics (A) and SEM observation (B) of GNT-1 composed of homogeneous nanotwinned components Ⓐ, Ⓑ, Ⓒ, and Ⓓ. In (B), gradient direction is indicated by the white arrow; white boxes outline the areas magnified in (C) to (F). (**C** to **F**) Magnified SEM and TEM images of components Ⓓ [(C) and (D)] and Ⓐ [(E) and (F)]. The insets in (D) show the SAED pattern and high-resolution TEM image of component Ⓓ. (**G**) Distribution of grain size and twin thickness along the sample thickness. (**H**) Distribution of hardness along the sample thickness. The vertical error bars in (G) and (H) represent standard deviations from statistical analyses with 500 (grain size), 1000 (twin thickness) and 5 to 10 (hardness) independent measurements, respectively. The horizontal error bars in (G) indicate the range of sample thickness of the measured samples.

an order of magnitude larger than that of mode I dislocations in the bundles (Table 2).

The observation of numerous BCDs containing both mode I and mode II dislocations in the grain interiors is distinct from the single activation of mode II dislocations that leads to strain localization near grain boundaries (GBs) in homogeneous nanotwinned samples at ε = 1% (fig. S5). In the latter case, only a negligible amount of mode I dislocations are detectable near the GBs, and mode II dislocations are randomly distributed in the columnar grains. Moreover, the average density of mode II dislocations in GNT-4 is measured to be about 1.7 × 10^{14} to 2.8 × 10^{14} m^{-2}, which is 4 to 8 times the densities observed in homogeneous nanotwinned samples (Table 2).

Atomistic simulations of extra strengthening in GNT Cu

To understand the underlying deformation mechanisms of the extra strengthening in GNT Cu, we performed large-scale fully three-dimensional atomistic simulations on samples that mimic the experimental setup. The largest simulation involves about 130 million atoms. A typical GNT simulation sample, GNT-ⒶⒷ, was created by combining two $[1\bar{1}1]$-textured homogeneous nanotwinned components (NT-Ⓐ and NT-Ⓑ) (Fig. 4A). We selected numbers of grains as n_A = 9 and n_B = 4 with average grain sizes of d_A = 30 nm and d_B = 50 nm, and twin thicknesses as λ_A = 0.83 nm and λ_B = 6.23 nm, respectively. By changing the thickness of the nongradient nanotwinned components, we tuned the structural gradient and created two different GNT samples with structural gradients of s = 6.2 MPa/nm (GNT-ⒶⒷ) and 12.4 MPa/nm (GNT-ⒶⒷⒶⒷ) (fig. S6, A and B). We measured the stress-strain curves of the GNT samples under uniaxial tensile loading along the x axis and compared them with the tensile responses of homogeneous samples NT-Ⓐ and NT-Ⓑ (Fig. 4B). Both gradient samples exhibit enhanced yield strength (defined as the average flow stress in the strain range of 6 to 15%) relative to their homogeneous counterparts. The strength increases with increasing structural gradient. We also performed simulations on GNT samples with statistically equivalent microstructures and observed the same extra strengthening effect (fig. S7). Thus, both our simulations and experiments demonstrate an unexpected ultrahigh strength in GNT Cu samples, which is higher than that of the strongest component of the structure.

Table 1. Tensile properties of GNT samples and free-standing homogeneous nanotwinned components. σ_y, yield strength; σ_{uts}, ultimate tensile strength; δ_u, uniform elongation.

Sample	σ_y (MPa)	σ_{uts} (MPa)	δ_u (%)
GNT-1	364 ± 12	397 ± 11	10 ± 2
GNT-2	434 ± 12	456 ± 15	9 ± 1
GNT-3	437 ± 19	471 ± 18	9 ± 1
GNT-4	481 ± 15	520 ± 12	7 ± 1
NT-Ⓐ	446 ± 10	470 ± 11	1 ± 0.5
NT-Ⓑ	394 ± 6	421 ± 5	6 ± 0.5
NT-Ⓒ	325 ± 4	350 ± 7	10 ± 0.5
NT-Ⓓ	223 ± 9	272 ± 4	22 ± 0.5

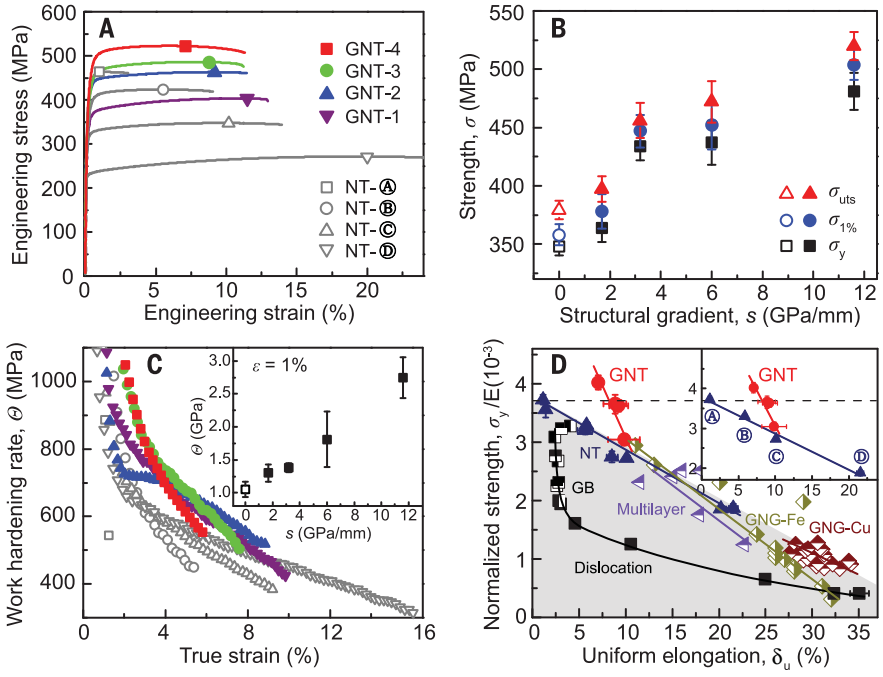


Fig. 2. Mechanical properties of GNT structures. (**A**) Tensile engineering stress-strain relations of GNT samples in comparison with those of homogeneous nanotwinned components. (**B**) Yield strength (σ_y), stress at ε = 1% ($\sigma_{1\%}$), and ultimate tensile strength (σ_{uts}) of GNT samples with various structural gradients. (**C**) Work-hardening rate and true strain relations of GNT samples compared with homogeneous components. The inset shows the variation of work-hardening rate measured at 1% strain with respect to structural gradient, s. Open symbols in (B) and (C) indicate average strength and work-hardening rate of the four homogeneous components. (**D**) Yield strength normalized by Young's modulus versus uniform elongation of GNT samples, in comparison with metals and alloys strengthened by dislocations (*43*), GBs (*44*, *45*), highly oriented nanotwins (*20*), and multilayered (*46*) and gradient nanograined (GNG) (*6*, *43*) structures. The inset shows the strengthening of GNT samples in comparison with homogeneous components. The error bars in (B) to (D) represent standard deviations from 6 to 12 independent tensile tests.

Formation mechanism of dislocation bundles in GNT Cu

The schematics (Fig. 4, C and D) illustrate typical deformation patterns of the GNT samples suggested by our experimental observations (Fig. 3). During deformation, a dislocation bundle forms in the interior of a nanotwinned columnar grain (colored pink) in the GNT structure (Fig. 4C). The bundle is composed of mode I dislocations intersecting with TBs and mode II dislocations gliding parallel to the TBs (Fig. 4D). We can show that the interactions between mode I and mode II dislocations in the dislocation bundles also lead to the formation of stair-rod dislocations (*22*).

We found from our simulations that upon yielding at ~4% strain, mode I dislocations (black triangles in Fig. 4, E and F, and fig. S8A) emit from the interface (*28*) toward nearby TBs. Meanwhile, mode II dislocations (black triangles in Fig. 4, G and H, and fig. S8A) are nucleated from lateral GBs. The simultaneous activation of mode I and II dislocations agrees with our experimental observations (Fig. 3). As the applied strain increased to 10%, a dislocation bundle was identified in the grain interior (fig. S8, B and C, and movie S1), where the dislocation density was measured as 2.2 × 10^{17} m^{-2}, almost an order of magnitude larger than that outside the dislocation bundle (fig. S8D). We found from our detailed analysis of the dislocation structures (Fig. 4I) that sessile stair-rod dislocations (in pink) account for about 15% of the total dislocation density. We attributed the formation of a typical stair-rod dislocation **δγ/BD** to the interaction between **γD** and **δB**, which are Shockley partial dislocations

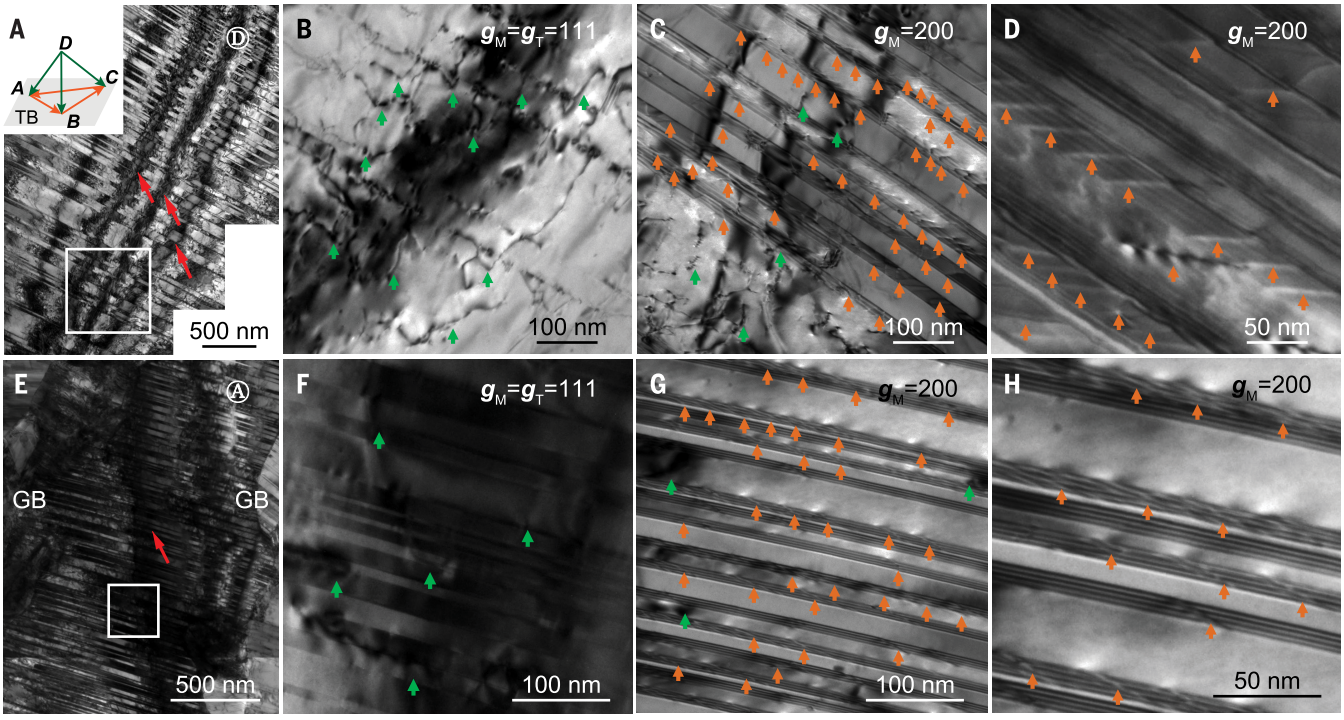


Fig. 3. Deformation microstructures of GNT-4 at 1% strain. (**A** and **E**) Bundles of concentrated dislocations (BCDs) are indicated by red arrows in components Ⓓ (A) and Ⓐ (E). (**B** and **C**) TEM images of the BCD region indicated by the white square in (A) identified by two-beam diffraction using vectors of $\mathbf{g}_M = \mathbf{g}_T$ = 111 (B) and $\mathbf{g}_M$ = 200 (C), respectively. (**D** and **H**) Magnified TEM images of the dislocations observed in (C) and (G), respectively. (**F** and **G**) TEM images of the BCD region indicated by the white square box in (E) identified by two-beam diffraction using vectors of $\mathbf{g}_M = \mathbf{g}_T$ = 111 (F) and $\mathbf{g}_M$ = 200 (G), respectively. Mode I dislocations are indicated by green arrows in (B), (C), (F), and (G); mode II dislocations are indicated by orange arrows in (C), (D), (G), and (H).

dissociated from mode I and mode II dislocations, respectively (Fig. 4J and fig. S8E). The movement of Shockley partial dislocations in the bundle region was thus substantially suppressed, contributing to additional strength enhancement (movie S2). For comparison, we observed a single slip mechanism governed by mode II dislocations and no formation of BCDs in the grain interior in homogeneous nanotwinned samples (fig. S9 and movie S3).

Table 2. Dislocation density measurement of bundles of concentrated dislocations (BCDs) and grain interiors for components Ⓐ and Ⓓ in GNT-4 and under free-standing homogeneous conditions at 1% strain.

Dislocation type	GNT-4, ρ (10^{13} m^{-2})				Homogeneous, ρ (10^{13} m^{-2})	
	Ⓐ		Ⓓ		Ⓐ	Ⓓ
	BCD	Average	BCD	Average		
Mode I	4.3	2.9	3.0	1.4	A few near GBs	
Mode II	36.7	28.3	20.6	17.4	7.2	2.3

Discussion

The storage of dislocations causes work hardening in crystals (*29*). For uniform plastic deformation, dislocations accumulate by trapping one another in a random way, giving rise to the storage of the so-called statistically stored dislocations (SSDs) (*9*). We found that as a result of the plastic strain gradients, the GNT components are not equally deformed during uniaxial tension (*30*). Component Ⓓ with lower yield strength plastically deforms before component Ⓒ. In a similar fashion, Ⓒ plastically deforms prior to Ⓑ. Component Ⓐ is the last to yield, leading to a gradient of plastic deformation in the GNT structure (fig. S10). In such circumstances, geometrically necessary dislocations (GNDs) (*9*–*13*) develop to accommodate the deformation gradients and allow compatible deformation among different components.

The concept of GNDs has been widely applied to explain size-dependent flow properties of crystals under nonuniform loading conditions including bending, torsion, indentation, and deformation at crack tips where stress and strain gradients are large (*10*, *11*, *31*, *32*). GNDs also explain the strengthening of polycrystalline metals (*9*), multilayer/multiphase metals (*33*–*35*), and gradient metals (*4*–*6*) due to inhomogeneous microstructures. However, the complex multiple slip systems in conventional materials make it challenging to distinguish GNDs from SSDs (*10*, *11*, *36*, *37*).

We took advantage of the simple deformation patterns of highly oriented coherent twins and found that the structural gradient in the GNT samples leads to the formation of a unique pattern of GNDs in the grain interiors (Fig. 3). These are in the form of BCDs composed of a majority of mobile mode II dislocations, which we never observed in free-standing nongradient nanotwinned components (fig. S5).

The local dislocation density we measured in the dislocation bundles is nearly 30% larger than the average dislocation density of the grain (Table 2), demonstrating the intense storage of GNDs in BCDs. We measured lattice rotation with a misorientation angle of ~5.2° for BCDs in component Ⓓ of GNT-4 (Fig. 3A) using an electron procession diffraction technique (*38*) in TEM, furthering demonstrating the concentration of GNDs in BCDs. Moreover, these concentrated GNDs may act as strong obstacles to slip and

accelerate the rate of statistical storage of dislocations in the material via long-range back stress (*35*, *39*, *40*), which increases substantially with structural gradient in the GNT samples (fig. S11). The GNT structures thus work-harden much faster than the nongradient nanotwinned samples (Fig. 2C). The average dislocation densities of GNT samples upon yielding are substantially elevated relative to those measured in the free-standing components (Table 2). For instance, at a small strain of 1%, the average dislocation density measured in GNT-4 is about 1.9 × 10^{14} to 3.1 × 10^{14} m^{-2}, which is an exceptionally high density of dislocations rarely reported in experiments (*41*). Interestingly, this value is also

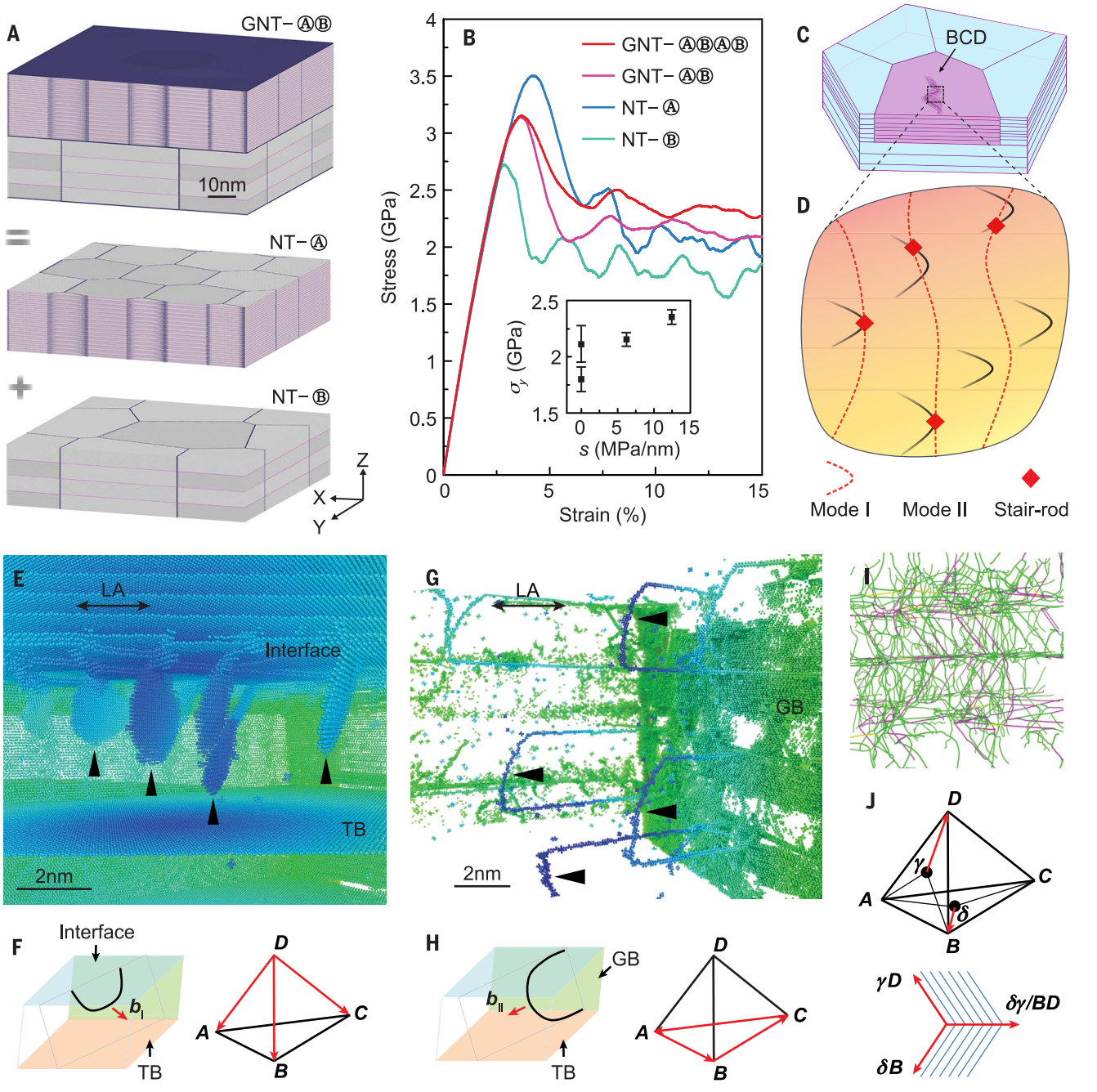


Fig. 4. GNT strengthening and associated deformation mechanisms by large-scale atomistic simulations. (**A**) Construction of a typical GNT structure, GNT-ⒶⒷ, through combinations of nanotwinned components NT-Ⓐ and NT-Ⓑ. (**B**) Stress-strain curves of GNT samples. The inset shows the relation between yield strength σ_y (the averaged flow stress from the strain range of 6 to 15%) and structural gradient *s*. Error bars represent the standard deviation from statistical analyses with 9000 sampling points. (**C**) Formation of a BCD in a nanotwinned grain (pink) in the GNT structure. (**D**) Zoom-in view of the BCD where interactions between mode I and mode II dislocations lead to stair-rod dislocations. (**E**) Emission of mode I dislocations (indicated by black triangles) from interface. (**F**) Schematic of mode I dislocations with Burgers vectors of **DA**, **DB**, and **DC**. (**G**) Emission of mode II dislocations (indicated by black triangles) from GBs. (**H**) Schematics of mode II dislocations with Burgers vectors of **AB**, **BC**, and **CA**. (**I**) High density of dislocations in the BCD composed of stair-rod dislocations (pink) and Shockley partial dislocations (green). (**J**) Schematics illustrating the formation of a stair-rod dislocation (**δγ/BD**).

two orders of magnitude higher than the density of GNDs (1.6×10^{12} m^{-2}) estimated from the plastic strain gradient along the thickness of the GNT structure, following Ashby's original work (*9*).

We can estimate the extra strengthening effect of the BCDs with Taylor's relation (*10*),

$$\Delta\sigma = M\alpha\mu b\left(\sqrt{\rho^{GNT}} - \sqrt{\rho^{NT}}\right) \quad (1)$$

where M = 3.17 is the Taylor factor of parallel loading with respect to TBs (*18*), μ = 42 GPa is the shear modulus of pure Cu, $b = 2.55 \times 10^{-10}$ m is the magnitude of the Burgers vector of dislocations, and α = 0.47 is the empirical constant for non-cell dislocations (*42*). We measured the total density of dislocations in components Ⓐ and Ⓓ of GNT-4 at 1% strain to be $\rho^{GNT} = 31.2 \times 10^{13}$ m^{-2} and 18.8×10^{13} m^{-2}, respectively; the corresponding values of the two components in their free-standing and homogeneous state are $\rho^{NT} = 7.2 \times 10^{13}$ m^{-2} and 2.3×10^{13} m^{-2}, respectively (Table 2). We support these values with our measurement that the back stresses of the GNT samples increase rapidly at low strains (fig. S11). The extra strength estimated from Eq. 1 is 146 MPa in component Ⓐ and 142 MPa in component Ⓓ. This agrees well with the extra yield strength of 134 MPa of GNT-4 (Fig. 2B).

The formation of BCDs also helps to delocalize plastic deformation as they are uniformly distributed inside the grains, unlike the strain concentration at GBs in nongradient components (fig. S5). Most important, our atomistic simulations show that most of the dislocations in BCDs are mobile Shockley partial dislocations (Fig. 4I), which are expected to accommodate large amounts of plastic deformation without a noticeable sacrifice in the ductility of the material (Fig. 2).

We compared the remarkable strength-ductility synergy of the GNT samples with various strengthening strategies reported in the literature (*6*, *43*–*46*) (Fig. 2D). We have normalized the strength by the Young's modulus of the materials. The gradient nanotwinned structure brings out an extra strengthening effect relative to homogeneous highly oriented nanotwins or multilayered materials strengthened by threading dislocations (*20*). In gradient nanograined structures, the density of GNDs is thought to be related to the plastic strain gradient (*30*). GNDs prefer to accumulate in the vicinity of GBs (*9*, *30*), inducing GB migration and grain coarsening or softening during plastic deformation (*4*). In contrast, by introducing a structural gradient in GNT structures, we obtained a high GND density in grain interiors (Table 2) and an exceptionally high strength exceeding that of the strongest component. When the structural gradient is sufficiently large, the strengthening effect of gradient nanotwins is superior to most existing gradient metals and alloys (*4*, *6*, *20*, *46*). The classical theory of strain gradient plasticity (*9*, *11*) only considers the plastic strain gradient due to incompatible plastic deformation in constituent components and may substantially underestimate the density of GNDs and the structural gradient strengthening in GNT structures.

Our observations point out the importance of designing gradient microstructures at multiple length scales in pursuit of high structural gradients for enhanced mechanical performance. The GNT strengthening strategy proposed in this work not only should help to advance the fundamental understanding of the processing-structure-property relationships of existing gradient-structured metals, but also suggests new challenges and opportunities for experimental, computational, and theoretical studies of next-generation metals and alloys.

Materials and methods

Materials

By means of direct-current electrodeposition technique, GNT Cu samples were fabricated by adjusting the electrolyte temperature in the range of 20° to 35°C. The current density was 30 mA/cm^2. The electrolyte was made with a $CuSO_4$ concentration of 90 g/liter. The pH value was ~1. The stirring speed was 1500 rpm and the additive concentration was 2 mg/liter. A commercial-purity Ti sheet was used as the cathode and a highly purified (99.995 wt %) Cu sheet was used as the anode. In the setup, the distance between the electrodes was kept as ~10 cm.

Various profiles of electrolyte temperature were used to synthesize GNT Cu samples (fig. S2B). The temperature was elevated stepwise (20°, 25°, 30°, 35°C) and preserved for 4 hours at each temperature to create GNT-1 with a spatial distribution of ⒶⒷⒸⒹ. The temperature rate, defined as temperature variation per hour, was 0.94°C/hour for GNT-1. Higher temperature rates of 1.88°C/hour, 3.75°C/hour, and 7.5°C/hour were used to fabricate GNT-2 (ⒶⒷⒸⒹⒹⒸⒷⒶ), GNT-3 (2×ⒶⒷⒸⒹⒹⒸⒷⒶ), and GNT-4 (4×ⒶⒷⒸⒹⒹⒸⒷⒶ), respectively (fig. S2, A and C). During the fabrication of the GNT samples, the total electrodeposition time was controlled to be 16 hours (i.e., 4 hours for each temperature) to ensure a sample thickness of about 400 μm and a volume fraction of 25% for each component.

For comparison, free-standing homogeneous nanotwinned Cu samples Ⓐ, Ⓑ, Ⓒ, and Ⓓ were constructed by electrodeposition at 20°, 25°, 30°, and 35°C, respectively. A total deposition time of 16 hours resulted in a sample thickness of about 400 μm.

Uniaxial tensile tests

For tensile tests, dog bone–shaped flat specimens with a gauge length of 5 mm and width of 2 mm were cut from the as-deposited GNT and nanotwinned Cu sheets using an electric spark machine. Before tensile tests, all tensile specimens were mechanically and electrochemically polished to minimize surface roughness. Uniaxial quasi-static tensile tests were carried out in an Instron 5848 micro tester with a 2-kN load cell at a strain rate of 5×10^{-3} s^{-1} at ambient temperature. During tensile tests, a contactless MTS LX300 laser extensometer was used to calibrate and measure the strain in the gauge. To confirm reproducibility, we repeated tensile tests 6 to 12 times for each structure, from which average values and standard errors were calculated for yield strength (at 0.2% offset strain), stress at 1% strain, ultimate tensile strength, and uniform elongation.

Micro-hardness measurements

Cross-sectional micro-hardness measurements were performed along the thickness of the GNT samples on a Qness Q10A+ micro-hardness tester with a load of 50 g and a loading time of 10 s. For each depth, 5 to 10 independent measurements were conducted.

Microstructure observations

Cross-sectional microstructures of GNT samples were characterized in a FEI NovaSEM 430 field emission gun SEM using backscattering electron imaging with a high-contrast (vCD) detector. The average grain size obtained from the SEM images was measured over 500 grains for each depth of the GNT samples. Samples were mechanically polished and electrochemically polished in a solution of phosphoric acid (25%), alcohol (25%), and deionized water (50%) at ambient temperature before SEM observations.

Cross-sectional microstructures of as-deposited and deformed GNT samples were also examined using a FEI Tecnai G2 F20 TEM operated at 200 kV. The average twin thickness was obtained from TEM images with edge-on TBs and was measured over 1000 twin lamellae for each depth of the GNT samples. A two-beam diffraction technique (*26*) was used to identify the type of dislocations (*18*, *26*) in the deformed GNT and free-standing nanotwinned samples. The TEM foils of the deformed samples were sliced from the gauge section along the direction parallel to the tensile axis. Each TEM foil was attached to a Cu ring 3 mm in diameter and thinned by twin-jet electrochemical polishing in a solution of phosphoric acid (25%), alcohol (25%), and deionized water (50%) at about −10°C.

Computational GNT Cu samples

Three GNT Cu samples were constructed for atomistic simulations (fig. S6). We started by building two homogeneous columnar-grained nanotwinned components (Fig. 4A), NT-Ⓐ and NT-Ⓑ, each having a $[1\bar{1}\bar{1}]$ our-of-plane texture and zero structural gradient (i.e., s = 0). By varying the thickness of the individual components and combining them in different stacking sequences (e.g., ⒶⒷ or ⒶⒷⒶⒷ), GNT samples with various structural gradients can be generated.

The homogeneous components NT-Ⓐ and NT-Ⓑ used to generate GNT-ⒶⒷ (fig. S6A) and GNT-ⒶⒷⒶⒷ (fig. S6B) have grain numbers of n_A = 9 and n_B = 4, average grain sizes of d_A = 30 nm and d_B = 50 nm, and twin thicknesses of λ_A = 0.83 nm and λ_B = 6.3 nm, respectively. The thicknesses of NT-Ⓐ and NT-Ⓑ are 25 nm in GNT-ⒶⒷ and 12.5 nm in GNT-ⒶⒷⒶⒷ. The volume fractions of NT-Ⓐ and NT-Ⓑ are equal in both GNT-ⒶⒷ and GNT-ⒶⒷⒶⒷ. Both GNT-ⒶⒷ and GNT-ⒶⒷⒶⒷ contain ~42,000,000 atoms and have dimensions of ~$100 \times 100 \times 50$ nm^3.

The homogeneous components NT-Ⓐ and NT-Ⓑ used to generate GNT-ⒶⒷ-2 (fig. S6C) have thicknesses of h_A = 6.3 nm and h_B = 33.2 nm, grain numbers of n_A = 400 and n_B = 4, average grain sizes of d_A = 10 nm and d_B = 100 nm, and twin thicknesses of λ_A = 0.63 nm and λ_B = 8.3 nm, respectively. The entire sample contains ~130,000,000 atoms and has dimensions of ~200 × 200 × 40 nm^3. The much larger "in-plane" size of GNT-ⒶⒷ-2 allows us to capture the formation process of a BCD in the interior of the nanotwinned grains. The large number of triple junctions at the interface between components NT-Ⓐ and NT-Ⓑ are potential dislocation sources during deformation (*28*).

The same relaxation, tension, and visualization methods were adopted for the simulated samples. Specifically, the system was initially equilibrated for 500 ps at 300 K, followed by uniaxial tensile loading along the x axis to a total strain of 15% at a constant strain rate of 2×10^8 s^{-1}. Throughout the simulation, periodic boundary conditions were applied in all three directions. Constant temperature and zero pressure in the nonstretching directions (i.e., y and z axes) were controlled by Nose-Hoover thermostatting and barostatting (*47*, *48*). The embedded atom method potential was used to compute the interatomic forces (*49*), and the integration time step was fixed at 1 fs. The common neighbor analysis (CNA) method (*50*) was used to identify defects that emerge during plastic deformation. A position-based coloring scheme (*19*) was used to visualize defects in which colors are assigned according to the spatial coordinates of atoms.

Calculation of dislocation density of GNDs from plastic strain gradient

Following Ashby's original work (*9*), the density of GNDs, ρ^{GNT}, was calculated by

$$\rho^{GNT} = \frac{\eta}{b} \quad (2)$$

where b = 2.55 × 10^{-10} m is the magnitude of the Burgers vector of dislocations, and η is the average plastic strain gradient in GNT calculated through

$$\eta = \frac{\varepsilon_p^D - \varepsilon_p^A}{t_{AD}} \quad (3)$$

where ε_p^A and ε_p^D are the plastic strains of components Ⓐ and Ⓓ, respectively. The spacing between components Ⓐ and Ⓓ, t_{AD}, increases with the structural gradient s. The difference in the plastic strain ($\varepsilon_p^D - \varepsilon_p^A$), originating from the different yield strengths of components Ⓐ and Ⓓ, reaches a maximum value of ~0.2% and remains constant after component Ⓐ yields. The density of GNDs was calculated to be 1.6 × 10^{12} m^{-2} in GNT-4, assuming that it is proportional to the plastic strain gradient.

Back stress measurement

A gauging system including a Navitar Zoom 6000 optic lens with a magnification of 2 and a high-resolution charge-coupled device (CCD) camera (2088 × 1024 pixels) (*51*) was applied to measure the strain during loading-unloading tensile tests. The back stress was obtained from unloading curves with an offset plastic strain of 0.01% and a slope equal to the Young's modulus E according to the Dickson method (*38*, *51*, *52*).

Misorientation mapping of BCDs

Misorientation of BCDs in GNT samples was measured by means of an electron procession diffraction technique in TEM using NanoMEGAS (Brussels, Belgium) hardware and ASTAR system. The BCD region was scanned by a nanoscale beam with a step of 3.6 nm, and the electron diffraction pattern at each scanned point was collected using a fast CCD camera, providing the misorientation mapping of BCDs. Details about the electron procession diffraction technique can be found in (*38*, *53*).

REFERENCES AND NOTES

1. K. Lu, L. Lu, S. Suresh, Strengthening materials by engineering coherent internal boundaries at the nanoscale. *Science* **324**, 349–352 (2009). doi: 10.1126/science.1159610; pmid: 19372422
2. Z. Liu, M. A. Meyers, Z. Zhang, R. O. Ritchie, Functional gradients and heterogeneities in biological materials: Design principles, functions, and bioinspired applications. *Prog. Mater. Sci.* **88**, 467–498 (2017). doi: 10.1016/j.pmatsci.2017.04.013
3. S. Suresh, Graded materials for resistance to contact deformation and damage. *Science* **292**, 2447–2451 (2001). doi: 10.1126/science.1059716; pmid: 11431558
4. T. H. Fang, W. L. Li, N. R. Tao, K. Lu, Revealing extraordinary intrinsic tensile plasticity in gradient nano-grained copper. *Science* **331**, 1587–1590 (2011). doi: 10.1126/science.1200177; pmid: 21330487
5. Y. Wei *et al*., Evading the strength-ductility trade-off dilemma in steel through gradient hierarchical nanotwins. *Nat. Commun.* **5**, 3580 (2014). doi: 10.1038/ncomms4580; pmid: 24686581
6. X. Wu, P. Jiang, L. Chen, F. Yuan, Y. T. Zhu, Extraordinary strain hardening by gradient structure. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7197–7201 (2014). doi: 10.1073/pnas.1324069111; pmid: 24799688
7. S. Zhao *et al*., Generating gradient germanium nanostructures by shock-induced amorphization and crystallization. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9791–9796 (2017). doi: 10.1073/pnas.1708853114; pmid: 28847926
8. R. Thevamaran *et al*., Dynamic creation and evolution of gradient nanostructure in single-crystal metallic microcubes. *Science* **354**, 312–316 (2016). doi: 10.1126/science.aag1768; pmid: 27846562
9. M. F. Ashby, The deformation of plastically non-homogeneous materials. *Philos. Mag.* **21**, 399–424 (1970). doi: 10.1080/14786437008238426
10. N. A. Fleck, G. M. Muller, M. F. Ashby, J. W. Hutchinson, Strain gradient plasticity: Theory and experiment. *Acta Metall. Mater.* **42**, 475–487 (1994). doi: 10.1016/0956-7151(94)90502-9
11. H. Gao, Y. Huang, W. D. Nix, J. W. Hutchinson, Mechanism-based strain gradient plasticity I. Theory. *J. Mech. Phys. Solids* **47**, 1239–1263 (1999). doi: 10.1016/S0022-5096(98)00103-3
12. M. E. Gurtin, L. Anand, A theory of strain-gradient plasticity for isotropic, plastically irrotational materials. Part I: Small deformations. *J. Mech. Phys. Solids* **53**, 1624–1649 (2005). doi: 10.1016/j.jmps.2004.12.008
13. N. A. Fleck, M. F. Ashby, J. W. Hutchinson, The role of geometrically necessary dislocations in giving material strengthening. *Scr. Mater.* **48**, 179–183 (2003). doi: 10.1016/S1359-6462(02)00338-X
14. T. Roland, D. Retraint, K. Lu, J. Lu, Fatigue life improvement through surface nanostructuring of stainless steel by means of surface mechanical attrition treatment. *Scr. Mater.* **54**, 1949–1954 (2006). doi: 10.1016/j.scriptamat.2006.01.049
15. J. Li, S. Chen, X. Wu, A. Soh, J. Lu, The main factor influencing the tensile properties of surface nano-crystallized graded materials. *Mater. Sci. Eng. A* **527**, 7040–7044 (2010). doi: 10.1016/j.msea.2010.07.064
16. X. Wu *et al*., Synergetic strengthening by gradient structure. *Mater. Res. Lett.* **2**, 185–191 (2014). doi: 10.1080/21663831.2014.935821
17. L. Lu, Y. Shen, X. Chen, L. Qian, K. Lu, Ultrahigh strength and high electrical conductivity in copper. *Science* **304**, 422–426 (2004). doi: 10.1126/science.1092905; pmid: 15031435
18. Z. You *et al*., Plastic anisotropy and associated deformation mechanisms in nanotwinned metals. *Acta Mater.* **61**, 217–227 (2013). doi: 10.1016/j.actamat.2012.09.052
19. Q. Pan, H. Zhou, Q. Lu, H. Gao, L. Lu, History-independent cyclic response of nanotwinned metals. *Nature* **551**, 214–217 (2017). pmid: 29088707
20. Z. S. You, L. Lu, K. Lu, Tensile behavior of columnar grained Cu with preferentially oriented nanoscale twins. *Acta Mater.* **59**, 6927–6937 (2011). doi: 10.1016/j.actamat.2011.07.044
21. I. J. Beyerlein, X. Zhang, A. Misra, Growth twins and deformation twins in metals. *Annu. Rev. Mater. Res.* **44**, 329–363 (2014). doi: 10.1146/annurev-matsci-070813-113304
22. Z. X. Wu, Y. W. Zhang, D. J. Srolovitz, Dislocation–twin interaction mechanisms for ultrahigh strength and ductility in nanotwinned metals. *Acta Mater.* **57**, 4508–4518 (2009). doi: 10.1016/j.actamat.2009.06.015
23. J. Wang *et al*., Detwinning mechanisms for growth twins in face-centered cubic metals. *Acta Mater.* **58**, 2262–2270 (2010). doi: 10.1016/j.actamat.2009.12.013
24. W. D. Nix, Yielding and strain hardening of thin metal films on substrates. *Scr. Mater.* **39**, 545–554 (1998). doi: 10.1016/S1359-6462(98)00195-X
25. A. Misra, J. P. Hirth, R. G. Hoagland, Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites. *Acta Mater.* **53**, 4817–4824 (2005). doi: 10.1016/j.actamat.2005.06.025
26. Q. Lu, Z. You, X. Huang, N. Hansen, L. Lu, Dependence of dislocation structure on orientation and slip systems in highly oriented nanotwinned Cu. *Acta Mater.* **127**, 85–97 (2017). doi: 10.1016/j.actamat.2017.01.016
27. H. T. Wang, N. R. Tao, K. Lu, Architectured surface layer with a gradient nanotwinned structure in a Fe–Mn austenitic steel. *Scr. Mater.* **68**, 22–27 (2013). doi: 10.1016/j.scriptamat.2012.05.041
28. R. J. Asaro, S. Suresh, Mechanistic models for the activation volume and rate sensitivity in metals with nanocrystalline grains and nano-scale twins. *Acta Mater.* **53**, 3369–3382 (2005). doi: 10.1016/j.actamat.2005.03.047
29. U. F. Kocks, H. Mecking, Physics and phenomenology of strain hardening: The FCC case. *Prog. Mater. Sci.* **48**, 171–273 (2003). doi: 10.1016/S0079-6425(02)00003-8
30. Z. Zeng *et al*., Gradient plasticity in gradient nano-grained metals. *Extreme Mech. Lett.* **8**, 213–219 (2016). doi: 10.1016/j.eml.2015.12.005
31. J. G. Swadener, E. P. George, G. M. Pharr, The correlation of the indentation size effect measured with indenters of various shapes. *J. Mech. Phys. Solids* **50**, 681–694 (2002). doi: 10.1016/S0022-5096(01)00103-X
32. S. S. Chakravarthy, W. A. Curtin, Stress-gradient plasticity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15716–15720 (2011). doi: 10.1073/pnas.1107035108; pmid: 21911403
33. M. Kouzeli, A. Mortensen, Size dependent strengthening in particle reinforced aluminium. *Acta Mater.* **50**, 39–51 (2002). doi: 10.1016/S1359-6454(01)00327-5
34. X. Ma *et al*., Mechanical properties of copper/bronze laminates: Role of interfaces. *Acta Mater.* **116**, 43–52 (2016). doi: 10.1016/j.actamat.2016.06.023
35. X. Wu *et al*., Heterogeneous lamella structure unites ultrafine-grain strength with coarse-grain ductility. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14501–14505 (2015). doi: 10.1073/pnas.1517193112; pmid: 26554017
36. G. A. Malygin, Dislocation mechanism of dynamic polygonization of a crystal caused by its bending. *Phys. Solid State* **44**, 1249–1253 (2002). doi: 10.1134/1.1494626
37. D. A. Hughes, N. Hansen, D. J. Bammann, Geometrically necessary boundaries, incidental dislocation boundaries and geometrically necessary dislocations. *Scr. Mater.* **48**, 147–153 (2003). doi: 10.1016/S1359-6462(02)00358-5
38. D. Viladot *et al*., Orientation and phase mapping in the transmission electron microscope using precession-assisted diffraction spot recognition: State-of-the-art results. *J. Microsc.* **252**, 23–34 (2013). doi: 10.1111/jmi.12065; pmid: 23889078
39. X. Feaugas, On the origin of the tensile flow stress in the stainless steel AISI 316L at 300 K: Back stress and effective stress. *Acta Mater.* **47**, 3617–3632 (1999). doi: 10.1016/S1359-6454(99)00222-0
40. H. Mughrabi, Deformation-induced long-range internal stresses and lattice plane misorientations and the role of geometrically necessary dislocations. *Philos. Mag.* **86**, 4037–4054 (2006). doi: 10.1080/14786430500509054

41. T. Ungár, E. Schafler, P. Hanák, S. Bernstorff, M. Zehetbauer, Vacancy production during plastic deformation in copper determined by in situ X-ray diffraction. *Mater. Sci. Eng. A* **462**, 398–401 (2007). doi: 10.1016/j.msea.2006.03.156
42. T. Ungár, A. D. Stoica, G. Tichy, X. L. Wang, Orientation-dependent evolution of the dislocation density in grain populations with different crystallographic orientations relative to the tensile axis in a polycrystalline aggregate of stainless steel. *Acta Mater.* **66**, 251–261 (2014). doi: 10.1016/j.actamat.2013.11.012
43. K. Lu, Gradient nanostructured materials. *Acta Metall. Sin.* **51**, 1–10 (2015). doi: 10.11900/0412.1961.2014.00395
44. P. Zhang, S. X. Li, Z. F. Zhang, General relationship between strength and hardness. *Mater. Sci. Eng. A* **529**, 62–73 (2011). doi: 10.1016/j.msea.2011.08.061
45. K. Edalati, T. Fujioka, Z. Horita, Microstructure and mechanical properties of pure Cu processed by high-pressure torsion. *Mater. Sci. Eng. A* **497**, 168–173 (2008). doi: 10.1016/j.msea.2008.06.039
46. X. L. Ma *et al.*, Strain hardening and ductility in a coarse-grain/nanostructure laminate material. *Scr. Mater.* **103**, 57–60 (2015). doi: 10.1016/j.scriptamat.2015.03.006
47. W. G. Hoover, Constant-pressure equations of motion. *Phys. Rev. A* **34**, 2499–2500 (1986). doi: 10.1103/PhysRevA.34.2499; pmid: 9897546
48. S. Nosé, A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984). doi: 10.1063/1.447334
49. Y. Mishin, M. J. Mehl, D. A. Papaconstantopoulos, A. F. Voter, J. D. Kress, Structural stability and lattice defects in copper: *Ab initio*, tight-binding, and embedded-atom calculations. *Phys. Rev. B* **63**, 224106 (2001). doi: 10.1103/PhysRevB.63.224106
50. D. Faken, H. Jonsson, Systematic analysis of local atomic structure combined with 3D computer graphics. *Comput. Mater. Sci.* **2**, 279–286 (1994). doi: 10.1016/0927-0256(94)90109-0
51. H. Wang, Z. You, L. Lu, Kinematic and isotropic strain hardening in copper with highly aligned nanoscale twins. *Mater. Res. Lett.* **6**, 333–338 (2018). doi: 10.1080/21663831.2018.1455752
52. J. I. Dickson, J. Boutin, L. Handfield, A comparison of two simple methods for measuring cyclical internal and effective stresses. *Mater. Sci. Eng.* **64**, L7–L11 (1984). doi: 10.1016/0025-5416(84)90083-1
53. W. Xu, X. Liu, K. Lu, Strain-induced microstructure refinement in pure Al below 100 nm in size. *Acta Mater.* **152**, 138–147 (2018). doi: 10.1016/j.actamat.2018.04.014

ACKNOWLEDGMENTS

We thank N. R. Tao, T. Zhu, N. Hansen, and X. X. Huang for inspirational discussions, and S. Jin for samples preparation.

Funding: Supported by NSFC grants 51471172 and U1608257, the Key Research Program of Frontier Science, CAS and the National Basic Research Program of China (973 Program, 2012CB932202) (L.L.); NSFC International collaboration grant 51420105001 (L.L. and H.G.); and NSF grant DMR-1709318 and XSEDE grant MS090046 (H.Z. and H.G.). **Author contributions:** L.L. and H.G. conceived and led the project; Z.C., Q.L., and L.L. performed the experiments; H.Z. performed the simulations; and all authors wrote and approved the submitted manuscript. **Competing interests:** None declared. **Data and materials availability:** The data generated in this study are included in the manuscript and its supplementary materials. Atomistic simulations were performed using LAMMPS, which is freely available at http://lammps.sandia.gov/.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/eaau1925/suppl/DC1
Figs. S1 to S11
Tables S1 and S2
Movies S1 to S3

16 May 2018; accepted 7 September 2018
10.1126/science.aau1925

REPORT

NANOMATERIALS

A general synthesis approach for supported bimetallic nanoparticles via surface inorganometallic chemistry

Kunlun Ding[1]*, David A. Cullen[2], Laibao Zhang[1], Zhi Cao[3,4], Amitava D. Roy[5], Ilia N. Ivanov[6], Dongmei Cao[7]

The synthesis of ultrasmall supported bimetallic nanoparticles (between 1 and 3 nanometers in diameter) with well-defined stoichiometry and intimacy between constituent metals remains a substantial challenge. We synthesized 10 different supported bimetallic nanoparticles via surface inorganometallic chemistry by decomposing and reducing surface-adsorbed heterometallic double complex salts, which are readily obtained upon sequential adsorption of target cations and anions on a silica substrate. For example, adsorption of tetraamminepalladium(II) [$Pd(NH_3)_4^{2+}$] followed by adsorption of tetrachloroplatinate [$PtCl_4^{2-}$] was used to form palladium-platinum (Pd-Pt) nanoparticles. These supported bimetallic nanoparticles show enhanced catalytic performance in acetylene selective hydrogenation, which clearly demonstrates a synergistic effect between constituent metals.

Heterogeneous catalysts that contain bimetallic nanoparticles (NPs) are used in many petrochemical processes, including reforming (*1*), selective hydrogenation (*2*, *3*) and dehydrogenation (*4*), and acetoxylation (*5*). In recent years, bimetallic NPs have been used in biomass conversions (*6*), electrocatalysis (*7*–*10*), and many other catalytic processes (*11*–*13*). The difference in catalytic properties of bimetallic NPs compared with their parent metals originates from their distinct geometric and electronic structures as well as the synergistic effects between the two metals (*6*, *14*–*16*). In particular, synthetic protocols have been demonstrated to be crucial to the structural, electronic, and, hence, catalytic performance of bimetallic NPs. The conventional impregnation method usually results in ill-defined bimetallic NPs with inhomogeneous particle sizes and compositions (*17*, *18*). New strategies for the synthesis of bimetallic NPs include colloidal synthesis (*19*), surface organometallic chemistry (*20*), atomic layer deposition (*21*), coadsorption and coreduction of metal cations (*18*), and carbothermal shock synthesis (*22*). Nevertheless, the synthesis of ultrasmall (<3 nm in diameter) supported bimetallic NPs with well-defined stoichiometry and intimacy between constituent metals remains challenging.

We explored the preparation of supported bimetallic NPs by reducing heterometallic double complex salt (DCS) precursors (e.g., $[Pd(NH_3)_4][PtCl_4]$), which differ from conventional precursors in that the stoichiometry and intimacy of two metals are already established. Although DCSs have received tremendous interest in the study of inorganometallic coordination compounds (*23*, *24*), their poor solubility greatly hinders their applications in many fields (*23*, *25*–*27*). Limited success has been achieved in the synthesis of supported bimetallic catalysts using DCS precursors. Boellaard *et al.* (*28*) described a synthesis of supported Ni-Fe catalysts from DCSs by mixing two precursors in solution in the presence of a support. Owing to the lack of interaction between bulk DCSs and the catalyst support, the deposition of DCSs on the surface of the support is less controllable. Rather than depositing DCSs from solution, we show that heterometallic DCSs can be synthesized directly on a substrate by sequential adsorption of complex metal cations and anions (Fig. 1A). The supported DCSs with precisely paired metal cations and anions through electrostatic interactions can be converted to well-defined supported bimetallic NPs upon reduction. This approach eliminates the issue of competitive adsorption of different metal cations in the coadsorption and coreduction methods (*18*); on the other hand, it also overcomes the solubility issues of DCSs for the bimetallic catalyst synthesis. It can be applied to a large variety of bimetallic NPs, including several bulk immiscible systems. The bimetallic NPs are extremely small (1 to 3 nm in diameter) and have narrow size distributions (±25%), and their bimetallic nature was confirmed by single-NP elemental analysis. These bimetallic catalysts outperform their parent metals in selective hydrogenation of acetylene.

The adsorption of metal cations was carried out on silica (*18*, *29*) (see the supplementary materials). A porous silica support (fig. S1) was first dispersed in water, and the pH of the solution was adjusted to 9 to obtain a negatively charged silica surface. After the adsorption of complex metal cations, e.g., $Pd(NH_3)_4^{2+}$, silica was then separated from the mother solution, washed with copious water, and dried before the adsorption of metal anions. Our initial attempt for the subsequent anion adsorption using an aqueous solution led to the formation of either large aggregates (fig. S2) or monometallic complexes (fig. S3). This failure was likely caused by nonnegligible solubility of DCSs in water, which led to dissolution, ripening, or leaching of the DCSs. Alternatively, we adopted an aprotic solvent dichloromethane and introduced quaternary ammonium cations to assist dissolution (figs. S4 to S6). Complex metal anions in dichloromethane can precisely target the multivalent metal cations preadsorbed on silica to form supported heterometallic DCSs. Further reducing these silica-supported DCSs under a flow of H_2/N_2 (v/v = 1/9) mixture at 400°C afforded the desired silica-supported bimetallic NPs.

Figure 1B shows high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) images of 10 types of supported bimetallic NPs synthesized by this approach. Large-area transmission electron microscopy and HAADF-STEM images and size distributions of these samples are shown in figs. S7 to S16. These NPs are mostly smaller than 3 nm in diameter and uniformly dispersed on the silica support. These NPs are smaller than most bimetallic NPs from colloidal synthesis (*19*) but slightly bigger than those synthesized by the coadsorption and coreduction method (*18*). The polydispersities are similar to the latter. The bimetallic nature of these NPs was confirmed by single-NP energy-dispersive x-ray (EDX) spectroscopy analysis. The composition of several individual NPs was also obtained for each sample, as well as large-area EDX spectroscopy analysis to obtain an average composition of each sample (quantification results are shown in tables S1 to S10). In all cases, the single-NP analysis confirmed that the NPs are bimetallic. Furthermore, the average composition from the single-NP analysis was generally consistent with the average composition provided by large-area EDX spectroscopy analysis, suggesting that the supported DCS precursors were evenly decomposed into bimetallic NPs.

The bulk phase diagrams of the Cu-Ir, Pd-Ir, and Pt-Ir systems indicate that these metal pair are immiscible below 800°C (*30*). Interestingly, these systems still form bimetallic NPs with metal ratios near 1. Aberration-corrected HAADF-STEM images of Cu-Ir/SiO_2 and Pd-Ir/SiO_2 show uneven contrast within single NPs (figs. S17 and S18), suggesting the occurrence of intraparticle phase

[1]Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA 70803, USA. [2]Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [3]State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan, Shanxi 030001, People's Republic of China. [4]Synfuels China Co., Ltd., Beijing 100195, People's Republic of China. [5]Center for Advanced Microstructures and Devices, Louisiana State University, Baton Rouge, LA 70806, USA. [6]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [7]Shared Instrumentation Facility, Louisiana State University, Baton Rouge, LA 70803, USA.
*Corresponding author. Email: kunlunding@lsu.edu

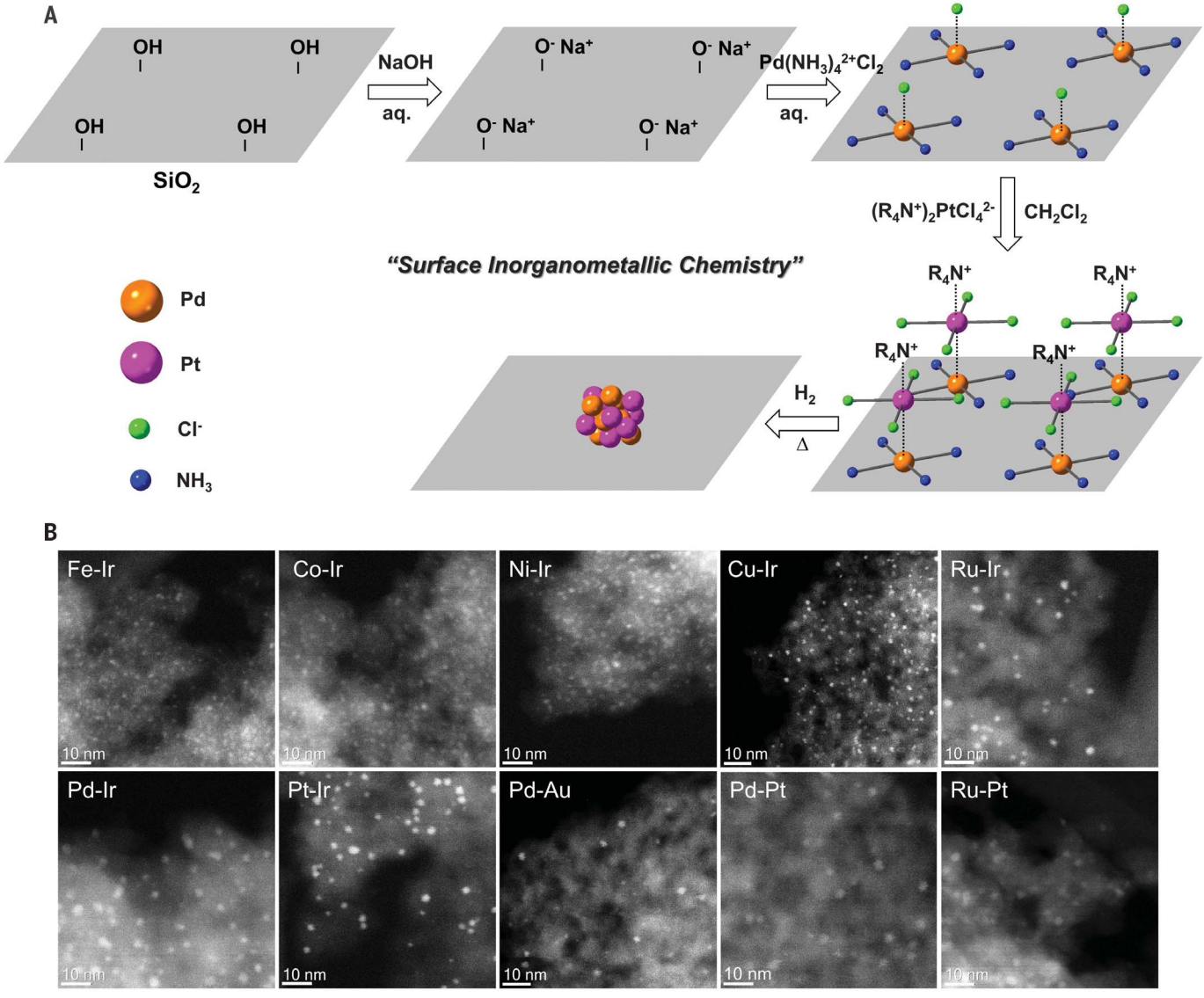


Fig. 1. Schematic illustration and electron microscopy analysis of the supported bimetallic NPs. (**A**) A schematic illustration of the surface inorganometallic chemistry for the synthesis of supported bimetallic NPs. (**B**) HAADF-STEM images of 10 types of supported bimetallic NPs synthesized by this approach. All scale bars are 10 nm.

segregation at a subnanometer level. This type of nanoscopic phase segregation has been reported in several bimetallic systems, including Cu-Ir (*31*), Ru-Pd (*32*), and Pt-Ni (*33*). The lack of macroscopic phase segregation in Cu-Ir, Pd-Ir, and Pt-Ir systems may be partially attributed to the extremely high melting point of Ir metal, which may prevent the nanoscopically phase segregated NPs from coarsening.

To validate our proposed surface inorganometallic chemistry for the bimetallic NPs synthesis, we used x-ray photoelectron spectroscopy (XPS) to study the structural evolution of the surface intermediates during the synthesis. The XPS results are shown in Fig. 2 and figs. S19 to S27. After the adsorption of complex metal cations [$Pd(NH_3)_4^{2+}$, $Pt(NH_3)_4^{2+}$, and $Ru(NH_3)_6^{2+}$], the XPS signals of metals and nitrogen were observed (Fig. 2 and figs. S19 and S23). Weak Cl 2p signals were also observed on these samples (fig. S25), indicating that the divalent complex metal cations require an extra Cl^- ion to balance the excess positive charge. After the adsorption of complex metal anions ($PtCl_4^{2-}$, $IrCl_6^{2-}$, and $AuCl_4^-$), the intensity of the Cl 2p signal increased along with the appearance of Pt, Ir, and Au signals, confirming the successful adsorption of complex metal anions.

In addition to the N 1s signal from the NH_3 ligand (400 eV), a new N 1s signal was observed at 402 eV on the nonreduced $[Pd(NH_3)_4][PtCl_4]/SiO_2$, $[Pd(NH_3)_4][IrCl_6]/SiO_2$, $[Pt(NH_3)_4][IrCl_6]/SiO_2$, and $[Ru(NH_3)_4][IrCl_6]/SiO_2$ samples (Fig. 2 and fig. S23) from the N atom in the quaternary ammonium cation, which was used to assist the dissolution. The adsorption of quaternary ammonium cations was required to balance the excess negative charge of the divalent complex anions. No quaternary ammonium cation was needed when a monovalent anion, $AuCl_4^-$, was used, and in this case, the N 1s signal at 402 eV was absent for the nonreduced $[Pd(NH_3)_4][AuCl_4]/SiO_2$ sample (Fig. 2).

Because the adsorption of complex metal anions was carried out in an aprotic solvent, dichloromethane, ion pairs should mostly stay associated and form dipoles. Therefore, the adsorption of complex metal anions may not be solely driven by electrostatic interactions. Additionally, dipole-dipole and dipole-induced dipole interactions between the anionic complex and the surface may also play important roles. The surface chemistry described here works well for a silica surface but may not be simply extended to other types of support. For instance, alumina has a higher point of zero charge compared with silica (*18*). Thus, complex metal anions are normally

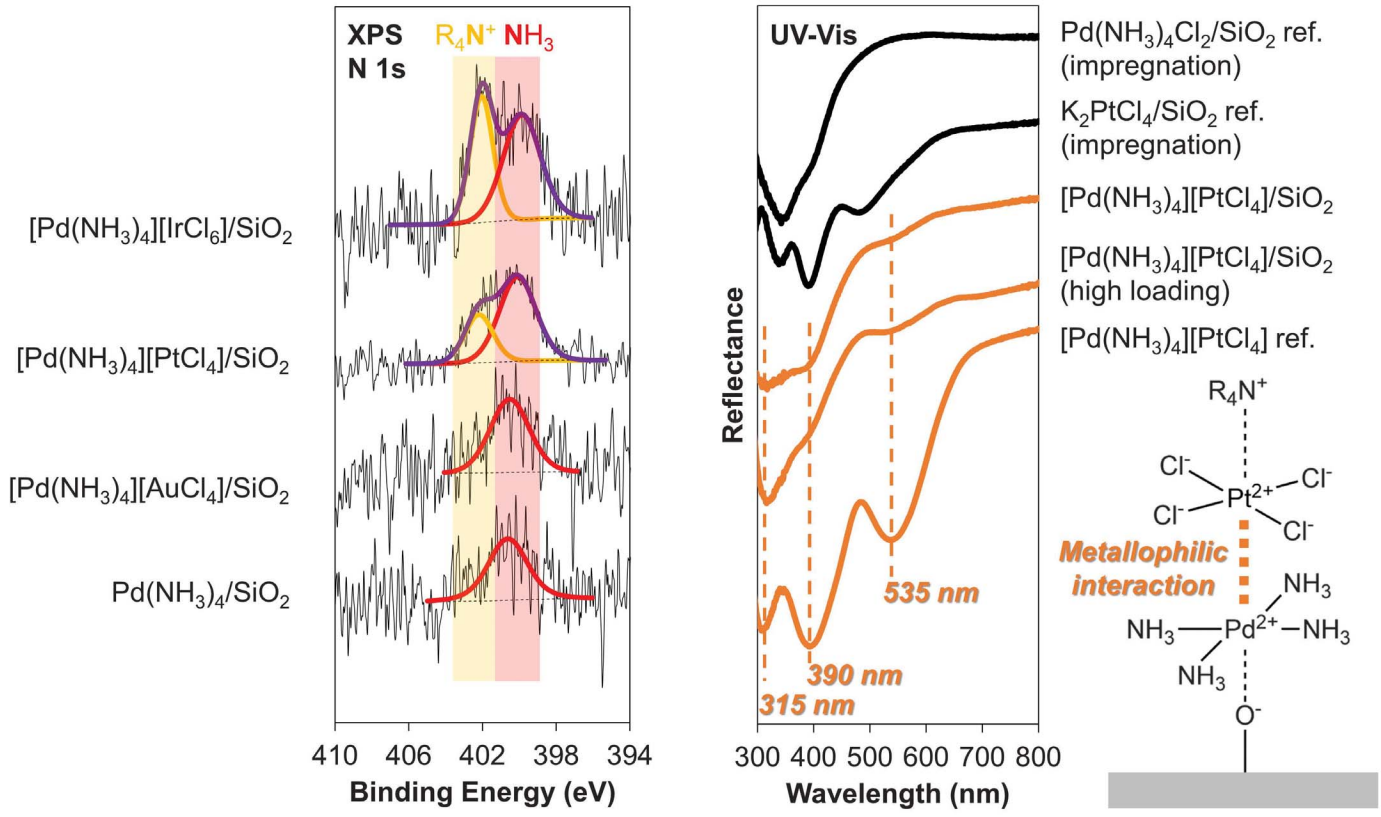


Fig. 2. XPS and UV-vis spectra of supported DCSs and reference compounds. The structure of $[Pd(NH_3)_4][PtCl_4]/SiO_2$ is illustrated on the right. The absorption bands associated with the metallophilic interaction between Pd and Pt metal centers are marked by dashed lines.

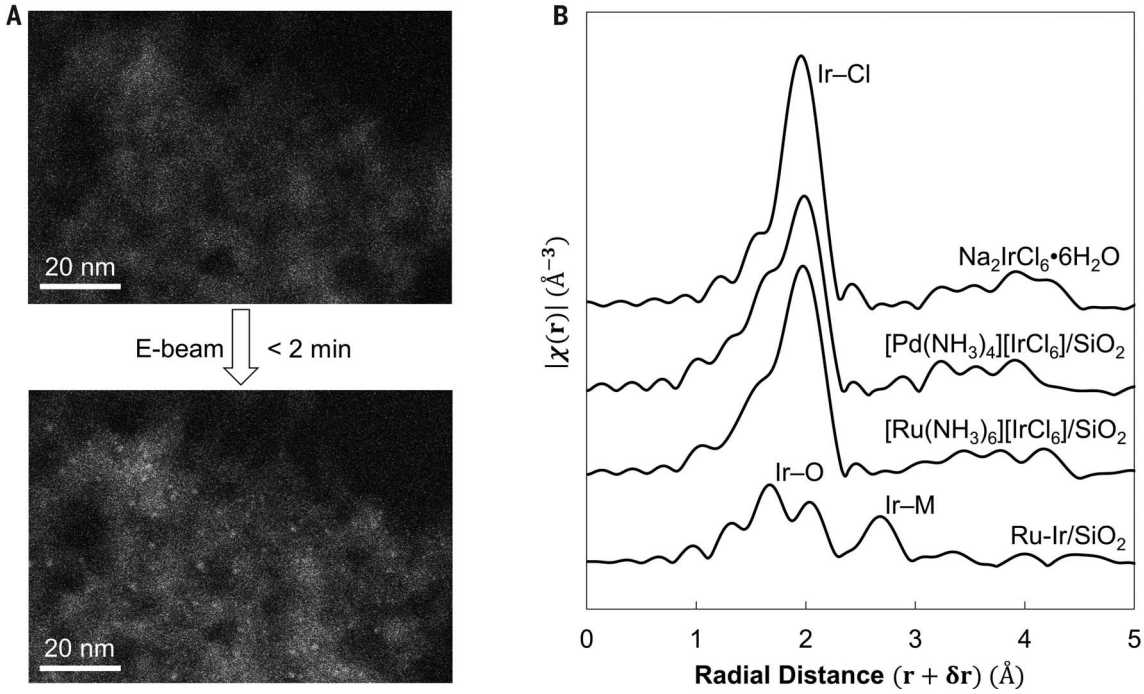


Fig. 3. Structural evolution of supported DCSs upon reduction. (**A**) HAADF-STEM images of nonreduced $[Ru(NH_3)_6][IrCl_6]/SiO_2$ synthesized by sequential adsorption. E-beam, electron beam. (**B**) Fourier transforms of EXAFS spectra of various Ir-containing samples at the Ir L_{III} edge: $Na_2IrCl_6 \cdot 6H_2O$ reference; nonreduced $[Ru(NH_3)_6][IrCl_6]/SiO_2$ and $[Pd(NH_3)_4][IrCl_6]/SiO_2$; and reduced $Ru\text{-}Ir/SiO_2$ synthesized by sequential adsorption. χ(r), EXAFS function.

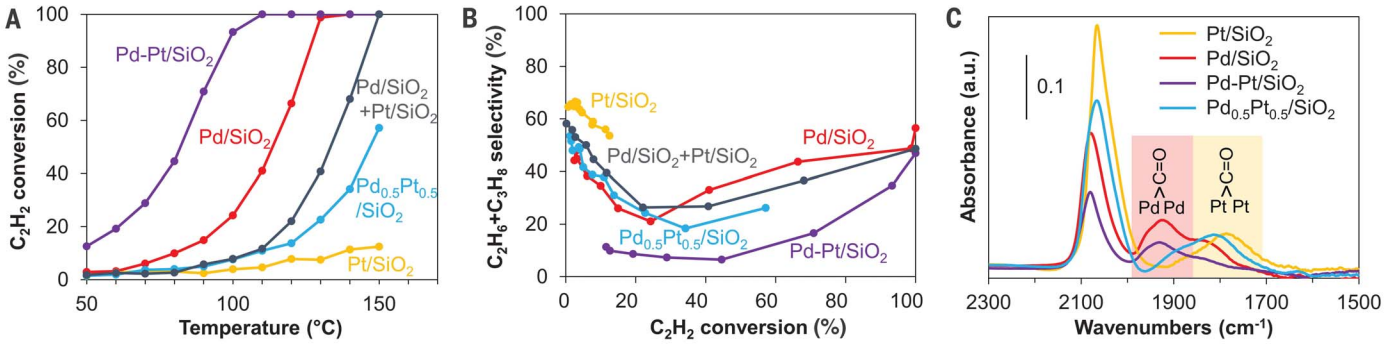


Fig. 4. Catalysis and IR studies of the supported bimetallic NPs. (**A** and **B**) Catalytic performance of supported monometallic and bimetallic NPs in acetylene hydrogenation. (**C**) IR spectra of CO molecule adsorbed on various catalysts. Pd-Pt/SiO_2 and $Pd_{0.5}Pt_{0.5}$/SiO_2 are synthesized by sequential adsorption (this work) and coadsorption (*18*) methods, respectively. All catalysts were reduced at 400°C. Experimental details are provided in the supplementary materials. a.u., arbitrary units.

adopted for the synthesis of alumina-supported noble metal catalysts. Extending our synthesis strategy to alumina support may require substantial efforts toward solvent engineering of metal precursors. Furthermore, chloride binds strongly to the alumina surface and may change the surface properties.

The metallophilic interaction between the adsorbed complex metal cations and anions was confirmed by ultraviolet-visible (UV-vis) diffuse reflectance spectroscopy, indicating the formation of supported DCSs. DCSs usually exhibit distinct colors originating from their metal-metal interactions; many DCSs are named after their colors, e.g., Magnus' green salt $[Pt(NH_3)_4][PtCl_4]$ and Vauquelin's pink salt $[Pd(NH_3)_4][PdCl_4]$ (*23*). Because of the relatively low loadings, the supported complex species obtained by our sequential adsorption approach do not exhibit noticeable colors. Nevertheless, the supported $[Pd(NH_3)_4][PtCl_4]/SiO_2$ synthesized by sequential adsorption shows identical UV-vis absorption bands (315, 390, and 535 nm) with bulk DCS $[Pd(NH_3)_4][PtCl_4]$, which exhibits a pink color (*34*, *35*). By contrast, the UV-vis spectra of the impregnated $Pd(NH_3)_4Cl_2/SiO_2$ and K_2PtCl_4/SiO_2 were different from the spectra of $[Pd(NH_3)_4][PtCl_4]/SiO_2$.

The amorphous nature of supported DCSs was confirmed by electron microscopy studies. No NPs were observed on the supported DCSs initially (Fig. 3A). However, within 2 min of electron beam exposure, many NPs in the 1- to 2-nm size range were observed, indicating the in situ decomposition of DCSs associated with the electron beam damage. Extended x-ray absorption fine structure (EXAFS) measurements of the Ir L_{III} edge were performed to study the structural evolution of coordination sphere of Ir atoms. As shown in Fig. 3B, the EXAFS spectra of nonreduced $[Ru(NH_3)_6][IrCl_6]/SiO_2$ and $[Pd(NH_3)_4][IrCl_6]/SiO_2$ were similar to that of the $Na_2IrCl_6 \cdot 6H_2O$ reference. We associated the major peak between 1.5 and 2.5 Å with the Ir–Cl bond in $IrCl_6^{2-}$. After the reduction of $[Ru(NH_3)_6][IrCl_6]/SiO_2$ at 400°C under a hydrogen atmosphere, a new peak appeared between 2.5 and 3 Å, corresponding to Ir–M (M = Ru or Ir) bonds. XPS analysis of the reduced DCSs/SiO_2 samples confirmed the reduction of metal atoms into metallic states (figs. S20 and S22). Further, the disappearance of N 1s and Cl 2p signals indicated the successful removal of ligands (figs. S24, S26, and S27).

The catalytic performance of the supported bimetallic NPs was evaluated for acetylene hydrogenation under noncompetitive conditions (figs. S28 to S30). The bimetallic catalysts (Pd-Ir/SiO_2, Pd-Pt/SiO_2, Pd-Au/SiO_2, Pt-Ir/SiO_2, and Ru-Pt/SiO_2) generally showed higher catalytic activity and lower alkane selectivity compared to their parent metals. Ru-Ir/SiO_2 was an exception in that it showed the same negligible activity as its parent metals. Among the group VIII metals, Ru and Ir often showed the strongest binding strengths for many types of adsorbates (*36*), which might explain the low activity of Ru/SiO_2, Ir/SiO_2, and Ru-Ir/SiO_2 in acetylene hydrogenation. The enhanced activity and suppressed ethane selectivity of bimetallic Pd-M/SiO_2 and Pt-M/SiO_2 compared to their parent metals, could be partly attributed to the relatively weaker binding strengths of acetylene and ethylene on these bimetallic NPs (*2*, *3*).

We further studied Pd-Pt bimetallic NPs with different compositions in acetylene hydrogenation under competitive conditions (Fig. 4, A and B, and figs. S31 to S33). The compositions of supported Pd-Pt bimetallic NPs were adjusted by coadsorption of complex metal cations followed by adsorption of complex metal anions. For instance, equal molar $Pd(NH_3)_4^{2+}$ and $Pt(NH_3)_4^{2+}$ cations were coadsorbed on SiO_2, followed by adsorption of $PdCl_4^{2-}$ or $PtCl_4^{2-}$ anions and subsequent reduction, yielding $(Pd_{0.5}Pt_{0.5})$-Pd/SiO_2 or $(Pd_{0.5}Pt_{0.5})$-Pt/SiO_2, respectively (fig. S34). Compared to the monometallic Pd/SiO_2 and Pt/SiO_2, the three bimetallic catalysts prepared by sequential adsorption method—$(Pd_{0.5}Pt_{0.5})$-Pd/SiO_2, $(Pd_{0.5}Pt_{0.5})$-Pt/SiO_2, and Pd-Pt/SiO_2—all showed enhanced catalytic activity and suppressed alkane selectivity in acetylene hydrogenation in the presence of high-concentration propylene. Moreover, three bimetallic catalysts were prepared by coadsorption and coreduction of $Pd(NH_3)_4^{2+}$ and $Pt(NH_3)_4^{2+}$ cations (fig. S35) (*18*)—$Pd_{0.25}Pt_{0.75}/SiO_2$, $Pd_{0.5}Pt_{0.5}/SiO_2$, and $Pd_{0.75}Pt_{0.25}/SiO_2$—which showed lower activity than a physical mixture of Pd/SiO_2 and Pt/SiO_2, and their alkane selectivity is greater than those prepared by sequential adsorption method.

Infrared (IR) spectroscopy with CO as a probe molecule was used to study the surface properties of these monometallic and bimetallic catalysts (Fig. 4C). We found that the bridge CO peak was very sensitive to the type of metals (1920 and 1790 cm^{-1} for Pd and Pt surfaces, respectively). Interestingly, the IR spectra of the $Pd_{0.5}Pt_{0.5}/SiO_2$ catalyst prepared by the coadsorption method exhibits a Pt-like feature, whereas the Pd-Pt/SiO_2 catalyst prepared by the sequential adsorption method exhibits a Pd-like feature. This suggests that the surfaces of the bimetallic NPs prepared by coadsorption and sequential adsorption methods are Pt- and Pd-rich, respectively. The superior intrinsic activity of Pd and the electronic modification by Pt underneath might be the origin of the improved catalytic performance of the Pd-Pt/SiO_2 catalyst prepared by the sequential adsorption method.

REFERENCES AND NOTES

1. J. H. Sinfelt, *Acc. Chem. Res.* **20**, 134–139 (1987).
2. G. Kyriakou *et al.*, *Science* **335**, 1209–1212 (2012).
3. F. Studt *et al.*, *Science* **320**, 1320–1322 (2008).
4. J. J. H. B. Sattler, J. Ruiz-Martinez, E. Santillan-Jimenez, B. M. Weckhuysen, *Chem. Rev.* **114**, 10613–10653 (2014).
5. M. Chen, D. Kumar, C. W. Yi, D. W. Goodman, *Science* **310**, 291–293 (2005).
6. D. M. Alonso, S. G. Wettstein, J. A. Dumesic, *Chem. Soc. Rev.* **41**, 8075–8098 (2012).
7. B. Lim *et al.*, *Science* **324**, 1302–1305 (2009).
8. C. Chen *et al.*, *Science* **343**, 1339–1343 (2014).
9. X. Q. Huang *et al.*, *Science* **348**, 1230–1234 (2015).
10. L. Bu *et al.*, *Science* **354**, 1410–1414 (2016).
11. J. K. Edwards *et al.*, *Science* **323**, 1037–1041 (2009).
12. S. J. Freakley *et al.*, *Science* **351**, 965–968 (2016).
13. N. Agarwal *et al.*, *Science* **358**, 223–227 (2017).
14. N. Toshima, T. Yonezawa, *New J. Chem.* **22**, 1179–1201 (1998).
15. R. Ferrando, J. Jellinek, R. L. Johnston, *Chem. Rev.* **108**, 845–910 (2008).
16. M. Sankar *et al.*, *Chem. Soc. Rev.* **41**, 8099–8139 (2012).
17. O. S. Alexeev, B. C. Gates, *Ind. Eng. Chem. Res.* **42**, 1571–1587 (2003).
18. A. Wong, Q. Liu, S. Griffin, A. Nicholls, J. R. Regalbuto, *Science* **358**, 1427–1430 (2017).

19. K. D. Gilroy, A. Ruditskiy, H. C. Peng, D. Qin, Y. Xia, *Chem. Rev.* **116**, 10414–10472 (2016).
20. J. P. Candy, B. Didillon, E. L. Smith, T. B. Shay, J. M. Basset, *J. Mol. Catal.* **86**, 179–204 (1994).
21. J. Lu *et al*., *Nat. Commun.* **5**, 3264 (2014).
22. Y. Yao *et al*., *Science* **359**, 1489–1494 (2018).
23. J. Bremi *et al*., *Chem. Mater.* **11**, 977–994 (1999).
24. L. H. Doerrer, *Dalton Trans.* **39**, 3543–3553 (2010).
25. D. I. Potemkin *et al*., *Chem. Eng. J.* **207-208**, 683–689 (2012).
26. D. I. Potemkin *et al*., *Catal. Today* **235**, 103–111 (2014).
27. S. T. Thompson, H. H. Lamb, *J. Catal.* **350**, 111–121 (2017).
28. E. Boellaard, A. M. van der Kraan, J. W. Geus, *Appl. Catal. A Gen.* **224**, 1–20 (2002).
29. L. Jiao, J. R. Regalbuto, *J. Catal.* **260**, 329–341 (2008).
30. ASM International, *Alloy Phase Diagrams* (ASM Handbook Series, ASM International, 1992), vol. 3.
31. F. Wang *et al*., *Angew. Chem. Int. Ed.* **57**, 4505–4509 (2018).
32. K. Kusada *et al*., *J. Am. Chem. Soc.* **136**, 1864–1871 (2014).
33. Z. Niu *et al*., *Nat. Mater.* **15**, 1188–1194 (2016).
34. J. R. Miller, *J. Chem. Soc.* **0**, 4452–4457 (1961).
35. J. R. Miller, *J. Chem. Soc.* **0**, 713–720 (1965).
36. F. Abild-Pedersen *et al*., *Phys. Rev. Lett.* **99**, 016105 (2007).

ACKNOWLEDGMENTS

We thank J. Spivey for the access to the IR instrument and K. Dooley for the assistance in thermogravimetric and nitrogen sorption analysis. T. Blanchard is acknowledged for inductively coupled plasma analysis. **Funding:** K.D. acknowledges the startup funding from Louisiana State University. Electron microscopy and UV-vis spectroscopy were conducted at the Center for Nanophase Materials Sciences, which is a DOE Office of Science User Facility. XPS was performed at the Shared Instrumentation Facility (SIF) at Louisiana State University. The Ketek seven-element detector array used in this study for x-ray measurements at CAMD was acquired with funds provided by The Louisiana Board of Regents [LEQSF(2016-17)-ENH-TR-07]. **Author contributions:** K.D. conceived the concept, designed the experiments, synthesized and characterized the nanoparticles, performed catalytic tests, and wrote the paper; D.A.C. performed HAADF-STEM and single-NP EDX spectroscopy characterization; L.Z. carried out IR, thermogravimetric analysis, x-ray diffraction, and nitrogen sorption measurements; Z.C. assisted with data analysis and manuscript preparation; A.D.R. carried out EXAFS measurements; I.N.I. carried out UV-vis measurements; and D.C. performed XPS characterization. **Competing interests:** K.D. has filed provisional patent application no. 62/721,690 regarding the synthesis of supported bimetallic nanoparticles via double complex salts. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/560/suppl/DC1
Materials and Methods
Figs. S1 to S35
Tables S1 to S11
References (*37*–*39*)

11 June 2018; accepted 6 September 2018
10.1126/science.aau4414

ORGANIC CHEMISTRY

Catalytic palladium-oxyallyl cycloaddition

Barry M. Trost*, Zhongxing Huang, Ganesh M. Murhade

Exploration of intermediates that enable chemoselective cycloaddition reactions and expeditious construction of fused- or bridged-ring systems is a continuous challenge for organic synthesis. As an intermediate of interest, the oxyallyl cation has been harnessed to synthesize architectures containing seven-membered rings via (4+3) cycloaddition. However, its potential to access five-membered skeletons is underdeveloped, largely due to the thermally forbidden (3+2) pathway. Here, the combination of a tailored precursor and a Pd(0) catalyst generates a Pd-oxyallyl intermediate that cyclizes with conjugated dienes to produce a diverse array of tetrahydrofuran skeletons. The cycloaddition overrides conventional (4+3) selectivity by proceeding through a stepwise pathway involving a Pd-allyl transfer and ring closure sequence. Subsequent treatment of the (3+2) adducts with a palladium catalyst converts the heterocycles to the carbocyclic cyclopentanones.

Cycloaddition reactions are of fundamental importance to organic synthesis, as they provide access to cyclic motifs in an efficient and convergent manner (*1*). Particularly, the capability of cycloaddition reactions to rapidly access fused- and bridged-ring systems has substantially streamlined the synthesis of complex natural products, pharmaceuticals, and agrochemicals. The development of cycloaddition reactions hinges heavily on the design and use of new precursors that can be activated to produce transient and highly reactive intermediates for the subsequent cyclization with acceptors. Among these reactive species, an intermediate of enduring interest is the oxyallyl cation, which consists of a positively charged allyl motif and a negatively charged oxygen atom or, in an alternative resonance structure, a carbocation attached to an enolate (Fig. 1A) (*2*). These diverse structural features have made oxyallyl cations potentially versatile intermediates for cycloaddition, as bond formation can take place at either the two terminal carbons, to afford carbocycles, or one of the carbons and the oxygen site, to produce cyclic ethers.

Conventionally, oxyallyl cations are generated from ketones or enol ethers with α-halo or sulfonyl substituents by using stoichiometric amounts of bases, acids, or reductants. Oxyallyl cations from these precursors were shown to readily react with conjugated dienes to yield seven-membered rings (*3*, *4*). In comparison, cycloadditions between oxyallyl cations and 2π acceptors to afford cyclopentanone or tetrahydrofuran motifs are underdeveloped despite these five-membered rings being ubiquitous in bioactive molecules, functional materials, and their synthetic precursors. The underdevelopment of (3+2) cycloaddition with the oxyallyl cation is largely attributed to the forbidden concerted pathway due to unmatched frontier molecular orbitals, and only a limited number of reactions that operate through stepwise mechanisms have been developed to date (*5*). For example, Noyori and co-workers have pioneered the (3+2) cycloaddition between styrene- or enamine-type olefins with α,α′-dihaloketones to cyclopentanone derivatives by using low-valent iron carbonyl as the reductant (*6*–*8*). A similar cyclopentanone synthesis has also been achieved by the Kuwajima group, with electron-rich olefins, where the oxyallyl cation was generated with aluminum-based Lewis acids from α-acetoxy silyl enol ether (*9*). In both cases, stoichiometric amounts of metal-based reagents were required to generate oxyallyl intermediates.

For the past several decades, palladium catalysts have been found efficient in promoting a wide range of cycloaddition reactions in a chemo- and stereoselective fashion (*10*, *11*). The development and wide application of palladium-stabilized zwitterions (*12*), especially palladium-trimethylenemethane (TMM) in cycloaddition reactions (*13*, *14*), prompted us to consider a Pd-oxyallyl intermediate for the catalytic synthesis of important cyclopentanone and tetrahydrofuran structures. Thus, we devised a bifunctional precursor **1** for the oxyallyl generation that incorporates both an allyl carbonate and a silyl enol ether moiety (Fig. 1B). We hypothesized that once exposed to Pd(0), the allyl carbonate would first be ionized to produce Pd-allyl **2**, and the alkoxide released (together with CO_2) in that process would subsequently remove the silyl group of **2** and generate the proposed Pd-oxyallyl ***3***. However, our and other laboratories' early attempts to access five-membered rings with various 2π acceptors were unsuccessful (Fig. 1C). Instead, Pd-oxyallyl intermediates generated from different precursors were all found to react only with norbornene-type strained alkenes to yield unexpected cyclopropanation products **4** (*15*–*18*). It is proposed that after the migratory insertion of Pd-oxyallyl to norbornene, the reductive elimination of the adducts **5** or **6** toward (3+2) products is intercepted by a proton shift to produce four-membered palladacycle **7**, which upon the C–C reductive elimination, affords the cyclopropane product.

A potential approach to switch the chemoselectivity to the desired (3+2) cycloadditions would be the inhibition of the proton transfer process that leads to the cyclopropanation (Fig. 2A). We envisioned that an additional electron-withdrawing group on the oxyallyl intermediate would substantially enhance the acidity of the α proton in **10**, which, in turn, would drive the equilibrium toward the six-membered palladacycles **8** and **9** that lead to the cyclopentanone or tetrahydrofuran products. Thus, a tailored oxyallyl precursor **11** with an ester motif as the electron-withdrawing group was readily synthesized from the corresponding β-keto ester and subsequently examined with a wide range of unsaturated acceptors using palladium catalysts. We found that the desired (3+2) cycloaddition could indeed take place between the oxyallyl precursor and 1,3-cyclohexadiene to produce the bicyclic tetrahydrofuran product **12a** with CpPd(cinnamyl)/tris(*o*-methoxyphenyl) phosphine as the catalyst (Fig. 2B). The cycloaddition turned out to be highly chemoselective, as neither the cyclopropanation product nor the cyclopentanone product from the (3+2) cycloaddition (**22e**, vide infra) was generated. This reaction stands in stark contrast to the conventional oxyallyl cation cycloaddition, in which reactions with conjugated dienes often proceed through (4+3) pathways to produce seven-membered rings (*19*). We propose that the (3+2) cycloaddition described here proceeds through a migratory insertion between the conjugated diene and the Pd-oxyallyl intermediate to generate a zwitterionic intermediate **14** with a newly formed palladium π-allyl motif (*20*, *21*), followed by the reductive elimination to form the five-membered ether (Fig. 2C, bottom). The formal Pd-allyl transfer process is sharply distinct from cycloaddition reactions of analogous Pd-TMM intermediates (Fig. 2C, top), in which a general cyclization pathway involves the addition of Pd-TMM to electron-deficient olefins, generating a new nucleophilic carbanion **13**.

Control reactions using different oxyallyl precursors indicated that the electron-withdrawing substituent was indeed essential for the cycloaddition, as precursors with a phenyl group or without any substitution both failed to yield the (3+2) product (Fig. 2B, entries 1 and 2). The combination of TBS (*tert*-butyl-dimethylsilyl) enol ether and carbonate leaving group was also shown to be important. Replacement of the carbonate leaving group with acetate largely decreased the yield, probably owing to the slower desilylation step with less-basic acetate anion (entry 3). On the other hand, oxyallyl precursors with TIPS

Department of Chemistry, Stanford University, Stanford, CA 94305-5080, USA.
*Corresponding author. Email: bmtrost@stanford.edu

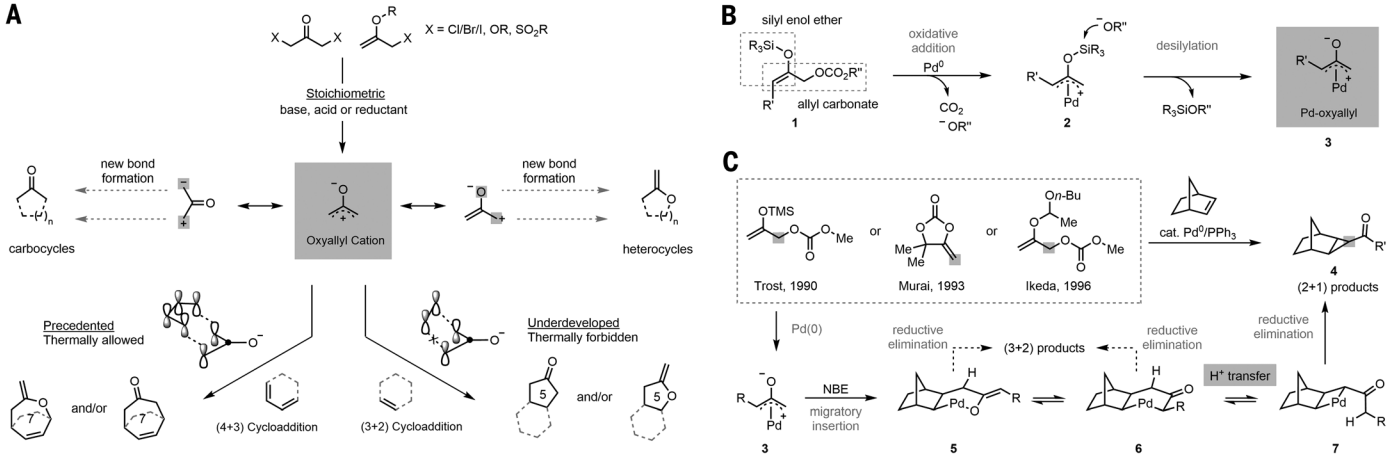


Fig. 1. Oxyallyl cycloaddition. (**A**) Oxyallyl cation cycloaddition. R, aryl or silyl groups. (**B**) Bifunctional precursor for Pd-oxyallyl intermediate. (**C**) Unexpected (2+1) cyclopropanation. TMS, trimethylsilyl; NBE, norbornene; Me, methyl; Bu, butyl.

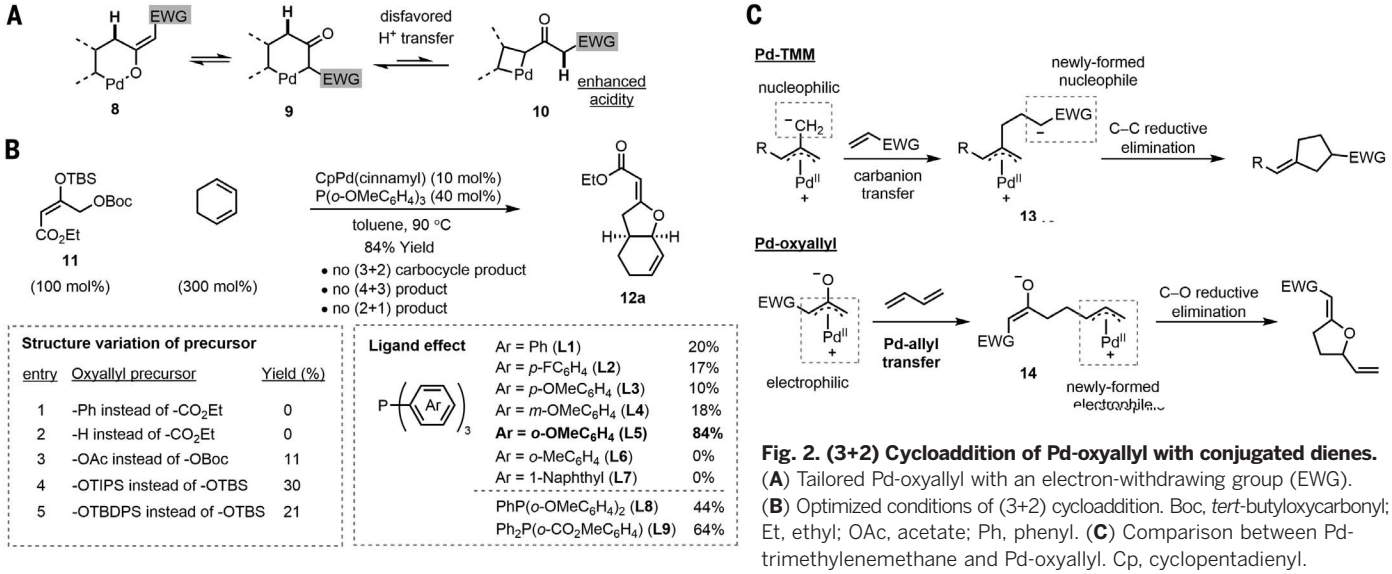


Fig. 2. (3+2) Cycloaddition of Pd-oxyallyl with conjugated dienes. (**A**) Tailored Pd-oxyallyl with an electron-withdrawing group (EWG). (**B**) Optimized conditions of (3+2) cycloaddition. Boc, *tert*-butyloxycarbonyl; Et, ethyl; OAc, acetate; Ph, phenyl. (**C**) Comparison between Pd-trimethylenemethane and Pd-oxyallyl. Cp, cyclopentadienyl.

(triisopropylsilyl) or TBDPS (*tert*-butyl-diphenylsilyl) groups both afforded the desired cycloaddition product, albeit in lower yields compared with the TBS group (entries 4 and 5).

Ligand screening demonstrated that the (3+2) cycloaddition reactions were promoted by monodentate triarylphosphine ligands (Fig. 2B, right column). Whereas simple triphenylphosphine (**L1**), as well as its meta- and para-substituted derivatives with different electronic properties (**L2** to **L4**), afforded the product in low yields, an *ortho*-methoxy substituent (**L5**) was found to substantially boost the efficiency of the cycloaddition reaction. Similarly improved yields were also observed with other analogous ligands with *ortho*-methoxy (**L8**) and *ortho*-carboxylic ester groups (**L9**). Nevertheless, ligands with noncoordinating ortho-substituents (**L6** and **L7**) were completely ineffective for the cycloaddition, excluding the increased steric hindrance of **L5** as a contributing factor for the improved reactivity. In addition, bidentate phosphine ligands were also incompatible with the reaction (fig. S1). On the basis of these results, we propose that the *ortho*-methoxy and ester groups of the ligand stabilize the palladium intermediates and facilitate the cycloaddition by providing secondary coordination, though their Lewis basicity is not strong enough to block the reaction site by chelation (*22*).

The cycloaddition reaction was readily scalable, and only a slight decrease of yield was observed when the catalyst loading was halved (Fig. 3A, **12a**). Different types of esters, including benzyl (**12b**) and *tert*-butyl esters (**12c**), could all facilitate the (3+2) cycloaddition reaction as electron-withdrawing groups. The cycloaddition reaction was also found to be compatible with an oxyallyl precursor bearing a Weinreb amide motif (**12d**), which can be further transformed to various ketones with corresponding metal-based nucleophiles.

When cyclohexadienes with a terminal substituent were used as acceptors, the cycloaddition reactions occurred chemo- and regioselectively at the less-hindered olefin with C–C bond formation at the terminal carbon (Fig. 3B). The bicyclic tetrahydrofuran product from 1-methylcyclohexadiene (**12e**) shares the skeleton with a number of monoterpene lactones and thus would be a potential synthetic intermediate to access these bioactive natural products (*23*). Cyclohexadienes with a wide range of other functional groups at the terminal position, such as phenyl (**12f**), acetate

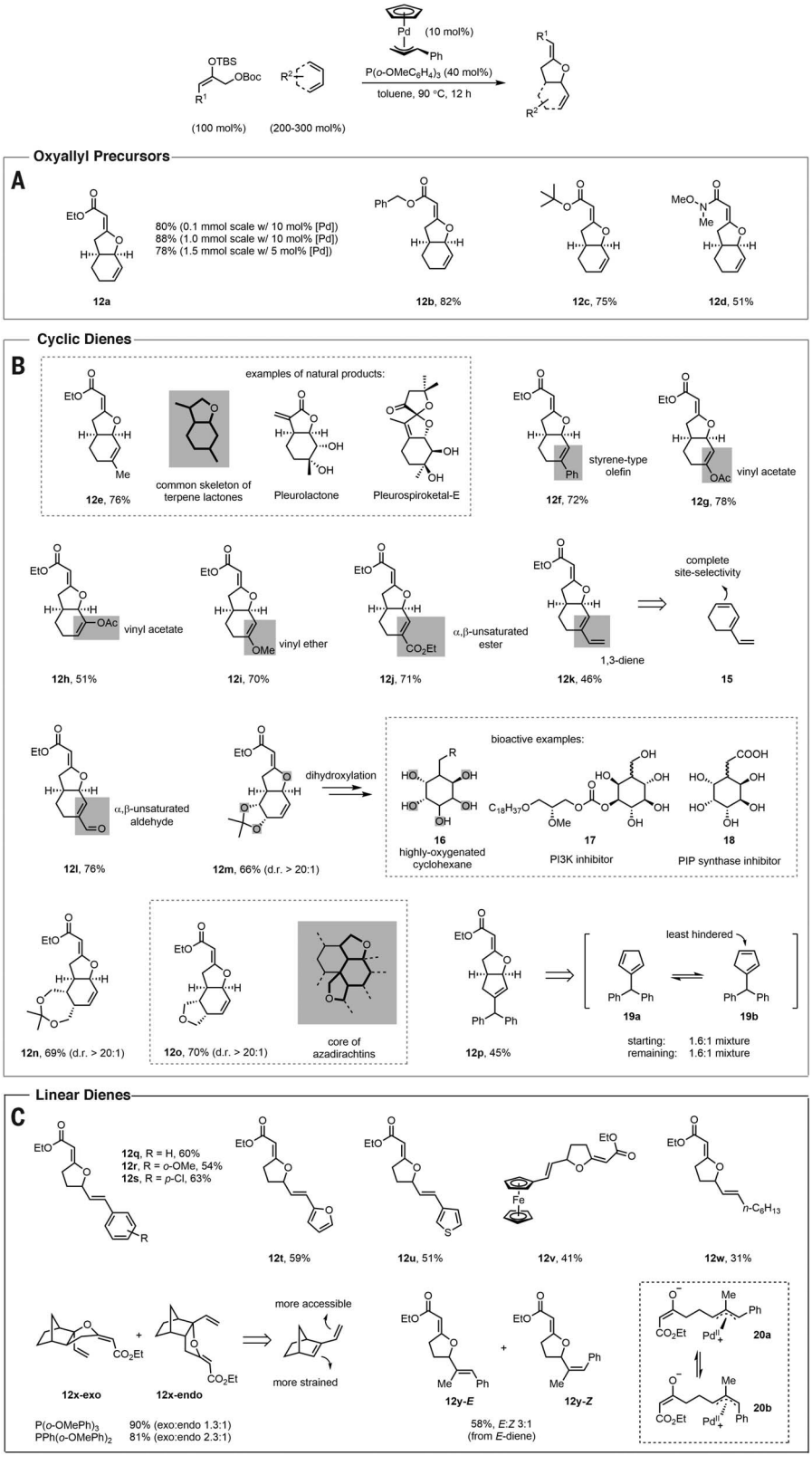


Fig. 3. Scope of the (3+2) cycloaddition of Pd-oxyallyl with conjugated dienes. (**A**) Cycloaddition with different oxyallyl precursors. (**B**) Substrate scope of cyclic dienes. PI3K, phosphatidylinositol 3-kinase; PIP, phosphatidylinositol phosphate; d.r., diastereomeric ratio. (**C**) Substrate scope of linear dienes. Unless otherwise noted, cycloaddition reactions were run with 10 mol % CpPd(cinnamyl), 40 mol % tris(*o*-methoxyphenyl)phosphine, 0.1 mol of oxyallyl precursors, and 200 to 300 mol % dienes at 90°C for 12 hours. The data are reported as isolated yields.

(**12g**), methoxy (**12i**), and ester (**12j**), were also suitable substrates and afforded cyclic olefins with diverse structural and electronic properties that can serve as versatile handles for further modification of the bicyclic ring systems. The extraordinary chemoselectivity was also manifested in the case of the substrate bearing an aldehyde, which produced adduct **12l** in a good yield. Cyclohexadiene with a C-2 acetate substituent also participated in the cycloaddition smoothly to afford the corresponding bicycle (**12h**), albeit in a lower yield compared with its C-1–substituted counterpart. 1-Vinylcyclohexadiene **15**, a conjugated triene substrate with mono-, di-, and trisubstituted olefins, cyclized with oxyallyl selectively at the cyclic and less-substituted olefin instead of the least sterically accessible site (**12k**), probably driven by the more-efficient orbital overlap of the cyclic diene. Additionally, we found that the (3+2) cycloaddition reaction was able to build a variety of complex fused-ring structures rapidly by using easily available cyclohexadienes with multiple functional groups. For example, the 5-6-5 fused-ring system (**12m**) could be readily assembled with high diastereoselectivity through the cycloaddition with the bicyclic diene that originates from the microbial dearomative dihydroxylation of benzene (*24*). The trans relationship between the acetal ring and the tetrahydrofuran ring supports an inner-sphere cycloaddition pathway in which the new bonds form at the same and, in this case, the less-hindered face of the diene coordinating the palladium. The remaining olefin in **12m** could potentially be further dihydroxylated to produce pentaoxygenated derivatives **16**, which would be an advanced synthetic intermediate toward a number of bioactive compounds, especially enzyme inhibitors related to phosphatidylinositols (e.g., **17** and **18**) (*25*). The (3+2) cycloaddition also enabled the synthesis of a cyclohexene derivative with two fused tetrahydrofuran rings (**12o**), a motif that resembles the core structure of azadirachtins and related tetranortriterpenoids (*26*). In addition to cyclohexadienes, cyclopentadiene derivatives can also participate in the (3+2) cycloaddition reaction. When a dynamic mixture of cyclopentadienes **19a** and **19b** reduced from 6,6-diphenylfulvene was employed, the cis-fused 5-5 ring **12p** was obtained as the only product, indicating a preference for the less-hindered olefin.

Linear dienes could also undergo the (3+2) cycloaddition reaction, producing 2-alkylidene-5-vinyl-tetrahydrofuran products (Fig. 3C). In particular, terminally substituted aromatic dienes with different substituents on the arene all afforded the corresponding cycloaddition products in good yields (**12q** to **12s**). Dienes with heterocycles, including furan and thiophene, as well as a ferrocene moiety, were also compatible with the (3+2) cycloaddition (**12t** to **12v**). Nevertheless, alkyl-substituted linear dienes showed much lower reactivity compared with their cyclic analogs, probably due to the lack of the confined *s*-cis conformation that lowers the entropy of activation for the transition state of

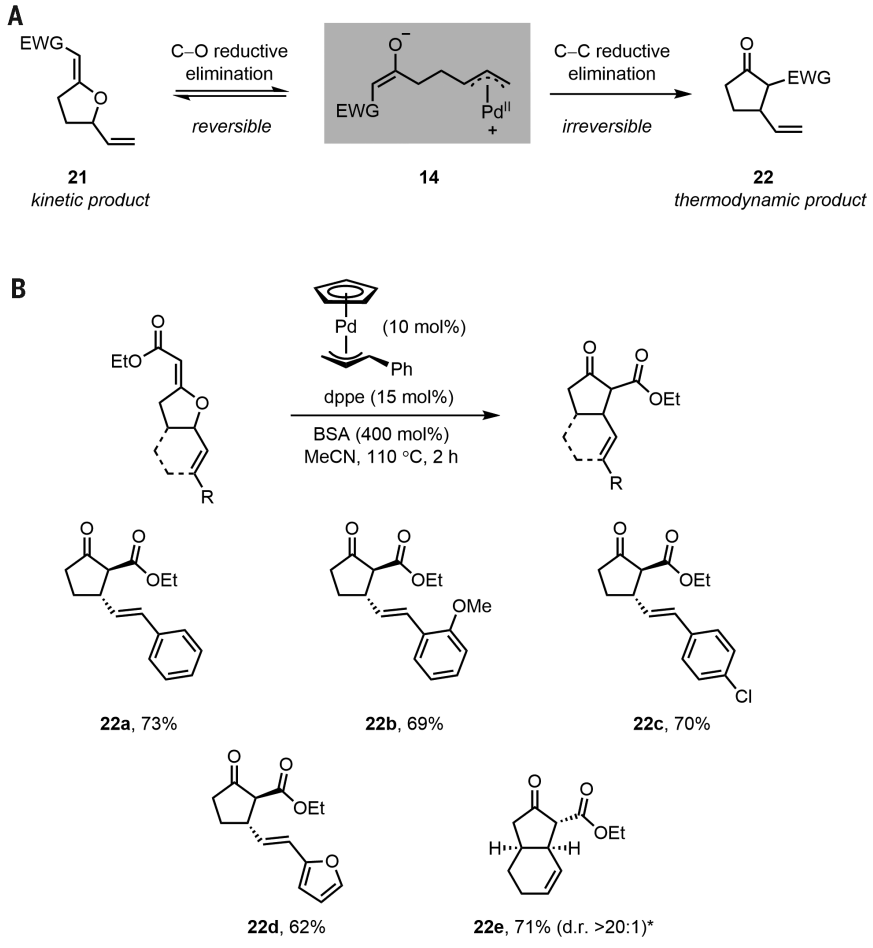


Fig. 4. Isomerization of tetrahydrofuran to cyclopentanone products. (**A**) Pathways to different (3+2) cycloaddition products. (**B**) Scope of [1,3] O-to-C rearrangement. Unless otherwise noted, the isomerization reactions were run using 10 mol % CpPd(cinnamyl), 15 mol % dppe, 400 mol % *N*,*O*-bis(trimethylsilyl)acetamide (BSA), and 0.1 mol of tetrahydrofuran substrate in acetonitrile at 110°C. The data are reported as isolated yields. The asterisk indicates a reaction performed at 80°C in the absence of BSA and with 20 mol % of RuPhos instead of dppe.

π-allyl palladium generation. On the other hand, the absence of extra stabilization from adjacent arenes for the linear alkyl dienes also leads to their underperformance compared with dienes with aromatic substituents. When 1-vinylnorbornene was submitted to the reaction, (3+2) cycloaddition occurred site-selectively at the more-strained cyclic olefin instead of the terminal one (**12x**), and cyclopropane products from (2+1) cycloaddition were not obtained. Both endo and exo cycloaddition products were formed. The use of a less-reactive bis(*o*-methoxyphenyl)phenylphosphine ligand resulted in a lower yield but a higher selectivity toward the exo isomer. Reaction of 1-phenyl-2-methylbutadiene with exclusive E-geometry yielded a mixture of E and Z isomers as the cycloaddition product (**12y**). This observation is consistent with the formation of a new π-allyl palladium (i.e., **20a** and **20b**) after the addition of Pd-oxyallyl to the diene substrate, as the bond rotation during the π-σ-π isomerization of Pd-allyl would alter the olefin geometry and form the Z-product.

Although the current cycloaddition reactions produce tetrahydrofuran rings exclusively, it is equally important to devise an expeditious synthesis of cyclopentanone derivatives using the (3+2) cycloaddition of Pd-oxyallyl intermediate. We envisioned that the product-forming C–O bond reductive elimination should be a reversible process (Fig. 4A), as the β-keto ester anion is weakly basic, well-stabilized, and thus a good leaving group for the palladium π-allyl formation. Therefore, we anticipated that an isomerization reaction from the tetrahydrofuran **21** to more thermodynamically stable cyclopentanone **22** via intermediate **14** should be feasible by using palladium catalysts that can facilitate the irreversible C–C bond reductive elimination (*27*–*29*). Indeed, the desired [1,3] O-to-C rearrangement of cycloaddition products from linear dienes was found to proceed and afford α,β-disubstituted cyclopentanones (**22a** to **22d**) using CpPd(cinnamyl)/1,2-bis(diphenylphosphino)ethane (dppe) as the catalyst under elevated temperature. The bicyclic tetrahydrofuran product **12a** was also suitable for the isomerization, in which 2-dicyclohexylphosphino-2′,6′-diisopropoxybiphenyl (RuPhos), a prominent ligand for promoting palladium-mediated oxidative addition and reductive elimination (*30*), was employed and fused cyclopentanone **22e** was obtained in good yield and high diastereoselectivity.

Together, the synthetic methods described here toward these two important classes of five-membered cyclic motifs are expected to serve as useful tools for synthesis of complex molecules, particularly those with fused-ring systems. The presence of the ketoester motif especially provides a powerful intermediate for extensive structural elaboration.

REFERENCES AND NOTES

1. S. Kobayashi, K. A. Jørgensen, Eds., *Cycloaddition Reactions in Organic Synthesis* (Wiley, 2001).
2. J. Mann, *Tetrahedron* **42**, 4611–4659 (1986).
3. H. M. R. Hoffmann, *Angew. Chem. Int. Ed. Engl.* **23**, 1–19 (1984).
4. R. Noyori, Y. Hayakawa, *Org. React.* **29**, 163–344 (1983).
5. H. Li, J. Wu, *Synthesis* **47**, 22–33 (2015).
6. R. Noyori *et al.*, *J. Am. Chem. Soc.* **94**, 7202–7204 (1972).
7. Y. Hayakawa, K. Yokoyama, R. Noyori, *J. Am. Chem. Soc.* **100**, 1791–1799 (1978).
8. R. Noyori, *Acc. Chem. Res.* **12**, 61–66 (1979).
9. K. Masuya, K. Domon, K. Tanino, I. Kuwajima, *J. Am. Chem. Soc.* **120**, 1724–1731 (1998).
10. M. Lautens, W. Klute, W. Tam, *Chem. Rev.* **96**, 49–92 (1996).
11. H.-W. Frühauf, *Chem. Rev.* **97**, 523–596 (1997).
12. B. D. W. Allen, C. P. Lakeland, J. P. A. Harrity, *Chemistry* **23**, 13830–13857 (2017).
13. B. M. Trost, *Angew. Chem. Int. Ed. Engl.* **25**, 1–20 (1986).
14. E. Nakamura, *Org. React.* **61**, 1–217 (2003).
15. B. M. Trost, S. Schneider, *J. Am. Chem. Soc.* **111**, 4430–4433 (1989).
16. B. M. Trost, H. Urabe, *Tetrahedron Lett.* **31**, 615–618 (1990).
17. K. Ohe *et al.*, *J. Org. Chem.* **58**, 1173–1177 (1993).
18. I. Ikeda, A. Ohsuka, K. Tani, T. Hirao, H. Kurosawa, *J. Org. Chem.* **61**, 4971–4974 (1996).
19. B. Lo, P. Chiu, *Org. Lett.* **13**, 864–867 (2011).
20. R. Hughes, J. Powell, *J. Am. Chem. Soc.* **94**, 7723–7732 (1972).
21. Y. Takahashi, S. Sakai, Y. Ishii, *J. Organomet. Chem.* **16**, 177–183 (1969).
22. B. P. Fors, D. A. Watson, M. R. Biscoe, S. L. Buchwald, *J. Am. Chem. Soc.* **130**, 13552–13554 (2008).
23. S. Wang *et al.*, *J. Agric. Food Chem.* **61**, 5122–5129 (2013).
24. R. A. Johnson, *Org. React.* **63**, 117–264 (2004).
25. H. Morii *et al.*, *J. Biochem.* **153**, 257–266 (2013).
26. H. B. Broughton, S. V. Ley, A. M. Z. Slawin, D. J. Williams, E. D. Morgan, *J. Chem. Soc. Chem. Commun.* **1986**, 46–47 (1986).
27. C. G. Nasveschuk, T. Rovis, *Org. Biomol. Chem.* **6**, 240–254 (2008).
28. B. M. Trost, T. A. Runge, *J. Am. Chem. Soc.* **103**, 7550–7559 (1981).
29. J. C. R. Brioche, T. A. Barker, D. J. Whatrup, M. D. Barker, J. P. A. Harrity, *Org. Lett.* **12**, 4832–4835 (2010).
30. D. S. Surry, S. L. Buchwald, *Chem. Sci.* **2**, 27–50 (2011).

ACKNOWLEDGMENTS

We thank C. Kalnmals for proofreading the manuscript and providing several synthetic intermediates. **Funding:** We thank the NSF (CHE-1360634) and the NIH (GM033049) for support of our programs. We are also grateful to the Tamaki Foundation for financial support.

G.M.M. acknowledges the SERB Indo-U.S. Postdoctoral Fellowship for additional financial support. **Author contributions:** B.M.T. conceived of and designed the project, Z.H. and G.M.M. carried out the experiments, B.M.T. and Z.H. wrote the manuscript, and all authors analyzed the data and discussed the results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data, including experimental details, spectroscopic data, and additional information about the optimization of reaction conditions, are in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/564/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
NMR Spectra
References (*31*–*49*)

14 June 2018; accepted 5 September 2018
10.1126/science.aau4821

QUANTUM OPTICS

Topological protection of biphoton states

Andrea Blanco-Redondo[1]*, Bryn Bell[1]†, Dikla Oren[2], Benjamin J. Eggleton[1], Mordechai Segev[2]

The robust generation and propagation of multiphoton quantum states are crucial for applications in quantum information, computing, and communications. Although photons are intrinsically well isolated from the thermal environment, scaling to large quantum optical devices is still limited by scattering loss and other errors arising from random fabrication imperfections. The recent discoveries regarding topological phases have introduced avenues to construct quantum systems that are protected against scattering and imperfections. We experimentally demonstrate topological protection of biphoton states, the building block for quantum information systems. We provide clear evidence of the robustness of the spatial features and the propagation constant of biphoton states generated within a nanophotonics lattice with nontrivial topology and propose a concrete path to build robust entangled states for quantum gates.

Discoveries about topological states of matter have emerged over the past decade, affecting fields beyond condensed matter, such as electromagnetism (*1*) and photonics (*2*–*4*), acoustics (*5*), cold atoms (*6*, *7*), and mechanics (*8*). The reasons for the extensive research interest are twofold: The underlying concepts are fundamental and universal to many wave systems in nature, and the topological features of these systems offer the possibility of topological protection of transport for classical and quantum waves.

In photonics specifically, even though multiphoton quantum states have played a major role in testing quantum mechanics, quantum computing, and communications, the scalability to large quantum optical devices is limited by loss and fabrication imperfections (*9*). With the major progress in topological photonics, experiments examining the transport of quantum edge states with the use of single photons have recently been carried out (*10*–*12*) and are of clear interest for quantum simulation and sensing. These single-photon experiments investigated the physics of topologically protected bound states (*10*) and topological transitions (*11*) in photonic quantum walks, as well as demonstrating an interface between a quantum emitter and a photonic topological edge state (*12*). However, quantum information systems rely on multiphoton states, thus highlighting the importance of demonstrating topological robustness of multiphoton states. It has been suggested theoretically that topological features can provide robustness to the transport of biphoton states (*13*–*15*), and recent experiments have shown that topology can lead to robustness in the spectral correlations between photon pairs (*16*).

Our experimental system is a one-dimensional (1D) array of silicon nanowires (Fig. 1A) comprising alternating short and long gaps that modulate the coupling strength between adjacent nanowires, creating a Su-Schrieffer-Heeger (SSH) lattice (*17*, *18*). A long-long defect in the middle of the lattice acts as a topological interface between two mirror-reflection versions of the SSH lattice, yielding a topological edge state at the interface between both versions (*19*, *20*). Given the topological nature of the SSH lattice, two key features of this mode are protected against any disorder that preserves the chiral symmetry of the system: the propagation constant (analogous to the energy in condensed matter systems) and the zero amplitude of the wave function in every other site.

Intense picosecond pump pulses from a mode-locked laser (MLL) at 1550 nm are focused into the waveguide at the center of the lattice (Fig. 1A). As the pulses propagate in the lattice, they efficiently generate correlated signal and idler photons over a broad bandwidth via spontaneous four-wave mixing (SFWM) (*21*). At the output of the chip, we filter signal and idler photons at 1545 and 1555 nm, respectively. We detect individual photons by using superconducting single-photon detectors (SSPDs) and identify matching arrival times by using a time correlation circuit (TCC). In this way, we measure the spatial profile of the wave function and the biphoton correlations as a function of the lattice position.

Generally, photon pairs generated within a lattice have entanglement between the spatial modes of the lattice (in addition to frequency entanglement) (*22*). In our lattice, the nontrivial topology provides selection rules for the generation of the biphoton states (*23*) in addition to the phase-matching condition typical of trivial quantum walks of correlated photons using SFWM. Because the pump beam excites the topological defect mode, it stays confined to the defect. The biphotons thus strongly overlap with the wave function of the topological defect mode. This yields topological protection of several features of the biphototon wave function and, at the same time, makes the wave function close to—but not exactly—separable.

This underlies an important difference between the current experiments with biphoton states and previous topological photonic studies that used classical light. Although the eigenmodes of the structure are the same as those for classical states, the wave functions of multiphoton states exist in a higher-dimensional Hilbert space—in particular, the current experiments with two nonidentical photons in m waveguides exist in an m^2 dimensional Hilbert space. The presence of the topological defect state localizes the biphoton to a near-separable state, but an m^2 dimensional Hilbert space is still required to describe the state because the signal and idler cannot be described independently, even if the photons are nonidentical (*24*). In this context, previous topological experiments using single-photon states (*10*–*12*), which live in an m^1 dimensional Hilbert space, are similar to those using the corresponding classical states.

Figure 1A shows a scanning electron micrograph (SEM) of the fabricated topological lattice of silicon nanowires on a silica substrate (Fig. 1A), where height h = 220 nm, width w = 450 nm, short gap g_s = 173 nm, and long gap g_l = 307 nm (*21*). The lattice consists of 203 nanowires with length L = 381 μm. The center nanowires fan out at the output of the array to allow the collection of photons via grating coupling to a fiber array. For comparison, we also fabricated a nontopological lattice of equidistant nanowires with a wider center waveguide (Fig. 1B), giving rise to a localized, topologically trivial defect mode. Choosing a gap g = 100 nm and w_s = 465 nm, we achieved the same confinement at the five center waveguides for the topologically and trivially localized defect modes.

To experimentally demonstrate the robustness of the topological biphoton state against disorder, we fabricated and measured structures without disorder and with deliberately introduced random disorder in the position of the waveguides and thus in the coupling constants. In our nanophotonics platform, randomizing the positions of the center of each nanowire within a range of ±14 nm (±43 nm) results in disorder of 20% (60%) of the original coupling constant.

We measured biphoton correlations (Fig. 2) over 120 s at the output of the five center waveguides of the topological (Fig. 2, A to C) and the nontopological (Fig. 2, D to F) lattice for increasing levels of disorder. In this context, we measure individual structures exemplifying a particular instance of random disorder. The measurements agree with the quantum states resulting from the propagation simulations at each level of disorder

[1]Institute of Photonics and Optical Science (IPOS), The Sydney Nano Institute, School of Physics, The University of Sydney, Sydney, New South Wales 2006, Australia. [2]Physics Department and Solid State Institute, Technion-Israel Institute of Technology, Haifa 32000, Israel.
*Corresponding author. Email: andrea.blancoredondo@sydney.edu.au
†Present address: Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, UK.

(movies S1 and S2). The vertical scale is distinct for each measured structure and shows the true signal and idler coincidence counts between waveguides, with the background level of accidental coincidences subtracted. The coincidence-to-accidental ratio (CAR) is well above 10 for all significant points in the correlation matrix.

For the biphoton quantum state in the topological lattice, the strong center peak in Fig. 2, A to C, shows that, with high probability, both correlated photons are in waveguide 0 even for very high levels of disorder. This is evidence of the strong localization of the quantum state at the topological defect. However, the most crucial feature in this topological quantum state is that the counts remain zero in the odd waveguides. The robustness of this quantum state stems from its strong overlap with the topological edge mode of the lattice, whose wave function is immune to disorder (Fig. 3A). The Schmidt number (K) of the measured topological biphoton states, which essentially represents the mode entanglement, is not protected against disorder, as shown in Fig. 2. Nevertheless, K remains close to 1 because the topology guarantees the presence of a single localized mode, even with disorder, and the biphotons are well overlapped with this mode.

These measurements beg the question: Are the quantum correlations playing a crucial role, or is the robustness observed here just a logical consequence of the topological protection of classical light (*21*) or of single photons (*12*, *13*)

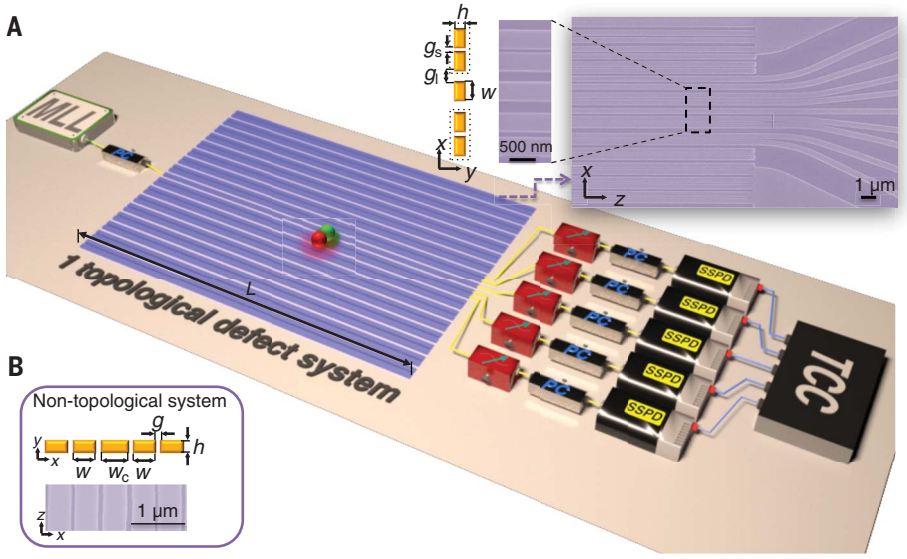


Fig. 1. Our nanophotonics topological system and experimental setup. (**A**) An MLL, a dimer chain of silicon nanowires with one topological defect, polarization controllers (PCs), tunable filters, SSPDs, and a TCC. (Inset) SEM of the topological defect. (**B**) SEM of the center waveguides of the structure providing nontopological (index-guiding) localization.

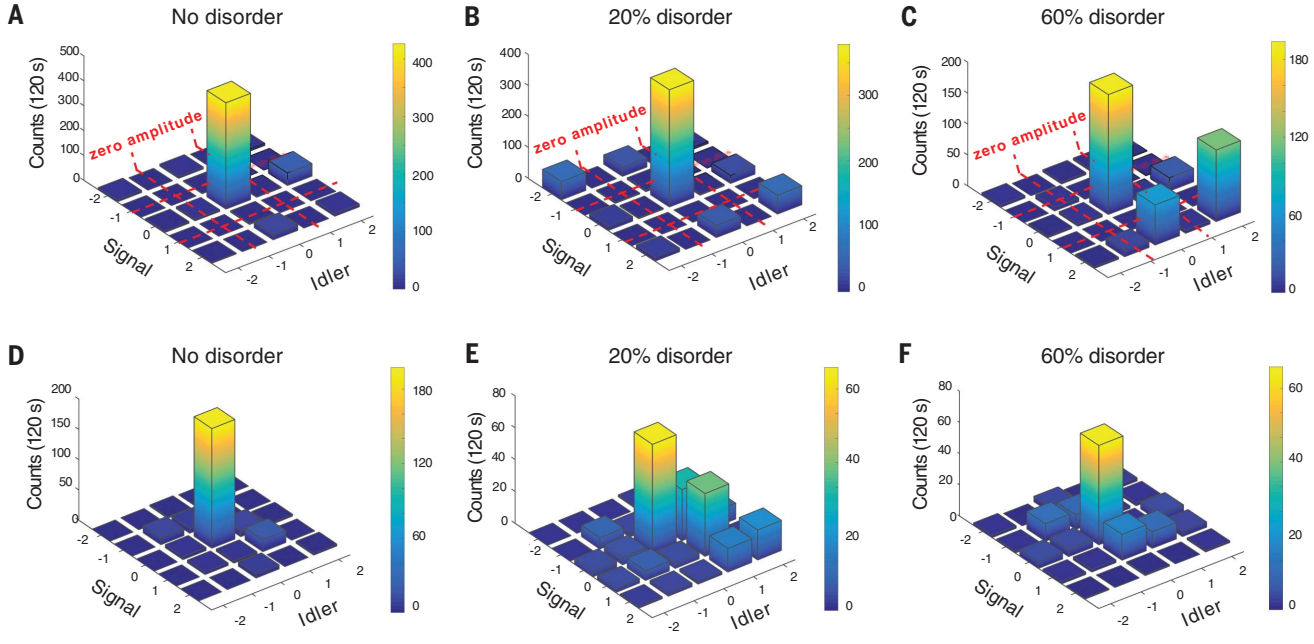


Fig. 2. Measured biphoton states with and without disorder. (**A** to **C**) Detected biphoton correlations of the topological quantum state for random disorder in a range of 0, 20, and 60% in the coupling constants. The probability amplitude remains zero in the odd waveguides. (**D** to **F**) Detected biphoton correlations of the nontopological quantum state under the same conditions of disorder. The disorder causes all coincidence amplitudes to vary, and no feature is preserved.

ILLUSTRATIONS [SEM INSET AND (B)]: G. REN/RMIT UNIVERSITY

in the SSH model? We calculate the value of the quantity

$$\Gamma_{ij}^{\Delta,si} = \left\langle n_i^s n_j^i \right\rangle - \left\langle n_i^s \right\rangle \left\langle n_j^i \right\rangle$$

where $\langle n_i^s n_j^i \rangle$ denotes the signal (s)–idler (i) correlations and $\langle n_i^s \rangle \langle n_j^i \rangle$ denotes the product of the signal and idler intensities in waveguides i and j, respectively. Our calculation shows that $\Gamma_{ij}^{\Delta,si}$ is almost entirely determined by the biphoton correlation (*21*). Therefore, the measurements in Fig. 2 reveal an unexplored regime beyond the classical topological protection of electromagnetic states, with new quantities coming into play and new aspects of the wave function being protected.

The topological edge states in the SSH model are protected against any kind of disorder that preserves the chirality. In our nanophotonics SSH platform, this includes disorder in the position of the waveguides, environmental changes resulting in a z-axis uniform refractive index change, and wavelength fluctuations. However, a small amount of disorder in the width of the individual waveguides, which does not respect the lattice chirality, is inevitable because of fabrication precision limits. This is the cause of the only minor outlier in our measurements, the detection of a small but nonvanishing number of coincidences between waveguides 2 and –1 in Fig. 2C. Because of these fabrication imperfections, waveguide –1 is slightly wider than the rest in this particular device, leading to slight index localization.

For comparison, we repeated the same measurements on the nontopological lattice described above. The measured biphoton states in Fig. 2, D to F, show localization of the biphotons around the center waveguide but reveal severe delocalization even at low levels of disorder. More notably, no experimental observable is conserved, even though the pump is launched at a defect mode. Because the defect in Fig. 2, D to F, is nontopological, the emerging biphoton state does not exhibit any feature that is conserved in the presence of disorder (Fig. 3B).

The other prominent signature of topological protection in the SSH model is associated with the zeroes of the wave function: The energy of the topological edge mode (red dots in Fig. 3C) remains pinned at zero for all levels of disorder. In practice, this means that the propagation constant (k_z) of the topological mode does not change with disorder, which implies that the phase imparted by the lattice on the quantum state is robust. In contrast, the propagation constant of the index-guided mode (red dots in Fig. 3D) varies appreciably even for low levels of disorder, leading to phase errors that accumulate into very large phase variations at the lattice output. This highlights the importance of the results presented here in the context of quantum information systems with entanglement between different modes.

Simulations show that the nanophotonic platform for topologically protected waveguiding of biphoton quantum states demonstrated here can be used as the key building block of an entangled system. We put together two such building blocks in a two-topological-defect system (Fig. 4A). This yields two topological defect modes that are uncoupled from each other, provided that there is enough separation between defects *A* and *B*. The entanglement between these two modes can be generated in the same lattice, or alternatively, one can use entangled states generated outside the chip as the input to the topological modes. We simulated the generation of an entangled state between the two topological defect modes on-chip by using two intense pumps at the center of each topological defect and SFWM (*22*). This topological two-mode system can serve as the building block for a variety of quantum gates in a quantum information system. As an example, we demonstrate how it could be used for a quantum gate operating on 2 qubits.

The interference between the quantum states emerging from the two topological defect modes

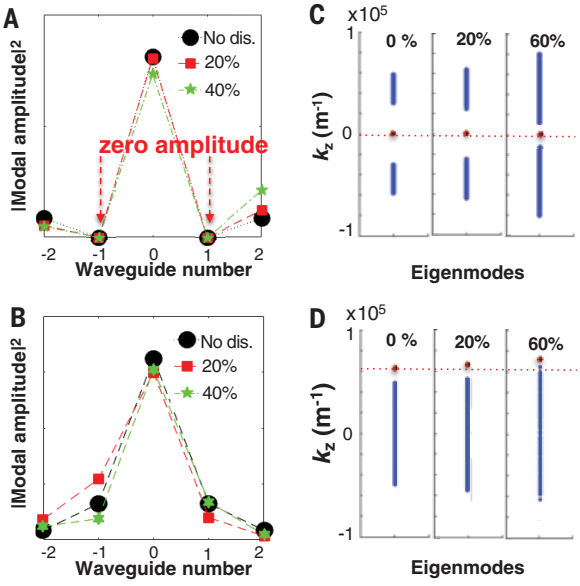


Fig. 3. Modal amplitudes and energies in the presence of lattice disorder. (**A**) The calculated amplitude squared of the topological defect mode with 0, 20, and 40% disorder. The modal amplitude remains zero in the odd waveguides for all levels of disorder. (**B**) Similar to (A) but for the trivial mode. Here, no feature is preserved. (**C** and **D**) Mode propagation constants (k_z), equivalent to mode energies, for topological and trivial systems, respectively. The red dot is k_z of the defect mode, which remains unchanged under disorder in the topological system but changes in the nontopological system.

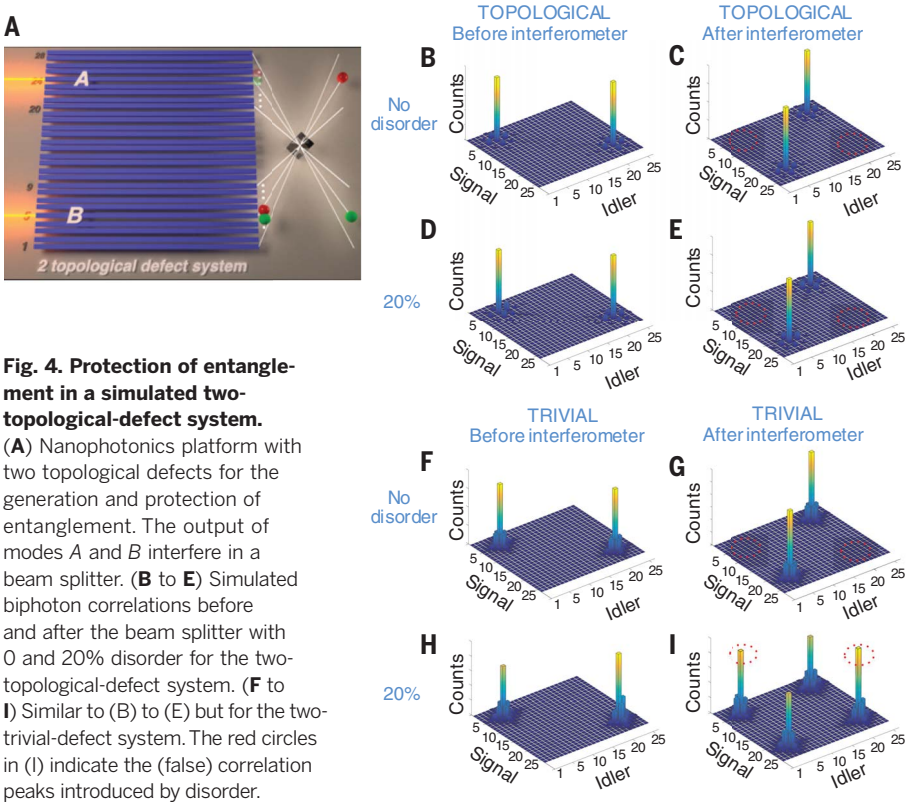


Fig. 4. Protection of entanglement in a simulated two-topological-defect system. (**A**) Nanophotonics platform with two topological defects for the generation and protection of entanglement. The output of modes *A* and *B* interfere in a beam splitter. (**B** to **E**) Simulated biphoton correlations before and after the beam splitter with 0 and 20% disorder for the two-topological-defect system. (**F** to **I**) Similar to (B) to (E) but for the two-trivial-defect system. The red circles in (I) indicate the (false) correlation peaks introduced by disorder.

of the lattice is robust against disorder. We simulated the biphoton generation and propagation with and without disorder and the interference of the defect modes A and B at the output (Fig. 4A) by using a beam splitter. This interference generates four peaks, $|A\rangle_s|A\rangle_i$, $|B\rangle_s|B\rangle_i$, $|A\rangle_s|B\rangle_i$, and $|B\rangle_s|A\rangle_i$, representing the presence of the signal and idler in defect modes A and B. At the output of the lattice but before the interferometer, we have $|A\rangle_s|A\rangle_i$, $|B\rangle_s|B\rangle_i$, corresponding to the entangled state in Fig. 4B. After the interferometer, these two peaks disappear because of destructive interference, and the peaks $|A\rangle_s|B\rangle_i$, $|B\rangle_s|A\rangle_i$ appear, experiencing constructive interference (Fig. 4C). This is similar to reverse Hong-Ou-Mandel interference, where entangled pairs of photons at the beam-splitter inputs are split and appear as one photon at each output.

In the presence of 20% disorder in the coupling constant, the propagation constants of the topological defect modes A and B, $k_{z,A}$ and $k_{z,B}$, remain unchanged, and thus the entanglement is preserved (Fig. 4, D and E). This robustness should be compared with an equivalent system of two trivial (nontopological) defects, such as the ones measured above. In the absence of disorder, our simulations show that the two peaks before (Fig. 4F) and after (Fig. 4G) the interferometer are similar to the topological case. However, in the presence of disorder, $k_{z,A}$ and $k_{z,B}$ change and the output phases are randomized, which leads to all four peaks being present in the biphoton correlation after the beam splitter (Fig. 4I). When the topological protection is absent, the entangled state takes on vastly different forms in the presence of disorder, such that on average (over multiple realizations of disorder), entanglement is lost. The concept of topology in this system, however, goes beyond making this platform more tolerant to fabrication defects. It offers a degree of freedom to generate new entangled quantum states in a multidimensional Hilbert space in a complementary metal-oxide semiconductor–compatible platform.

The results presented here provide definitive experimental evidence of the protection that the underlying topology of the structure can provide to the spatial features of states living in high-dimensional Hilbert spaces. We have shown how the topological nanophotonic platform can be used to enable scalable entangled quantum information systems robust to fabrication disorder. The multidimensionality of the wave function highlights new directions in topological physics, as new variables become relevant and can be affected by topology. This is also a versatile platform to investigate phenomena combining topological edge modes with nonlinear optics, with the ability to generate and manipulate quantum correlations. Lastly, the topological protection of the biphoton states demonstrated here for the 1D SSH lattice serves as an experimental proof of concept for any other topological photonic quantum system, including 2D lattices such as those described in (*2*–*4*), which should exhibit topologically protected transport around the perimeter of the lattice, with protection against any disorder that does not break time-reversal symmetry. These findings lead to many new avenues in conjunction with boson sampling (*25*, *26*), quantum simulations (*27*, *28*), quantum computing (*9*), and a variety of exciting possibilities related to the evolution of quantum states in photonic structures (*29*).

REFERENCES AND NOTES

1. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, *Nature* **461**, 772–775 (2009).
2. M. C. Rechtsman *et al*., *Nature* **496**, 196–200 (2013).
3. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, *Nat. Photonics* **7**, 1001–1005 (2013).
4. X. Cheng *et al*., *Nat. Mater.* **15**, 542–548 (2016).
5. Z. Yang *et al*., *Phys. Rev. Lett.* **114**, 114301 (2015).
6. M. Aidelsburger *et al*., *Nat. Phys.* **11**, 162–166 (2014).
7. G. Jotzu *et al*., *Nature* **515**, 237–240 (2014).
8. R. Süsstrunk, S. D. Huber, *Science* **349**, 47–50 (2015).
9. A. Politi, J. Matthews, M. G. Thompson, J. L. O'Brien, *IEEE J. Sel. Top. Quantum Electron.* **15**, 1673–1684 (2009).
10. T. Kitagawa *et al*., *Nat. Commun.* **3**, 882 (2012).
11. F. Cardano *et al*., *Nat. Commun.* **7**, 11439 (2016).
12. S. Barik *et al*., *Science* **359**, 666–668 (2018).
13. M. C. Rechtsman *et al*., *Optica* **3**, 925 (2016).
14. S. Mittal, V. V. Orre, M. Hafezi, *Opt. Express* **24**, 15631–15641 (2016).
15. M. A. Gorlach, A. N. Poddubny, *Phys. Rev. A* **95**, 053866 (2017).
16. S. Mittal, M. H. Hafezi, Topologically robust generation of correlated photon pairs. arXiv:1709.09984 [physics.optics] (28 September 2017).
17. W. P. Su, J. R. Schrieffer, A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
18. N. Malkova, I. Hromada, X. Wang, G. Bryant, Z. Chen, *Opt. Lett.* **34**, 1633–1635 (2009).
19. A. Blanco-Redondo *et al*., *Phys. Rev. Lett.* **116**, 163901 (2016).
20. A. Blanco-Redondo *et al*., *Phys. Rev. Lett.* **117**, 129901 (2016).
21. Supplementary materials.
22. A. Peruzzo *et al*., *Science* **329**, 1500–1503 (2010).
23. D. Leykam, A. S. Solntsev, A. A. Sukhorukov, A. S. Desyatnikov, *Phys. Rev. A* **92**, 033815 (2015).
24. A. S. Solntsev, A. A. Sukhorukov, D. N. Neshev, Y. S. Kivshar, *Opt. Express* **20**, 27441–27446 (2012).
25. J. B. Spring *et al*., *Science* **339**, 798–801 (2013).
26. M. A. Broome *et al*., *Science* **339**, 794–798 (2013).
27. M. S. Rudner, L. S. Levitov, *Phys. Rev. Lett.* **102**, 065703 (2009).
28. J. M. Zeuner *et al*., *Phys. Rev. Lett.* **115**, 040402 (2015).
29. C. Reimer *et al*., *Science* **351**, 1176–1180 (2016).

ACKNOWLEDGMENTS

Funding: We gratefully acknowledge financial support from The Technion Society of Australia (NSW) and the NSW Department of Industry (The University of Sydney and the Technion collaborative photonics research project), the School of Physics of the University of Sydney (Professor Harry Messel research fellowship), and the Australian Research Council (CE110001018, FL120100029). **Author contributions:** All authors contributed to all aspects of this work. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/568/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
References (*30*–*32*)
Movies S1 and S2

9 June 2018; accepted 10 September 2018
10.1126/science.aau4296

PHYSICS

Ultrafast disordering of vanadium dimers in photoexcited VO_2

Simon Wall[1]*†, Shan Yang[2]†, Luciana Vidas[1], Matthieu Chollet[3], James M. Glownia[3], Michael Kozina[4], Tetsuo Katayama[5], Thomas Henighan[4], Mason Jiang[4], Timothy A. Miller[1], David A. Reis[4,6,7], Lynn A. Boatner[8], Olivier Delaire[2,9]*, Mariano Trigo[4,6]*

Many ultrafast solid phase transitions are treated as chemical reactions that transform the structures between two different unit cells along a reaction coordinate, but this neglects the role of disorder. Although ultrafast diffraction provides insights into atomic dynamics during such transformations, diffraction alone probes an averaged unit cell and is less sensitive to randomness in the transition pathway. Using total scattering of femtosecond x-ray pulses, we show that atomic disordering in photoexcited vanadium dioxide (VO_2) is central to the transition mechanism and that, after photoexcitation, the system explores a large volume of phase space on a time scale comparable to that of a single phonon oscillation. These results overturn the current understanding of an archetypal ultrafast phase transition and provide new microscopic insights into rapid evolution toward equilibrium in photoexcited matter.

Manipulating the structure and function of materials remains one of the ultimate challenges of modern science and technology (*1*). Ultrafast pulses can create new phases not accessible in equilibrium by transiently modifying the interatomic potential (*2*–*4*) and the atomic positions (*5*–*7*). However, probing the short-lived transient structures of 10^{23} atoms is a challenge, and the dynamics are often simplified by assuming they evolve on a potential energy surface described by a few degrees of freedom with a well-defined spatial periodicity (*3*, *8*–*11*). Although ultrafast electron (*8*, *12*, *13*) and x-ray diffraction (*2*, *10*, *14*–*17*) are sensitive to the structure with atomic length scale and femtosecond time resolution, they measure averages over many unit cells, and thus are less sensitive to random displacements. To overcome this limitation, we used femtosecond total x-ray scattering (*18*) to study the dynamics of the structural transition of bulk VO_2 at all length scales. We observed that the structural transition proceeds by uncorrelated disordering of the vanadium ions from their initial dimerized distribution, rather than the previously proposed synchronized motion along an optical phonon mode (*8*, *12*, *14*, *19*). Supported by ab initio molecular dynamics, our results show that the highly anharmonic, flat potential energy surface allows the quasi-rutile structure in the photoexcited state to develop on femtosecond time scales, by disrupting the vanadium pairs and populating a continuum of modes that enables the system to reach ergodicity within 150 fs. More generally, these results question the common interpretation of many ultrafast measurements in terms of motion along a well-defined reaction coordinate, or order parameter, on a potential energy surface (*5*, *8*, *10*, *12*, *16*, *20*) and provide a new level of understanding regarding how complex systems can reach ergodicity on an ultrafast time scale.

Ultrafast structural transitions are usually described in terms of a few phonon modes with well-defined wave vectors, and it is assumed that the system evolves along a potential surface connecting the initial and final structures that also determines the speed of the photoexcited transition (*2*–*4*, *14*, *19*). However, this assumes a coordinated motion along a well-defined reaction coordinate and ignores the role of disorder and entropy, which are thought to develop on slower time scales, even though entropy may be the driving factor in the thermodynamic transition (*21*). To date, all attempts to measure the "molecular movie" of the dynamics of a phase transition at the atomic scale have exploited Bragg scattering, which can only measure the average motions of atoms and is insensitive to deviations from the average reaction coordinate. Using femtosecond pulses from a free-electron laser (FEL), we measured total x-ray scattering, containing discrete Bragg peaks as well as the diffuse continuum between them, and show that the lattice disorder in photoexcited VO_2, manifested as diffuse intensity across momentum space, reaches close to its equilibrium value on a time scale comparable to that of a single atomic oscillation.

Vanadium dioxide (VO_2) can be switched from an insulating to a metallic state with ultrafast pulses. The insulator-to-metal transition (IMT) is accompanied by a structural change from a monoclinic (M1) to a rutile (R) structure (Fig. 1, A and B), and numerous prior efforts have focused on elucidating the evolution of both the electronic and lattice degrees of freedom via ultrafast pump-probe measurements (*4*, *8*, *9*, *12*, *14*, *21*). Generally, a solid-solid phase transition where the symmetry of the crystal is raised (from M1 to R in VO_2) as the temperature increases can be dominated by either a displacive process or an order-disorder process (Fig. 1, C and D). The distortion at the IMT in VO_2 has widely been interpreted as a displacive transition corresponding to a phonon soft mode of the high-symmetry metallic phase, based on symmetry considerations and diffraction measurements (*8*–*12*), as well as lattice dynamics simulations (*22*–*25*). In a displacive transition, the atoms collaboratively reshuffle their positions under the effect of one or few well-defined, spatially coherent vibrational modes, of amplitude ξ, that link the low- and high-symmetry structures. This is illustrated for the dimerized vanadium pairs of the monoclinic (M1) phase of VO_2 in Fig. 1C, with wave vector at the R point in the rutile Brillouin zone, $\mathbf{k}^{(R)} = (0, \frac{1}{2}, \frac{1}{2})_R$ [here, subscripts denote the Bravais basis convention (M1 or R), while the superscript (R) represents the special point in the Brillouin zone]. Below the critical temperature, T_c = 340 K, the system is locked near an energy minimum at a large amplitude of ξ, yielding the structural distortion and additional superstructure diffraction peaks, owing to the doubling of the unit cell volume, represented by the red dots in Fig. 1E. Above T_c, the onset of a displacive mechanism would show characteristic soft phonons at the associated wave-vector (red crosses at $\mathbf{k}^{(R)}$ in Fig. 1F, corresponding to additional Brillouin zone centers in Fig. 1E) (*26*–*28*). Alternatively, in an order-disorder transition, the atoms move from the low- to the high-symmetry structure in a spatially incoherent manner with no characteristic correlation length (*26*, *28*) and without soft modes at any wave vector (Fig. 1D). The two possible types of transformation can be distinguished by probing for diffuse intensity extended across reciprocal space, which is a hallmark of large-amplitude motions due to the uncorrelated disorder (orange shaded areas, Fig. 1G). Notably, in both cases, the average atomic position may follow the same trajectory, and thus an ultrafast diffraction measurement probing the Bragg peaks alone cannot distinguish between displacive and order-disorder mechanisms (Fig. 1, H and I).

Optical probes of photoexcited VO_2 show that ~6-THz coherent optical phonons, corresponding to the motion of the vanadium ions, are induced for low-intensity excitation (*4*, *9*, *19*). At the same time, structural probes (*8*, *12*) have shown that

[1]Institut de Ciències Fotòniques (ICFO), The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain. [2]Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA. [3]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [4]Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [5]Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198, Japan. [6]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [7]Department of Applied Physics, Stanford University, Stanford, CA, USA. [8]Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [9]Department of Physics, Duke University, Durham, NC 27708, USA.
***Corresponding author. Email: simon.wall@icfo.eu (S.W.); olivier.delaire@duke.edu (O.D.); mtrigo@slac.stanford.edu (M.T.)**
†These authors contributed equally to this work.

the long-range order of the V dimerization is lost in <300 fs, but these measurements lacked the sub–100 fs time resolution required to probe the transition on the time scale of relevant vibrational periods. Furthermore, these experiments lacked sensitivity to deviations from average lattice periodicity at arbitrary wave vectors, which are a signature of disorder in the transition mechanism. Here, we used femtosecond total x-ray scattering (~50-fs time resolution) with an x-ray FEL to obtain a more complete picture of the dynamics during the photoinduced structural transition in VO_2. We detected a rapid disordering of the low-temperature structure that starts as soon as 50 fs after photoexcitation, as observed through a rapid increase in diffuse x-ray intensity and an equally fast decrease in the intensity of the Bragg peaks of the low-symmetry structure. Furthermore, the diffuse intensity distribution in reciprocal space at 150 fs resembles the rutile equilibrium distribution and suggests that the lattice reaches a near-ergodic state in that time scale. This unequivocally shows that the nonequilibrium photoinduced transition is of the order-disorder type, driven by an ultrafast change in the lattice potential that suddenly unlocks the V dimers and yields large-amplitude uncorrelated motions, rather than occurring along a coherent displacive coordinate. Our ab initio molecular dynamics (AIMD) simulations based on density functional theory (DFT) potentials corroborate the observed ultrafast time scale of the disordering. The DFT calculations show that the ionic potential in the photoexcited phase is highly anharmonic and flat, which explains the observed large V displacements, and is also the origin of the high phonon entropy of the equilibrium metallic phase (*21*).

Total x-ray scattering patterns were collected with 9.5-keV photons at the XPP end station at the Linac Coherent Light Source (LCLS) from high-quality single crystals of VO_2 [see methods for details (*29*)]. The difference in the equilibrium scattering for $T = 353$ K (R) and $T = 295$ K (M1) is shown in Fig. 2A [see fig. S2 for M1 and R raw patterns (*29*)]. The change is dominated by the suppression of three strong Bragg peaks of the M1 phase labeled $(-1, 0, -2)_{M1}$, $(-1, 2, 0)_{M1}$, and $(-1, 2, 2)_{M1}$. In addition, a rectangular-shaped diffuse feature strongly increases in intensity. Our simulations of the change in the total scattering (Fig. 2B), based on the equilibrium force constants obtained from DFT [see methods (*29*)], reproduce the diffuse signature well and indicate that it originates from soft transverse acoustic phonons of the rutile phase (*21*). This enables the V atoms to adopt a large distribution of displacements about their average positions in the rutile phase.

In Fig. 2C, we show the photoinduced changes in the total scattering 350 fs after photoexcitation for an absorbed pump fluence of 20 mJ/cm^2, in the saturation regime for the photoinduced phase transition (*19*). The photoinduced changes in the total scattering show a marked resemblance to the thermal changes in Fig. 2A. The intensities of the Bragg peaks are strongly suppressed as a result of the change in the crystalline symmetry, whereas the extended diffuse scattering strongly increases. This shows that on this very fast time scale, not only has the long-range, average symmetry of the crystal changed, as indicated by the fast suppression of the M1 peaks, but also the V atoms have already adopted a rutile-like displacement distribution.

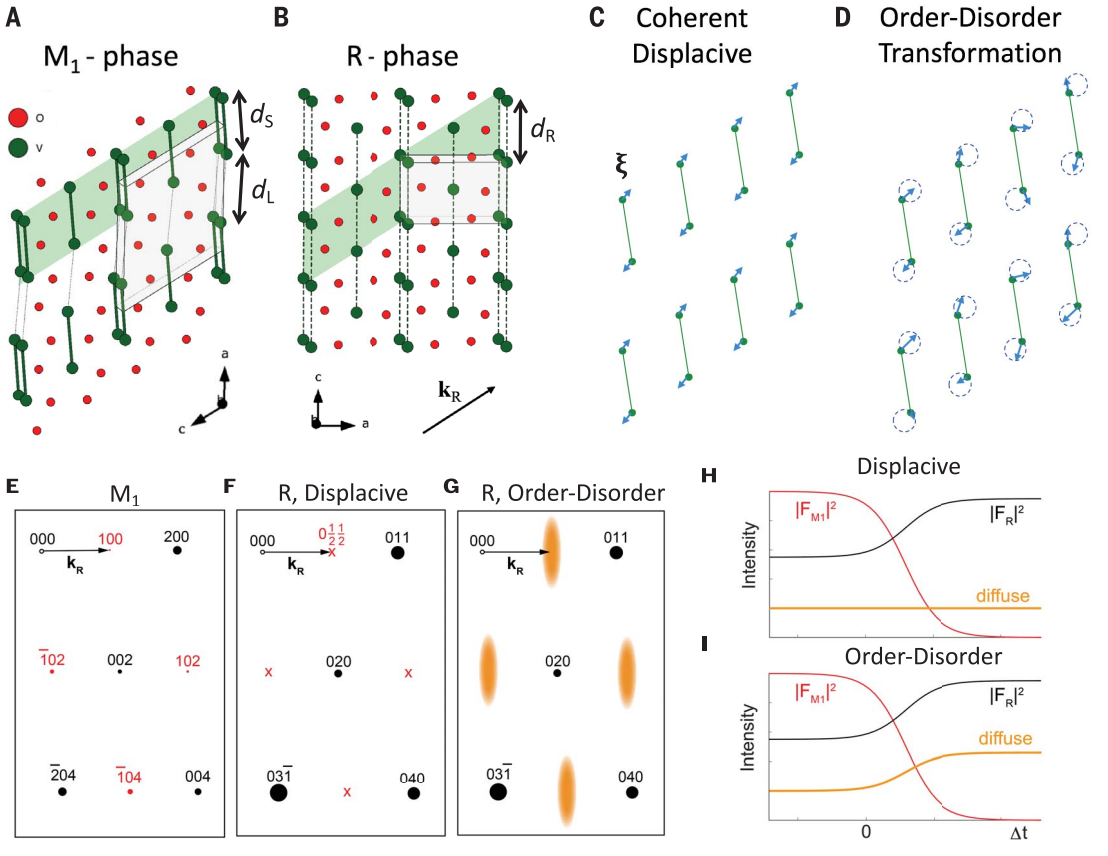


Fig. 1. VO_2 structural distortion in real and reciprocal space. (**A** and **B**) Correspondence between the M1 and R phases, highlighting the pattern of dimerization of V pairs along the rutile (R) *c* axis, leading to diagonal layers of tilted dimers (green bands), along the wave vector corresponding to the modulation of the R phase, $\mathbf{k}^{(R)} = (½, 0, ½)$ or (0, ½, ½) reciprocal lattice units. Gray boxes represent the unit cell of each structure. (**C** and **D**) Alternate possible transition mechanisms between the M1 phase and the R phase upon photoexcitation. The spatially coherent distortion (C) with coordinate ξ is a symmetry-allowed transformation path between R and M1, which matches low-energy phonon excitations of the rutile phase at the R point, localized at $\mathbf{k}^{(R)}$ in reciprocal space. The disordering mechanism (D) has no well-defined wave vector and causes extended diffuse scattering in **k**-space. Diffractograms for the (**E**) M1 phase and (**F** and **G**) rutile-like phases resulting from either displacive distortion (F) or a disordering transition (G). Note the appearance in M1 of superstructure Bragg peaks at $\mathbf{k}^{(R)}$ points of the rutile phase (crosses at coordinates with half integers in R phase). The shaded orange regions in (G) represent a diffuse scattering intensity that is more extended than the exact $\mathbf{k}^{(R)}$ modulation wave vector. (**H** and **I**) Schematic time evolutions of the intensity of Bragg peaks for M1 and R phases and diffuse signal, for a displacive transition (H) or an order-disorder transition (I). Zero time corresponds to photoexcitation.

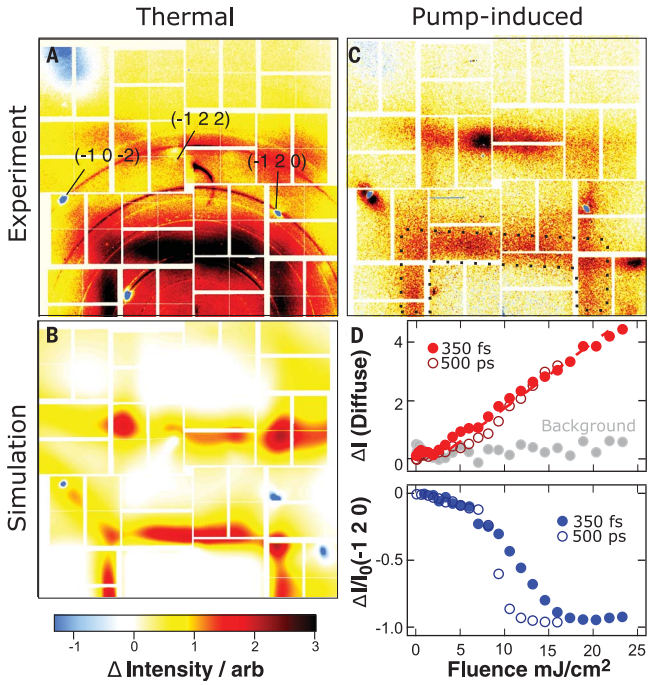


Fig. 2. Diffuse scattering from the thermal and photoinduced phase transitions in VO_2. Saturated false-color plot of the difference in the thermal diffuse scattering intensity between the R (353 K) and the M1 phases (295 K) [$\Delta I(\mathbf{Q}) = I(\mathbf{Q}, 353\ K) - I(\mathbf{Q}, 295\ K)$], (**A**) measured at LCLS and (**B**) calculated on the basis of the equilibrium force constants. The positions of several Bragg peaks are indicated. The rings in (A) arise from x-ray scattering at damaged edges of the sample, and the central semicircle-like feature corresponds to laser scatter [see methods and figs. S1 and S2 for details (*29*)]. (**C**) Pump-induced change in the scattering signal when excited at 20 mJ/cm^2 at a fixed delay of 350 fs [$\Delta I(\mathbf{Q}) = I(\mathbf{Q}, 350\ fs) - I(\mathbf{Q}, -11\ ns)$], showing changes very similar to those of the thermal response. (**D**) Pump fluence dependence of the $(-1, 2, 0)_{M1}$ peak amplitude and mean diffuse scattering as integrated over the boxed area at 350 fs and 500 ps. The gray dots represent the background, taken as the mean intensity at low scattering angles. A linear fit (dashed line) to the high-fluence part of the diffuse scattering gives a threshold of $F_c \sim 5$ mJ/cm^2.

Before turning to the time scale in more detail, we examine the fluence dependence of the photoinduced phase transition. Figure 2D shows the fluence dependence of the $(-1, 2, 0)_{M1}$ Bragg peak and the integrated diffuse scattering at ~350 fs and 500 ps after excitation. The Bragg peak shows a linear decrease in intensity with absorbed fluence for fluences below a threshold of $F_c \sim 5$ mJ/cm^2, after which the change is more rapid until saturation and complete suppression occurs. The dynamics of the Bragg peak also show distinct changes below and above threshold, with coherent phonons observed to modulate the Bragg intensity below threshold, whereas the coherent signal disappears above threshold (fig. S11), in agreement with optical measurements (*4*). The diffuse scattering also shows threshold-like behavior, rising abruptly above $F_c \sim 5$ mJ/cm^2, indicating that the V ions only adopt the broad, rutile-like distribution above a critical fluence.

Figure 3A shows the change in intensity $\Delta I(\mathbf{Q}, t)$ at delays of 50 fs, 100 fs, and 2 ps for excitation in the saturation regime. It is clear that changes in the diffuse scattering already begin to occur over extended portions of reciprocal space within 50 fs of photoexcitation, and the **Q** dependence of the scattering shows little subsequent change after reaching its full intensity at about 150 fs. To examine this time scale in more detail, we follow the temporal evolution of the intensity for specific regions of reciprocal space. Figure 3B shows the dynamics of multiple Bragg peaks and of the diffuse scattering intensity in the high-fluence regime. All curves clearly show that the time scale for the structural transition is at least twice as fast as the 300 fs reported from previous diffraction measurements (*8*, *12*). Notably, this time scale is above the ~50-fs experimental time resolution. The time scale for the rise of the integrated diffuse intensity is independent of fluence above the threshold, as shown in fig. S12. However, below threshold and at long delays, the Bragg peaks can be affected by strain, which leads to slow changes in intensity at the few-ps time scale.

The dynamics of both the Bragg peaks and diffuse scattering observed here indicate a direct structural transition between the M1 and R phase, without evidence of passing through intermediate structures previously reported (*12*). M1 superstructure peaks [e.g., $(-1, 0, -2)_{M1}$, $(-1, 2, 0)_{M1}$] are not present in the rutile structure, and all rapidly drop to zero intensity. Peaks labeled R are present in both phases and can initially increase or decrease, depending on their structure factor response to the average distortion, $F_{hkl}(\xi)$. Notably, the diffuse scattering signal changes on the same time scale as that of the Bragg peaks, showing that the disordering occurs as fast as the loss of M1 long-range order. This indicates that the phase transition is dominated by disordering of the V ions rather than a coherent atomic motion along the ξ coordinate.

To gain greater insight into the phase-change dynamics, we performed AIMD simulations of the photoinduced phase transition. The lattice potential was obtained from DFT, including electronic correlation effects [see (*29*) for additional information], and the effect of photoexcitation was modeled as a sudden increase in the electronic temperature, T_{el} [see supplementary materials for details (*29*)]. The relaxed ground-state structure in the M1 phase shows a clear threshold behavior in T_{el} above which the V dimerization disappears (fig. S4), leading to a rutile-like phase, similar to the trend observed with increasing hole concentration (*30*) and in agreement with the experimental observation of a threshold laser fluence. The time-dependent Bragg intensities derived from the ionic positions after electronic excitation are plotted in Fig. 3B, together with the measured diffraction signals. Excellent agreement is obtained both in terms of the time scale for the dynamics and the amplitude of the change. The agreement is particularly noteworthy considering that the only adjustable parameter was the global temporal offset between the data and the simulation.

Figure 4A shows the AIMD evolution of short dimerized bonds (d_S) and long V-V bonds (d_L) along the rutile *c* axis, after a sudden increase in T_{el} at $t = 0$, starting with a thermal M1 structure at 300 K. Bond-length distributions, averaged over specified time windows, are shown in Fig. 4B. The initial distribution is bimodal ($d_S = 2.50$ Å, $d_L = 3.16$ Å), reflecting the monoclinic dimerization pattern (Fig. 1A). After the prompt transition, it quickly collapses to the unimodal distribution expected for the rutile phase, $d_S = d_L = 2.81$ Å = d_R. Already within the first 75 fs, the tails of the distributions show some vanadium pairs with the rutile bond length, but the means only cross at about 150 fs and slightly overshoot the ideal rutile bond length at $t \sim 200$ fs before rebounding, resulting in a damped oscillation converging to the equidistant configuration. This overshoot is responsible for the single-cycle overdamped oscillation in the $(-1, 0, -2)_{M1}$ and $(-1, -1, 3)_{M1}$ peaks. Similar dynamics are observed in the evolution of the twist angle, and the frequency of the damped oscillation is about 3 THz (fig. S5), but is doubled (~6 THz) in the diffraction data shown in Fig. 3B because $|F_{hkl}(\xi)|^2 \sim \xi^2$ when $d_L = d_S = d_R$.

Our measurements, confirmed by simulations, show that the changes in bond length and angle occur concomitantly and within a few hundred femtoseconds, in contrast to previous scenarios based on analysis of time-resolved electron Bragg scattering alone (*8*). The vibrational modes at 3 THz are a factor of 2 slower than the 6-THz mode previously suggested to limit the phase transition (*19*), but agrees well with the frequency of V-dominated flat transverse acoustic modes of the R phase, as seen in our first-principles simulations of the phonon density of states (Fig. 4D) and dispersions (fig. S6d), showing that no single wave vector is responsible for the transition. In addition, the dynamics of the standard deviation of the bond lengths is consistent with

the time scale of the diffuse scattering (solid line and symbols in Fig. 3B, top trace), further demonstrating that a broad distribution of modes is required to describe the evolution of the distribution of displacements in Fig. 4, A and B. The AIMD simulations thus indicate that photoexcitation drives an instantaneous change in the lattice potential, leading to rutile-like dynamics of the V ions. The effect of changing the electronic temperature on the potential energy surface along the distortion ξ (M1-to-R distortion in Fig. 1C) is shown in Fig. 4C. As can be seen, the electronic excitation results in the loss of the double-well potential characteristic of the M1 phase and instead stabilizes the flat, strongly anharmonic quartic-like potential of the rutile phase.

These observations allow us to build a radically new description of the photoinduced phase transition in VO_2. To date, most theoretical models and interpretations of experimental data have assumed a transition driven coherently by phonon modes of well-defined wave vector (typically $\mathbf{k}_R$) and frequency (*8*, *9*, *12*, *14*, *19*), and have ignored the role of fluctuations and disorder. Our combined ultrafast total scattering measurements and first-principles simulations of dynamics show that ultrafast disordering is a key aspect of the photoexcited state of VO_2. Photoexcitation instantaneously creates a new lattice potential, which can be described by a hot electron distribution within the DFT framework. The new potential quickly alters the dynamics of V ions, effectively severing the intradimer spring constant, causing them to rapidly assume the new high-symmetry configuration. We note that these V-V interactions in the rutile phase yield a flat, dispersiveless transverse acoustic phonon branch (*21*). These features of the potential cause the transformation to proceed without spatial coherence and no well-defined wave vector.

The time scale of this disordering process is another particularly unexpected observation. The rapid excitation of phonons in a broad range of wave vectors, observed through the diffuse scattering and occurring within 150 fs, is faster than other disorder-driven phase transitions (*13*, *17*). Whereas the speed of a soft mode–driven transition is governed by the curvature of the potential surface along the ξ coordinate (*3*), the nonequilibrium order-disorder transition time scale reflects the time for the displacements to become uncorrelated and reach a quasi-equilibrium distribution. The dynamics observed here are fast because the average distance between M1 and rutile is relatively short (ξ_0 = 0.24 Å, the position of the minimum of the potential in Fig. 4C). Indeed, the average vanadium velocity at room temperature is 3.8 Å/ps, so the V atoms could traverse the distance ξ_0 in ~60 fs. Although this is a simplified model, the flatness of the photoexcited potential in Fig. 4C suggests that the dynamics are partially inertial and that the speed of the transition is mainly limited by the initial thermal distribution of velocities in the M1 phase and the amplitude of the M1-to-R distortion. The change in the potential produces an ultrafast activation of a large number of low-energy V-dominated phonon modes spread out in reciprocal space, as seen in the broadly distributed diffuse scattering signal, and in the phonon dispersions in fig. S6. The rapid evolution toward ergodicity is thus enabled by the large phase space with low-frequency V-dominated transverse acoustic modes, which was also noted in (*21*) to yield a large phonon entropy gain stabilizing the rutile phase and the strong damping of phonons in the rutile phase. Our current observations thus reveal an interesting parallel to the thermal transition.

Our results show that the structural transformation in VO_2 occurs along many concurrent degrees of freedom spanning a large phase space, and thus a complete description must extend beyond that of a single-order parameter or discrete set of displacements. The rapid, large-amplitude disordering observed here is caused by the flatness of the photoexcited potential and is intimately related to the high entropy of the rutile phase (*21*); but the implications of our findings extend well beyond the VO_2 system. The role of disorder has been neglected in many ultrafast solid-solid phase transitions, even though the ultimate control of properties on demand relies on a reversible pathway between the two states. Our findings suggest that disorder may play an important role in some materials and that a description in terms of a single degree of freedom is incomplete. This disorder presents both opportunities as well as challenges to ultrafast material control. Disorder ultimately prevents coherent control of phase transitions and complete control over the phase transition pathway. However, transient disorder can become a new knob by which materials can be controlled, just as disordered solids (glasses) can show marked differences compared with their crystalline counterparts

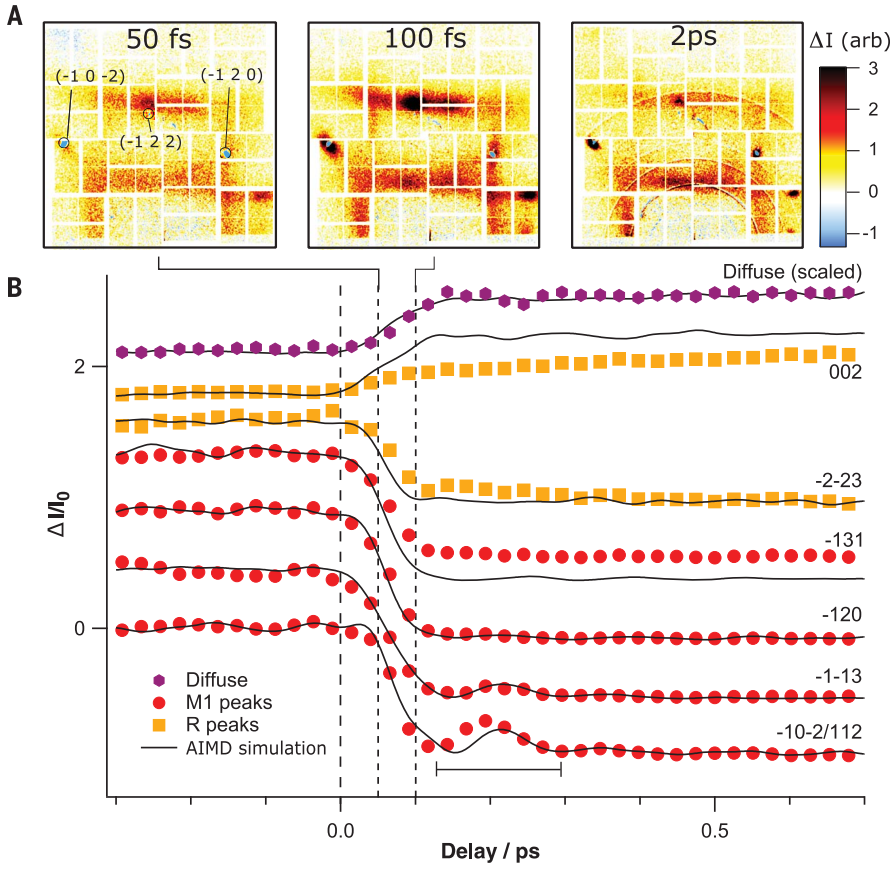


Fig. 3. Time dependence of Bragg peaks and diffuse scattering for above-threshold pumping of 22 mJ/cm^2. (**A**) Snapshots of the increase in the diffuse intensity $\Delta I(\mathbf{Q}, t)$ at several delays after laser excitation. (**B**) Temporal evolution of selected Bragg peaks and integrated diffuse scattering signal compared to AIMD simulations (solid lines). With the exception of the diffuse scattering signal, all data are normalized to the pre–time zero amplitude for both experiments and simulations with no further scale factor. The diffuse scattering data are compared to the variance of the vanadium positions obtained from the AIMD simulations and are thus separately normalized. The horizontal bar indicates half of the period of the relevant 3-THz acoustic modes of the rutile phase (see text for details). The additional Bragg peaks that do not appear in (A) were obtained in a different experimental geometry, shown in fig. S3.

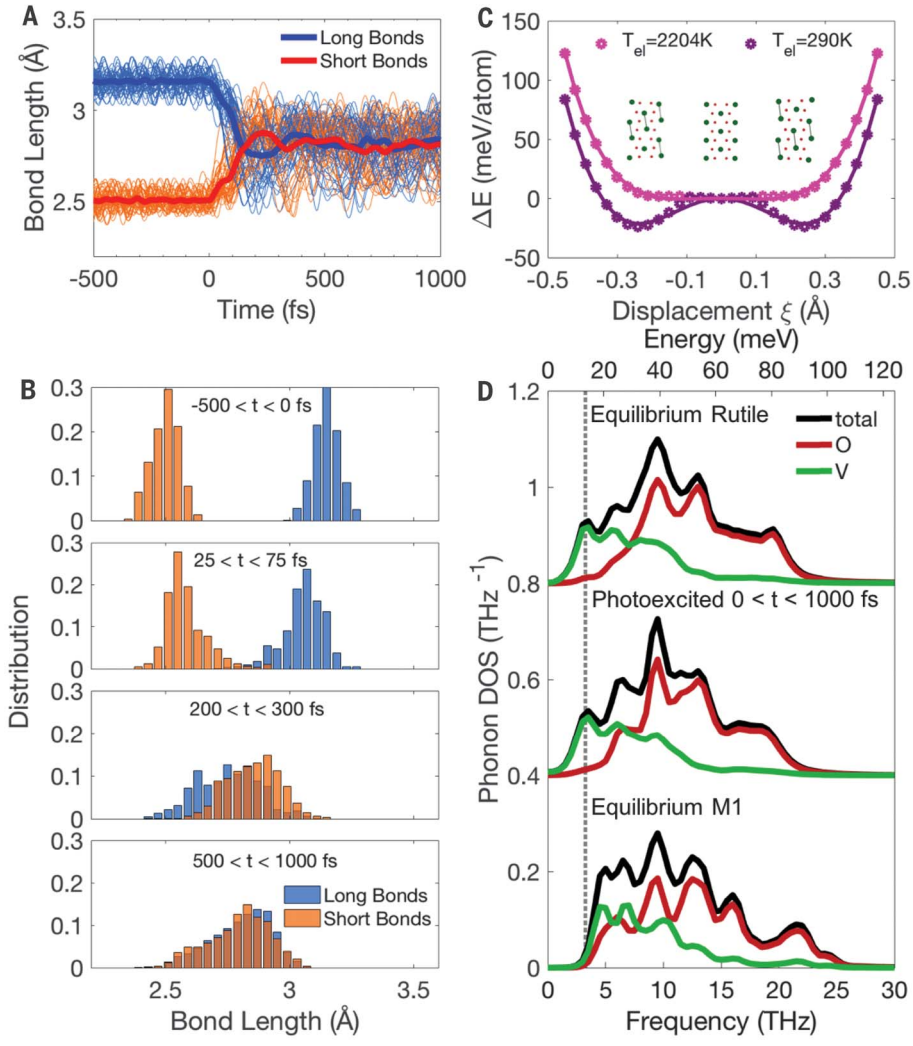


Fig. 4. Simulated VO_2 response under photoexcitation, based on first-principles DFT simulations. (**A**) Time evolution of dimer bond lengths from ab initio molecular dynamics. Short (d_S) and long (d_L) bonds are indicated in Fig. 1A. (**B**) Distribution of dimer bond lengths from (A), integrated over several time intervals before and after photoexcitation. The merging of the distributions after a brief overshoot is evident, indicating the evolution toward the rutile phase with chains of equidistant V atoms. (**C**) Effect of photoexcitation on the potential for the distortion coordinate ξ. Rutile corresponds to ξ =0 Å and M1 to ξ = ±0.24 Å. The photoexcitation lifts the two M1 minima of the double-well potential to stabilize the rutile phase in a flat, strongly quartic potential. Markers are DFT energies, and curves are polynomial fits. (**D**) Phonon density of states, and elemental decomposition on V and O atoms, obtained from AIMD trajectories for the equilibrium M1 phase before photoexcitation (bottom) and after photoexcitation (middle). The appearance of the pronounced V vibration peak at 3 THz (dashed line) after photoexcitation is characteristic of the rutile phase equilibrium phonon dynamics, shown in the top curves (T = 380 K).

in near-equilibrium conditions. Understanding whether disorder plays a general role in vibrationally excited solids (*7*) could ultimately provide a new perspective on how to control matter, suggesting a new direction in which light-induced superconductivity may be achieved (*5*).

REFERENCES AND NOTES

1. D. N. Basov, R. D. Averitt, D. Hsieh, *Nat. Mater.* **16**, 1077–1088 (2017).
2. D. M. Fritz *et al.*, *Science* **315**, 633–636 (2007).
3. T. Huber *et al.*, *Phys. Rev. Lett.* **113**, 026401 (2014).
4. S. Wall *et al.*, *Nat. Commun.* **3**, 721 (2012).
5. R. Mankowsky *et al.*, *Nature* **516**, 71–73 (2014).
6. I. Katayama *et al.*, *Phys. Rev. Lett.* **108**, 097401 (2012).
7. M. Först *et al.*, *Nat. Phys.* **7**, 854–856 (2011).
8. P. Baum, D.-S. Yang, A. H. Zewail, *Science* **318**, 788–792 (2007).
9. C. Kübler *et al.*, *Phys. Rev. Lett.* **99**, 116401 (2007).
10. P. Beaud *et al.*, *Nat. Mater.* **13**, 923–927 (2014).
11. S. Hellmann *et al.*, *Nat. Commun.* **3**, 1069 (2012).
12. V. R. Morrison *et al.*, *Science* **346**, 445–448 (2014).
13. L. Waldecker *et al.*, *Nat. Mater.* **14**, 991–995 (2015).
14. A. Cavalleri *et al.*, *Phys. Rev. Lett.* **87**, 237401 (2001).
15. M. Hada, K. Okimura, J. Matsuo, *Phys. Rev. B* **82**, 153401 (2010).
16. H. Ichikawa *et al.*, *Nat. Mater.* **10**, 101–105 (2011).
17. A. M. Lindenberg *et al.*, *Science* **308**, 392–395 (2005).
18. M. Trigo *et al.*, *Nat. Phys.* **9**, 790–794 (2013).
19. A. Cavalleri, T. Dekorsy, H. H. W. Chong, J. C. Kieffer, R. W. Schoenlein, *Phys. Rev. B* **70**, 161102 (2004).
20. A. V. Kolobov, M. Krbal, P. Fons, J. Tominaga, T. Uruga, *Nat. Chem.* **3**, 311–316 (2011).
21. J. D. Budai *et al.*, *Nature* **515**, 535–539 (2014).
22. J. R. Brews, *Phys. Rev. B* **1**, 2557–2568 (1970).
23. M. Gupta, A. J. Freeman, D. E. Ellis, *Phys. Rev. B* **16**, 3338–3351 (1977).
24. H. Terauchi, J. B. Cohen, *Phys. Rev. B* **17**, 2494–2496 (1978).
25. F. Gervais, W. Kress, *Phys. Rev. B Condens. Matter* **31**, 4809–4814 (1985).
26. A. D. Bruce, *Adv. Phys.* **29**, 111–217 (1980).
27. R. A. Cowley, *Adv. Phys.* **29**, 1–110 (1980).
28. M. T. Dove, *Am. Mineral.* **82**, 213–244 (1997).
29. Materials and methods are available as supplementary materials
30. X. Yuan, W. Zhang, P. Zhang, *Phys. Rev. B* **88**, 035119 (2013).

ACKNOWLEDGMENTS

Preliminary static diffuse scattering patterns were collected at the 11-3 beamline of the Stanford Synchrotron Radiation Light Source (SSRL). **Funding:** Work at ICFO was supported by Spanish MINECO (Severo Ochoa grant SEV-2015-0522), Ramón y Cajal programme RYC-2013-14838, and Fondo Europeo de Desarrollo Regional FIS2015-67898-P (MINECO/FEDER), as well as by the European Union's Horizon 2020 research and innovation programme (grant agreement no. 758461) Fundació Privada Cellex, and CERCA Programme – Generalitat de Catalunya. First-principles simulations and analysis (S.Y., O.D.) were supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under the Early Career Award no. DE-SC0016166. M.K., T.H., M.T., and D.A.R. were supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences through the Division of Materials Sciences and Engineering under contract no. DE-AC02-76SF00515. Use of the Linac Coherent Light Source (LCLS), and the Stanford Synchrotron Radiation Lightsource at SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. Additional x-ray measurements were performed at BL3 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal no. 2016A8008). Ab initio calculations were performed using resources of the National Energy Research Scientific Computing Center, a U.S. DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. First principles simulations for this research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC05-00OR22725. **Author contributions:** L.V., M.C., J.M.G., M.K., T.K., T.H., M.J., T.A.M., D.A.R., and O.D. contributed to the data collection led by S.W. and M.T. Data analysis was performed by S.W. and M.T. Sample synthesis was performed by L.A.B. S.Y. performed DFT calculations under the supervision of O.D. S.W., O.D., and M.T. proposed, planned, and coordinated all aspects of the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/572/suppl/DC1
Materials and Methods
Figs. S1 to S12
Movies S1 and S2
References (*31*–*48*)

7 June 2018; accepted 5 September 2018
10.1126/science.aau3873

EPIDEMIOLOGY

Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes

Simon A. Babayan[1,2], Richard J. Orton[3], Daniel G. Streicker[1,3]*

Identifying the animal origins of RNA viruses requires years of field and laboratory studies that stall responses to emerging infectious diseases. Using large genomic and ecological datasets, we demonstrate that animal reservoirs and the existence and identity of arthropod vectors can be predicted directly from viral genome sequences via machine learning. We illustrate the ability of these models to predict the epidemiology of diverse viruses across most human-infective families of single-stranded RNA viruses, including 69 viruses with previously elusive or never-investigated reservoirs or vectors. Models such as these, which capitalize on the proliferation of low-cost genomic sequencing, can narrow the time lag between virus discovery and targeted research, surveillance, and management.

Preventing emerging viral infections—including Ebola, SARS, and Zika—requires identification of reservoir hosts and/or blood-feeding arthropod vectors that perpetuate viruses in nature. Current practice requires combining evidence from field surveillance, phylogenetics, laboratory experiments, and real-world interventions but is time consuming and often inconclusive (*1*). This creates prolonged periods of uncertainty that may amplify economic and health losses. We aimed to develop a general model to predict reservoir hosts and arthropod vectors across single-stranded RNA (ssRNA) viruses, the viral group most commonly implicated in zoonotic disease outbreaks (*2*), building on the modern expansion of low-cost viral sequence data (*3*).

We collected a single representative genome sequence per viral species or strain from 12 taxonomic groups (11 families and one order) of ssRNA viruses that can infect humans; that is, 80% of all human-infective groups (Fig. 1A). For each virus, we used extensive literature searches to determine currently accepted reservoir hosts (437 viruses, 11 reservoir groups), whether transmission involves an arthropod vector (527 viruses), and if so, the identity of arthropod vectors (98 viruses, four vector groups). To maximize predictive scope, reservoir and vector groups included the most-frequent sources of emerging human viruses as well as other common hosts in human-infective viral families (e.g., fish, plants, and insects) (*2*, *4*).

Because related viruses often have closely related hosts owing to cospeciation and preferential host switching among related host species, we first designed an algorithm to predict host associations from viral phylogenetic relatedness alone (*5*, *6*). This phylogenetic neighborhood (PN) model identified the reservoir hosts of 58.1 ± 0.07% (±SD) of viruses, whether or not 95 ± 0.24% of viruses were transmitted by an arthropod vector, as well as the vector identity of 67.2 ± 0.12% of arthropod-borne viruses. Biases in viral genome composition can also inform host-virus associations. Specifically, viral codon pair and dinucleotide biases are reported to mimic those of their hosts, representing either a genome-wide strategy for adaptation to specific host groups or genomic imprinting by the host cellular machinery that viruses co-opt for replication (*7*). In any case, genomic biases can coarsely discriminate viruses from different host groups within several well-studied viral families (*8*–*10*). However, whether genomic biases can predict hosts from smaller or less-studied groups of viruses remains unresolved (*11*). We quantified 4229 traits from the 536 viral genomes in our dataset, including all possible codon pair, dinucleotide, codon, and amino acid biases (*6*) (fig. S1). When all traits were weighted equally, dissimilarity-based clustering grouped viruses predominately by viral taxonomy; however, paraphyly of most viral groups implied selective forces on viral genomic biases that outweighed phylogenetic history (Fig. 1, B and C). Generalized linear mixed models further revealed that even after controlling for effects of viral taxonomy, some genomic biases of viruses were correlated with their reservoir and vector associations, suggesting host effects on viral genomes that transcend viral groups (figs. S2 to S7). We hypothesized that combining host-associated genomic biases with viral PNs could maximize prediction of reservoirs and vectors from viral sequence data.

We addressed this challenge by using supervised machine learning, a class of statistical models that can integrate multiple traits that carry a weak signal in isolation but build a strong signal when optimally weighted (*12*). Gradient boosting machines (GBMs) (*13*) outperformed seven alternative classifiers in predicting host associations from viral genomic biases and identified the most informative genomic traits for each aspect of viral ecology (figs. S8 to S12). GBMs combining selected genomic traits (SelGen) with viral PNs predicted reservoir hosts with up to 83.5% accuracy, distinguishing all 11 reservoir groups, including taxonomic divisions within the birds (i.e., Neoaves versus Galloanserae) and bats [i.e., Pteropodiformes ("Pterobat") versus Vespertilioniformes ("Vespbat")] (Fig. 2A). Reservoirs of arthropod-borne and non–arthropod-borne viruses were predicted equally well (χ^2 test, P = 0.5). Averaging predictions across observations of each virus in models trained on different data subsets (i.e., "bagging") improved prediction of most reservoir groups, such that the reservoirs of 71.9% of all viruses in the study were correctly assigned. GBMs lacking PN or SelGen misclassified the reservoirs of 33 and 22 more viruses, respectively (Fig. 2, B and C).

We trained two additional sets of models that focused on arthropod-borne transmission (*6*). The first nearly perfectly identified which viruses were transmitted by arthropod vectors. Combined GBMs were most accurate overall (bagged accuracy = 97.0%) (Fig. 2D and fig. S11). Only 5 out of 527 viruses were misclassified by all three GBMs (PN, SelGen, and combined), potentially reflecting uncertainty in some currently accepted transmission routes (supplementary text). The second set of models distinguished transmission by all four vector classes (bagged accuracy = 90.8%) (Fig. 2, E and F). Ranking traits according to their predictive power showed that midge and sandfly vectors were identified predominately from genomic biases, whereas mosquito and tick vectors were strongly correlated with viral phylogeny (fig. S12). Accuracy declined by 9.2 and 2.0 percentage points for GBMs lacking SelGen or PN (Fig. 2G). Thus, although phylogeny and genome-wide biases are partially correlated, algorithms successfully exploited independent information in each for all three prediction types.

All models misclassified some currently accepted hosts. We therefore analyzed whether attributes of predictions could help assess their veracity. Predictions with higher GBM probability [bagged prediction strength (BPS)] were correct more often than those diffused across multiple host groups (fig. S13, A to C). Furthermore, when models misclassified viruses, the true host was most often the second-ranked prediction, such that study-wide accuracy for reservoir and vector prediction rose to 81 and 95.9%, respectively, when considering the top two predictions as plausible (Fig. 2, C and G, and fig. S13, D and E). Consequently, BPS provides a confidence metric, such that weaker predictions imply that alternative hosts should be considered in order of their relative support.

We next used our trained models to predict the natural epidemiology of viruses with previously unknown hosts (hereafter "orphan" viruses). As expected from the accuracy of our models on viruses with known hosts, model-projected reservoirs and vectors often matched those suspected from epidemiological investigations (Fig. 3 and

[1]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, Scotland, UK. [2]The Moredun Research Institute, Pentlands Science Park, Penicuik EH26 0PZ, Scotland, UK. [3]MRC-University of Glasgow Centre for Virus Research, Glasgow G61 1QH, Scotland, UK.
*Corresponding author. Email: daniel.streicker@glasgow.ac.uk

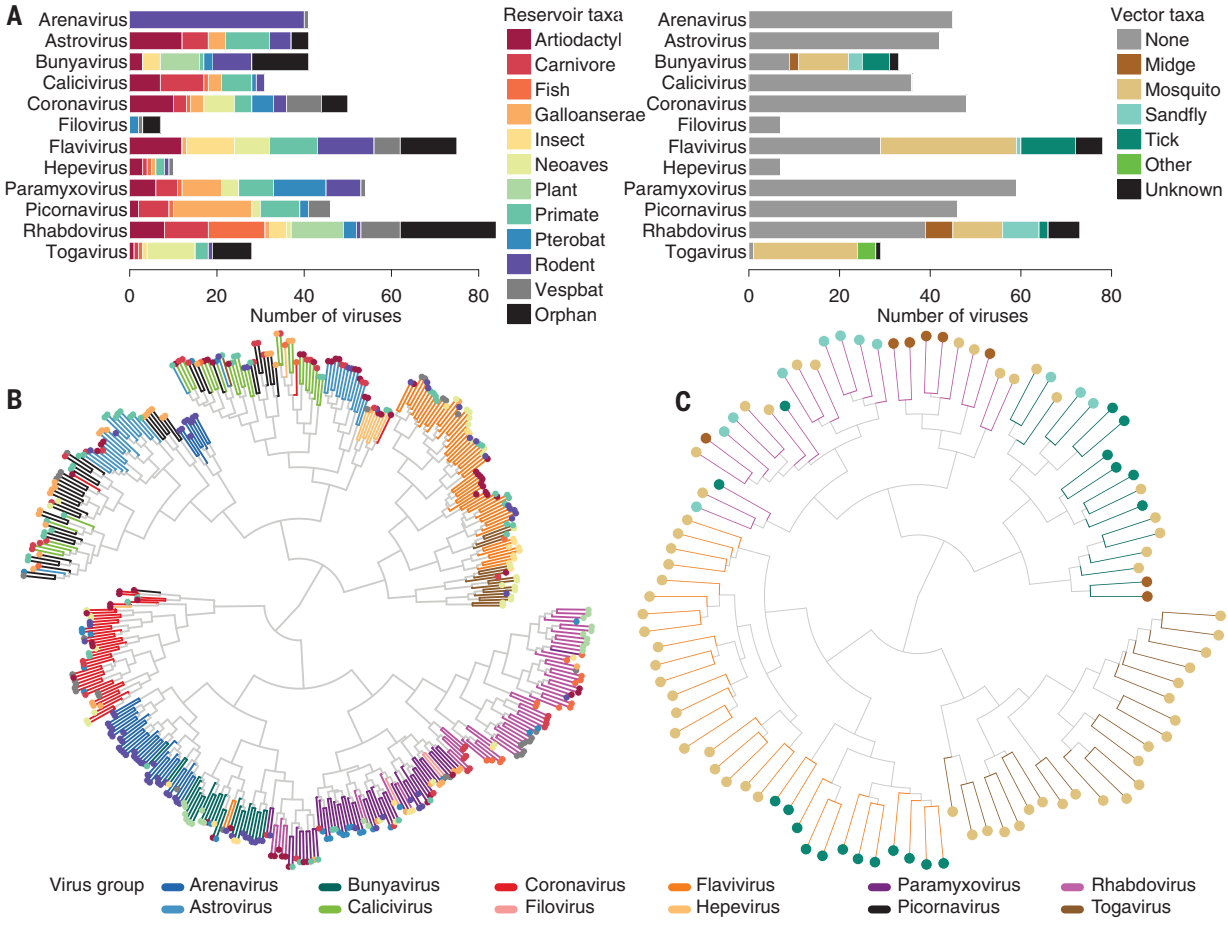


Fig. 1. Distribution and hierarchical clustering of reservoir host and arthropod vector associations across viral taxonomic groups. (**A**) Bar plots show the number of viruses in the dataset from each reservoir host and vector class and the number of orphan viruses in each viral group. The order Artiodactyla (even-toed ungulates) includes the Bovidae, Camelidae, Suidae, Antilocapridae, and Giraffidae families. Galloanserae (ducks, fowl) and Neoaves (most other modern birds) are superorders within the class Aves (birds). (**B** and **C**) Dendrograms of 437 viruses with known reservoir hosts and 98 viruses with known arthropod vectors, estimated by hierarchically clustering 4229 biases calculated from viral genomes. Colors of tip symbols indicate reservoir or vector associations. Branch colors denote viral taxonomic groups. Branch lengths are log(n + 1) transformed for visualization. (B) Trait models with true viral taxonomic group associations were favored over those with randomly shuffled viral groups [change in Akaike information criterion (ΔAIC) = −1690.6] but also clustered significantly by reservoir (ΔAIC = −540.7). (C) Arboviruses clustered by both viral taxonomy (ΔAIC = −238.1) and vector group (ΔAIC = −61.5). ΔAIC values are from models comparing true associations to the mean AIC from 500 tip trait randomizations.

figs. S14 to S16). For example, we predicted an artiodactyl reservoir for human enteric coronavirus 4408, a suspected spillover infection from cows into humans; a primate reservoir of O'nyong-nyong virus, for which humans are the presumed reservoir; and that outbreaks of Tembusu virus in domestic ducks follow cross-species transmission from wild Neoaves (*14*–*16*). Other results pointed to unexpected reservoirs. For example, all four orphan ebolaviruses had greater support for the commonly accepted Pterobat (suborder Pteropodiformes) than for Vespbat (suborder Vespertilioniformes) reservoirs, but surprisingly, Bundibugyo and Tai Forest ebolaviruses had equal or stronger support for primate reservoirs. This indicates that signals learned from primate viruses from divergent viral families occurred in these ebolavirus genomes. Neither species of ebolavirus has been detected in bats (*17*), and the slow evolution of genomic biases in filoviruses implied that the observed signal could not have evolved during short chains of transmission in primates (fig. S17). The possibility of an undiscovered primate ebolavirus reservoir therefore deserves empirical validation. For viruses without conjectured reservoirs or vectors, we generate candidates for prioritized surveillance. For example, Bas-Congo virus caused an outbreak of hemorrhagic fever in the Democratic Republic of the Congo and was detected in humans only (*18*). Our models predicted an artiodactyl reservoir, a high probability of arthropod-borne transmission, and midges as the likely vector of this emerging disease (Fig. 3, A and C). Such predictions may ultimately support earlier interventions targeting appropriate reservoirs or vectors that interrupt the critical early phases of outbreaks or limit future reemergence. Likewise, our models can provide ecological insights for virus discovery programs (Fig. 3B).

By virtue of using slowly evolving biases spread across viral genomes, our models predict taxa that maintain long-term viral circulation rather than bridge hosts that sustain insufficient chains of transmission to imprint evolutionary signals in viral genomes (e.g., pig hosts of bat-borne Nipah virus). Similarly, sustained transmission by divergent hosts may create conflicting signals that obscure model predictions (supplementary text). Finally, models predict only the reservoir and vector groups used for training and will erroneously assign a host from these same categories if applied to viruses from host groups that were too rare to include (fig. S18). As virus discoveries expand databases, evaluating predictive accuracy for additional host groups will be an important improvement.

In summary, we created a machine learning framework that leverages traits from individual viruses with network-derived information from

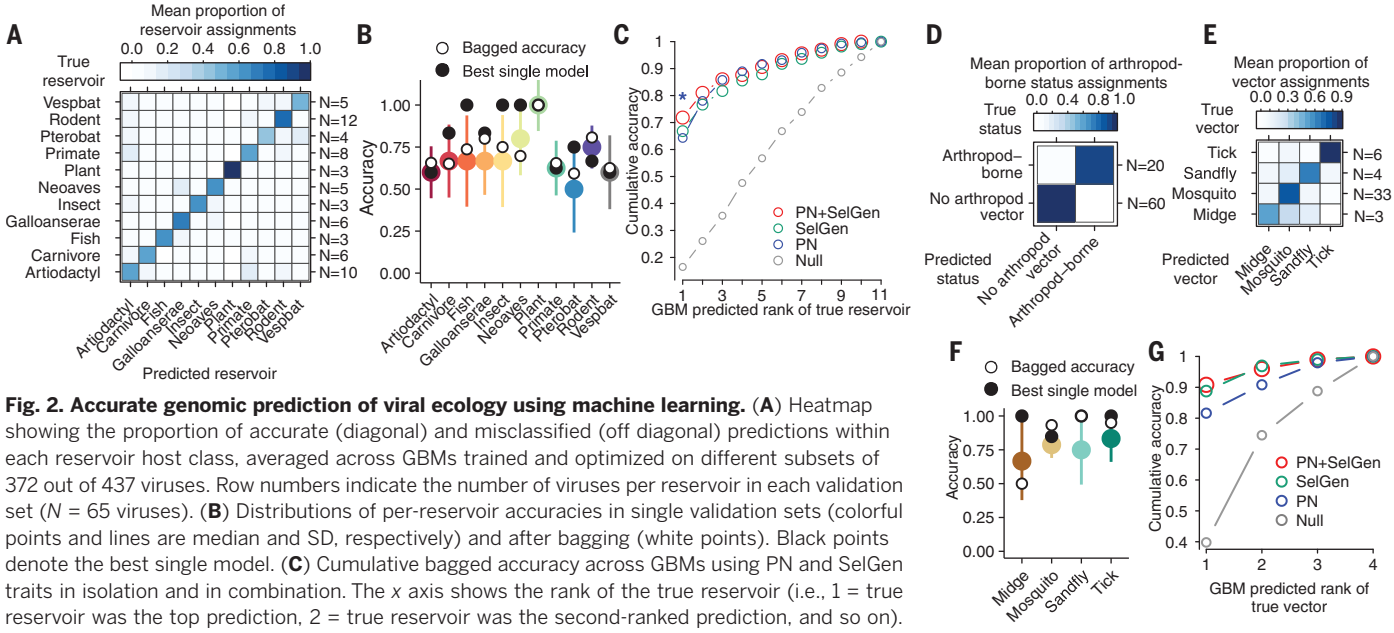


Fig. 2. Accurate genomic prediction of viral ecology using machine learning. (**A**) Heatmap showing the proportion of accurate (diagonal) and misclassified (off diagonal) predictions within each reservoir host class, averaged across GBMs trained and optimized on different subsets of 372 out of 437 viruses. Row numbers indicate the number of viruses per reservoir in each validation set (N = 65 viruses). (**B**) Distributions of per-reservoir accuracies in single validation sets (colorful points and lines are median and SD, respectively) and after bagging (white points). Black points denote the best single model. (**C**) Cumulative bagged accuracy across GBMs using PN and SelGen traits in isolation and in combination. The x axis shows the rank of the true reservoir (i.e., 1 = true reservoir was the top prediction, 2 = true reservoir was the second-ranked prediction, and so on). The y axis shows accuracy when considering increasing numbers of predictions as plausible. The asterisk indicates significantly higher accuracy in the combined model (χ^2 test: $P < 0.05$). Cumulative null model accuracy was estimated by training GBMs on 50 randomly generated traits that were simulated from normal distributions ranging from 0 to 2 and randomly assigned to viruses. (**D** and **E**) Heatmaps showing the average proportion of accurate predictions of arthropod-borne status and vector identity (N = 80 and 46 viruses per validation set, respectively). (**F**) Distributions of per-vector accuracies as in (B). (**G**) Cumulative bagged accuracy in vector prediction across models as in (C).

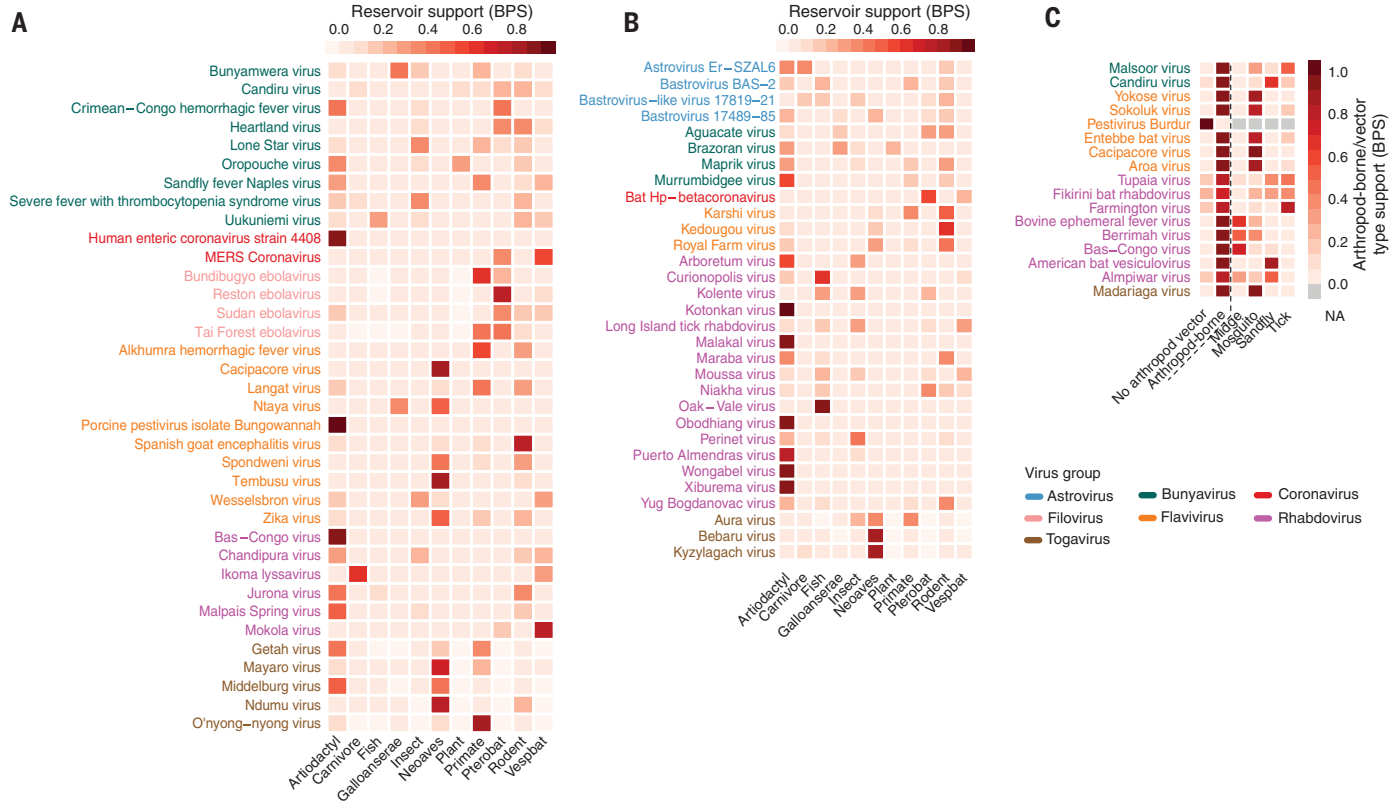


Fig. 3. Reservoir hosts and arthropod vectors of orphan viruses predicted from their genome sequences. (**A**) Predicted reservoirs for 36 viruses that emerged from unknown sources. (**B**) Thirty-one viruses discovered by active surveillance of wildlife or blood-feeding arthropods. (**C**) Predictions of arthropod-borne status for 17 viruses (left of dashed line) and vector identities (rightmost four columns, when applicable). Color gradients show the BPS (bagged prediction strength) for each class from the top 25% of models from each set of GBMs. Figs. S14 to S16 show the full probability distributions of predictions.

their relatives to predict: (i) the reservoir hosts of 12 key groups of RNA viruses, (ii) whether their transmission involves an arthropod vector, and (iii) the identity of that vector. Our models make these predictions, supply quantitative measures of confidence, and provide relative support for alternatives from single genome sequences, with no requirement for experiments, longitudinal surveillance, or genomes of candidate reservoirs or vectors. As viral genomes are now produced within hours of detection (*19*), algorithms that rapidly generate field-testable hypotheses from sequence data narrow the gap between virus discovery and actionable understanding of virus ecology.

REFERENCES AND NOTES

1. M. Viana *et al*., *Trends Ecol. Evol.* **29**, 270–279 (2014).
2. M. Woolhouse, E. Gaunt, *Crit. Rev. Microbiol.* **33**, 231–242 (2007).
3. R. R. Kao, D. T. Haydon, S. J. Lycett, P. R. Murcia, *Trends Microbiol.* **22**, 282–291 (2014).
4. K. J. Olival *et al*., *Nature* **546**, 646–650 (2017).
5. J. L. Geoghegan, S. Duchêne, E. C. Holmes, *PLOS Pathog.* **13**, e1006215 (2017).
6. Materials and methods are available as supplementary materials.
7. M. A. Martínez, A. Jordan-Paiz, S. Franco, M. Nevot, *Trends Microbiol.* **24**, 134–147 (2016).
8. B. D. Greenbaum, A. J. Levine, G. Bhanot, R. Rabadan, *PLOS Pathog.* **4**, e1000079 (2008).
9. F. P. Lobo *et al*., *PLOS ONE* **4**, e6282 (2009).
10. A. Kapoor, P. Simmonds, W. I. Lipkin, S. Zaidi, E. Delwart, *J. Virol.* **84**, 10322–10328 (2010).
11. F. Di Giallonardo, T. E. Schlub, M. Shi, E. C. Holmes, *J. Virol.* **91**, e02381-16 (2017).
12. M. K. K. Leung, A. Delong, B. Alipanahi, B. J. Frey, *Proc. IEEE* **104**, 176–197 (2016).
13. B. J. H. Friedman, *Ann. Stat.* **29**, 1189–1232 (2001).
14. M. G. Han, D.-S. Cheon, X. Zhang, L. J. Saif, *J. Virol.* **80**, 12350–12356 (2006).
15. A. D. LaBeaud *et al*., *PLOS Negl. Trop. Dis.* **9**, e0003436 (2015).
16. Y. Tang *et al*., *Transbound. Emerg. Dis.* **60**, 152–158 (2013).
17. K. J. Olival, D. T. S. Hayman, *Viruses* **6**, 1759–1788 (2014).
18. G. Grard *et al*., *PLOS Pathog.* **8**, e1002924 (2012).
19. J. Quick *et al*., *Nature* **530**, 228–232 (2016).

ACKNOWLEDGMENTS

We thank R. Biek, B. Mable, M. Viana, D. Haydon, P. Johnson, S. Altizer, B. Brennan, A. Szemiel, M. Palmarini, and three anonymous reviewers for helpful feedback. **Funding:** S.A.B. was supported by a Glasgow University Research Fellowship and the BBSRC (BB/M012956/1). D.G.S. was supported by a Sir Henry Dale Fellowship, jointly funded by the Wellcome Trust and Royal Society (102507/Z/13/Z). Additional funding was provided from the Medical Research Council (MC_UU_12014/12). **Author contributions:** D.G.S. conceived of the research; D.G.S. and R.J.O. collected the data; D.G.S., R.J.O., and S.A.B. analyzed the data; and D.G.S. and S.A.B. wrote and revised the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and code reported in this paper are available at https://github.com/DanielStreicker/ViralHostPredictor.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/577/suppl/DC1
Materials and Methods
Supporting Text
Figs. S1 to S18
References (*20*–*43*)
Data S1
Appendix S1

14 September 2017; resubmitted 21 May 2018
Accepted 12 September 2018
10.1126/science.aap9072

PALEOCLIMATOLOGY

East Asian hydroclimate modulated by the position of the westerlies during Termination I

Hongbin Zhang[1,2], Michael L. Griffiths[3]*, John C. H. Chiang[4], Wenwen Kong[4], Shitou Wu[5,6], Alyssa Atwood[4,7], Junhua Huang[2], Hai Cheng[8,9], Youfeng Ning[8], Shucheng Xie[1,10]*

Speleothem oxygen isotope records have revolutionized our understanding of the paleo East Asian monsoon, yet there is fundamental disagreement on what they represent in terms of the hydroclimate changes. We report a multiproxy speleothem record of monsoon evolution during the last deglaciation from the middle Yangtze region, which indicates a wetter central eastern China during North Atlantic cooling episodes, despite the oxygen isotopic record suggesting a weaker monsoon. We show that this apparent contradiction can be resolved if the changes are interpreted as a lengthening of the Meiyu rains and shortened post-Meiyu stage, in accordance with a recent hypothesis. Model simulations support this interpretation and further reveal the role of the westerlies in communicating the North Atlantic influence to the East Asian climate.

Speleothem oxygen isotope records ($\delta^{18}O_c$) from China have revealed the most detailed insights into past Asian summer monsoon variability of any climate proxy to date (*1*–*3*). A recent study by Cheng *et al.* (*3*) was the first to extend the Chinese cave records back to the dating limits of uranium-thorium (~640,000 years), providing critical new insight into the nature and cause(s) of Asian summer monsoon variability, and importantly, its connection with glacial terminations. Moreover, numerous studies have also shown a strong link between $\delta^{18}O_c$ and millennial-scale perturbations of the last glacial and deglaciation, such as Heinrich events, characterized by large freshwater discharges into the North Atlantic that corresponded with higher $\delta^{18}O_c$ in Chinese speleothems, interpreted as a weakening of the monsoon (*1*, *4*).

However, we still lack critical understanding of what exactly $\delta^{18}O_c$ represents in terms of large-scale shifts in Asian summer monsoon circulation, and more "local-scale" changes in precipitation. The prevailing literature presents a number of competing interpretations. The traditional interpretation links $\delta^{18}O_c$ as a proxy for East Asian summer monsoon (EASM) "strength" or "intensity," defined in terms of large seasonal differences in precipitation oxygen isotopes ($\delta^{18}O_p$), where higher EASM rainfall corresponds with more depleted $\delta^{18}O_c$ (*1*, *2*). By contrast, others have interpreted $\delta^{18}O_c$ as reflecting changes in the fraction of water vapor rained out along the moisture trajectory between tropical sources and the cave site (*4*, *5*); subsequent isotope-enabled general circulation model (GCM) simulations support this latter interpretation, although they emphasized the "upstream depletion" rather than "local depletion" at the cave site (*6*, *7*). More recent interpretations attempt to reconcile these two end-members, calling for $\delta^{18}O_c$ to represent EASM intensity via changes in the strength of the southerly monsoon flow and associated heterogeneous expression of EASM rainfall, with lower $\delta^{18}O_c$ implying higher upstream rainout integrated between different moisture sources en route to the cave sites, but not local rainfall (*3*, *8*).

The second challenge arises from consideration of the observed seasonal evolution of the EASM (*9*, *10*). Unlike other monsoons that have only one rainfall stage, the seasonal rainfall evolution over East Asia undergoes a number of quasi-stationary intraseasonal stages with abrupt transitions in between (*10*) (Fig. 1A). During spring, persistent rainfall in southern China is followed by substantial convection over the South China Sea during the pre-Meiyu stage in mid-May. By mid-June, the Meiyu begins and rainfall shifts to central China, and around mid-July, the rain band shifts again to be located over northeast China, marking the onset of the midsummer stage, which terminates around mid-August. Moreover, dynamical arguments link these stages to the configuration of the westerlies impinging on the Tibetan Plateau; as the peak westerlies migrate from the south to the north of the Plateau and off it, the downstream circulation changes and thus determines

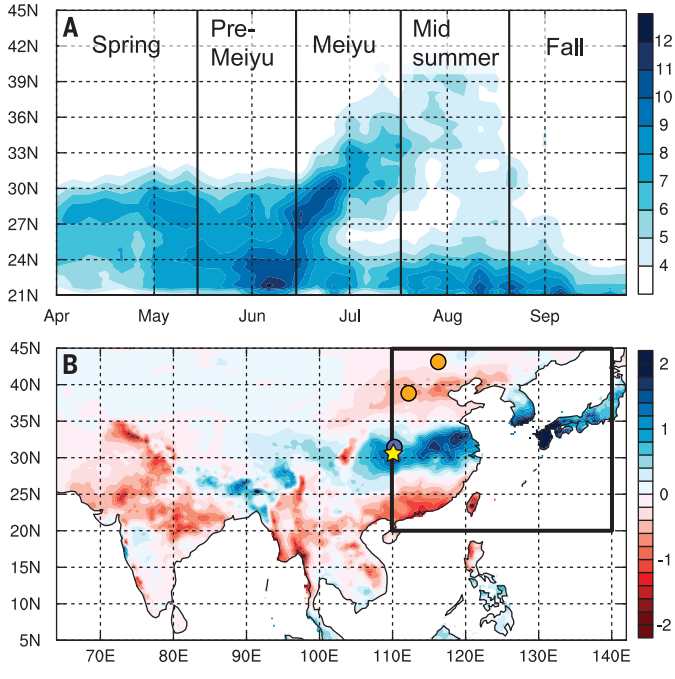


Fig. 1. Cave location and climatology of East Asia. (**A**) Hovmöller diagram of the precipitation climatology (unit: mm/day) over eastern China (110° to 120°E). Approximate timings of the various stages are shown. (**B**) Spatial pattern of the dominant tripole mode of interannual variability in rainfall over East Asia. Chiang *et al.* (*27*) showed the first empirical orthogonal function (EOF) and the normalized first principle component (PC1) of July-August mean precipitation over East Asia (110° to 145°E, 20° to 45°N) [area indicated by the black box in (B); see figures 1, b and c, of Chiang *et al.* (*27*)]. Here we show the regression of the PC1 of Chiang *et al.* (*27*) onto the July-August rainfall across Asia. We use the APHRODITE dataset spanning 1951 to 2007. Units in (B) are mm/day per standard deviation. The yellow star denotes Haozhu Cave (this study), the blue circle shows Dajiuhu peatland (*23*), and orange circles represent Gonghai Lake (*24*) and Dali Lake (*25*) as plotted in Fig. 2.

[1]State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, 430074, China. [2]State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan, 430074, China. [3]Department of Environmental Science, William Paterson University, Wayne, NJ 07470, USA. [4]Department of Geography and Berkeley Atmospheric Sciences Center, University of California, Berkeley, CA 94720, USA. [5]Department of Geochemistry, Faculty of Geoscience and Geography, Georg-August University Goettingen, Goettingen 37077, Germany. [6]Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, 100029, China. [7]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA. [8]Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an, 710049, China. [9]Department of Earth Science, University of Minnesota, Minneapolis, MN 55455, USA. [10]Department of Geography, School of Earth Sciences, China University of Geosciences, Wuhan, 430074, China.
***Corresponding author. Email: xiecug@163.com (S.X.); griffithsm@wpunj.edu (M.L.G.)**

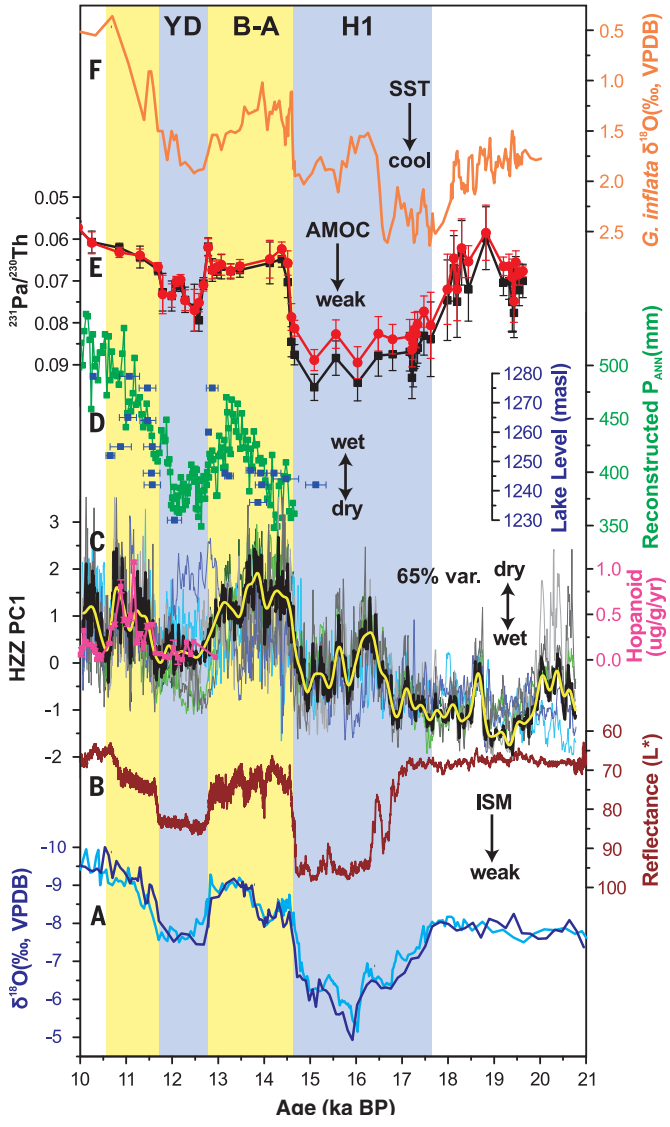


Fig. 2. Central China hydroclimate and high northern latitude teleconnections during the last deglaciation. (**A**) $\delta^{18}O$ records of HZZ11 and HZZ27 (*14*). VPDB, Vienna Pee Dee belemnite. (**B**) Arabian Sea sediment total reflectance (L*), a proxy for Indian summer monsoon (ISM) strength (*18*). (**C**) Leading PC (HZZ PC1; black curve) of HZZ11 and HZZ27 Mg/Ca, Sr/Ca, and Ba/Ca records and hopanoid flux (pink curve), a proxy for water-table depth from nearby Dajiuhu peatland (*23*). Overlapped thin lines show the standardized Sr/Ca, Mg/Ca, and Ba/Ca records. Yellow line shows HZZ PC1 smoothed with a 100-year loess filter. (**D**) Reconstructed annual precipitation (green) and lake level (blue) in North China from Gonghai Lake (*24*) and Dali Lake (*25*), respectively. (**E** and **F**) North Atlantic sediment core GGC5 $^{231}Pa/^{230}Th$ (proxy for AMOC) and planktonic foraminifera $\delta^{18}O$ [proxy for sea surface temperature (SST)] (*22*).

the various intraseasonal stages (*11*–*13*). In other words, the EASM undergoes a complex spatiotemporal evolution, and so it is unclear what "intensity" means in this context.

To address these uncertainties, here we present a new multiproxy record of EASM variability from two absolute-dated speleothems from central eastern China (*14*), which extends through the last deglaciation [~10 to 21 thousand years (ka) before the present (B.P.)]. The two specimens (HZZ11 and HZZ27) were recovered from Haozhu Cave (109°59′ E, 30°41′ N, 1017 m), located in the Hubei Province of central China (Fig. 1B). We examined both $\delta^{18}O_c$ and trace elements (Mg^{2+}, Sr^{2+}, Ba^{2+}) in two speleothems to help deconvolve large-scale EASM circulation [inferred from $\delta^{18}O_c$ (*3*, *8*)] from local hydrology (inferred from the trace elements) during Termination I (supplementary materials). The isotopic and elemental profiles were anchored by 22 U-Th dates by using linear interpolation between each successive date and further refined by aligning major $\delta^{18}O_c$ excursions within their respective age uncertainties (fig. S4). Although only representing a minor contribution to speleothem paleoclimate studies (relative to $\delta^{18}O_c$), research has increasingly shown that trace elements can faithfully record environmental (particularly hydrological) changes operating in and between the cave interior and soil zone (*15*). Hence, these elemental ratios reflect more local hydrological changes happening at the site of deposition (*16*, *17*), and thus when paired with oxygen isotopes, can add valuable information to proxy interpretations.

Results show that the HZZ11 and HZZ27 $\delta^{18}O_c$ time series largely parallel those from other distal cave sites in China and India, showing enriched $\delta^{18}O_c$ values during North Atlantic cooling events [i.e., Heinrich event 1 (H1) and the Younger Dryas (YD)] and more depleted $\delta^{18}O_c$ values during the Bølling-Allerød (B-A) interstadial (Fig. 2A and fig. S2). In addition, these millennial-scale $\delta^{18}O_c$ fluctuations are also consistent with a record of terrigenous sediment flux (*18*) (Fig. 2B)—a proxy for the Indian summer monsoon—into the Arabian Sea. Recent hydroclimate reconstructions of the tropical Indian Ocean, combined with model simulations, demonstrate that a weakened Indian monsoon circulation during North Atlantic stadials was likely caused by a reorganization of the Hadley circulation with a southward shift of the ascending branch—the intertropical convergence zone (ITCZ) (*19*). Taken together, these records are consistent with the interpretation that North Atlantic stadial events led to an overall weakening of the Asian summer monsoon via declines in hemispheric surface temperatures and subsequent meridional displacements of the ITCZ (*1*, *8*, *18*, *19*).

Our speleothem trace-element results, however, offer a more complex story. Indeed, analyses of Sr-Mg-Ba/Ca (hereafter X/Ca) ratios in both speleothems reveal that during H1 and YD, X/Ca values decreased significantly, whereas during the B-A interstadial, values abruptly increased (Fig. 2C and fig. S6). Significant correlations amongst the various trace element–to–calcium ratios (fig. S7) suggest that similar processes controlled their variability, which can be summarized as follows: Higher values indicate overall drier conditions when reduced infiltration within the karst fracture network favors CO_2 degassing and prior calcite precipitation (PCP) in the epikarst, leaving the cave drip-water, and hence speleothems, enriched in the trace elements relative to calcium; the opposite occurs during wetter periods (*15*, *20*). However, drier conditions may have prolonged water-rock interaction (WRI) in the aquifer, which would have favored selective leaching of Mg and Sr from the dolomite host rock (*21*), causing enrichment in Mg and Sr (fig. S8) (see supplementary text for more details). Nonetheless, whichever mechanism dominated, it is apparent that drier (wetter) intervals would have favored higher (lower) X/Ca values in HZZ11 and HZZ27 during the deglaciation.

To effectively combine the X/Ca records for both speleothems into one composite hydroclimate record, we calculated the leading principal component ("HZZ PC1") on the six trace-element time series, where lower values indicate wetter conditions and higher values, drier conditions. Examination of HZZ PC1 reveals that the region became wetter during the H1 and YD stadials (Fig. 2C), when the abrupt input of fresh water to the North Atlantic led to declines in the Atlantic Meridional Overturning Circulation (AMOC) (Fig. 2E) and resultant lowering of SSTs (*22*) (Fig. 2F). This finding is compelling because it shows that, despite a weakened Asian summer monsoon (Fig. 2, A and B), central China was in fact wetter during these stadial events, while similar to southern Asia, northern China was drier (*8*, *23*–*25*) (Fig. 2, C and D). The HZZ PC1 results also suggest that central eastern China was wetter during the Last Glacial Maximum (LGM; ~19 ka B.P.) and experienced an overall drying trend until the early Holocene; this trend is supported by speleothem $\delta^{13}C$ and initial uranium isotope ratios from the same specimens, which we interpret to reflect local hydrology (figs. S5 and S6 and supplementary text).

At face value, these phase relationships in the EASM region suggest antiphasing of rainfall between

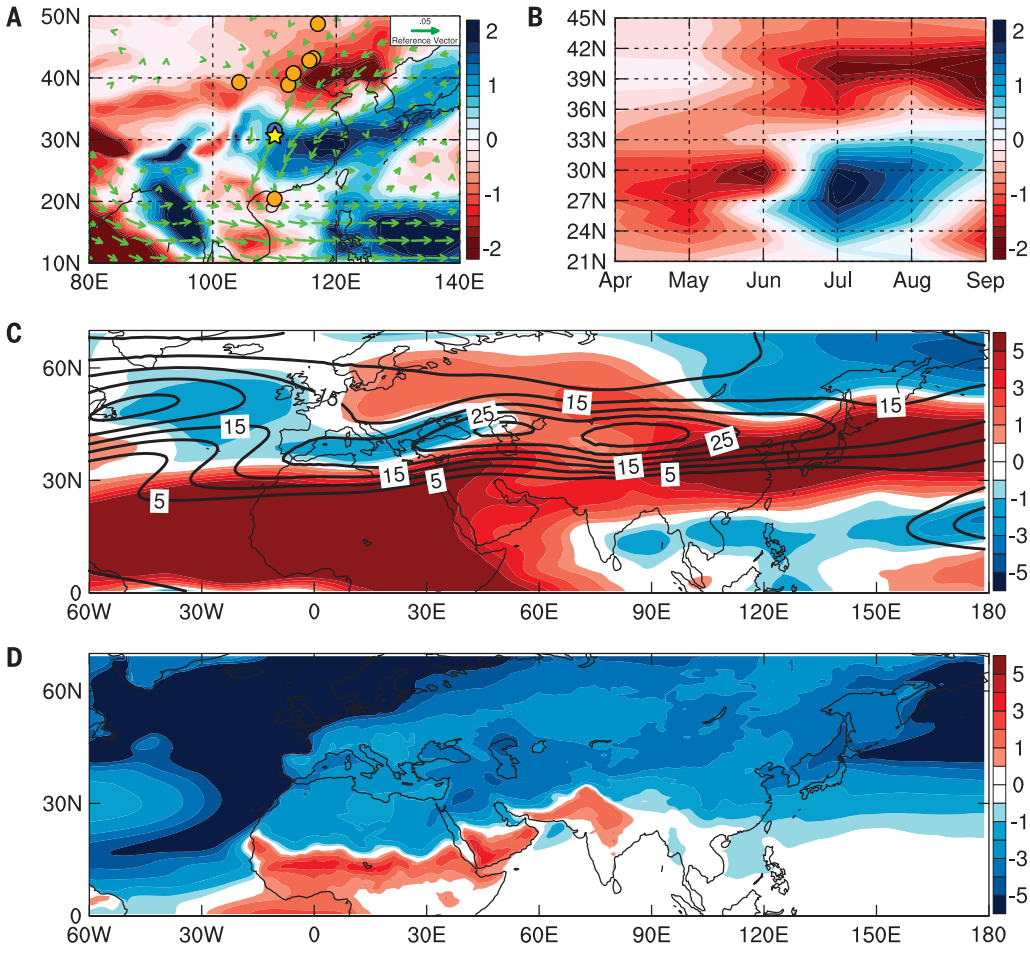


Fig. 3. CESM 1.0.5 simulated precipitation and Asian westerlies anomalies between 1-Sv hosing experiment and control simulation. (**A**) July-August rainfall anomalies in East Asia (shading; unit: mm/day) superimposed with moisture flux anomalies at 850 mb (vectors; unit: kg/kg·m/s). The yellow star and blue circle show sites [Haozhu cave (star, this study) and Dajiuhu peatland (circle, (23)] indicating wet conditions during H1 and/or YD, and orange circles show sites indicating dry conditions during H1 and/or YD (see fig. S1 for list of sites). (**B**) Hovmöller diagram of rainfall anomalies zonally averaged between 110° and 120°E (unit: mm/day). This shows the increase in rainfall over central eastern China in July-August and decreased rainfall to the north. (**C**) July-August zonal wind at 200 mb from the control run (contour; interval 5 m/s) and anomalies from the 1-Sv hosing run (shading), showing the equatorward displacement of the jet in the hosing simulation. (**D**) July-August air-temperature anomalies at 2 m above ground (unit: K). One hundred years climatology of the control run and 80 years climatology of the hosing run after spin-up were used to generate the anomalies.

central and northeastern China. Liu *et al.* (*8*) first noted an inhomogeneity in the simulated East Asian rainfall response to climate forcing over the last 21,000 years, specifically tying a negative $\delta^{18}O_p$ signal over China with enhanced monsoon rainfall restricted to northeastern China. However, the antiphased relationship has a more direct analog to the observed modern interannual variability. The dominant mode of summer rainfall variability over East Asia is characterized by a meridional "tripole" pattern of precipitation, with increased rainfall over northeastern and southeastern China and decreased rainfall over central eastern China and Japan (*12*, *26*, *27*) (Fig. 1B). Moreover, this pattern is also known to covary with strengthened Indian summer monsoon rainfall in South Asia (*28*). Hence, this modern-day analog appears relevant to understanding the East Asian rainfall response during the H1 and YD events. However, the question remains why East Asian climate responds this way during North Atlantic stadials. To help answer this, we analyzed a freshwater "hosing" simulation using the Community Earth System Model (CESM) version 1.0.5 of the National Center for Atmospheric Research (NCAR). Specifically, the control run was perturbed with 1 Sv (1 Sv = 10^6 m^3 s^{-1}) of freshwater injection into the North Atlantic to simulate a rapid shutdown of the AMOC akin to H1 and the YD.

Our model results are consistent with HZZ PC1 and other proxy records (*8*, *23*–*25*), which show an increase in summer rainfall in central eastern China and a decrease in northern China during stadial events of the deglaciation (Fig. 3A). A recent hypothesis—the "jet transition hypothesis" (*10*)—argues that rainfall changes over East Asia arise through changes in the transition timing and duration of the EASM intraseasonal stages (i.e., spring, pre-Meiyu, Meiyu, and midsummer), which are in turn tied to the south-north displacement of the westerlies relative to the Tibetan Plateau. In particular, the tripole pattern of the interannual variability arises from variation in the termination date of Meiyu; an earlier termination shortens the Meiyu stage and lengthens the midsummer stage, leading to less rain over central eastern China and more over northeastern China. A paleoclimate modeling study found support for this hypothesis in precession-driven EASM changes during the Holocene (*29*). More precisely, the simulated decreased pole-equator temperature gradient during the early Holocene resulted in an earlier northward positioning of the westerlies relative to the Tibetan Plateau, leading to an earlier termination of the Meiyu stage and prolonged midsummer stage. The hypothesis also provided a simple and physically consistent explanation for the "the asynchronous Holocene optimum" (*30*), where the Holocene peak in East Asian hydroclimate was observed to occur earlier for northeastern China but later for central and southeastern China.

We invoke a similar mechanism to explain the observed antiphasing of hydroclimate across China during H1 and the YD. Indeed, the simulated seasonal evolution of rainfall over central eastern China shows the distinct characteristics of a longer Meiyu and shortened midsummer stage (Fig. 3B), with increased rainfall over central eastern China and reduced rainfall over northeastern China during July and August (note that climatological midsummer stage is between mid-July through end of August). Moreover, the westerlies during July and August were displaced southward (Fig. 3C), consistent with a delayed northward migration. This delayed migration occurs because of the colder conditions in the northern extratropics (Fig. 3D) as a result of the AMOC slowdown, which increases the equator-to-pole temperature gradient. Thus, the jet transition hypothesis (*10*) provides a simple and consistent explanation for both the observed antiphasing of rainfall between north and central eastern China and the simulated large-scale response of the East Asian hydroclimate to AMOC slowdown representative of the YD and H1. It

is also possible that this hypothesis, particularly as it relates to shifts in the equator-to-pole temperature gradient, can explain the wetter conditions in central eastern China during the LGM, although this will need to be further tested with additional proxy records and model simulations.

Critically, our results can be combined with previous isotope-enabled model results to provide a more comprehensive view of the paleo EASM system. In particular, Liu *et al.* (*8*) find model evidence that speleothem $\delta^{18}O_c$ reflects changes to the strength of the southerly monsoonal flow over East Asia—specifically, heavier $\delta^{18}O_c$ values indicate reduced monsoonal flow that translates to reduced rainfall over northeastern China. Our simulations do indeed show weaker southerlies and reduced northeastern China moisture transport with a reduced AMOC (Fig. 3A and fig. S9), thus providing a link between circulation changes that influence $\delta^{18}O_p$ and rainfall variability inferred in our simulations. Indeed, as elaborated in Chiang *et al.* (*10*), the meridional jet position is seen to limit the northward extent of the southerly monsoon flow penetrating into East Asia (figs. S9 and S10), with the northward migration of the jet allowing for a more northward penetration of the monsoon flow. This view nicely integrates the jet transition hypothesis with the hypothesis provided by Liu *et al.* (*8*) (i.e., the "moisture transport hypothesis") and provides the link between our results and $\delta^{18}O_c$ in lieu of our own isotope-enabled model simulation.

The results of this study support the strong teleconnections between East Asia and the North Atlantic during the deglaciation, but notably, lend credence to the jet transition hypothesis (*10*). Our results also suggest that a stronger meridional temperature gradient [i.e., during stadial events (Fig. 3D)] and resultant southward-shifted and strengthened westerlies (Fig. 3C and fig. S10) can lead to increased convection and uplift of moisture on the leeward side of the Tibetan Plateau during the summer months in central eastern China. This pattern, invoking wetter conditions to the south but drier conditions to the north, calls for a reevaluation of the regional heterogeneity of East Asian hydroclimate during abrupt climate perturbations of the last glacial cycle (*8*). These findings highlight the potential for abrupt and prolonged redistributions of moisture across East Asia under varying boundary conditions, which has important implications for future projections of rainfall in this densely populated region, particularly in light of the "north drought–south flood" pattern that has persisted in China since the 1970s (*31*).

REFERENCES AND NOTES

1. Y. J. Wang *et al.*, *Science* **294**, 2345–2348 (2001).
2. H. Cheng *et al.*, *Science* **326**, 248–252 (2009).
3. H. Cheng *et al.*, *Nature* **534**, 640–646 (2016).
4. D. Yuan *et al.*, *Science* **304**, 575–578 (2004).
5. C. Y. Hu *et al.*, *Earth Planet. Sci. Lett.* **266**, 221–232 (2008).
6. A. LeGrande, G. Schmidt, *Clim. Past* **5**, 441–455 (2009).
7. F. S. R. Pausata, D. S. Battisti, K. H. Nisancioglu, C. M. Bitz, *Nat. Geosci.* **4**, 474–480 (2011).
8. Z. Y. Liu *et al.*, *Quat. Sci. Rev.* **83**, 115–128 (2014).
9. Y. Ding, J. C. L. Chan, *Meteorol. Atmos. Phys.* **89**, 117–142 (2005).
10. J. C. Chiang *et al.*, *Quat. Sci. Rev.* **108**, 111–129 (2015).
11. T. C. Yeh, S. J. Dao, M. T. Li, [The abrupt change of circulation over the Northern Hemisphere during June and October] in *The Atmosphere and the Sea in motion*, B. B., Ed. (Rockefeller Inst. Press, New York, 1959), pp. 249-267.
12. H. H. Hsu, X. Liu, *Geophys. Res. Lett.* **30**, 1182–1200 (2003).
13. P. Molnar, W. R. Boos, D. S. Battisti, *Annu. Rev. Earth Planet. Sci.* **38**, 77–102 (2010).
14. H. Zhang *et al.*, *Earth Planet. Sci. Lett.* **453**, 243–251 (2016).
15. I. J. Fairchild, P. C. Treble, *Quat. Sci. Rev.* **28**, 449–468 (2009).
16. Y. H. Liu *et al.*, *Nat. Geosci.* **6**, 117–120 (2013).
17. M. L. Griffiths *et al.*, *Nat. Commun.* **7**, 11719 (2016).
18. G. Deplazes *et al.*, *Nat. Geosci.* **6**, 213–217 (2013).
19. M. Mohtadi *et al.*, *Nature* **509**, 76–80 (2014).
20. K. R. Johnson, C. Y. Hu, N. S. Belshaw, G. M. Henderson, *Earth Planet. Sci. Lett.* **244**, 394–407 (2006).
21. I. J. Fairchild *et al.*, *Chem. Geol.* **166**, 255–269 (2000).
22. J. F. McManus, R. Francois, J. M. Gherardi, L. D. Keigwin, S. Brown-Leger, *Nature* **428**, 834–837 (2004).
23. S. C. Xie *et al.*, *Geology* **41**, 827–830 (2013).
24. F. Chen *et al.*, *Sci. Rep.* **5**, 11186 (2015).
25. Y. Goldsmith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1817–1821 (2017).
26. H. H. Hsu, S. M. Lin, *J. Clim.* **20**, 4443–4458 (2007).
27. J. C. H. Chiang, L. M. Swenson, W. Kong, *Geophys. Res. Lett.* **44**, 3788–3795 (2017).
28. J. A. Day, I. Fung, C. Risi, *J. Clim.* **28**, 4330–4356 (2015).
29. W. Kong, L. M. Swenson, J. C. Chiang, *J. Clim.* **30**, 3343–3365 (2017).
30. Z. An *et al.*, *Quat. Sci. Rev.* **19**, 743–762 (2000).
31. R. Yu, B. Wang, T. Zhou, *Geophys. Res. Lett.* **31**, 217–244 (2004).

ACKNOWLEDGMENTS

We gratefully acknowledge two anonymous reviewers for their constructive comments and C. Hu for the personal communication about cave monitoring work in Heshang Cave. We thank D. S. Battisti and C. M. Bitz for their assistance in developing the model simulations. **Funding:** This work was supported by the Key R&D Project of Ministry of Science and Technology (grant no. 2016YFA0601100), the Chinese National Natural Science Foundation (grants 41330103 and 41773135), the 111 Program (National Bureau for Foreign Experts and Ministry of Education of China; grant B08030), and the U.S. National Science Foundation (grant AGS-1405479). **Author contributions:** H.Z., M.L.G., S.X., and J.H. designed the project. H.Z. and J.H. performed the fieldwork and sampling. H.Z., H.C., and Y.N. carried out the uranium-thorium dating. H.Z. and J.H. contributed to the oxygen-isotope measurements. S.W. carried out the trace element measurements. J.C.H.C. and W.K. analyzed the modeling simulations, which were conducted by A.A. H.Z., M.L.G., S.X., J.C.H.C., and W.K. wrote the manuscript with contributions from other authors. All authors discussed the results and implications and commented on the manuscript at all stages. **Competing interests:** All authors declare no competing interests. **Data and materials availability:** All data reported in this paper are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/580/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
References (*32*–*75*)

20 April 2018; accepted 25 September 2018
10.1126/science.aat9393

NEUROSCIENCE

Efficient cortical coding of 3D posture in freely behaving rats

Bartul Mimica*†, Benjamin A. Dunn*, Tuce Tombaz, V. P. T. N. C. Srikanth Bojja, Jonathan R. Whitlock†

Animals constantly update their body posture to meet behavioral demands, but little is known about the neural signals on which this depends. We therefore tracked freely foraging rats in three dimensions while recording from the posterior parietal cortex (PPC) and the frontal motor cortex (M2), areas critical for movement planning and navigation. Both regions showed strong tuning to posture of the head, neck, and back, but signals for movement were much less dominant. Head and back representations were organized topographically across the PPC and M2, and more neurons represented postures that occurred less often. Simultaneous recordings across areas were sufficiently robust to decode ongoing behavior and showed that spiking in the PPC tended to precede that in M2. Both the PPC and M2 strongly represent posture by using a spatially organized, energetically efficient population code.

More than a century of clinical observations have implicated the posterior parietal cortex (PPC) and related networks as essential for maintaining awareness of the spatial configuration of the body, or "body schema" (*1*, *2*). Consistent with this notion, neurophysiological investigations in head-fixed subjects have identified key roles for the PPC and frontal motor cortices in controlling the positioning of individual effectors, such as the eye, arm, or hand (*3*–*10*). Parallel studies in rodents have demonstrated ostensibly similar functions for the PPC and frontal motor cortex region M2 in spatial orienting (*11*), movement planning (*12*–*14*), and navigation (*15*–*17*), but the field still lacks a quantitative understanding of how the cortex represents posture in freely behaving individuals.

We therefore tracked the heads and backs of 11 rats in three dimensions while recording neural ensembles with dual microdrives targeting deep (>500 μm) layers of the PPC and M2, which exhibit thalamic, cortical, and subcortical connections similar to those of the PPC and premotor areas across mammals (*16*–*18*). We recorded 729 well-isolated single units in the PPC and 808 units in M2 during 20-min foraging sessions in a 2-m octagonal arena (fig. S1 and movie S1).

By measuring Euler angles (pitch, azimuth, and roll) of the head, pitch and azimuthal flexion of the back, and neck elevation in an egocentric reference frame (Fig. 1A and methods), we found robust tuning curves for all postural features in the PPC and M2, with peak rates often >5 standard deviations (SD) from the shuffled distribution (Fig. 1, B to G). The majority of cells with tuning peaks exceeding the 99th percentile of the shuffled data (fig. S3, A and B) were stable across recording sessions (mean of 56.4% in the PPC and 57.8% in M2) (Fig. 1, B to G, and table S1). Postural tuning was also stable across light and dark sessions (fig. S4), indicating its independence from allocentric landmarks and visually oriented attention (*18*).

Cells in the PPC and M2 frequently responded to conjunctive postures involving the head, back, or whole body (Fig. 2, A and C, and movies S2 to S6), prompting us to build a generalized linear model (GLM) (methods) to identify features best explaining neural activity. We utilized a forward-search procedure in which egocentric posture variables and their derivatives, as well as allocentric features including head direction, running direction, and spatial location, were added until the cross-validated model performance no longer improved significantly (*19*) (methods).

This approach indicated that the largest fractions of cells in the PPC (n = 237, 32.5% of 729 cells) and M2 (n = 316, 39.1% of 808 cells) were driven by postural features of the head, including interactions (e.g., between pitch and azimuth), conjunctions of head posture and neck height, and movement (Fig. 2, B and D). Substantial fractions of cells were also tuned to back posture or movement (n = 69, 9.5% in the PPC; n = 75, 9.3% in M2), as well as elevation or movement of the neck (n = 43, 5.9% in the PPC; n = 84, 10.4% in M2) (Fig. 2, B and D, and table S2). Smaller percentages of cells exhibited whole-body tuning, being driven by combinations of head, neck, and back angles [n = 29, 4.0%, Z = 7.9, P < 0.001 in the PPC (large-sample binomial test with expected null probability P_0 of 0.01); n = 27, 3.3%, Z = 6.5, P < 0.001 in M2].

Running speed (n = 38 cells, 5.2% in the PPC; n = 26 cells, 3.2% in M2) and self-motion (n = 15 cells, 2.1%, Z = 2.7, P < 0.01 in the PPC; n = 4 cells, 0.5%, Z = −1.3, P > 0.95 in M2) (Fig. 2, B and D; fig. S5; and table S2) accounted for considerably less of the population than posture (fig. S5). Weaker still were allocentric signals, including head and running direction (Z = −0.67, P > 0.85 in the PPC; Z = −0.56, P > 0.81 in M2) and spatial location (Z = −2.16, P > 0.99 in the PPC; Z = 0.86, P > 0.19 in M2), which did not reach significance in either area (Fig. 2, B and D, and table S2).

The statistical model indicated that the main features driving cells in the PPC and M2 related to posture (46.2% in the PPC; 58.7% in M2) as opposed to movement (5.6% in the PPC; 3.6% in M2). We tested this further by splitting recording sessions on the basis of movement velocity or posture and found that tuning curves for posture remained virtually identical regardless of movement status, whereas tuning to movement varied unreliably when split by posture (fig. S6). Postural tuning was thus expressed independently of movement, but not vice versa.

Previous studies showed anatomical organization for body and facial movement in parietal and motor areas in various mammalian species (*20*–*23*), so we assessed whether postural tuning was also topographical. Head representation in M2 was concentrated at anterior [$\chi^2(4)$ = 57.1, P < 0.001; Yates corrected χ^2 test] (Fig. 3A) and medial [$\chi^2(4)$ = 110.6, P < 0.001] locations, whereas back posture predominated at the posterior [$\chi^2(4)$ = 98.1, P < 0.001] and lateral [$\chi^2(4)$ = 105, P < 0.001] poles (Fig. 3, A and B). In the PPC, anteromedial sites adjacent to M2 showed the strongest back tuning [$\chi^2(3)$ = 29.9, P < 0.001, anterior-posterior gradient; $\chi^2(4)$ = 12.5, P < 0.05, medial-lateral], whereas posterior-lateral regions responded primarily to head posture [$\chi^2(4)$ = 47.5, P < 0.001, anterior-posterior; $\chi^2(4)$ = 52.4, P < 0.001, medial-lateral], producing a coarse mirroring of head and back representation across the PPC and M2 (Fig. 3A).

Because the PPC and frontal motor cortices form an extended network supporting spatial movement planning and decision-making (*24*–*27*), we asked whether structured correlations existed between spikes recorded simultaneously across areas (n = 5 rats). We screened for cells with significant interregional signal correlations (methods) and identified 1017 positively and 182 negatively correlated pairs in one recording session (n = 15 sessions) and 758 positively and 141 negatively correlated pairs in a second session (n = 14 sessions) the same day. The average normalized positive cross correlations indicated a consistent peak, with the PPC preceding M2 by 50 ms across sessions [−72 ms, −32 ms, bootstrapped 99% confidence interval (CI) for session one; −71 ms, −31 ms for session two] and negative correlations peaking at −85 ms [−190 ms, +16 ms] in session one and −25 ms [−140 ms, +82 ms] in session two (Fig. 3, C and D).

To next address whether population activity was sufficient to reconstruct behavior, we reduced the behavioral dataset from six dimensions (three axes for the head, two for the back, and one for neck elevation) to two by using Isomap (*28*). This rendered posture for the head, back, and neck on a two-dimensional (2D) surface, or "posture map," with each pixel corresponding to a particular bodily configuration (fig. S7). We chose a session with 37 PPC and 22 M2 neurons recorded simultaneously to train a uniform prior decoder to

Kavli Institute for Systems Neuroscience, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway.
*These authors contributed equally to this work.
†Corresponding author. Email: bartul.mimica@ntnu.no (B.M.); whitlock@ntnu.no (J.R.W.)

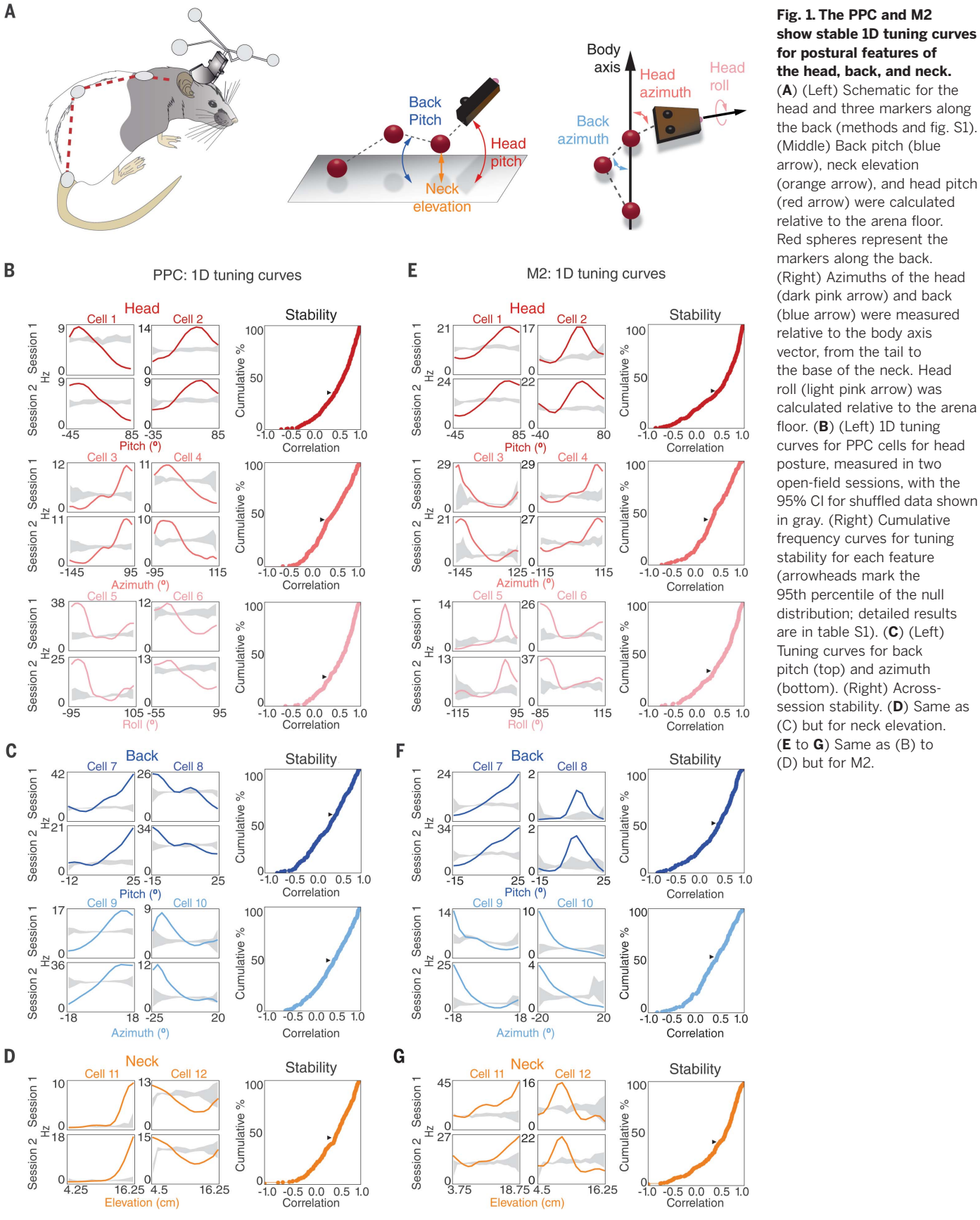


Fig. 1. The PPC and M2 show stable 1D tuning curves for postural features of the head, back, and neck. (**A**) (Left) Schematic for the head and three markers along the back (methods and fig. S1). (Middle) Back pitch (blue arrow), neck elevation (orange arrow), and head pitch (red arrow) were calculated relative to the arena floor. Red spheres represent the markers along the back. (Right) Azimuths of the head (dark pink arrow) and back (blue arrow) were measured relative to the body axis vector, from the tail to the base of the neck. Head roll (light pink arrow) was calculated relative to the arena floor. (**B**) (Left) 1D tuning curves for PPC cells for head posture, measured in two open-field sessions, with the 95% CI for shuffled data shown in gray. (Right) Cumulative frequency curves for tuning stability for each feature (arrowheads mark the 95th percentile of the null distribution; detailed results are in table S1). (**C**) (Left) Tuning curves for back pitch (top) and azimuth (bottom). (Right) Across-session stability. (**D**) Same as (C) but for neck elevation. (**E** to **G**) Same as (B) to (D) but for M2.

predict the animal's dynamic position on the posture map on withheld segments (Fig. 4A and movies S7 and S8). Decoder performance, on average, exceeded the shuffled distribution by >45 SD (Fig. 4B).

Cumulative occupancy on the posture map was dominated by epochs when the animal was on all fours with its head lowered (i.e., foraging) (Fig. 4C, left). We found significantly fewer cells tuned to these high-occupancy, or "default," postures (Fig. 4C, dashed oval), whereas less-visited postures were represented more densely by the ensemble (t_{10} = 4.82, P < 0.01, Welch's two-sided t test) (Fig. 4D, left). The same pattern was observed across animals (t_{10} = 7.74, P < 0.001) (Fig. 4D, right, and fig. S8), suggesting that receptive fields were distributed on the basis of occupancy. Despite this anisotropy in representation, decoder performance was significantly better than chance for all postures, with smaller error for high- than for low-occupancy postures (t_{74} = 6.21, P < 0.001) (Fig. 4E).

The finding that cell populations in parietal and frontal motor cortices represent 3D posture

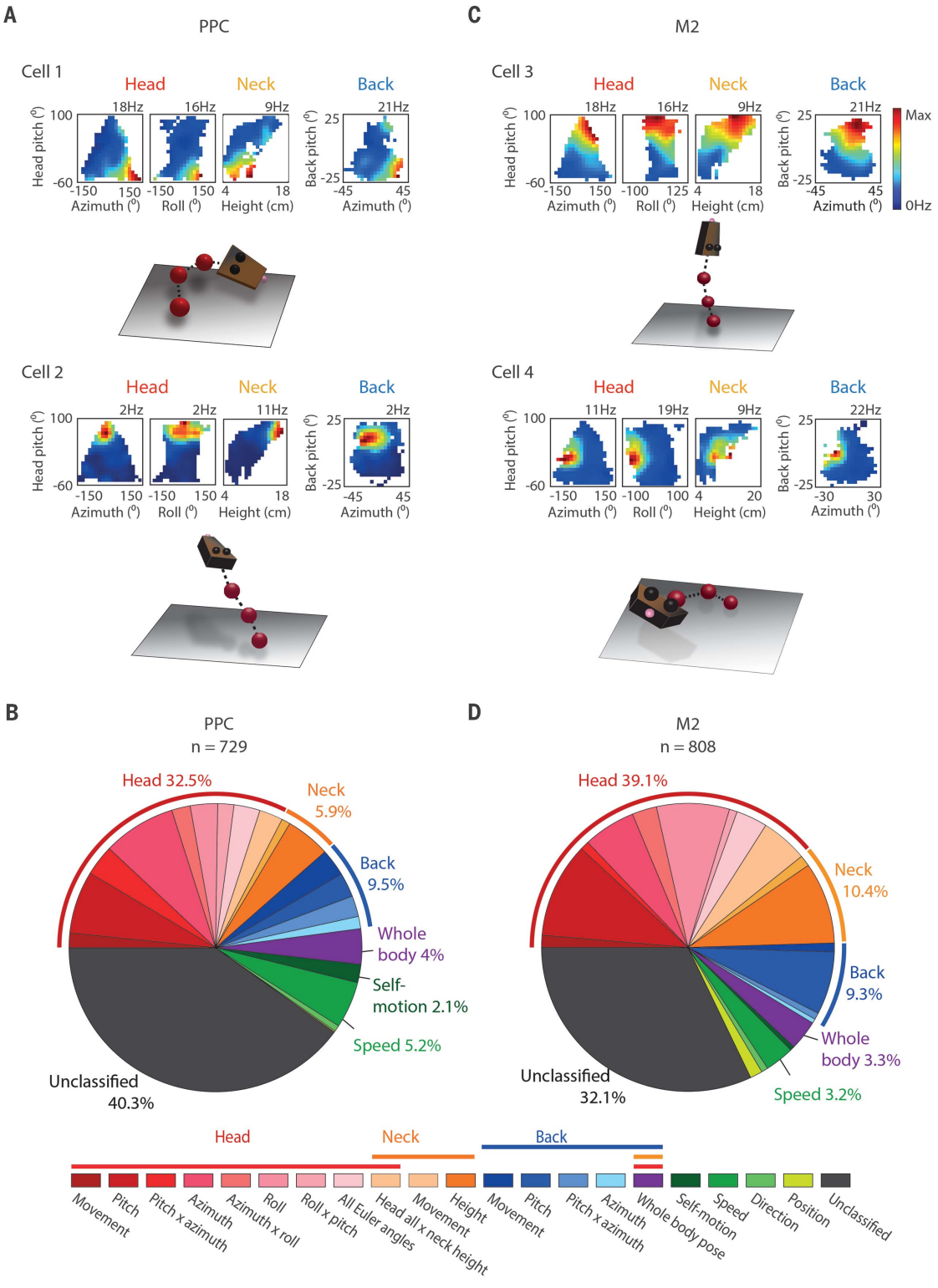


Fig. 2. The PPC and M2 are tuned to combinations of head, back, and neck positions. (**A**) Example PPC cells tuned to combinations of head, back, and neck positions. Conjunctive representations produce single-firing fields in the 2D rate maps; maximal firing rates (in hertz) are indicated above each map (top). 3D animal models (bottom) depict postures to which cells were tuned. Cell 1 preferred whole-body flexion and head roll to the right; cell 2 fired during rearing, with firing driven by the interaction of head pitch with neck elevation. (**B**) Distribution of behavioral tuning in the PPC as determined by the GLM (see the color-coded legend and table S2 for detailed results). (**C**) Examples of postural tuning in M2 cells. Cell 3 (top right) fired when the head, back, and neck were raised vertically; cell 4 was tuned to leftward head roll and back flexion during sharp turns. (**D**) Distribution of coding properties for 808 M2 cells.

robustly and in larger proportions than other behavioral features complements and extends decades of study on the positional coding of single effectors in stationary animals [e.g., (*5*, *29*–*31*)]. The predominance of postural tuning in our data may reflect the myriad kinematic computations that must be solved to coordinate whole-body movement during free behavior. It is also consistent with a functional division of labor in which higher cortical areas specify body position and goals (*32*–*34*) whereas descending motor pathways and subcortical nuclei control movement dynamics more directly (*35*–*39*).

The topographical distribution of postural tuning for the head and back appeared to follow

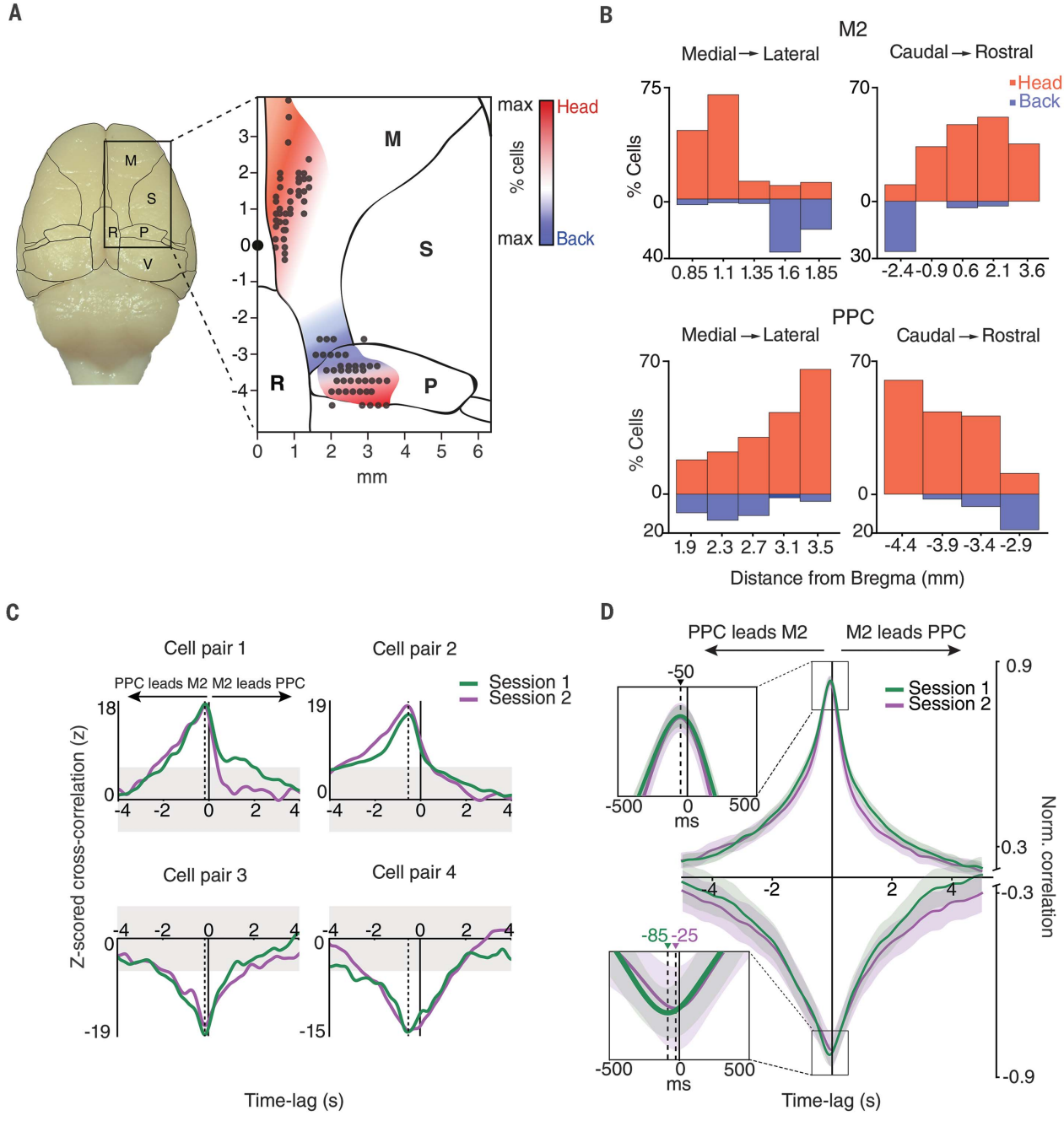


Fig. 3. Head and back posture were organized topographically across the PPC and M2, and the PPC led cross correlations between areas. (**A**) Dorsal view of the cortex with boundaries delineating primary and secondary motor cortices (M) and somatosensory (S), retrosplenial (R), posterior parietal (P), and visual (V) cortices. The magnified view (right) shows recording locations (gray dots; 41 sites in M2, 40 sites in the PPC), and shading indicates tuning for the head (red) and back (blue). The black dot represents the bregma, with distance marked in millimeters. (**B**) Percentages of cells in M2 (top) and the PPC (bottom) driven by head and back positions. For all comparisons, the actual distribution of tuning differed significantly from theoretical distributions that assumed a constant proportion of tuned cells across bins. (**C**) Four cell pairs in the PPC and M2 showing stable *z*-scored cross correlations, with the PPC preceding M2. Dashed and solid lines represent the temporal offset of the cross-correlation peaks and time zero, respectively. Gray-shaded areas indicate ±6 SD of the shuffled data. (**D**) The normalized cross correlation for all cell pairs shows a negative peak for the PPC relative to M2 for positive and negative correlations. Shading indicates the 99% CI.

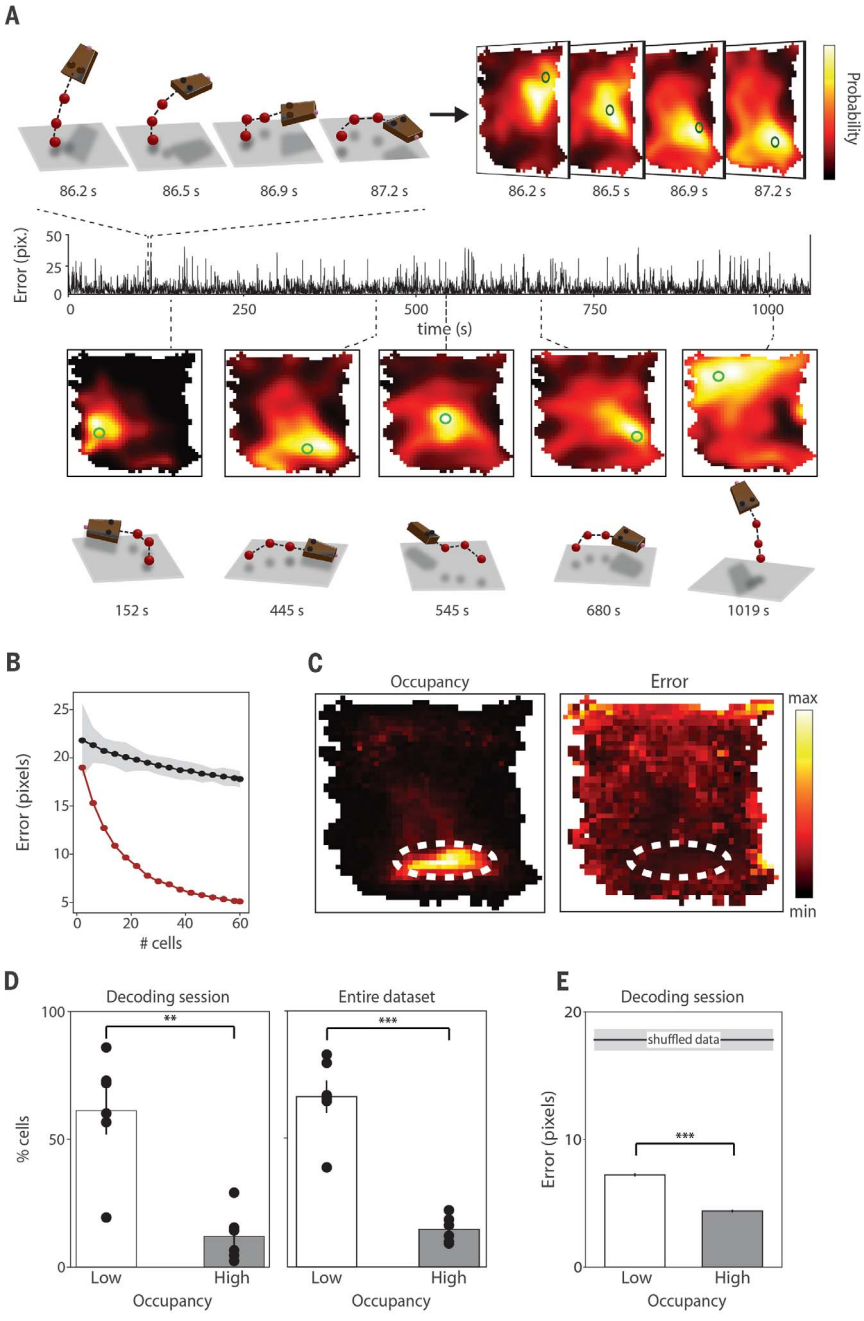


Fig. 4. Ensemble decoding of posture in the PPC and M2 reveals a nonuniform distribution of tuning. (**A**) (Top left) Four snapshots taken within 1 s as the animal came down from rearing and bent rightward. (Top right) Corresponding posture maps illustrate the log posterior distribution of the animal's posture estimated by using PPC and M2 cell ensembles. Actual posture is marked with a green "O," and the maximum likelihood is color coded from yellow to black. (Middle) Timeline indicating error, in Isomap pixels (pix.), over a 20-min recording. (Bottom) Five examples of distinct poses (time points are listed below) and attendant Isomaps illustrating real and decoded postures. (**B**) Decoder accuracy as a function of cell sample size (red dots), with the null distribution above (black dots). The shaded area indicates ±3 SD. (**C**) (Left) Cumulative occupancy on the Isomap showed that the longest dwell times were in the low center of the map, corresponding to foraging. The dashed oval delineates the high-occupancy area where the animal spent >50% of the session. (Right) The average decoding error on the Isomap was smaller in the low center of the map than elsewhere. The dashed oval is the same as in the left panel. (**D**) (Left) The percentage of cells representing the six postural features (black dots) was significantly higher for low- versus high-occupancy bins in the decoding session. (Right) The same analysis for all animals in the study. (**E**) Decoder error was below chance for low- and high-occupancy regions of the posture map and was significantly smaller for the high- than for the low-occupancy area. The line near the top of the graph indicates the mean ± 3 SD for shuffled data. Bar graphs in (D) and (E) indicate mean ± SEM. $^{**}P < 0.01$; $^{***}P < 0.001$.

a functional organization identified in earlier microstimulation studies in anaesthetized animals (*22*, *23*). We found mainly head and back representation in the PPC and M2, but it is possible that posture for the entire body overlays the cortical surface, including primary somatosensory and motor cortices. More broadly, it remains to be established whether postural signals are generated in the cortex specifically or whether they are inherited from other regions. Our cross-correlation analyses also suggest that a temporal structure exists for postural representations across areas, with the PPC operating upstream from M2, though such an ordering could shift in the context of different tasks (*26*).

Our use of 3D tracking additionally revealed that speed and self-motion tuning in the PPC (*13*, *14*, *40*) were likely overestimated in previous studies using 2D tracking of rodents, owing to insufficient resolution to disambiguate posture from movement. Tracking the back allowed us to detect neural tuning to flexion of the trunk, indicating that vestibular signaling (*41*) alone could not explain the postural coding in our recordings. For both the back and the head, the arrangement of postural tuning peaks was notably nonuniform and appeared to be optimized for the duration for which postures were occupied (Fig. 4D and figs. S3, A and B, and S8). Previous theoretical works considering optimal coding strategies in sensory systems (*42*, *43*) suggested that the range of the stimulus spectrum visited most should be encoded by more cells with narrower tuning widths, but this was not the case in our data. Rather, we found that proportionately less of the network was dedicated to default states in which the animals spent more time. This arrangement allowed high-fidelity decoding of the entire range of postures while minimizing metabolic demand on the cells, making it both precise and efficient. Together, our results strongly support the notion that the PPC-M2 network plays a key role in representing the dynamic organization of the body in space, or body schema, postulated more than 100 years ago (*1*, *2*).

REFERENCES AND NOTES

1. H. Head, G. Holmes, *Brain* **34**, 102–254 (1911).
2. M. Critchley, *The Parietal Lobes* (Williams and Wilkins, 1953).
3. V. B. Mountcastle, J. C. Lynch, A. Georgopoulos, H. Sakata, C. Acuna, *J. Neurophysiol.* **38**, 871–908 (1975).
4. J. Tanji, E. V. Evarts, *J. Neurophysiol.* **39**, 1062–1068 (1976).
5. R. A. Andersen, V. B. Mountcastle, *J. Neurosci.* **3**, 532–548 (1983).
6. A. P. Georgopoulos, R. E. Kettner, A. B. Schwartz, *J. Neurosci.* **8**, 2928–2937 (1988).
7. J. Wessberg *et al.*, *Nature* **408**, 361–365 (2000).
8. R. A. Andersen, C. A. Buneo, *Annu. Rev. Neurosci.* **25**, 189–220 (2002).
9. M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, J. P. Donoghue, *Nature* **416**, 141–142 (2002).
10. M. Hauschild, G. H. Mulliken, I. Fineman, G. E. Loeb, R. A. Andersen, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17075–17080 (2012).
11. B. Kolb, R. J. Sutherland, I. Q. Whishaw, *Behav. Neurosci.* **97**, 13–27 (1983).
12. J. C. Erlich, M. Bialek, C. D. Brody, *Neuron* **72**, 330–343 (2011).
13. B. L. McNaughton *et al.*, *Cereb. Cortex* **4**, 27–39 (1994).
14. J. R. Whitlock, G. Pfuhl, N. Dagslott, M. B. Moser, E. I. Moser, *Neuron* **73**, 789–802 (2012).
15. D. A. Nitz, *Neuron* **49**, 747–756 (2006).

16. C. D. Harvey, P. Coen, D. W. Tank, *Nature* **484**, 62–68 (2012).
17. A. A. Wilber, B. J. Clark, T. C. Forster, M. Tatsuno, B. L. McNaughton, *J. Neurosci.* **34**, 5431–5446 (2014).
18. E. B. Cutrell, R. T. Marrocco, *Exp. Brain Res.* **144**, 103–113 (2002).
19. K. Hardcastle, N. Maheswaranathan, S. Ganguli, L. M. Giocomo, *Neuron* **94**, 375–387.e7 (2017).
20. J. Hyvärinen, *Brain Res.* **206**, 287–303 (1981).
21. I. Stepniewska, P. C. Fang, J. H. Kaas, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4878–4883 (2005).
22. R. D. Hall, E. P. Lindholm, *Brain Res.* **66**, 23–38 (1974).
23. M. Brecht *et al*., *J. Comp. Neurol.* **479**, 360–373 (2004).
24. S. P. Wise, D. Boussaoud, P. B. Johnson, R. Caminiti, *Annu. Rev. Neurosci.* **20**, 25–42 (1997).
25. G. Rizzolatti, L. Fogassi, V. Gallese, *Curr. Opin. Neurobiol.* **7**, 562–567 (1997).
26. B. Pesaran, M. J. Nelson, R. A. Andersen, *Nature* **453**, 406–409 (2008).
27. T. D. Hanks *et al*., *Nature* **520**, 220–223 (2015).
28. J. B. Tenenbaum, V. de Silva, J. C. Langford, *Science* **290**, 2319–2323 (2000).
29. P. R. Brotchie, R. A. Andersen, L. H. Snyder, S. J. Goodman, *Nature* **375**, 232–235 (1995).
30. A. P. Batista, C. A. Buneo, L. H. Snyder, R. A. Andersen, *Science* **285**, 257–260 (1999).
31. E. Salinas, P. Thier, *Neuron* **27**, 15–21 (2000).
32. G. H. Mulliken, S. Musallam, R. A. Andersen, *J. Neurosci.* **28**, 12913–12926 (2008).
33. L. Fogassi *et al*., *J. Neurophysiol.* **76**, 141–157 (1996).
34. T. M. Pearce, D. W. Moran, *Science* **337**, 984–988 (2012).
35. G. M. Shepherd, *Nat. Rev. Neurosci.* **14**, 278–291 (2013).
36. G. Cui *et al*., *Nature* **494**, 238–242 (2013).
37. J. J. Wilson, N. Alexandre, C. Trentin, M. Tripodi, *Curr. Biol.* **28**, 1744–1755.e12 (2018).
38. M. S. Esposito, P. Capelli, S. Arber, *Nature* **508**, 351–356 (2014).
39. J. E. Markowitz *et al*., *Cell* **174**, 44–58.e17 (2018).
40. X. Chen, G. C. Deangelis, D. E. Angelaki, *Neuron* **80**, 1310–1321 (2013).
41. F. Klam, W. Graf, *Eur. J. Neurosci.* **18**, 995–1010 (2003).
42. N. S. Harper, D. McAlpine, *Nature* **430**, 682–686 (2004).
43. D. Ganguli, E. P. Simoncelli, *Neural Comput.* **26**, 2103–2134 (2014).
44. B. Mimica, Efficient cortical coding of 3D posture in freely behaving rats – data and GUI, Norstore (2018); https://doi.org/10.11582/2018.00028.

ACKNOWLEDGMENTS

We thank B. McNaughton and E. Moser for helpful comments on the manuscript; K. Haugen, K. Jenssen, E. Kråkvik, H. Obenhaus, R. Gardner, T. Feyissa, K. Hovde, H. Kleven, M. Gianatti, and H. Waade for technical and IT assistance; J. Jeon, J. Adams, and NeuroNexus for assistance in drive design; G. Olsen and M. Witter for assistance with anatomical delineations; and S. Eggen for veterinary oversight. **Funding:** This study was supported by research grants from the European Research Council ("RAT MIRROR CELL," starting grant agreement 335328), the Research Council of Norway (FRIPRO Young Research Talents, grant agreement 239963), the Kavli Foundation, and the Center of Excellence scheme of the Research Council of Norway (Center for Neural Computation). **Author contributions:** J.R.W. and B.A.D. designed the experiments. B.M. and T.T. conducted the experiments. B.A.D. designed the analyses. B.A.D., B.M., V.P.T.N.C.S.B., and J.R.W. performed the analyses. J.R.W. and B.M. wrote the paper with assistance from T.T. and B.A.D. **Competing interests:** No competing interests declared. **Data and materials availability:** Datasets validating the main findings and conclusions of the paper, as well as the 3D tracking graphical user interface and support folders, are available in the Norwegian national research data archive (*44*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/584/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 and S2
References (*45*–*51*)
Movies S1 to S8

18 May 2018; accepted 14 September 2018
10.1126/science.aau2013

MICROBIOTA

Experimental evolution of a fungal pathogen into a gut symbiont

Gloria Hoi Wan Tso[1], Jose Antonio Reales-Calderon[1], Alrina Shin Min Tan[1], XiaoHui Sem[1]*, Giang Thi Thu Le[1]†, Tze Guan Tan[1], Ghee Chuan Lai[1], K. G. Srinivasan[1], Marina Yurieva[1]‡, Webber Liao[1], Michael Poidinger[1], Francesca Zolezzi[1], Giulia Rancati[2], Norman Pavelka[1]§

Gut microbes live in symbiosis with their hosts, but how mutualistic animal-microbe interactions emerge is not understood. By adaptively evolving the opportunistic fungal pathogen *Candida albicans* in the mouse gastrointestinal tract, we selected strains that not only had lost their main virulence program but also protected their new hosts against a variety of systemic infections. This protection was independent of adaptive immunity, arose as early as a single day postpriming, was dependent on increased innate cytokine responses, and was thus reminiscent of "trained immunity." Because both the microbe and its new host gain some advantages from their interaction, this experimental system might allow direct study of the evolutionary forces that govern the emergence of mutualism between a mammal and a fungus.

Symbiotic relationships are ubiquitous in nature and vary on a continuum from parasitism to mutualism (*1*). Evolution plays a critical role in the establishment of these interactions and in moving them along the parasite-mutualist axis. In the case of host-microbe interactions, the mode of host-to-host transmission dictates whether the microbe will evolve toward higher pathogenicity, commensalism, or mutualism (*2*). Experimental evolution via serial passaging of pathogens in a new host is a powerful strategy to study these dynamic changes in real time. For instance, passaging parasites in new hosts via the systemic infection route most often selects for increased virulence against the new host (*3*). However, experimental systems to study the evolutionary emergence of mutualistic animal-microbe interactions are lacking (*4*).

The mammalian gastrointestinal (GI) tract harbors a large and diverse microbial community, whose members interact with the host primarily in commensal or mutualistic ways (*5*). This raises the question of how the mammalian host establishes these neutral or beneficial interactions while purging potential pathogens. We hereby hypothesize that the gut environment selects microbes on the basis of their pathogenicity and moves them along the parasitism-mutualism axis via evolutionary processes.

[1]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove Immunos #04, Singapore 138648, Singapore. [2]Institute of Medical Biology, A*STAR, 8A Biomedical Grove, Immunos #05, Singapore 138648, Singapore.
*Present address: Foundation for Innovative New Diagnostics (FIND), Campus Biotech, Chemin des Mines 9, 1202 Geneva, Switzerland. †Present address: Singapore Bioimaging Consortium (SBIC), A*STAR, 11 Biopolis Way, #01-02 Helios, Singapore 138667, Singapore. ‡Present address: The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032, USA.
§Corresponding author. Email: norman_pavelka@immunol.a-star.edu

To investigate how the gut environment shapes the evolution of a microbe, we developed an experimental system, based on long-term GI colonization of antibiotic-treated mice by the fungus *Candida albicans* coupled with serial fecal transplants from colonized to naïve hosts. Several variations of the protocol were tested (Fig. 1A), and all fecal transplants resulted in successful colonization of the recipient animals. Clonal isolates harvested after 8 or 10 weekly serial passages (w8 or w10 strains), but not after a 1-week passage (w1 strains), showed a significantly increased intra-GI competitive fitness (Fig. 1B), which was at least as high as that of strains deficient of the enhanced filamentous growth 1 (*EFG1*) gene (*efg1*$^{-/-}$ strains), which were previously known to be hyperfit in the antibiotic-treated mouse gut (*6*–*8*). Overall, these results demonstrate that a microbe can be experimentally induced to increase its competitive fitness in the gut of an unnatural host by means of adaptive evolution.

Host adaptive immunity appeared to play little or no role in this adaptation process, because all results presented herein were qualitatively similar whether the evolution experiments were performed in wild-type (WT) or Recombination activating 1–deficient (*Rag1*$^{-/-}$) mice, which lack functional T and B cells. To determine if antibiotic treatment played any role in this evolutionary process, we repeated the above evolution experiments in untreated mice. Without the use of antibiotics, however, and consistent with current knowledge (*9*), we were unable to establish a long-term colonization in the adult animals' guts, and the experiment had to be aborted after 3 weeks (Fig. 1A). We thus resorted to an alternative model involving neonatal mice, which have long been known to allow *C. albicans* GI colonization even in the absence of antibiotics (*10*). Specifically, we intragastrically inoculated 2-week-old infant mice with the same ancestral *C. albicans* used for the previous evolution experiments, with seven evolution lines maintained on oral antibiotics and eight lines on control water (Fig. 1A). Notably, individual colonies isolated after 5 weeks of evolution displayed a significantly increased GI competitive fitness in infant mice only if they had been evolved in antibiotic-treated, but not untreated, mice (Fig. 1B), suggesting that the presence of an intact bacterial microbiota limits the adaptation of *C. albicans* in the mouse GI tract.

Because *C. albicans* strains that are hyperfit in the antibiotic-treated or germ-free mouse gut are typically deficient in hyphal morphogenesis (*6*, *8*, *11*), we next examined cellular morphologies and noted that over the course of 10 weeks of evolution in infant mice, all *C. albicans* populations progressively lost their ability to respond to hyphal-inducing stimuli, but only if they had been evolved in the presence of antibiotics (Fig. 1C). Screening several independent clones across all evolution lines, we confirmed that in the presence of antibiotics, virtually all *C. albicans* cells lost their ability to form true hyphae by the end of each experiment (Fig. 1, D and E). Importantly, in vitro daily serial passaging of *C. albicans* for 5 weeks in the presence of antibiotics failed to yield such phenotypes (fig. S1). Though experimental evolution in germ-free mice would be the ideal model to definitively prove this point, these data strongly suggest that it was the absence of microbiota and not antibiotics per se that selected for hyphal-defective strains in our in vivo evolution experiments. Finally, nonfilamentous w10 strains and *efg1*$^{-/-}$ cells, although being hyperfit in the antibiotic-treated gut, actually colonized the GI tract of untreated mice less efficiently than WT cells (fig. S2). Together, these results indicate that, although the hyphal morphogenesis program is required for competition of *C. albicans* with resident gut bacteria, it otherwise represents a fitness burden and is thus rapidly lost when the gut microbiota are absent or chronically perturbed by long-term antibiotic treatment.

In search of the genetic basis underlying these evolved phenotypes, we performed high-depth, high-coverage whole-genome sequencing on four ancestral *C. albicans* strains used as inocula for the evolution experiments, four w1 strains, and 28 w5, w8, or w10 strains selected from 16 independent evolution lines across three different serial passaging protocols (table S1). A total of 34 verified open-reading frames were identified as carrying ≥1 de novo nonsynonymous mutation in ≥1 evolved w5, w8, or w10 strain (Fig. 2A and table S2). On the basis of Gene Ontology enrichment analysis, 20 of these genes were related to filamentous growth ($P = 5.5 \times 10^{-10}$, hypergeometric test, Bonferroni correction), 10 encoded proteins located in the cell wall ($P = 3.5 \times 10^{-7}$, hypergeometric test, Bonferroni correction), and seven functioned as transcription factors (TFs) ($P = 2.2 \times 10^{-3}$, hypergeometric test, Bonferroni correction). Specifically, six genes (e.g., *EFG1*) encoded TFs involved in hyphal gene expression regulation (table S3).

With 22 of 28 sequenced evolved strains (or 11 out of 16 evolution lines) carrying ≥1 de novo mutation in *FLO8*, which encodes a TF required for hyphal development (Fig. 2C) (*12*), this gene was the most frequently mutated (Fig. 2A). Interestingly, all identified *FLO8* mutations lead to premature stop codons or frame shifts (Fig. 2B), implicating these as loss-of-function mutations. In accordance, many stop codons were located near amino acid position 654, where truncations were shown to result in null phenotypes (*13*). To test if *FLO8* inactivation is sufficient to increase the fitness of *C. albicans* in the mouse GI tract, we generated heterozygous (*flo8*$^{+/-}$) and homozygous (*flo8*$^{-/-}$) *FLO8*-deletion strains not previously exposed to the mouse gut environment. Consistent with our predictions, *flo8*$^{-/-}$, but not *flo8*$^{+/-}$, cells were nonfilamentous (Fig. 2D) and outcompeted WT *C. albicans* as efficiently as the *efg1*$^{-/-}$ strain or any of the gut-evolved *C. albicans* strains (Fig. 2E). To test if the single-point mutations identified in the w10 strains were indeed required for increased competitive fitness, we restored one *FLO8* allele back to its ancestral sequence in strains w7 and L11.w10.c3 and observed reversion of filamentation (Fig. 2D) and a significant loss of intra-GI tract fitness in both strains (Fig. 2E). Further demonstrating these point mutations to be fully recessive and not dominant-negative, ectopic expression of a WT *FLO8* under its own promoter fully restored filamentation in the tested w10 strains, whereas a mutated *FLO8* had no effect when ectopically expressed in a WT strain (fig. S3). Consistent with this notion, 19 of the 22 strains carrying *FLO8*-inactivating mutations carried them in homozygosity (Fig. 2B). These results show that adaptive evolution in the antibiotic-treated mouse GI tract reproducibly selects for inactivating mutations in central transcriptional regulators of the hyphal morphogenesis program of *C. albicans* and, in particular, in *FLO8*, yielding hyphal-defective

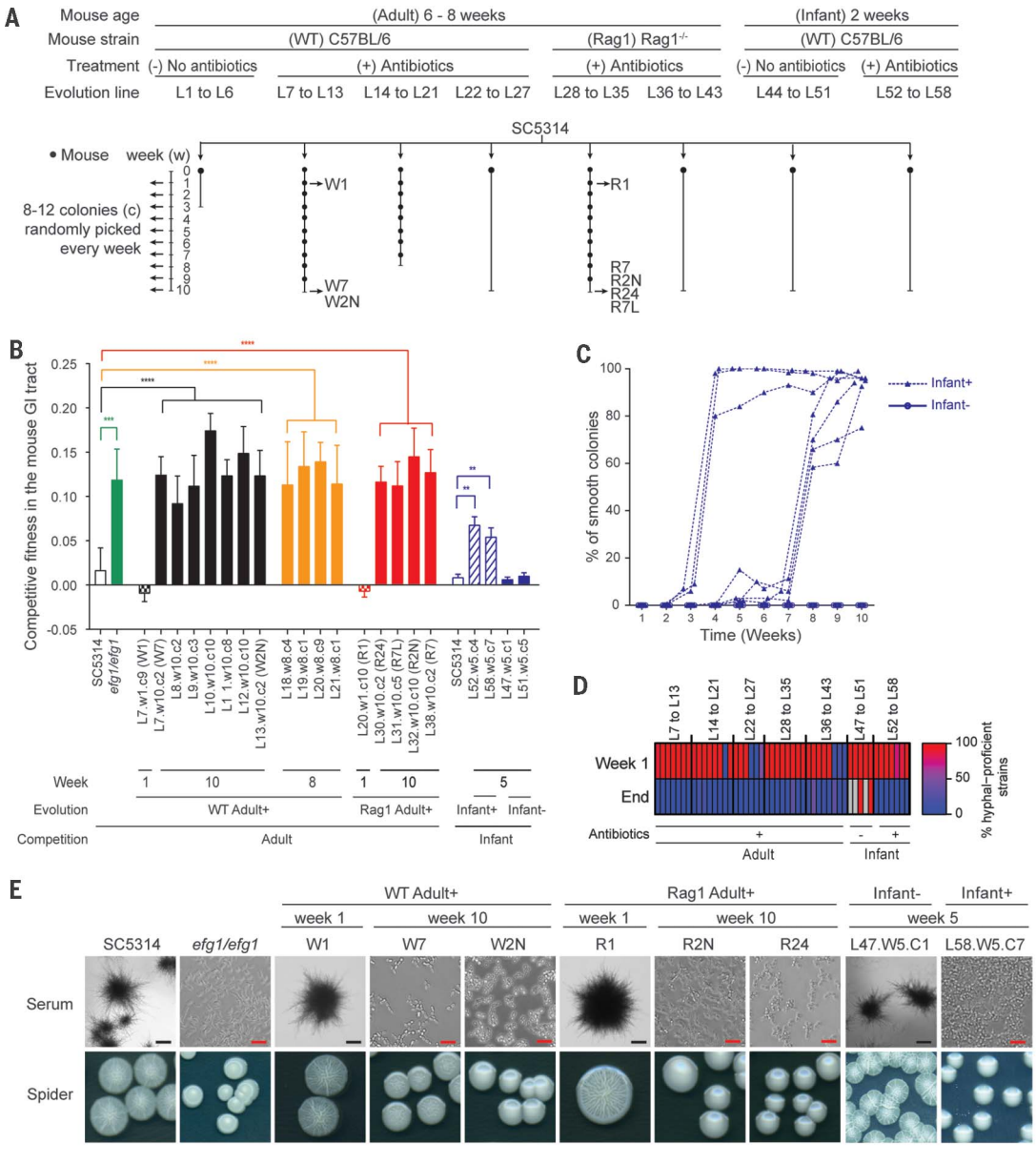


Fig. 1. Adaptive evolution of *C. albicans* in the mouse gut selects for hyphal-defective variants. (**A**) Schematic overview of the different evolution protocols. Individual strains are named according to line (L), week (w), and colony (c). Some specific strains are additionally identified by a shorter name (e.g., W2N or R24). (**B**) Similar to the *efg1*$^{-/-}$ (*efg1/efg1*) mutant, 10-week (w10) and 8-week (w8) *C. albicans* strains evolved in antibiotic-treated mice achieved increased competitive fitness in the mouse GI tract compared with WT (SC5314) and 1-week (w1) gut-evolved strains (W1 and R1). However, after 5 weeks of evolution in antibiotic-free infants, *C. albicans* strains failed to increase their competitive fitness when compared with strains evolved in antibiotic-treated infants. Data are means ± SD. $n = 4$ to 10 mice per group. Mann-Whitney test: $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. (**C**) Whole-stool populations from evolution lines L44 to L58 were plated on Spider agar and scored for smooth (indicative of yeast) and wrinkled (indicative of filamentous) colonies. Smooth colonies eventually appeared and often dominated the stool populations in all antibiotic-treated (Infant+), but not antibiotic-free (Infant−), lines. (**D**) Individual strains (clonal isolates) of each evolution line were scored for hyphal formation in response to serum. Percentages of hyphal-proficient strains were significantly lower at the end (week 8 or 10, depending on the protocol) than after 1 week of the evolution experiment. Gray indicates no data. (**E**) Representative images indicating that gut-evolved *C. albicans* strains are defective in hyphal formation in response to in vitro stimuli. Black scale bar, 200 μm; red scale bar, 20 μm.

mutants with increased intra-GI tract fitness. Notably, a spectrum of hyphal-formation defects, in part associated with homozygous nonsense mutations in key hyphal morphogenesis TFs, such as *EFG1*, are reported to also occur in human clinical isolates of *C. albicans* (*8*).

Beyond single-nucleotide changes, *C. albicans* is known to frequently undergo large-scale genomic mutations, such as short- and long-range loss of heterozygosity (LOH) and segmental or whole-chromosome aneuploidy, especially when under stress and during infection (*14*, *15*). In accordance, 24 of 28 w5, w8, and w10 evolved strains (15 of 16 evolution lines) underwent at least one LOH or aneuploidy event, with chromosome 6 (chr6) LOH (11 strains across six evolution lines) and chr7 trisomy (12 strains across 10 evolution lines) being the most frequently observed (fig. S4). Interestingly, *FLO8* is located on chr6, and all observed chr6 LOH events occurred in strains with homozygous *FLO8* mutations (Fig. 2A), suggesting that the LOH event followed the single-nucleotide change to allow expression of the recessive phenotype. Chr7 trisomy, instead, did not appear to contribute to the adaptive process, because an unevolved, independently obtained strain carrying this aneuploidy displayed neither filamentation defects nor an increased competitive fitness in the mouse GI tract (fig. S5).

As the hyphal morphogenesis program is a key virulence factor of *C. albicans* (*16*), we hypothesized that gut-evolved strains would be less damaging to host cells and be less pathogenic during infection. To test this, we first measured their cytotoxicity in cocultures with either murine macrophages or human colon epithelial cells. Similar to the *efg1*$^{-/-}$ or *flo8*$^{-/-}$ strain, we found that w10 strains induced significantly lower cell damage than WT or w1 strains (Fig. 3A). We next tested the virulence of gut-evolved strains in a mouse model of hematogenously disseminated candidiasis. As expected, using a lethal dose of 5×10^5 colony-forming units (CFUs) of WT *C. albicans*, we observed formation of hyphae in the kidneys of WT mice 2 days postinfection (dpi), as well as severe necrosis, moderate perivascular edema, and moderate diffuse pyogranulomatous renal capsulitis in most animals (Fig. 3B).

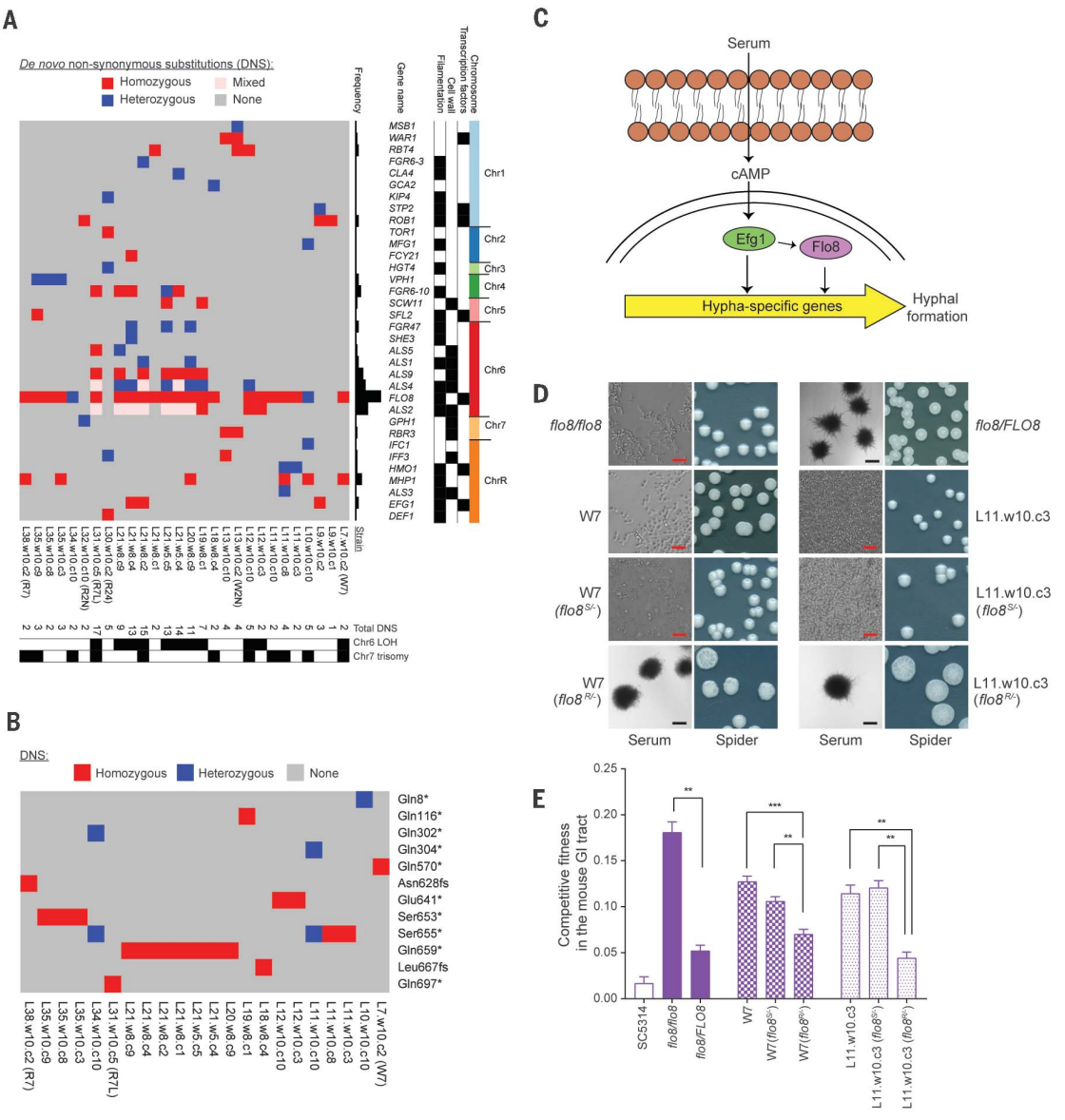


Fig. 2. Recurrent mutations in TFs regulating filamentous growth underlie increased competitive fitness in the mouse gut. (**A**) Clustering of gut-evolved *C. albicans* strains based on mutational pattern across 34 verified open-reading frames carrying de novo, nonsynonymous substitutions (DNSs) reveals recurrent mutations in *FLO8* and other TFs required for filamentous growth. Genes are ordered on the basis of chromosomal location. (**B**) Convergent acquisition of frameshift (fs) or nonsense (*) mutations in the *FLO8* gene across multiple independent evolution experiments, most often in homozygosity. (**C**) *FLO8* and *EFG1* act downstream of the cyclic adenosine 3′,5′-monophosphate (cAMP) pathway to regulate the expression of hypha-specific genes in response to environmental stimuli, such as serum (*28*). (**D**) The *flo8*$^{-/-}$ (*flo8/flo8*) mutant and two w10 evolved strains (W2N and L11.w10.c3) harboring *FLO8* loss-of-function mutations were unable to filament in hyphal-inducing media. By contrast, the *flo8*$^{+/-}$ (*flo8/FLO8*) mutant and the w10 evolved strains, in which one functional *FLO8* allele had been restored (*flo8*$^{R/-}$), but not control strains transformed only with the selection marker (*flo8*$^{S/-}$), were able to form hyphae. Black scale bar, 200 μm; red scale bar, 20 μm. (**E**) Although *flo8*$^{-/-}$ cells were significantly fitter than *flo8*$^{+/-}$ cells in the mouse GI tract, restoration of one *FLO8* allele to its ancestral sequence significantly decreased the GI fitness of strains W2N and L11.w10.c3. Data are means ± SD. $n = 5$ mice per group. Mann-Whitney test: $^{**}P < 0.01$ and $^{***}P < 0.001$.

Eventually, all mice infected with WT or w1 strains succumbed within 3 to 4 dpi (Fig. 3C). By contrast, at the same infection dose, w10 and *flo8*$^{-/-}$ strains strictly remained in the yeast form in kidneys at 2 dpi, whereas a few hyphae were observed with *efg1*$^{-/-}$ cells, and only mild to moderate necrosis was observed with the w10 strains, with only some animals displaying mild edema or capsulitis (Fig. 3B). Similar results were obtained when we infected *Rag1*$^{-/-}$ mice; mice infected by *efg1*$^{-/-}$, *flo8*$^{-/-}$, or w10 strains survived significantly longer than those infected by the WT or w1 strains (Fig. 3C). Again demonstrating a key role of the gut microbiome in this process, *C. albicans* strains evolved in the absence of, but not in the presence of, antibiotics retained their ability to kill mice (fig. S6). Notably, similar loss of virulence was observed both for w10 strains carrying and for those not carrying *FLO8* mutations (Fig. 3, A and C). However, the two w10 strains that were "cured" of their *FLO8*-inactivating mutations displayed a significantly increased virulence compared to their w10 counterparts (fig. S7), indicating that those point mutations were required for the loss of virulence in those evolved strains. Hence, this experimental system, unlike serial passaging via the systemic infection route or in cell culture (*17*–*19*), reproducibly yields *C. albicans* strains that are genetically locked in the yeast form and avirulent. And although *FLO8* appears to be a mutational hotspot with a clear implication in this adaptation mechanism, convergent evolution toward similar phenotypes was also observed in strains lacking *FLO8*-inactivating mutations.

Because gut colonization by commensal fungi protects hosts against infections (*20*), we next asked if hosts colonized by gut-evolved *C. albicans* strains carry a potential advantage over mice colonized with WT *C. albicans*. We first tested this hypothesis under conditions in which the evolved strains would express their increased competitive fitness, that is, in antibiotic-treated adult mice. Notably, animals colonized with the w10 strain R24 (which coincidentally does not harbor *FLO8* mutations), but not with the ancestral WT strain SC5314, showed a significantly increased survival upon a systemic challenge with a fully virulent *C. albicans* strain (Fig. 3D). Under these conditions, both the evolved microbe and its new host benefit from their interaction.

We then also repeated the colonization in non-antibiotic-treated 2-week-old pups, which were then systemically challenged with the fully virulent strains at 6 weeks of age. Although R24 is more rapidly lost than WT in untreated pups (fig. S2), >30% of R24-colonized mice survived until 35 dpi, whereas all SC5314-colonized animals died by ~20 dpi (Fig. 3D). We also tested gut-colonized pups for protection against the distantly related fungus *Aspergillus fumigatus*. Whereas most uncolonized mice or SC5314-colonized mice succumbed ≤8 or ≤15 dpi, respectively, >20% of R24-colonized mice survived at least until 35 dpi (Fig. 3E). Therefore, GI colonization with *C. albicans*, and especially with a gut-evolved strain, at a young age protects hosts against infections later in life, thereby conferring a potential advantage to the host.

The above-described protection is reminiscent of an innate memory-like mechanism termed "trained immunity" (*21*). To test if gut-evolved *C. albicans* strains are efficient immune trainers, we adopted a well-established mouse model based on intravenous priming of WT C57BL/6 mice followed by systemic challenge with a lethal

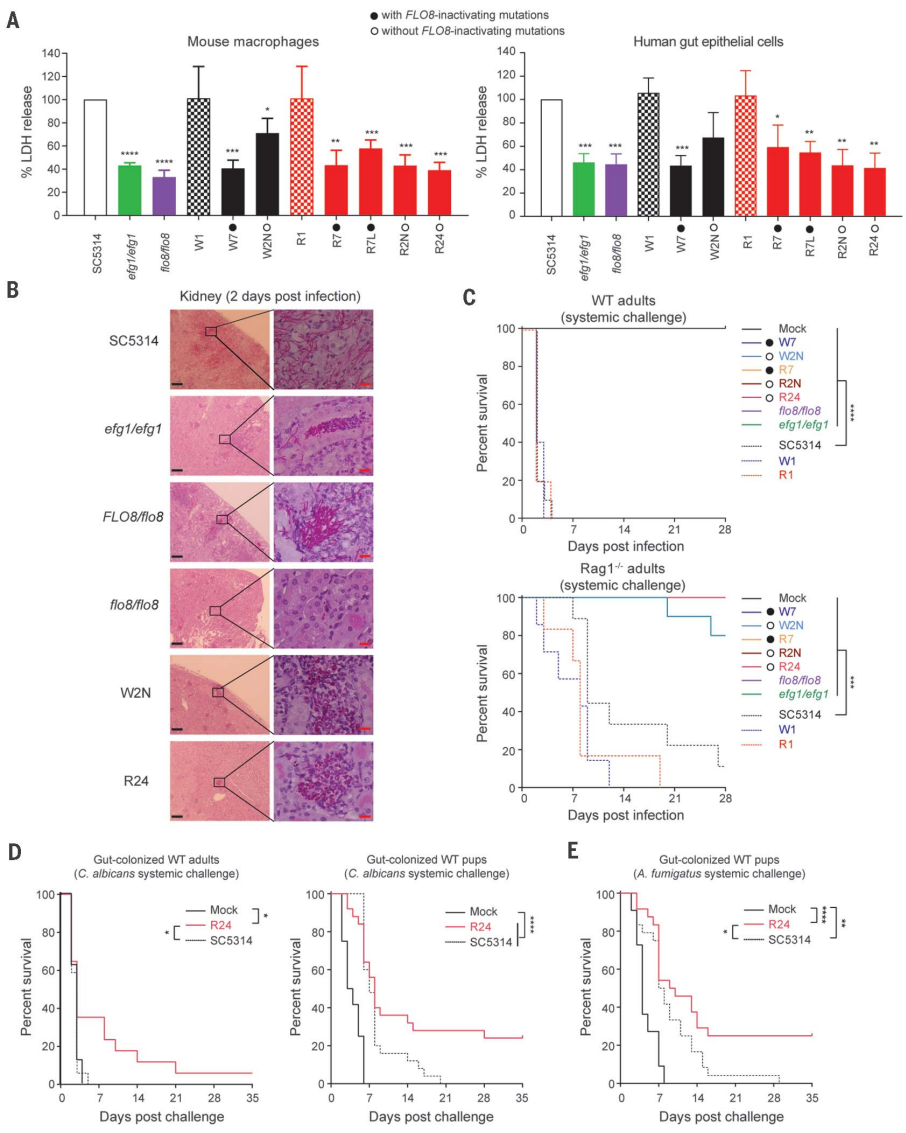


Fig. 3. Gut-evolved *C. albicans* strains are avirulent and protect gut-colonized hosts from systemic fungal infections. (**A**) Similar to the *efg1*$^{-/-}$ and *flo8*$^{-/-}$ mutants, *C. albicans* w10 evolved strains with (filled circles) or without (empty circles) *FLO8*-inactivating mutations had reduced in vitro cytotoxicity against J774A.1 mouse macrophages and HT-29 human gut epithelial cells compared with WT (SC5314) and w1 evolved strains (W1 and R1). LDH, lactate dehydrogenase. Data are means ± SD. $n = 3$ independent experiments. Unpaired Welch's t test: $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$, $^{****}P < 0.0001$. (**B**) One representative Periodic acid–Schiff–stained kidney sections of 3 to 5 mice per group infected with WT (SC5314), *efg1*$^{-/-}$, *flo8*$^{+/-}$, *flo8*$^{-/-+}$, and *C. albicans* w10 evolved strains (W2N or R24). Black scale bar, 200 μm; red scale bar, 20 μm. Images on the right are magnifications of the boxed areas to the left. (**C**) Similar to the *efg1*$^{-/-}$ and *flo8*$^{-/-}$ mutants, w10 evolved strains with (filled circles) or without (empty circles) *FLO8*-inactivating mutations are less virulent in WT and *Rag1*$^{-/-}$ mice compared with WT (SC5314) or w1 strains (W1 and R1). $n = 5$ to 10 mice per group. (**D** and **E**) Antibiotic-treated adults or antibiotic-free infants were colonized with either WT (SC5314) or a w10 strain (R24) and challenged systemically with either WT *C. albicans* (D) or *A. fumigatus* (E). In all tested cases, R24-colonized mice were qualitatively or quantitatively more protected from systemic fungal infections than SC5314-colonized mice. $n = 8$ to 25 mice per group. For (C) to (E), data were analyzed by log-rank test; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

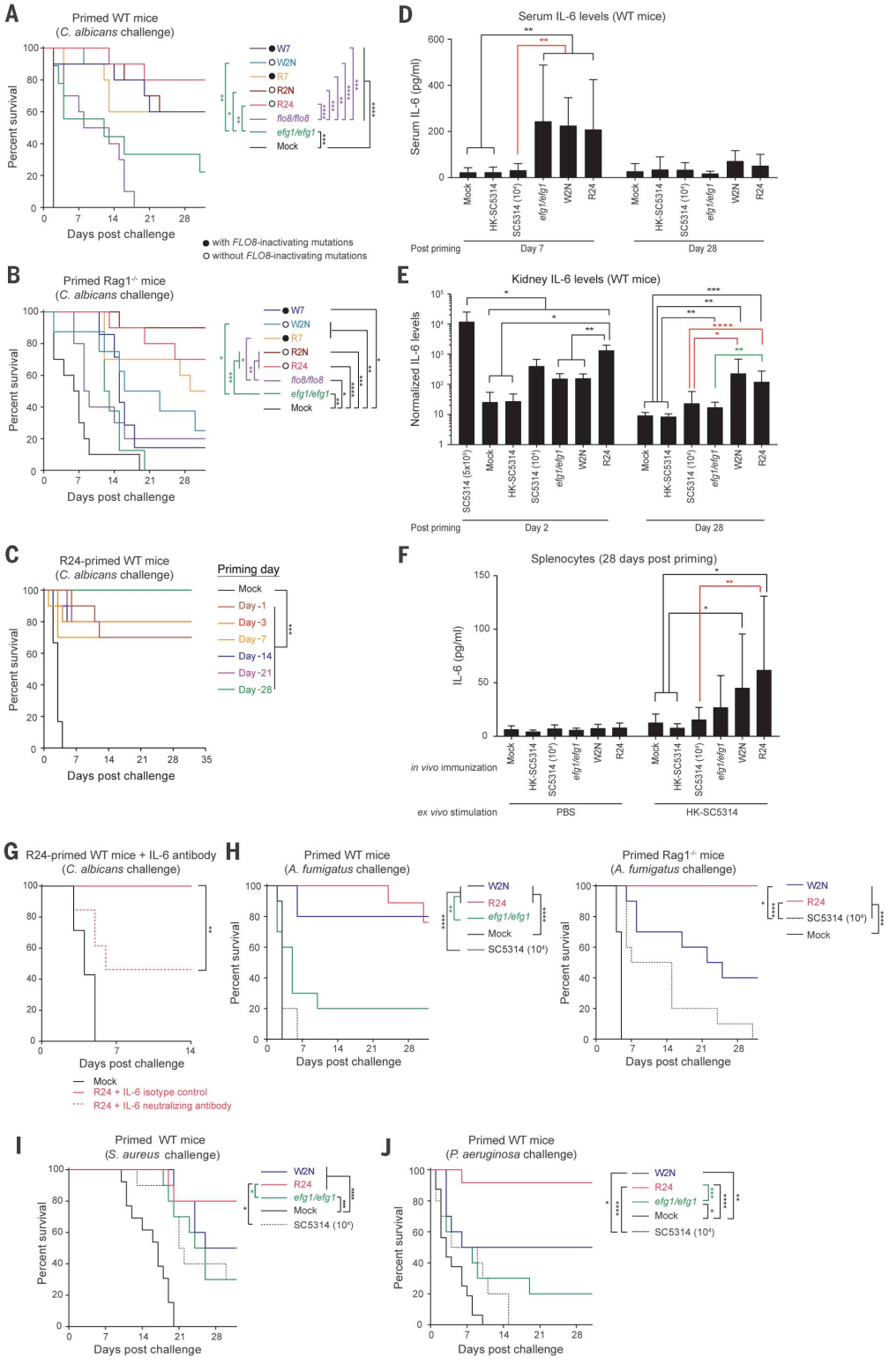


Fig. 4. Gut-evolved *C. albicans* strains confer hosts with broadly cross-protective innate immunity against a wide range of pathogens. (**A**) WT mice systemically primed with w10 evolved strains are significantly protected from systemic candidiasis. Survival is significantly higher in these mice than in mice primed with *efg1*$^{-/-}$ or *flo8*$^{-/-}$ mutants (shown here) or with a sublethal dose of WT *C. albicans* cells (fig. S8A). n = 10 mice per group. (**B**) *Rag1*$^{-/-}$ mice primed with most w10 evolved strains are significantly protected from systemic candidiasis. In most cases, survival is higher than that of mice primed with the *efg1*$^{-/-}$ mutant or with a sublethal dose of WT *C. albicans* cells (fig. S8B). n = 5 to 10 mice per group. In (A) and (B), w10 evolved strains harbored (filled circles) or did not harbor (empty circles) *FLO8*-inactivating mutations. (**C**) WT mice primed with a w10 evolved strain (R24) are significantly protected from systemic candidiasis as early as 1 dpp. n = 10 mice per group. (**D**) Serum IL-6 concentrations of WT mice infected with live w10 strains (W2N or R24) strains are increased at 7 dpp compared with mice infected with a sublethal dose of live (10^4 CFUs) or a full dose of heat-killed (HK) WT (SC5314). n = 4 to 8 mice per group. (**E**) Kidney IL-6 amounts, normalized based on organ weight, of WT mice are increased both at 2 and 28 dpp. n = 4 to 17 mice per group. (**F**) Splenocytes extracted from WT mice 28 dpp with a w10 strain (W2N or R24) produce higher amounts of IL-6 upon ex vivo stimulation with HK-WT SC5314. By contrast, the *efg1*$^{-/-}$ mutant fails to significantly train splenocytes. Cytokines concentrations were measured 48 hours poststimulation. n = 6 to 10 mice per group. Data are means ± SD. (**G**) 50% of mice primed with a w10 strain (R24) succumbed after systemic challenge by WT SC5314 when treated with an IL-6 neutralizing antibody but not a control antibody. n = 7 to 13 mice per group. (**H** to **J**) WT mice primed with a w10 strains (W2N or R24) are significantly protected from systemic challenge with *A. fumigatus* (H), *S. aureus* (I), or *P. aeruginosa* (J). n = 10 to 11 mice per group. In many cases, survival is significantly higher than that of mice primed with the *efg1*$^{-/-}$ mutant or a sublethal dose of WT SC5314. Data were analyzed by log-rank test [(A) to (C) and (G) to (J)] or Mann-Whitney test [(D) to (F)]. For (A) to (J), *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

dose of a fully virulent *C. albicans* strain (SC5314) 28 days later (*22*). As expected, all mock-vaccinated animals died within 3 to 4 days postchallenge (dpc) (Fig. 4A). Consistent with previous reports (*22*), priming with a sublethal dose (1×10^4 CFUs) of WT *C. albicans* SC5314 delayed host mortality, but eventually all animals succumbed to the challenge (fig. S8). By contrast, 60 to 80% of mice primed with a full dose of a w10 strain survived the secondary challenge (Fig. 4A). Importantly, afilamentous *efg1*$^{-/-}$ and *flo8*$^{-/-}$ strains also conferred a statistically significant protection, albeit not as efficiently as the gut-evolved strains. Comparing all strains at the lower dose, the evolved strain R24 protected WT and *Rag1*$^{-/-}$ mice against systemic candidiasis more efficiently than SC5314 (fig. S8, A and B). This enhanced immunity to a secondary challenge could be due to persistence of the primary avirulent strains in the mouse organs (fig. S9). However, the *efg1*$^{-/-}$ strain protected its host less efficiently against secondary infections even though it colonized the mouse kidneys at least as efficiently as gut-evolved strains. Taken together, these results suggest that filamentation loss increases the ability of *C. albicans* to boost host immunity but that gut-evolved strains likely carry additional modifications that further enhance this ability.

Because the increased protection observed in w10-primed mice over those immunized with sublethal doses of WT *C. albicans* correlated with increased total as well as anti–*C. albicans*–specific immunoglobulin G titers in the serum (fig. S10), we repeated above experiments in *Rag1*$^{-/-}$ mice to test the contribution of adaptive immunity to this protective mechanism. Similar to WT mice, *Rag1*$^{-/-}$ mice immunized with w10 strains were significantly more protected from systemic candidiasis than nonimmunized mice. Again, immunization with afilamentous *efg1*$^{-/-}$ or *flo8*$^{-/-}$ strains significantly delayed mortality, although not as efficiently as some of the w10 strains (Fig. 4B). Also gut-colonized *Rag1*$^{-/-}$ pups were significantly protected against systemic candidiasis once they reached adult age, with SC5314-colonized mice achieving ~30% survival and R24-colonized mice >40% survival at 35 dpc (fig. S11). Overall, these data indicate that gut-evolved *C. albicans* strains raise protective immune responses independently of B and T cells.

Whereas adaptive immunity typically requires a few weeks to mount long-lived memory responses, innate immune responses are characterized by a more rapid onset but shorter lifetimes (*23*). Consistent with an innate-like mechanism, significant protection against infection with a fully virulent *C. albicans* strain was achieved as early as 1 day postpriming (dpp) with the R24 gut-evolved *C. albicans* strain (Fig. 4C) and at 3 months postpriming, whereas most w10 strains still conferred WT mice with a significant protection against systemic candidiasis, mouse survival was significantly reduced compared to mice challenged 1 month postpriming with most w10 strains (fig. S8C). Moreover, *Rag1*$^{-/-}$ mice show some protection 3 months after immunization, although this effect was no longer significant for some w10 strains (fig. S8D), suggesting that the innate immune effect could last at least until this time point.

Cytokines play crucial roles in innate immune responses (*21*). Consistently, w10 strains R24 or W2N induced a significant increase in circulating interleukin-6 (IL-6) concentrations 7 dpp, similar to that induced by *efg1*$^{-/-}$ cells (Fig. 4D). Unlike *efg1*$^{-/-}$ cells, however, both gut-evolved strains also increased kidney IL-6 concentrations at 28 dpp (Fig. 4E), as well as IL-6 and tumor necrosis factor–α (TNF-α) production capacity of ex vivo restimulated splenocytes (Fig. 4F and fig. S12). Importantly, kidney IL-6 amounts, albeit increased compared to mock-infected animals, were at least 10-fold lower than those measured using a lethal dose of the WT strain (Fig. 4E), and IL-6 serum concentrations returned to physiological levels at 28 dpp (Fig. 4D), thereby limiting the potential collateral damage of chronic systemic inflammation. Moreover, injecting a neutralizing anti–IL-6 antibody significantly reduced R24-induced protection against systemic candidiasis (Fig. 4G), demonstrating a crucial role for this cytokine in this protective immune response.

The non-antigen-specific nature of innate immunity predicts that priming with a w10 strain should confer broad cross-protection against a wide range of pathogens. To test this, we first intravenously primed naïve mice with a w10 strain and then challenged them systemically with a lethal dose of the unrelated fungus *A. fumigatus*; a Gram-positive bacterium, *Staphylococcus aureus*; or a Gram-negative bacterium, *Pseudomonas aeruginosa*. In all cases, animals immunized with a w10 strain were significantly protected from infection compared with both naïve mice and those immunized with a sublethal dose of WT *C. albicans* (Fig. 4, H to J). Moreover, W2N and R24 protected mice significantly better than *efg1*$^{-/-}$ cells from systemic *A. fumigatus* challenge (Fig. 4H), and R24 protected them more efficiently than *efg1*$^{-/-}$ cells from *P. aeruginosa* infection (Fig. 4J). Confirming that this cross-protection was independent of adaptive immunity, *Rag1*$^{-/-}$ mice were also protected from systemic aspergillosis after immunization with R24 (Fig. 4H).

Altogether, gut-evolved w10 strains were able to raise protective immune responses characterized by rapid onset, short lifetime, increased innate cytokine responses, pathogen-aspecificity, and independence from T and B cells; thus, this protection mechanism resembles several critical aspects of trained immunity (*21*), although other innate immunity mechanisms cannot be discounted. This experimental system hence represents a highly efficient method to generate fungal strains that are not only less pathogenic but also highly immunogenic. These evolved *C. albicans* strains could therefore lead toward safe and effective universal vaccines with a broad spectrum of cross-protection (*24*) and with activity also in individuals with impaired adaptive immunity, such as HIV/AIDS patients or transplant recipients.

Taken together, these findings indicate that adaptive evolution of a microbe in the mammalian gut can alter a host-pathogen interaction into a mutually advantageous relationship, in which both the microbe and the host gain some benefit from their interaction. While the microbe enhances its competitive fitness in the host GI tract, the host augments (albeit temporarily) its resistance against a wide variety of pathogens, though long-term consequences on host fitness could not be investigated here. A similar observation was recently reported in laboratory worms (*25*); however, these organisms lack adaptive immunity and a complex gut microbiome. Our data demonstrate that, although adaptive immunity does not influence this innate selection mechanism, the presence of unperturbed bacterial microbiota interferes with this evolutionary process by shifting the balance between selective forces: Selection by gut bacteria outweighs that by the host, and the filamentation program upturns from a fitness burden to a selective advantage. This model explains why, in its natural niche—the human GI tract, where it is heavily outnumbered by commensal bacteria—*C. albicans* cannot afford to lose this important competition mechanism and why the "virulence" of this pathobiont often emerges only after antibiotic use (*26*), when this program is unintentionally redirected from the microbiota to the host.

REFERENCES AND NOTES

1. D. A. Relman, *Nat. Rev. Microbiol.* **6**, 721–724 (2008).
2. P. W. Ewald, *Ann. N. Y. Acad. Sci.* **503**, 295–306 (1987).
3. D. Ebert, *Science* **282**, 1432–1435 (1998).
4. K. L. Hoang, L. T. Morran, N. M. Gerardo, *Front. Microbiol.* **7**, 1444 (2016).
5. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, *Nat. Rev. Microbiol.* **14**, 20–32 (2016).
6. J. V. Pierce, C. A. Kumamoto, *mBio* **3**, e00117-12 (2012).
7. K. Pande, C. Chen, S. M. Noble, *Nat. Genet.* **45**, 1088–1091 (2013).
8. M. P. Hirakawa *et al*., *Genome Res.* **25**, 413–425 (2015).
9. A. Y. Koh, *Eukaryot. Cell* **12**, 1416–1422 (2013).
10. L. M. Pope, G. T. Cole, M. N. Guentzel, L. J. Berry, *Infect. Immun.* **25**, 702–707 (1979).
11. L. Böhm *et al*., *PLOS Pathog.* **13**, e1006699 (2017).
12. F. Cao *et al*., *Mol. Biol. Cell* **17**, 295–307 (2006).
13. J. Y. Liu *et al*., *Microbiol. Res.* **178**, 1–8 (2015).
14. A. Forche *et al*., *mBio* **2**, e00129-11 (2011).
15. A. Forche, P. T. Magee, A. Selmecki, J. Berman, G. May, *Genetics* **182**, 799–811 (2009).
16. H. J. Lo *et al*., *Cell* **90**, 939–949 (1997).
17. A. Lüttich, S. Brunke, B. Hube, I. D. Jacobsen, *PLOS ONE* **8**, e64482 (2013).
18. S. Brunke *et al*., *PLOS Pathog.* **10**, e1004478 (2014).
19. A. Wartenberg *et al*., *PLOS Genet.* **10**, e1004824 (2014).
20. T. T. Jiang *et al*., *Cell Host Microbe* **22**, 809–816.e4 (2017).
21. M. G. Netea *et al*., *Science* **352**, aaf1098 (2016).
22. J. Quintin *et al*., *Cell Host Microbe* **12**, 223–232 (2012).
23. S. Akira, S. Uematsu, O. Takeuchi, *Cell* **124**, 783–801 (2006).
24. H. S. Goodridge *et al*., *Nat. Rev. Immunol.* **16**, 392–400 (2016).
25. K. C. King *et al*., *ISME J.* **10**, 1915–1924 (2016).
26. M. S. Seelig, *Am. J. Med.* **40**, 887–917 (1966).
27. N. Pavelka, G. Tso, A. Tan, Raw data of 'Experimental evolution of a fungal pathogen into a gut symbiont'. Open Science Framework (2018).
28. P. E. Sudbery, *Nat. Rev. Microbiol.* **9**, 737–748 (2011).

ACKNOWLEDGMENTS

We thank A. Tan, Y. S. Lim, and N. Tay for technical assistance; Singapore Immunology Network (SIgN) Mouse Core for providing mice; IMCB Advanced Molecular Pathology

Laboratory for histology service; C. B. Ong for veterinary pathology evaluation; J. Berman for *C. albicans* strains; and M. G. Netea, Y. Wang, and G. De Libero for scientific discussions and feedback on the manuscript. Sequencing was performed by the Genome Institute of Singapore Genome Technology and Biology Group, Singapore. This study was supported by Agency for Science, Technology, and Research (A*STAR) Investigatorship awards JCO/1437a00117 to N.P. and JCO/1437a00119 to G.R. and by core funding from SIgN. **Author contributions:** G.H.W.T. and N.P. designed the study; K.G.S. and F.Z. prepared sequencing libraries; G.H.W.T., X.S., J.A.R.-C., A.S.M.T., T.G.T., G.T.T.L., and G.C.L. performed all other experiments; G.R. designed and supervised cloning experiments; M.Y., W.L., M.P., G.H.W.T., and N.P. analyzed sequencing data; G.H.W.T., J.A.R.-C., and N.P. analyzed all other data and wrote the manuscript; and N.P. supervised the project. **Competing interests:** G.H.W.T., X.S., and N.P. are inventors on Singapore patent application no. 10201702472T and International PCT application no. PCT/SG2018/050142 submitted by A*STAR that cover methods of generating attenuated fungi and uses thereof. **Data and materials availability:** The gut-evolved *C. albicans* strains are available from N.P. under a material transfer agreement with SIgN, A*STAR. Sequencing data has been deposited at NCBI under SRA accession number SRP116719. All other raw data and code can be accessed at (*27*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/589/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S6
References (*29*–*39*)

19 January 2018; accepted 12 September 2018
10.1126/science.aat0537

STRUCTURAL BIOLOGY

Structural basis of the nucleosome transition during RNA polymerase II passage

Tomoya Kujirai[1,2]*, Haruhiko Ehara[2]*, Yuka Fujino[1,3], Mikako Shirouzu[2], Shun-ichi Sekine[2]†, Hitoshi Kurumizaka[1,2,3]†

Genomic DNA forms chromatin, in which the nucleosome is the repeating unit. The mechanism by which RNA polymerase II (RNAPII) transcribes the nucleosomal DNA remains unclear. Here we report the cryo–electron microscopy structures of RNAPII-nucleosome complexes in which RNAPII pauses at the superhelical locations SHL(−6), SHL(−5), SHL(−2), and SHL(−1) of the nucleosome. RNAPII pauses at the major histone-DNA contact sites, and the nucleosome interactions with the RNAPII subunits stabilize the pause. These structures reveal snapshots of nucleosomal transcription, in which RNAPII gradually tears DNA from the histone surface while preserving the histone octamer. The nucleosomes in the SHL(−1) complexes are bound to a "foreign" DNA segment, which might explain the histone transfer mechanism. These results provide the foundations for understanding chromatin transcription and epigenetic regulation.

In eukaryotes, genomic DNA is packaged as chromatin, in which the nucleosome is the basic unit (*1*). In the nucleosome, a 145– to 147–base pair (bp) DNA segment is stably wrapped around a histone octamer, composed of two each of histone H2A-H2B and H3-H4 heterodimers (*2*). Because the nucleosome has a highly stable architecture, it strongly affects genomic DNA functions, such as transcription (*3*–*5*). RNA polymerase II (RNAPII), a multisubunit molecular machine (*6*), transcribes protein-coding genes, but the mechanism of nucleosome transcription remains obscure.

To address this problem, we performed cryo–electron microscopy (cryo-EM) single-particle analyses of transcribing RNAPII-nucleosome complexes. In cells, RNA elongation by RNAPII is stalled at multiple sites within a nucleosome up to the nucleosomal dyad (*7*). After passing through the nucleosomal dyad, RNAPII transcribes the DNA without obvious impediments (*7*), because the histones may be removed from the DNA ahead of the RNAPII. To avoid the histone dismantling by RNAPII, we designed a nucleosomal template with a T-less sequence region, by which transcription elongation is stalled at the superhelical location SHL(−1), a natural RNAPII pausing site located ~10 bp upstream of the nucleosomal dyad, in the presence of 3′–deoxyadenosine triphosphate (fig. S1A). Widom 601, an artificial sequence with high nucleosome positioning power, was used for the design (*8*). After reconstitution of the human nucleosome with the DNA, a linker DNA containing the transcription bubble was ligated to one end of the nucleosomal DNA (fig. S1, A to D). RNA elongation was performed by RNAPII from the yeast *Komagataella pastoris* (fig. S1E). RNAPII paused predominantly at the entry of the nucleosome [SHL(−5)], reflecting the rate-limiting step of the nucleosomal transcription in vitro (fig. S1F). By contrast, RNAPII elongated RNA until reaching SHL(−1) in the presence of the transcription elongation factor TFIIS (fig. S1E), which facilitates the nucleosomal transcription (fig. S1F) (*9*, *10*).

The obtained transcribing complexes, containing RNAPII, nucleosomes, and transcribed RNAs, were partially purified by the sucrose-glutaraldehyde gradient ultracentrifugation (GraFix) method (*11*). The RNAs within the RNAPII-nucleosome complex mixture were then analyzed by denaturing polyacrylamide gel electrophoresis (Fig. 1A). We observed substantial pausing of RNAPII near the entry [SHL(−5)] and before the dyad [SHL(−1)] of the nucleosome (Fig. 1A). These major pausing sites correspond well to the in vivo pausing sites within gene-body nucleosomes (*7*). In addition, we detected minor but clear RNA bands corresponding to the paused RNAPII at SHL(−6) and SHL(−2) (Fig. 1A).

We then collected cryo-EM images of the transcription reaction mixtures, and extensive three-dimensional classifications delineated seven distinct states of the RNAPII-nucleosome complexes, which belong to the complexes paused around SHL(−6), SHL(−5), SHL(−2), and SHL(−1) (Fig. 1, B to E; figs. S2 to S6; tables S1 and S2; and movie S1). These structures revealed the nucleosome transition during the RNAPII passage, whereas no major conformational change was observed in the RNAPII (see materials and methods). In these complexes, RNAPII is stalled at the major contact sites between the DNA and the core histones (Fig. 2). These sites are either the H2A-H2B contact site [SHL(−5)] or the H3-H4 contact sites [SHL(−6), SHL(−2), and SHL(−1)] (Fig. 2). These strong histone-DNA interactions should be the major cause of the RNAPII pausing and thus explain the previous observations that the histone mutations that weaken histone-DNA interactions relieve the RNAPII pausing (*12*). The nucleosome orientations relative to RNAPII are similar among these structures (fig. S7, A to E), in which the curved nucleosomal DNA fits on the RNAPII surface, interacting with the clamp head domain of the Rpb1 subunit (fig. S7F). These RNAPII-DNA interactions seem to stabilize the RNAPII stalled at each SHL.

In the RNAPII-nucleosome complexes stalled around SHL(−6) and SHL(−5), the RNAPII is located near the entrance of the nucleosome. In the SHL(−6) complex, the nucleosomal DNA is entirely wrapped around the histone octamer (Figs. 1B and 2A). In the SHL(−5) complex, the RNAPII has advanced by ~20 bp and the DNA end of the nucleosome is torn away from the N-terminal part of histone H3 (Figs. 1C and 2B). The H3 N-terminal region contains many modification sites, which may affect the DNA affinity for the histone surface (*13*–*15*). A similar 20-bp detachment reportedly occurs with nucleosome remodelers, such as CHD1 and INO80 (*16*–*18*). Therefore, the histone modifications and the nucleosome remodelers may modulate the nucleosome accessibility of RNAPII at the nucleosomal entry site.

Before the dyad, RNAPII is stalled at SHL(−2) and SHL(−1) of the nucleosome (Figs. 1, D and E, and 2, C and D). In the SHL(−2) and SHL(−1) complexes, ~50- and 60-bp DNA regions are peeled away from the histone surface by RNAPII, respectively. An H2A-H2B dimer is reportedly released from the nucleosome when a 40-bp stretch of the nucleosomal DNA is removed (*19*). However, the H2A-H2B dimer density is clearly observed in the SHL(−1) and SHL(−2) complexes (Fig. 3, A and B). Consistently, we observed no nucleosome particles lacking H2A-H2B during the classification processes. In the SHL(−2) complex, we observed a density connection between the Rpb2 lobe domain of RNAPII and the H2A-H2B dimer, and this interaction may retain the H2A-H2B dimer in the nucleosome (Fig. 3C).

The majority of the complexes stalled at SHL(−1) contained a ~60-bp DNA segment of unknown origin ("foreign" DNA), which was bound to the DNA-peeled region of the nucleosome (fig. S8, A to C). We also observed foreign DNA association in the non–cross-linked sample (fig. S9), indicating that the DNA binding is not due to cross-linking. Further analyses of these complexes yielded three distinguishable classes with the foreign DNA: the SHL(−1), tilted SHL(−1), and $SHL(-1)_{+1}$ complexes (fig. S8, A to C). In the $SHL(-1)_{+1}$ complex, RNAPII proceeded by 1 bp (fig. S8C) and rotated by ~36° (fig. S8D

[1]Laboratory of Chromatin Structure and Function, Institute for Quantitative Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. [2]RIKEN Center for Biosystems Dynamics Research, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [3]Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan.
*These authors contributed equally to this work.
†Corresponding author. Email: kurumizaka@iam.u-tokyo.ac.jp (H.K.); shunichi.sekine@riken.jp (S.S.)

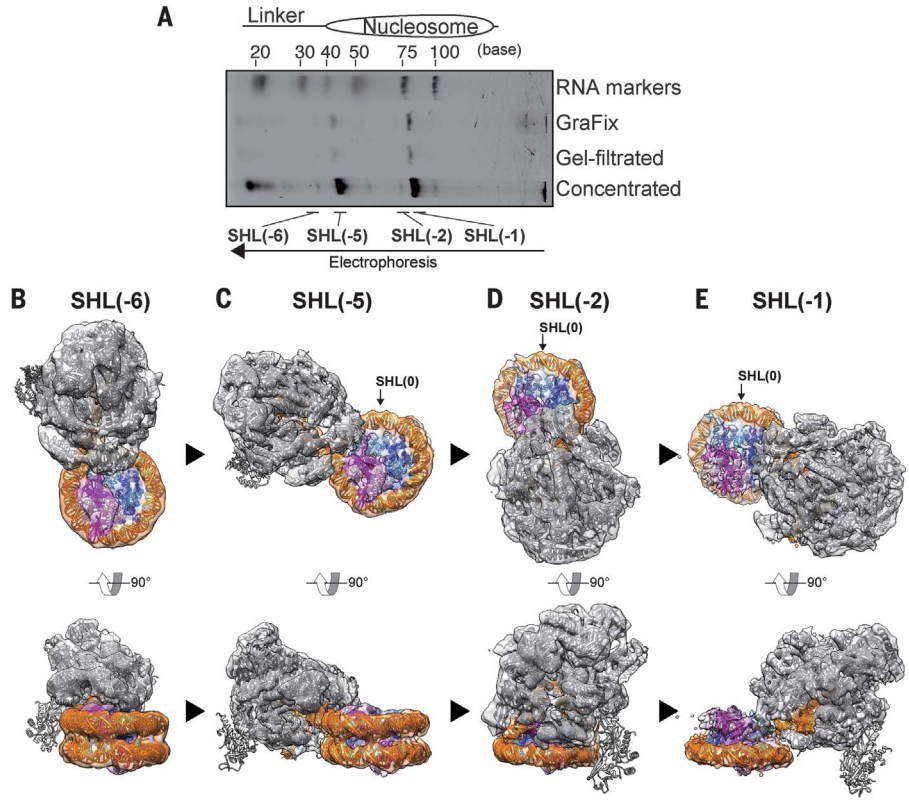


Fig. 1. Cryo-EM structures of the transcribing RNAPII-nucleosome complexes. (**A**) Elongated RNAs in the purified transcribing RNAPII-nucleosome complexes were analyzed by denaturing gel electrophoresis. After the transcription reaction, the complexes were fractionated by the GraFix method (GraFix lane), desalted on a gel filtration column (gel-filtrated lane), and concentrated (concentrated lane). (**B** to **E**) Cryo-EM densities of the RNAPII-nucleosome complexes paused at SHL(−6), SHL(−5), SHL(−2), and SHL(−1) with fitted structural models. The density for the Rpb4/Rpb7 stalk of RNAPII is weak, because of its inherent flexibility (*31*, *32*). Focused classification clearly reveals the stalk density (fig. S11). RNAPII, DNA, H2A, H2B, H3, and H4 are colored gray, orange, magenta, pink, light blue, and blue, respectively. The arrows indicate the SHL(0) positions. Arrowheads represent transitions between the complexes.

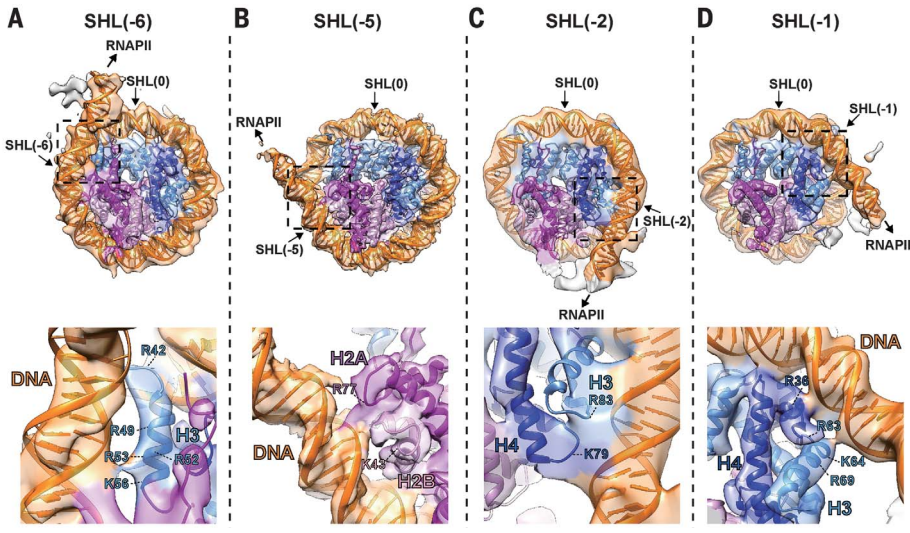


Fig. 2. Structural transition of the nucleosome. Improved density maps after focused refinement (see materials and methods) for the SHL(−6) (**A**), SHL(−5) (**B**), SHL(−2) (**C**), and SHL(−1) (**D**) complexes are shown with the atomic models. Top panels indicate overall nucleosome structures. These views are the same orientations as those in Fig. 1, B to E (top panels). Close-up views of the regions enclosed by the dashed squares in the top panels are presented in the bottom panels. The arginine (R) and lysine (K) residues that potentially contact the DNA at the RNAPII pausing sites are highlighted.

and movie S2) compared with the positions in the SHL(−1) complex. The tilted SHL(−1) complex seems to be an intermediate between the SHL(−1) and SHL$(-1)_{+1}$ complexes (fig. S8, B and D, and movie S2). In these complexes, RNAPII is stalled because of the DNA contacts with histones H3-H4. The acetylation of Lys^{64} in H3, which may reduce the H3-DNA contact at SHL(−1) (Fig. 2D), is reportedly accumulated at the transcription start sites of active genes (*20*). Therefore, SHL(−1) is probably another important site for transcriptional regulation coupled with histone modifications.

The foreign DNA is derived from a linker DNA region of other nucleosomes or disassembled nucleosomes. We observed the former particles in the raw cryo-EM images (fig. S10). This suggests that the DNA-peeled region (mainly H2A-H2B) has a propensity to associate with DNA to regenerate nucleosome-like structures. Therefore, the foreign DNA could serve as an intermediate for histone transfer to other DNA regions in cis or trans. It has been proposed that the histone is transferred from ahead of the transcribing RNA polymerase to behind it via a "template looping" intermediate (*21*–*24*). If a region of the upstream DNA behind RNAPII interacted with the DNA-peeled part of H2A-H2B, like the foreign DNA observed here, then it would function as a histone transfer intermediate.

The present study unveiled the sequential transitions of nucleosomes during RNAPII passage (Fig. 4). The transcribing RNAPII first encounters the nucleosome [step 1, SHL(−6)] and begins its invasion with the peeling of a helical pitch of the DNA segment from the histone surface [step 2, SHL(−5)]. The RNAPII elongates the RNA with continuous peeling of the nucleosomal DNA and stalls at SHL(−2) of the nucleosome (step 3).

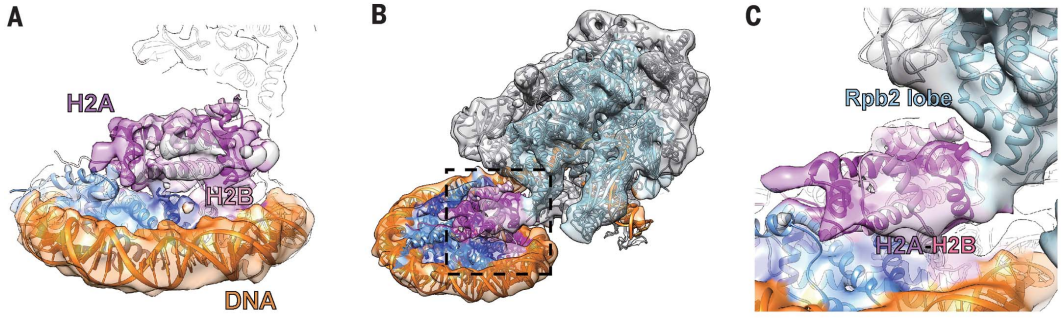


Fig. 3. Interaction between an H2A-H2B dimer and RNAPII. (**A**) A close-up view of an H2A-H2B dimer in the nucleosomes of the SHL(−1) complex. Cryo-EM density maps of the nucleosomes in the SHL(−1) complex, fitted with the structure, are shown. (**B**) The overall density map with the atomic model structure of the SHL(−2) complex. Rpb2 of RNAPII is colored pale blue. (**C**) The region enclosed by a dashed square in (B) is enlarged.

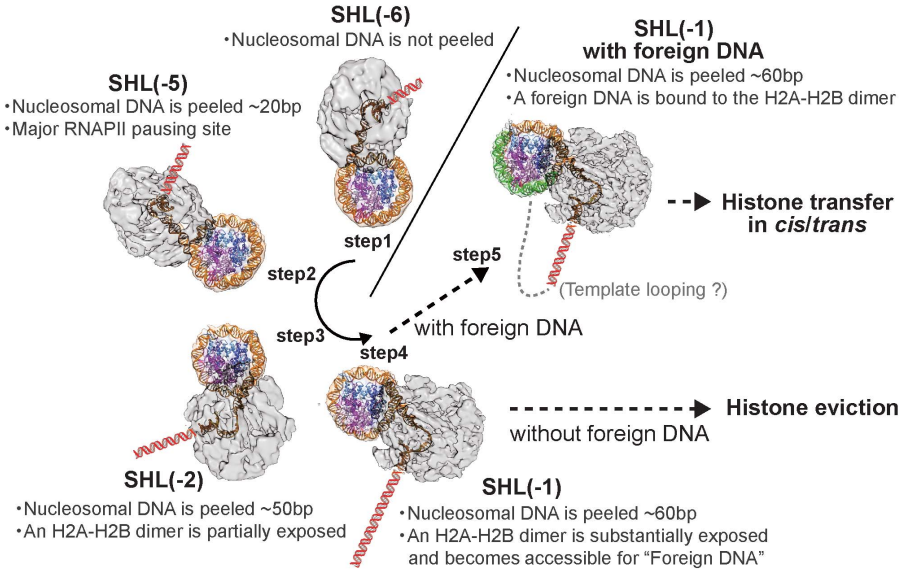


Fig. 4. Scheme of transcription elongation through a nucleosome by RNAPII. RNAPII encounters a nucleosome [SHL(−6), step 1] and then proceeds to SHL(−5), which is a major pausing site for RNAPII (step 2). After passing through SHL(−5), the RNAPII pauses at SHL(−2) (step 3) and proceeds to SHL(−1) (step 4). In the SHL(−2) and SHL(−1) complexes, an H2A-H2B dimer is exposed. Then foreign DNA is bound to the DNA-peeled region of the nucleosome in the complex (step 5). Histones may be removed from the region ahead of the RNAPII, by transfer in cis or trans or by eviction. The red DNA cartoons indicate possible orientations of the upstream DNAs.

Finally, RNAPII stalls at SHL(−1) (step 4), where the nucleosome interacts with the foreign DNA (step 5). At or beyond SHL(−1), transcription results in either histone transfer to the upstream DNA behind RNAPII or histone eviction from the DNA. We have revealed the structures of the RNAPII-nucleosome complexes formed in the presence of the transcription elongation factor TFIIS. However, in cells, several other elongation factors, including Spt4, Spt5, Spt6, PAF1C, and Elf1, interact with RNAPII (*9*, *25*–*27*) and function in proper transcription in chromatin (*28*–*30*). Further studies are needed to understand the mechanism of chromatin transcription by RNAPII.

The nucleosome-dependent transcriptional pause is important for gene regulation. The first (+1) nucleosome is implicated in the promoter-proximal pausing of transcription, which regulates nucleosomal entry by RNAPII (*7*). The partially DNA-peeled nucleosome within the paused complexes should be the target of histone chaperones, nucleosome remodelers, and modification enzymes. Our present RNAPII-nucleosome complex structures at the major pause sites shed new light on such regulation mechanisms.

REFERENCES AND NOTES

1. A. Wolffe, *Chromatin: Structure and Function* (Academic Press, ed. 3, 1998).
2. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **389**, 251–260 (1997).
3. S. S. Teves, C. M. Weber, S. Henikoff, *Trends Biochem. Sci.* **39**, 577–586 (2014).
4. Y. Lorch, R. D. Kornberg, *Q. Rev. Biophys.* **50** (e5), e5 (2017).
5. W. K. M. Lai, B. F. Pugh, *Nat. Rev. Mol. Cell Biol.* **18**, 548–562 (2017).
6. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863–1876 (2001).
7. C. M. Weber, S. Ramachandran, S. Henikoff, *Mol. Cell* **53**, 819–830 (2014).
8. P. T. Lowary, J. Widom, *J. Mol. Biol.* **276**, 19–42 (1998).
9. H. Ehara *et al.*, *Science* **357**, 921–924 (2017).
10. M. L. Kireeva *et al.*, *Mol. Cell* **18**, 97–108 (2005).
11. B. Kastner *et al.*, *Nat. Methods* **5**, 53–55 (2008).
12. F. K. Hsieh, M. Fisher, A. Ujvári, V. M. Studitsky, D. S. Luse, *EMBO Rep.* **11**, 705–710 (2010).
13. B. D. Strahl, C. D. Allis, *Nature* **403**, 41–45 (2000).
14. T. Kouzarides, *Cell* **128**, 693–705 (2007).
15. A. J. Bannister, T. Kouzarides, *Cell Res.* **21**, 381–395 (2011).
16. L. Farnung, S. M. Vos, C. Wigge, P. Cramer, *Nature* **550**, 539–542 (2017).
17. S. Eustermann *et al.*, *Nature* **556**, 386–390 (2018).
18. R. Ayala *et al.*, *Nature* **556**, 391–395 (2018).
19. Y. Arimura, H. Tachiwana, T. Oda, M. Sato, H. Kurumizaka, *Biochemistry* **51**, 3302–3309 (2012).
20. V. Di Cerbo *et al.*, *eLife* **3**, e01632 (2014).
21. V. M. Studitsky, D. J. Clark, G. Felsenfeld, *Cell* **76**, 371–382 (1994).
22. J. Bednar, V. M. Studitsky, S. A. Grigoryev, G. Felsenfeld, C. L. Woodcock, *Mol. Cell* **4**, 377–386 (1999).
23. C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, C. Bustamante, *Science* **325**, 626–628 (2009).
24. L. Bintu *et al.*, *Nat. Struct. Mol. Biol.* **18**, 1394–1399 (2011).
25. C. Bernecky, J. M. Plitzko, P. Cramer, *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).
26. Y. Xu *et al.*, *Nat. Commun.* **8**, 15741 (2017).
27. S. M. Vos *et al.*, *Nature* **560**, 607–612 (2018).
28. G. A. Hartzog, T. Wada, H. Handa, F. Winston, *Genes Dev.* **12**, 357–369 (1998).
29. D. Prather, N. J. Krogan, A. Emili, J. F. Greenblatt, F. Winston, *Mol. Cell. Biol.* **25**, 10122–10135 (2005).
30. J. Kim, M. Guermah, R. G. Roeder, *Cell* **140**, 491–503 (2010).
31. H. Ehara, T. Umehara, S. I. Sekine, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **487**, 230–235 (2017).
32. C. Bernecky, F. Herzog, W. Baumeister, J. M. Plitzko, P. Cramer, *Nature* **529**, 551–554 (2016).

ACKNOWLEDGMENTS

We thank K. Katsura and T. Yokoyama (RIKEN) for their help with the RNAPII preparation and cryo-EM analyses and Y. Iikura (University of Tokyo) for her assistance. **Funding:** This work was supported in part by the RIKEN Dynamic Structural Biology project (to M.S., S.S., and H.K.); JSPS KAKENHI grants JP18H05534 (to H.K.), JP25116002 (to H.K.), and JP15H04344 (to S.S.); JST CREST grant JPMJCR16G1 (to H.K.); and the Platform Project for Supporting Drug Discovery and Life Science Research (BINDS) from AMED under grants JP18am0101076 (to H.K.) and JP18am0101082 (to M.S.). **Author contributions:** T.K. and Y.F. prepared the RNAPII-nucleosome complexes and performed biochemical analyses. H.E., T.K., M.S., and S.S. performed cryo-EM analyses. S.S. and H.K. conceived, designed, and supervised all of the work. T.K., H.E., S.S., and H.K. wrote the paper. All of the authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The cryo-EM reconstructions and atomic models of the RNAPII-nucleosome complexes have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank

(PDB) under the following accession codes: EMD-6981 and PDB ID 6A5O for the RNAPII complex stalled at SHL(−6) of the nucleosome; EMD-6982 and PDB ID 6A5P for the RNAPII complex stalled at SHL(−5) of the nucleosome; EMD-6983 and PDB ID 6A5P for the RNAPII complex stalled at SHL(−2) of the nucleosome; EMD-6984 and PDB ID 6A5T for the RNAPII complex stalled at SHL(−1) of the nucleosome; EMD-6980 and PDB ID 6A5L for the RNAPII complex stalled at SHL(−1) of the nucleosome, with foreign DNA; EMD-6985 and PDB ID 6A5U for the RNAPII complex stalled at SHL(−1) of the nucleosome, with foreign DNA, tilt conformation; and EMD-6986 and PDB ID 6A5V for the RNAPII complex stalled at SHL$(-1)_{+1}$ of the nucleosome, with foreign DNA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/595/suppl/DC1

Materials and Methods
Figs. S1 to S11
Tables S1 and S2
References (*33*–*47*)
Movies S1 and S2

3 August 2018; accepted 19 September 2018
Published online 4 October 2018
10.1126/science.aau9904

ANTI-FLU THERAPY

Universal protection against influenza infection by a multidomain antibody to influenza hemagglutinin

Nick S. Laursen[1]*, Robert H. E. Friesen[2]†, Xueyong Zhu[1], Mandy Jongeneelen[3], Sven Blokland[3], Jan Vermond[4], Alida van Eijgen[4], Chan Tang[3], Harry van Diepen[4], Galina Obmolova[2], Marijn van der Neut Kolfschoten[3], David Zuijdgeest[3], Roel Straetemans[5], Ryan M. B. Hoffman[1], Travis Nieusma[1], Jesper Pallesen[1], Hannah L. Turner[1], Steffen M. Bernard[1], Andrew B. Ward[1], Jinquan Luo[2], Leo L. M. Poon[6], Anna P. Tretiakova[7]‡, James M. Wilson[7], Maria P. Limberis[7], Ronald Vogels[3], Boerries Brandenburg[3], Joost A. Kolkman[8]§, Ian A. Wilson[1,9]§

Broadly neutralizing antibodies against highly variable pathogens have stimulated the design of vaccines and therapeutics. We report the use of diverse camelid single-domain antibodies to influenza virus hemagglutinin to generate multidomain antibodies with impressive breadth and potency. Multidomain antibody MD3606 protects mice against influenza A and B infection when administered intravenously or expressed locally from a recombinant adeno-associated virus vector. Crystal and single-particle electron microscopy structures of these antibodies with hemagglutinins from influenza A and B viruses reveal binding to highly conserved epitopes. Collectively, our findings demonstrate that multidomain antibodies targeting multiple epitopes exhibit enhanced virus cross-reactivity and potency. In combination with adeno-associated virus–mediated gene delivery, they may provide an effective strategy to prevent infection with influenza virus and other highly variable pathogens.

Seasonal influenza epidemics cause worldwide morbidity and mortality (*1*), whereas the vast reservoir of influenza A viruses in aquatic birds represents continual pandemic threats (*2*–*4*). Vaccines remain essential for influenza prevention, but their efficacy is substantially reduced in the elderly, who are at increased risk of influenza-related complications (*3*, *5*, *6*). Annual selection of vaccine strains presents many challenges, and a poor match with circulating viruses can result in suboptimal effectiveness (*7*). Moreover, most vaccine-induced antibodies are directed against the highly variable head region of hemagglutinin (HA) and are strain specific. However, broadly neutralizing antibodies (bnAbs) targeting influenza HA have been isolated and characterized (*8*). Several bnAbs have entered clinical trials as therapeutic agents, but their use in influenza prophylaxis remains elusive because of the incomplete coverage against circulating human influenza A and B viruses, which necessitates administration of a bnAb cocktail, and the need for multiple, high-dose injections for protection throughout the entire influenza season. High serum bnAb levels are required because of poor distribution to the upper airways. We present an alternative strategy for long-lasting protection in which single-domain antibodies (sdAbs) (*9*) with influenza A or B reactivity are linked together into a multidomain antibody (MDAb) and expressed at the nasopharyngeal mucosa through the intranasal administration of a recombinant adeno-associated virus (AAV) vector (*10*, *11*) encoding the MDAb transgene.

Broadly neutralizing sdAbs were obtained by immunizing llamas with influenza vaccine and H7 and H2 recombinant HA (rHA) (*12*). HA cross-reactive sdAbs were isolated from the sdAb [the single variable domain of a heavy-chain-only camelid antibody (V_HH)] repertoires of the immunized llamas by phage display with various cross-selection strategies on rHAs from different influenza subtypes. We isolated two influenza A (SD36 and SD38) and two influenza B (SD83 and SD84) sdAbs and analyzed their in vitro neutralizing activity (Fig. 1). SD36 potently neutralized influenza A group 2 (H3, H4, H7, and

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]Janssen Research and Development, Spring House, PA 19002, USA. [3]Janssen Vaccines and Prevention, Archimedesweg 4-6, 2333 CN, Leiden, Netherlands. [4]Janssen Prevention Center, Archimedesweg 6, 2333 CN, Leiden, Netherlands. [5]Quantitative Sciences, Janssen Pharmaceutical Companies of Johnson and Johnson, Turnhoutseweg 30, 2340 Beerse, Belgium. [6]Center of Influenza Research and School of Public Health, The University of Hong Kong, Hong Kong SAR, China. [7]Gene Therapy Program, Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [8]Janssen Infectious Diseases, Turnhoutseweg 30, 2340, Beerse, Belgium. [9]Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.
*Present address: Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark. †Present address: Ablynx, a Sanofi company, Technologiepark 21, 9052 Zwijnaarde, Belgium. ‡Present address: Pfizer, Rare Disease Research Unit, 610 Main Street, Cambridge, MA 02139, USA.
§Corresponding author. Email: wilson@scripps.edu (I.A.W.); jkolkman@its.jnj.com (J.A.K.)

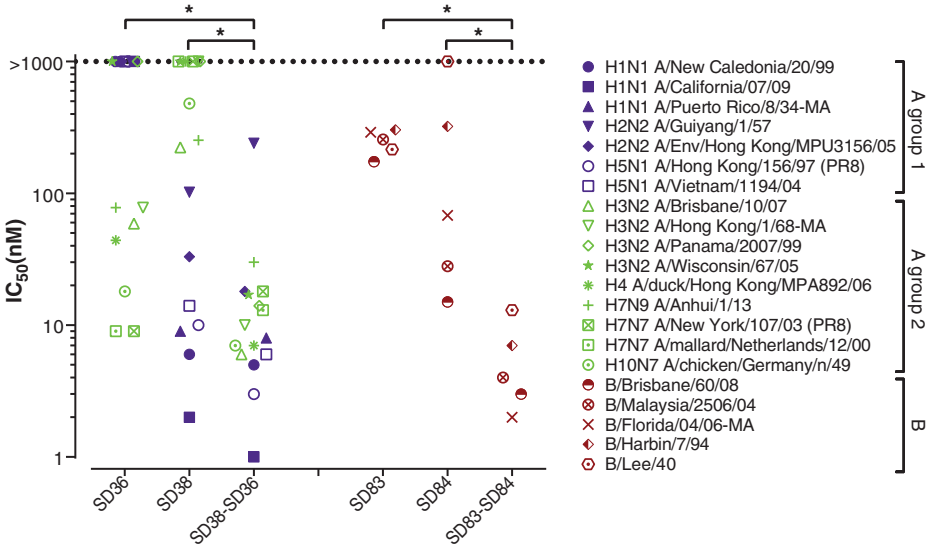


Fig. 1. In vitro neutralization of influenza A and B viruses by individual and genetically fused sdAbs. In vitro potencies of SD36, SD38, SD83, and SD84 and genetically fused sdAbs SD38-SD36 and SD83-SD84 against selected influenza A and B viruses. Both SD38-SD36 and SD83-SD84 are significantly more potent (*$P < 0.05$) than each of their individual components (comparisons are shown via brackets) (see materials and methods in the supplementary materials). Data are representative of at least three independent experiments performed in quadruplicate.

H10) but not group 1 (H1, H2, and H5) viruses, whereas SD38 potently neutralized group 1 (H1, H2, and H5) and some group 2 (H3, H7, and H10) viruses, albeit with lower potency. SD84 and SD83 neutralized representative viruses from both influenza B lineages.

To elucidate the molecular basis for the broad HA recognition, we determined the crystal structures of SD38 and SD83 to 2.0-Å resolution; SD36, SD38, and SD83 with different HAs to 2.2 to 2.8 Å; humanized SD84 (SD84h) to 0.94 Å; and SD84h with HA (B/Brisbane/60/08) and bnAb CR9114 to 4.1 Å (*13*) (Fig. 2 and table S1). A cryo–electron microscopy (EM) reconstruction of SD84 with B/Massachusetts/02/12 HA and CR9114 was determined to 7.1-Å resolution (fig. S1). All sdAbs except SD84 recognized the HA stem, with three sdAbs bound per trimer (Fig. 2). SD83 contacted conserved residues in the fusion subdomain (Fig. 2A), with complementarity-determining region 2 (CDR2) and CDR3 framing the epitope consisting of HA1 residues 30 to 32, Lys^{45}, Asp^{291}, Asn^{301}, and Pro^{305}; the HA2 A helix; and N-linked glycans at Asn^{301} and Asn^{330} (Fig. 2A). The SD83 epitope was highly conserved, with all contact residues being >99% identical in influenza B viruses (Fig. 2A and table S2). SD83 buried a relatively small surface area (588 Å^2) on HA, which potentially presented fewer possibilities for escape. In contrast, SD84 bound a conserved epitope in the HA head around the receptor binding site (Fig. 2B and table S2). SD36 and SD38 also recognized conserved epitopes that partially overlap with influenza A stem epitopes of bnAbs CR9114, CR6261, and FI6v3 and, to a lesser extent, those of CR8043 and CR8020 (*13*–*16*). SD36 CDR2 and CDR3 contacted the HA2 A helix, HA1 Asn^{291} and Thr^{318}, and HA1 Arg^{32} of the adjacent protomer (Fig. 2C and fig. S2). CDR1 and CDR3 interacted with HA2 residues 19 and 20, and Framework2 (FR2) interacted with HA1 Lys^{54} and Glu^{278}. In SD38, the 17-residue CDR3 dominated and bound to a shallow groove formed by the HA2 A helix, HA2 residues 18 to 21, and HA1 His^{38} and Thr^{318} (Fig. 2D and figs. S2 and S3) (*17*).

Thus, SD38 and SD36 made extensive contacts to the A helix and other highly conserved residues (tables S3 and S4). However, 33% of H3 viruses had an Asp^{46}→Asn (D46N) polymorphism and were not neutralized by SD36 (table S5). Asn^{46} would disrupt a salt bridge and hydrogen bond network with SD36 (fig. S4). SD36 bound H1 HA (A/Brisbane/59/07) with a 290 nM dissociation constant (K_d) (table S6) but did not neutralize the corresponding virus in vitro [median inhibitory concentration

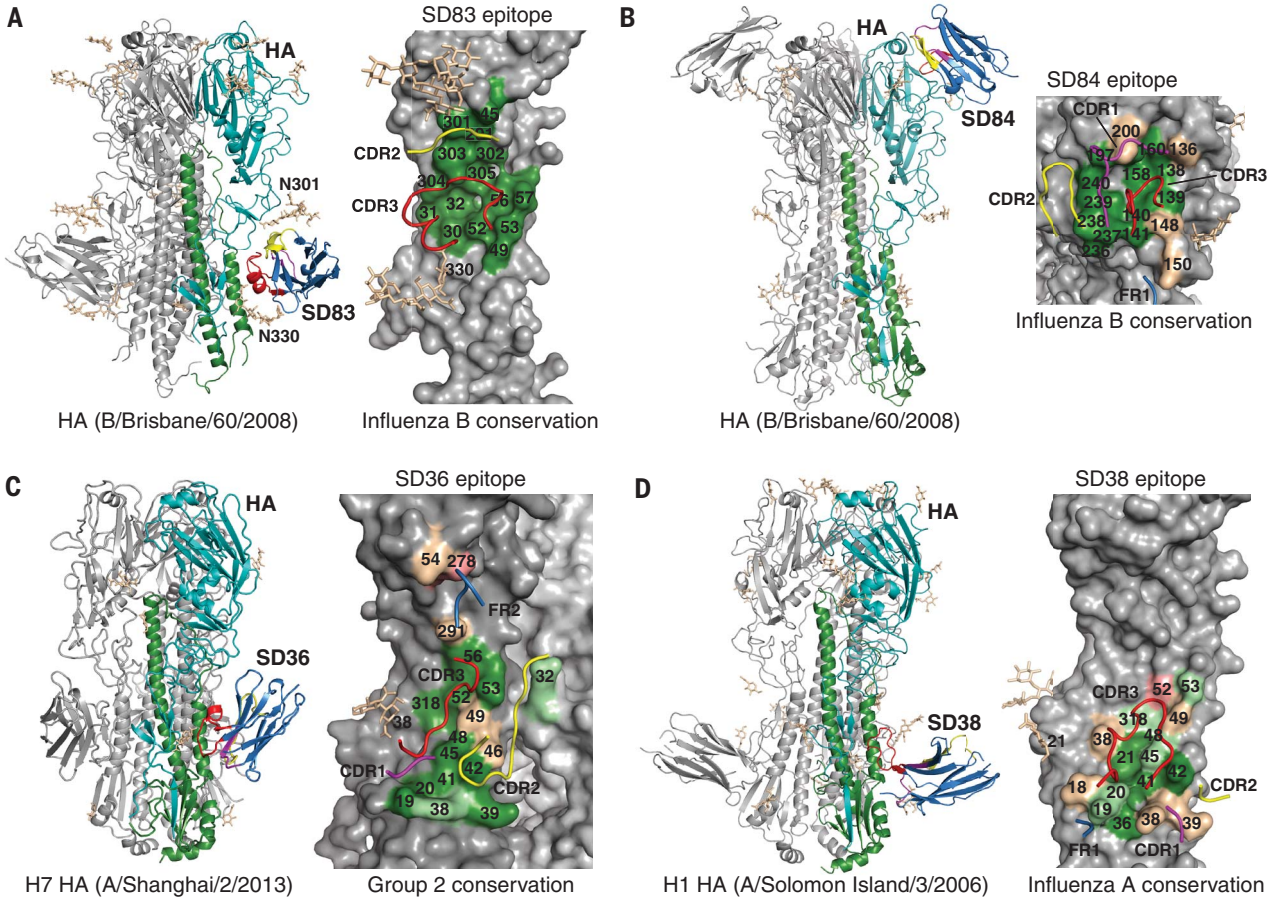


Fig. 2. Crystal structures of sdAbs in complex with HAs and conservation of their epitopes. (**A**) Crystal structure of SD83 with influenza B HA (B/Brisbane/60/08) at 2.2-Å resolution. One HA-SD83 protomer of the trimeric complex is colored with HA1 in cyan, HA2 in green, and SD83 in blue (left). CDR1 is colored in magenta, CDR2 in yellow, and CDR3 in red. The other HA protomers are colored in gray. N-linked glycans are shown in beige in stick representation. The epitope of SD83 is mapped onto the influenza B HA and colored by conservation across influenza B HAs (right). Only the CDR loops involved in the interaction are shown. N301 and N330, Asn^{301} and Asn^{330}. (**B**) Crystal structure of SD84h with influenza B HA (B/Brisbane/60/08) as well as with CR9114 Fab at 4.1-Å resolution (left). The epitope of SD84h is mapped onto the influenza B HA and colored by conservation across influenza B HAs (right). For clarity, CR9114 Fab, which binds to the conserved HA stem region, is not shown. (**C**) Crystal structure of SD36 with H7 HA (A/Shanghai/2/13) at 2.65-Å resolution (left). The epitope of SD36 is mapped onto the H7 HA and colored by conservation across influenza A group 2 HAs (right). (**D**) Crystal structure of SD38 with H1 HA (A/Solomon Islands/3/06) at 2.8-Å resolution (left). The epitope of SD38 is mapped onto the H1 HA and colored by conservation across influenza A HAs (right). For all panels, conservation color coding is as follows: dark green, >95% conserved; light green, 75 to 95% conserved; beige, 50 to 75% conserved; and pink, 35 to 50% conserved.

(IC_{50}) >1 μM]. As all group 1 viruses were glycosylated at HA1 N289, steric hindrance may have lowered the SD36 affinity below the neutralization threshold, whereas SD38 avoided this glycan (table S6). HA2 Val^{18} and Leu^{38} on the SD38 epitope periphery were the least conserved across H1 strains (table S3 and Fig. 2D), but natural mutations Gln^{38} or Ile^{18} and Gln^{38} did not affect neutralization (table S7).

We next rationalized that linking different sdAbs could generate MDAbs with increased potency and breadth. Thus, MDAbs were engineered by genetically fusing individual sdAbs with peptide linkers and then linking them to human immunoglobulin G1 (IgG1) Fc (Fig. 3). The SD38-SD36 fusion was generally more potent, with broader cross-reactivity, than SD36 and SD38 alone, especially against H3 viruses (Fig. 1). Notably, SD38-SD36 neutralized D46N-carrying H3N2 viruses A/Panama/2007/99 and A/Wisconsin/67/05, neither of which was neutralized by SD36 or SD38, thereby demonstrating a synergistic effect (table S8). SD83-SD84 also neutralized influenza B viruses much more potently, especially B/Lee/40 (Fig. 1). Notably, MD2407 (SD38-SD36-SD83-SD84) and MD3606 (MD2407 fused to human IgG1 Fc) neutralized all A (H1 to H12 and H14) and B viruses tested except for one avian H12 virus (*18*) and had much greater breadth and potency than the individual sdAbs or CR9114 (Fig. 3 and table S9). MD3606 also strongly bound avian H13, H15, and H16 and bat H17 and H18 HAs (table S6).

Depending on their epitopes, known bnAbs inhibit influenza infection by blocking viral attachment, HA proteolytic activation, membrane fusion, or viral egress (*19*–*21*). The HA head–binding SD84 and both MDAbs inhibited hemagglutination by influenza B (B/Florida/04/06-MA) virus (table S10). All three stem binders and both MDAbs blocked the low-pH HA rearrangements required for membrane fusion (fig. S5). Unlike CR8020, the sdAbs or MDAbs could not prevent HA proteolytic activation (HA0 to HA1 and HA2), reflecting their epitope location further up the HA stem (figs. S2 and S6). In contrast to SD83 and SD84, MD2407 and MD3606 inhibited egress of B/Malaysia/2506/04 virus (fig. S7), similar to some other bnAbs (*21*, *22*).

A shift from monovalency (Fab) to bivalency (IgG) can considerably increase the binding and neutralization breadth of influenza A antibodies, mainly to the HA head (*23*–*25*). We extended this concept with our MDAbs that targeted different HA conserved epitopes, resulting in greatly increased potency and unparalleled neutralization breadth. From our crystal and EM structures, we were able to exclude simultaneous binding of the MDAbs to epitopes on the same HA trimer (fig. S8A and table S11) (*26*). MDAbs could also increase avidity through cross-linking adjacent HA trimers on the viral surface. Inter–cross-linked HA complexes (85 to 96 Å) were indeed visible in negative-stain EM (fig. S8B). If this holds for HA trimers on the viral surface, it could partially explain the increased potency of MD2407 and MD3606 versus their individual sdAb components.

To assess in vivo protection, we compared the prophylactic efficacy of MD3606 with that of CR9114 and CR8071 in BALB/c mice challenged with H1N1, H3N2, H7N9, and B viruses by using intravenous administration (Fig. 4, A to D, and figs. S9 and S10). MD3606 at 1.7 mg per kilogram of body weight completely protected mice from a lethal dose of mouse-adapted H1N1 virus (A/Puerto Rico/8/34-MA) and was superior to CR9114 at 5 mg/kg. MD3606 at 5 mg/kg provided complete protection from H3N2 virus (A/Hong Kong/1/68-MA) infection, whereas at 1.7 mg/kg, seven of eight mice survived with MD3606, compared with five of eight with CR9114. With an H7N9 virus (A/Anhui/1/13), mice were completely protected with MD3606 at 5 mg/kg versus 15 mg/kg for CR9114. Finally, at 1 mg/kg, MD3606 protected all mice from influenza B virus (B/Florida/4/06-MA) infection and was superior to CR9114 and CR8071 (12.5 and 50% survival). MD3606 administered 1 day before challenge with H1N1 or B virus also resulted in a dose-dependent reduction in lung viral load (fig. S11).

Antibody-dependent cellular cytotoxicity (ADCC) can also contribute to in vivo bnAb efficacy (*27*). In an ADCC reporter assay, MD3606 activated FcγRIIIa similarly to CR9114 (fig. S12). MD3606 with a murine IgG2a (mIgG2a) Fc protected all mice from lethality with influenza B virus (B/Florida/4/06-MA) at 1 mg/kg intravenously (iv), similar to MD3606 with human IgG1 Fc (fig. S13). A human IgG1 LALA Fc mutant, which decreases binding to human and mouse Fcγ receptors (FcγRs), resulted in 37.5% survival at 1 mg/kg (iv) but 100% survival at 5 mg/kg. MD3606 with a mIgG2aσ Fc domain lacking FcγR binding (*28*) provided full protection at 5 mg/kg but none at 1 mg/kg. These FcγR-mediated effector functions will most likely extend to influenza A viruses, as binding of MD3606 to cells expressing H3 HA (A/Wisconsin/67/05) induced ADCC in vitro (fig. S12).

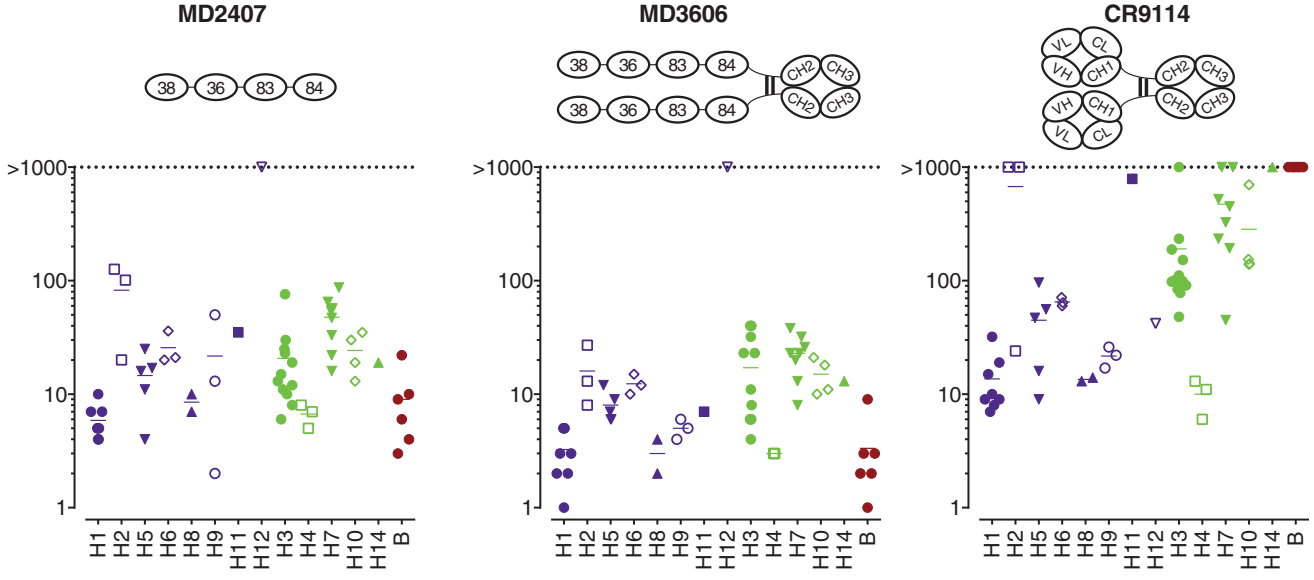


Fig. 3. In vitro neutralization of influenza A and B viruses by pan-influenza MDAbs MD2407 and MD3606 versus CR9114. In vitro potencies of MD2407, MD3606, and CR9114 against a panel of 60 influenza A group 1 (blue), A group 2 (green), and B (red) viruses. A list of the influenza virus strains tested, together with their IC_{50} values, is compiled in table S9. Both MD3606 and MD2407 are significantly more potent ($P < 0.05$) than CR9114 (see materials and methods in the supplementary materials). Data are representative of at least three independent experiments performed in quadruplicate. V_H, variable region of immunoglobulin heavy chain; C_H, constant region of immunoglobulin heavy chain; C_L, constant region of immunoglobulin light chain; V_L, variable region of immunoglobulin light chain.

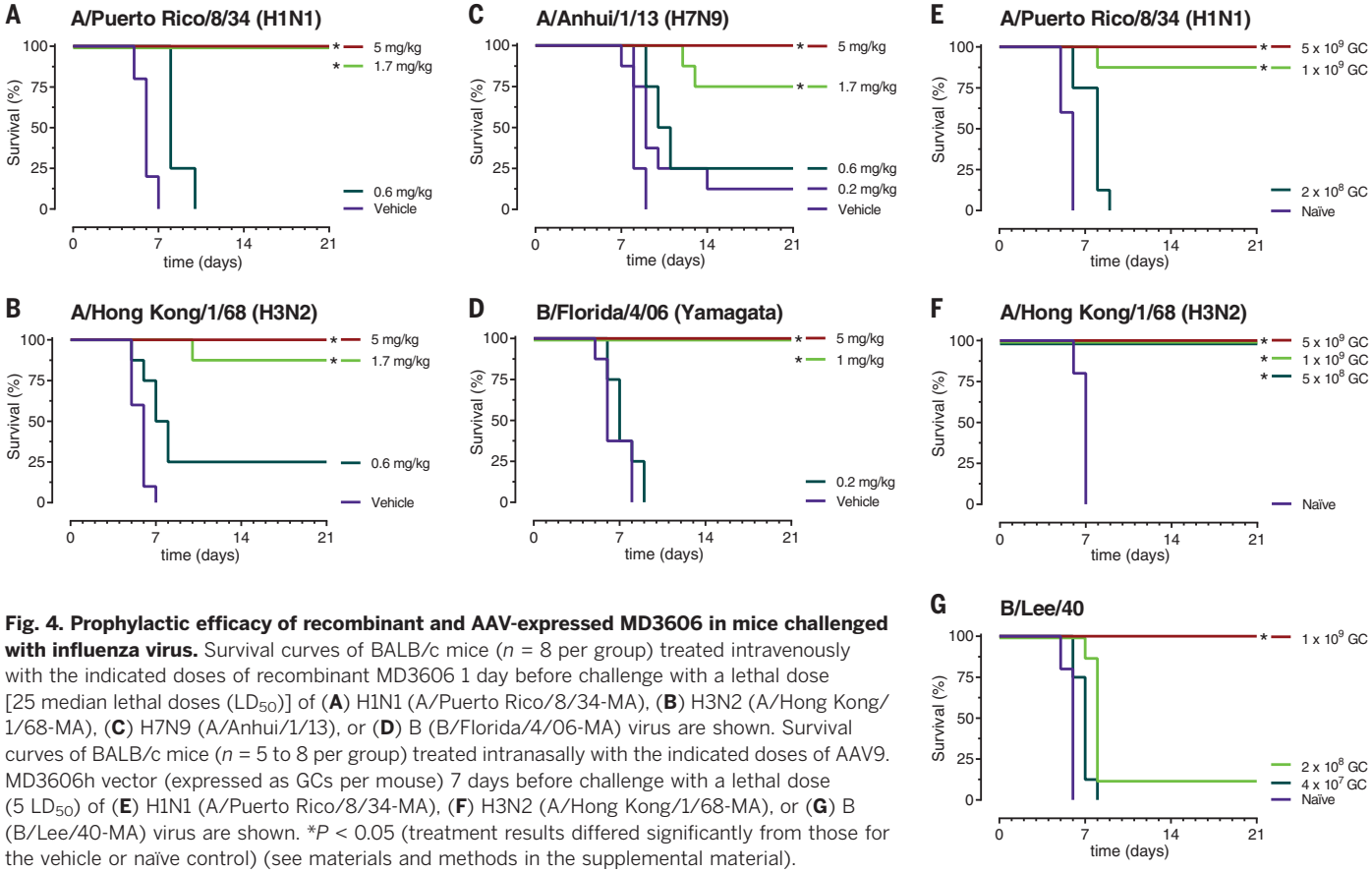


Fig. 4. Prophylactic efficacy of recombinant and AAV-expressed MD3606 in mice challenged with influenza virus. Survival curves of BALB/c mice (n = 8 per group) treated intravenously with the indicated doses of recombinant MD3606 1 day before challenge with a lethal dose [25 median lethal doses (LD_{50})] of (**A**) H1N1 (A/Puerto Rico/8/34-MA), (**B**) H3N2 (A/Hong Kong/1/68-MA), (**C**) H7N9 (A/Anhui/1/13), or (**D**) B (B/Florida/4/06-MA) virus are shown. Survival curves of BALB/c mice (n = 5 to 8 per group) treated intranasally with the indicated doses of AAV9.MD3606h vector (expressed as GCs per mouse) 7 days before challenge with a lethal dose (5 LD_{50}) of (**E**) H1N1 (A/Puerto Rico/8/34-MA), (**F**) H3N2 (A/Hong Kong/1/68-MA), or (**G**) B (B/Lee/40-MA) virus are shown. *P < 0.05 (treatment results differed significantly from those for the vehicle or naïve control) (see materials and methods in the supplemental material).

We then evaluated the prophylactic efficacy of a recombinant AAV9 vector encoding humanized MD3606 (*29*) (AAV9.MD3606h) in mice challenged with mouse-adapted influenza H1N1, H3N2, and B viruses (Fig. 4, E to G, and fig. S14). AAV9.MD3606h was administered intranasally 7 days before influenza challenge at vector doses ranging from 4×10^7 to 5×10^9 genome copies (GCs) per mouse. A dose of 5×10^9 GCs completely protected against lethal challenge with H1N1 virus (A/Puerto Rico/8/34-MA), whereas seven of eight mice survived with a dose of 1×10^9 GCs. Mice challenged with H3N2 (A/Hong Kong/1/68-MA) and B (B/Lee/40-MA) viruses were fully protected by intranasal administration of 5×10^8 GCs (the lowest dose tested) and 1×10^9 GCs, respectively. AAV9.MD3606h also conferred protection against H1N1 virus when administered 35 days before challenge (fig. S15). Preexisting serum-circulating AAV9-specific neutralizing antibodies did not affect the prophylactic efficacy of AAV9.MD3606h (fig. S16).

Limitations of seasonal influenza vaccines, together with the constant threat of a new influenza pandemic, have spurred the search for new influenza prevention strategies. One such strategy involves passive immunization by AAV-mediated delivery of genes encoding protective bnAbs (*11*). In preclinical mouse models, intranasal delivery of AAV9 encoding bnAb FI6 provided protection against H1N1 infection 3 days after vector administration (*11*). Notably, old and immunodeficient mice were also protected from a lethal H1N1 dose (*10*). Mucosal expression of the AAV9 transgene is durable [>9 months in mice (*30*) and 4 months in rhesus macaques (*11*)], and AAV9 vectors can be readministered in the airway without loss of efficiency (*30*). To be clinically useful, AAV-encoded bnAbs should neutralize both influenza A and B viruses, but all bnAbs to date lack sufficient influenza A and B cross-reactivity; the limited AAV packaging capacity (<5 kb) also precludes the expression of two individual bnAbs or a bispecific bnAb from a single vector. We used an alternative antibody platform with camelid-derived sdAbs to create two highly potent MDAbs, MD3606 and MD2407, which have near-universal activity against influenza A and B viruses and can both be expressed from a single AAV vector. Fc-containing MD3606 administered intravenously was more effective than state-of-the-art bnAb CR9114 against seasonal and pandemic influenza viruses, with a significant effect (P < 0.05) on survival in H7N9 and B models (table S12). Intranasal delivery of AAV9.MD3606h provided full protection in mice at doses as low as 5×10^8 GCs. If the above preclinical findings translate to humans, an annual intranasal administration of AAV9.MD3606h may provide passive protection for the entire influenza season and would be of particular benefit to the elderly and other high-risk groups. The rapid onset of protection, together with the unprecedented cross-reactivity of MD3606 to avian influenza strains, also offers the possibility of using this approach as a prophylactic immediately upon onset of an influenza pandemic, providing substantial advantages over vaccination.

REFERENCES AND NOTES

1. World Health Organization (WHO), http://www.who.int/mediacentre/factsheets/fs211/en.
2. J. Liu *et al.*, *Science* **309**, 1206 (2005).
3. H. Chen *et al.*, *Nature* **436**, 191–192 (2005).
4. R. Gao *et al.*, *N. Engl. J. Med.* **368**, 1888–1897 (2013).
5. M. T. Osterholm, N. S. Kelley, A. Sommer, E. A. Belongia, *Lancet Infect. Dis.* **12**, 36–44 (2012).
6. W. E. Beyer *et al.*, *Vaccine* **31**, 6030–6033 (2013).
7. H. Xie *et al.*, *Sci. Rep.* **5**, 15279 (2015).
8. N. C. Wu, I. A. Wilson, *J. Mol. Biol.* **429**, 2694–2709 (2017).
9. S. Krah *et al.*, *Immunopharmacol. Immunotoxicol.* **38**, 21–28 (2016).
10. V. S. Adam *et al.*, *Clin. Vaccine Immunol.* **21**, 1528–1533 (2014).
11. M. P. Limberis *et al.*, *Sci. Transl. Med.* **5**, 187ra72 (2013).
12. See supplementary materials.
13. C. Dreyfus *et al.*, *Science* **337**, 1343–1348 (2012).
14. D. Corti *et al.*, *Science* **333**, 850–856 (2011).
15. D. C. Ekiert *et al.*, *Science* **324**, 246–251 (2009).
16. J. Sui *et al.*, *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009).
17. CDR1 and CDR2 also made minor contacts to the A helix.
18. H12 HA has more mutations in the SD38 epitope than other HA subtypes, including those at HA1 residue 318 (Ile) and HA2 residues 48 (Met), 49 (Gln), and 52 (Leu), which are not present in other subtypes (table S3). However, MD2407 binds H12 HA in biolayer interferometry experiments but with fast on and fast off rates and a K_d value of >550 nM for the first of two sites in a 2:1 binding model, which is higher than those for other subtypes. Further studies will be needed to find out which H12 HA mutation(s) is responsible for the lack of neutralization and whether these resistance mutations can

emerge in HAs of human influenza viruses (e.g., H1N1 or H3N2) upon treatment with MD3606.
19. J. P. Julien, P. S. Lee, I. A. Wilson, *Immunol. Rev.* **250**, 180–198 (2012).
20. N. S. Laursen, I. A. Wilson, *Antiviral Res.* **98**, 476–483 (2013).
21. B. Brandenburg *et al.*, *PLOS ONE* **8**, e80034 (2013).
22. G. S. Tan *et al.*, *J. Virol.* **88**, 13580–13592 (2014).
23. P. S. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17040–17045 (2012).
24. A. Hultberg *et al.*, *PLOS ONE* **6**, e17665 (2011).
25. D. C. Ekiert *et al.*, *Nature* **489**, 526–532 (2012).
26. The 10-residue linkers in MD2407 were too short to allow SD36 and SD38, or SD84 and SD83, to bind the same HA trimer (fig. S8A). Furthermore, the potency of SD38-SD36 with an 18-residue linker was indistinguishable from that of constructs with 38- or 60-residue linkers, which should be more permissive for intratrimer binding (table S11).
27. D. J. DiLillo, G. S. Tan, P. Palese, J. V. Ravetch, *Nat. Med.* **20**, 143–151 (2014).
28. O. Vafa *et al.*, *Methods* **65**, 114–126 (2014).
29. Humanization of the four sdAb components of MD3606 (SD36, SD38, SD83, and SD84) is described in patent application WO/2016/124768 (Janssen Vaccines and Prevention, 2016). The humanized sdAbs share 92 to 98% of their framework identity with the closest human V_H germline sequences. Wild-type and humanized MD3606 were virtually indistinguishable in in vitro neutralization assays using different influenza A and B viruses.
30. M. P. Limberis, J. M. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12993–12998 (2006).

ACKNOWLEDGMENTS

We thank N. van Dijk for performing label-free binding assays, W. Yu for protein purification, H. Tien for automated crystal screening, and the staff of the Gene Therapy Program for their invaluable assistance with AAV-directed gene transfer studies. This is publication 29558 from The Scripps Research Institute. **Funding:** This work is supported in part by NIH grants R56 AI117675 and R56 AI127371 (to I.A.W.), by a grant from Janssen, and in part by the Defense Advanced Research Projects Agency, Department of Defense (grant 64047-LS-DRP.02 to J.M.W.) and the Theme-based Research Scheme, Research Grants Council of the Hong Kong (ref. T11-705/14N). X-ray diffraction data were collected at the Advanced Photon Source (APS) beamlines 23ID-B, 23ID-D, and 17-ID and at the Stanford Synchrotron Radiation Lightsource (SSRL) BL12-2. Use of the APS was supported by the U.S. Department of Energy (DOE), Basic Energy Sciences, Office of Science, under contract DE-AC02-06CH11357. GM/CA CAT is funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the NIGMS (Y1-GM-1104). Use of the SSRL, SLAC National Accelerator Laboratory, is supported by the U.S. DOE, Office of Science, Office of Basic Energy Sciences under contract DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the NIGMS (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS, NIAID, or NIH. **Author contributions:** J.A.K., I.A.W., R.H.E.F., R.V., and J.M.W. designed the project; N.S.L., X.Z., R.M.B.H., T.N., J.P., H.L.T., S.M.B., M.J., S.B., J.V., A.V.E., C.T., G.O., and D.Z. performed experiments; J.L. and R.S. performed data analysis; H.V.D., M.P.L., B.B., M.V.D.N.K., L.L.M.P., and A.P.T. designed experiments; and N.S.L., X.Z., A.B.W., B.B., J.A.K., and I.A.W. wrote the manuscript. **Competing interests:** Janssen Vaccines and Prevention has a pending patent application (WO/2016/124768) relating to certain molecules described in this article. **Data and materials availability:** Coordinates and structure factors are deposited in the Protein Data Bank as entries 6FYU, 6FYT, 6CK8, 6FYW, 6FYS, 6CNW, and 6CNV. The EM reconstructions have been deposited in the Electron Microscopy Data Bank under accession code EMD-9029. All other data needed to evaluate the conclusions in this paper are present either in the main text or in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6414/598/suppl/DC1
Material and Methods
Figs. S1 to S16
Tables S1 to S12
References (*31*–*52*)

27 September 2017; accepted 14 September 2018
10.1126/science.aaq0620

Dosing Pumps

The Modular Dosing System (MoDoS) provides highly accurate, continuous delivery of fluids in laboratory, pilot-plant, or continuous-production processes. Each tailor-made MoDoS modular pump system is fully equipped and ready-to-use, and incorporates an HNP Mikrosysteme internal gear pump, filters, sensors, valves, fluid connections, and controllers, all housed within a compact, sturdy aluminum frame. HNP Mikrosysteme's integral micro-annular gear pumps are chemically inert and particularly suited to handling aggressive and corrosive fluids. Their ranges of hermitically sealed pumps with a magnetic drive can be incorporated into the MoDoS dosing unit, ensuring increased operator safety, optimized production, no cross-contamination of pumped fluids, no wastage, and no environmental hazards due to leaks. MoDoS is ideal for fine chemical or pharmaceutical applications and can be specified for capacities ranging from 0.003 mL to 1152 mL/min, for differential pressures up to 80 bar, and for liquids with viscosities from 0.3 mPa to 1,000 mPa.

Michael Smith Engineers
For info: 800-316-7891
www.michael-smith-engineers.co.uk

Minicircle Technology

The MC-Easy System from AMS Biotechnology enables a simple, reproducible, and efficient way to produce high-quality minicircle DNA. Minicircles are circular DNA elements that no longer contain antibiotic resistance markers or the bacterial origin of replication. These small vectors can be used in vivo or in vitro, and provide for long-term transient expression of one or more transgenes without the risk of immunogenic responses caused by the bacterial backbone in standard plasmids. Minicircle DNA vectors allow sustained transgene expression in quiescent cells and tissues. The finely tuned growth and induction media produces minicircle DNA that is free of parental and genomic DNA contamination. The kit also includes an additional, innovative method for degrading any contaminating genomic DNA using an adenosine triphosphate (ATP)-dependent DNase reagent that will selectively remove genomic DNA without affecting minicircle DNA yield. This method produces clean, effective minicircles every time.

AMS Biotechnology
For info: 617-945-5033
www.amsbio.com/minicircle-gene-expression.aspx

Electronic Lab Notebook

The IDBS E-WorkBook Cloud is a complete, end-to-end, cloud-based R&D platform that supports internal, external, and hybrid data management and research needs. By integrating the best of lab-based informatics solutions (ELN, LIMS, LES, SDMS, and more), the E-WorkBook Cloud goes beyond traditional lab management software, providing cutting-edge data capture and analysis tools, job requesting and management, inventory management, and biology and chemistry functionality. With the E-WorkBook, you get flexible, scalable, and powerful enterprise-level software combined into one of the world's foremost R&D software platforms, enabling you to start building the lab of the future.

IDBS
For info: 800-881-9953
www.idbs.com/discover-e-workbook/eln

Reagent Reservoirs

Available in 96- and 384-well configurations, our reagent reservoirs are designed for simple integration into any automated liquid-handling system, are fully automation compatible, and are compliant with ANSI/SLAS standards. Reagent reservoirs provide an economical way to divide up reagents and wash stations on the decks of automated liquid-handling systems. By placing more than one reagent or wash station in each position, the overall size of the automation platform can be reduced, saving valuable bench space. Molded from high-purity polypropylene, each reservoir tray is highly resistant to heat and to most organic solvents, acids, and bases. Incorporating a low plate profile and a novel "pyramid bottom" design, Porvair reservoir trays provide easy access to all liquid contents, also ensuring a low reagent dead volume (down to <7 uL) for 96- and 384-well–filling applications.

Porvair Sciences
For info: 800-552-3696
www.porvair-sciences.com/reservoirs-1

Scanning Electron Microscope

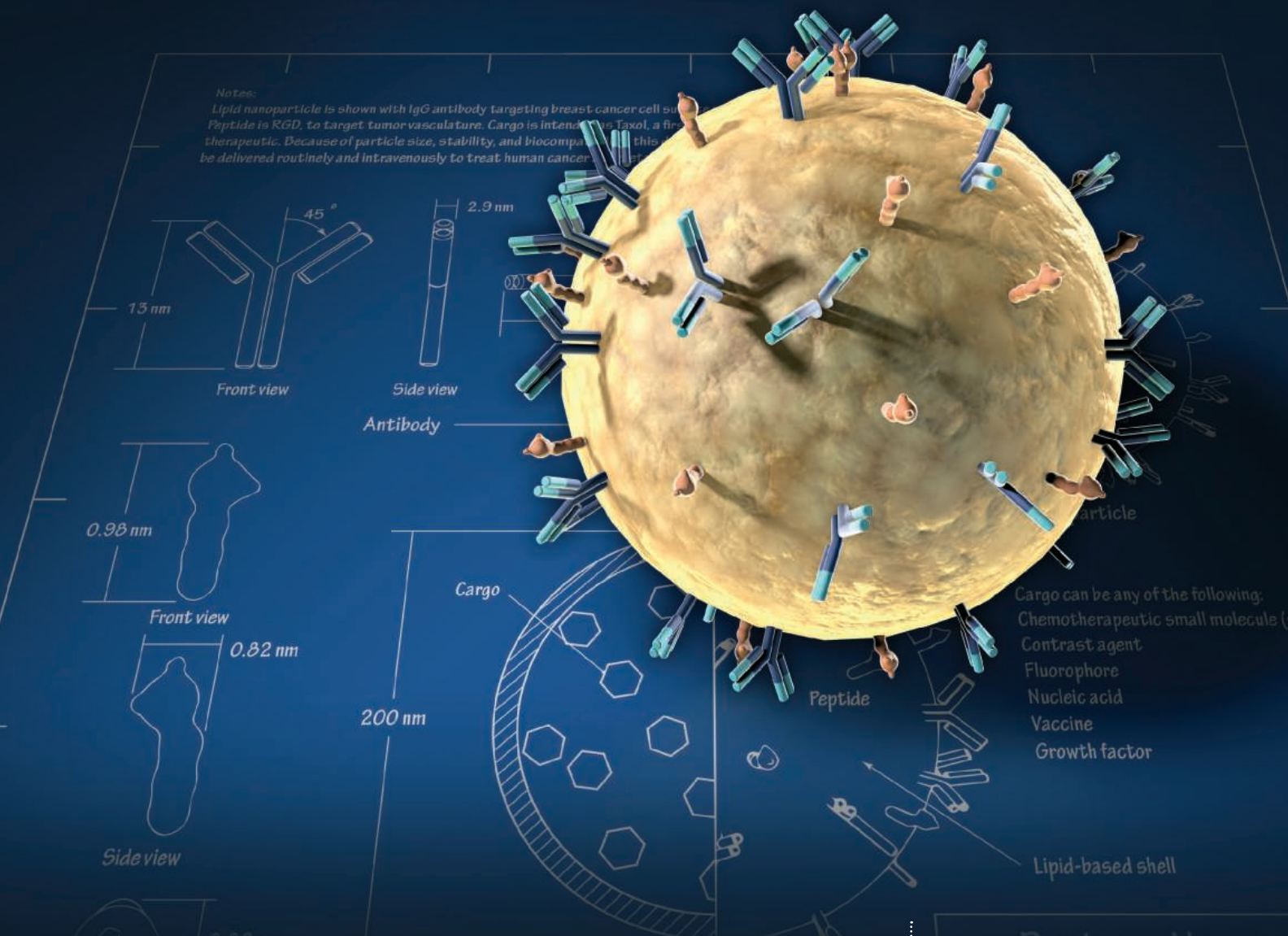
The Thermo Scientific Prisma E Scanning Electron Microscope platform incorporates extensive automation and a friendly user interface that makes it easy to learn and operate in routine industrial applications, while preserving the flexibility needed in a research or academic setting. It features advanced automation; robust support for analytics; a large, precise, and flexible stage; and a wide range of optional accessories. Labs with narrower dedicated needs will value its speed and simplicity, while multiuser, multiapplication facilities will appreciate the broad selection of optional imaging detectors and analytical tools available on a single integrated platform. The Prisma's sample chamber readily handles large, heavy samples and accommodates advanced energy-dispersive X-ray spectroscopy (EDS) configurations that deliver fast, accurate elemental microanalysis. The system can operate over a wide range of vacuum conditions, permitting observations of nonconductive, wet, dirty, or outgassing samples in their natural states without the need for cleaning, drying, or coating, with no risk of damage to the instrument.

Thermo Fisher Scientific
For info: 866-356-0354
www.thermofisher.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Make Your Research Hit the Target



Science Translational Medicine publishes peer-reviewed, cutting-edge biomedical research in the fields of cardiology, cancer, immunotherapy, infectious diseases and vaccines, bioengineering and devices, neurology and neurodegenerative diseases, obesity, diabetes and metabolic disorders, drug discovery, genomic medicine, imaging, stem cell therapy and regenerative medicine.

Submit your research today

Learn more at: **ScienceTranslationalMedicine.org**

Science Translational Medicine

AAAS

BIOLOGY ■ EARTH ■ ENVIRONMENT ■ LIFE ■ INTERDISCIPLINARY ■ PHYSICAL ■ MATERIAL ■ SOCIAL SCIENCES

Science Advances | AAAS

OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

OPEN ACCESS

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

10 ways that *Science* Careers can help advance your career

1. Register for a free online account on **ScienceCareers.org**.
2. Search thousands of job postings and find your perfect job.
3. Sign up to receive e-mail alerts about job postings that match your criteria.
4. Upload your resume into our database and connect with employers.
5. Watch one of our many webinars on different career topics such as job searching, networking, and more.
6. Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
7. Complete an interactive, personalized career plan at "my IDP."
8. Visit our Career Forum and get advice from career experts and your peers.
9. Research graduate program information and find a program right for you.
10. Read relevant career advice articles from our library of thousands.

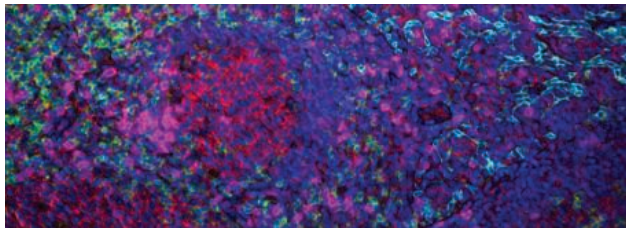
Visit **ScienceCareers.org** today — all resources are free

ScienceCareers

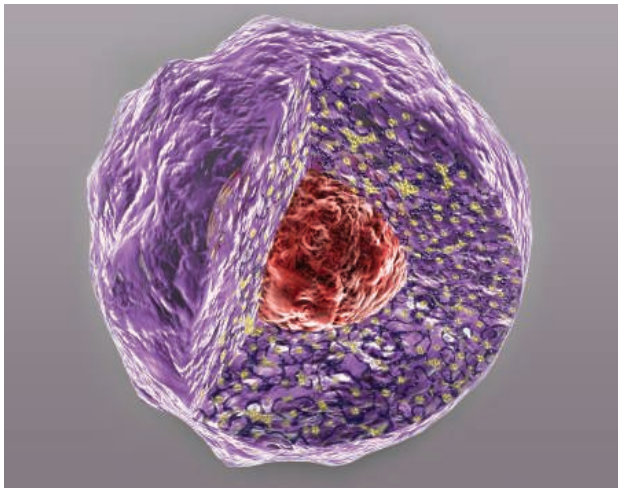
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

want new technologies?



antibodies
apoptosis
biomarkers
cancer
cytometry
data
diseases
DNA
epigenetics
genomics
immunotherapies
medicine
microbiomics
microfluidics
microscopy
neuroscience
proteomics
sequencing
toxicology
transcriptomics



watch our webinars

Learn about the latest break-throughs, new technologies, and ground-breaking research in a variety of fields. Our expert speakers explain their quality research to you and answer questions submitted by live viewers.

VIEW NOW!
webinar.sciencemag.org

Science
AAAS

Brought to you by the *Science*/AAAS Custom Publishing Office

@SciMagWebinars

ASSOCIATE PROFESSOR/PROFESSOR AND VICE CHAIR, DEPARTMENT OF BASIC MEDICAL SCIENCES

The Western University of Health Sciences (WesternU), College of Osteopathic Medicine of the Pacific (COMP), invites applicants for the position of Associate Professor/Professor and Vice Chair of the Department of Basic Medical Sciences (BMS) at its campus in Lebanon, OR (COMP-Northwest).

The Vice-Chair of the Basic Medical Sciences at COMP-Northwest will lead the development of an integrated program in Global and Environmental Health that is part of a major campus expansion in Lebanon. The Vice-Chair will work with the BMS Chair on the WesternU campus in Pomona, CA to develop a strategic plan to enhance research, our educational mission and operational efficiencies. Successful candidates will have a strong research program with current and past history of extramural funding, and administrative and educational experience.

Candidates will be expected to develop an active extramurally funded research program in the areas of Global and Environmental Health, and help develop and deliver the pre-clinical curriculum to medical students. Preference will be given to candidates with expertise in the following curricular disciplines: Biochemistry, Immunology, Physiology or Virology.

A competitive start-up package, research space and core facilities, including a vivarium, will be available. Interested individuals should submit a curriculum vitae, a two page outline of their research/scholarly work and teaching philosophy to posting number A00259 on **http://apptrkr.com/1320256**. Inquiries can be addressed to: Dr. Michelle Steinauer, msteinauer@westernu.edu, 541-259-0233.

Western University of Health Sciences is an Equal Opportunity/Affirmative Action Employer and is committed to a policy of equal employment to all applicants.

Tulane University

DIRECTOR TULANE HYPERTENSION AND RENAL CENTER OF EXCELLENCE (THRCE)

The Tulane University School of Medicine invites applications from outstanding scientists with strong and robust research programs funded by multiple major grants. Applicants should hold a MD or PhD or both and have a record of excellence in research and history of a rigorous sustained externally-funded research program. Applicants must also have a history of academic service in medical and graduate education. Research areas targeted for development can involve basic, translational, and or clinical levels determining cardiovascular, renal, endocrine and neural mechanisms of hypertension and associated cardiovascular diseases including diabetes. The THRCE Director will have responsibility for managing the Cores including the Molecular, Analytical and Imaging Core, and the Phenotyping Core. Further responsibilities include the recruitment and nurturing of new faculty.

Required application materials include a curriculum vitae, a brief statement of research interests, copies of 3 major scholarly papers and 3 letters of recommendation. Applicants must apply via *Interfolio* at **http://apply.interfolio.com/53517**. Applications will be accepted until the position is filled.

Tulane University is an Equal Opportunity/ Affirmative Action/Person with disabilities/ Veterans Employer committed to excellence through diversity. Tulane will not discriminate against individuals with disabilities or veterans. All eligible candidates are encouraged to apply.

DEPARTMENT OF IMMUNOLOGY
UNIVERSITY of WASHINGTON

ASSISTANT PROFESSOR FACULTY POSITION Tenure-Track

The Department of Immunology at the University of Washington seeks a highly qualified applicant for a full-time tenure-track faculty position. Candidates with a background in adaptive immunity are particularly encouraged to apply. Candidates for this position must hold a PhD and/or MD (or foreign equivalent) degree in immunology or related discipline and have a strong record of published research in immunology. The University of Washington faculty engage in teaching, research and service. The successful candidate will be expected to teach at both the undergraduate and graduate level and lead a strong research program. The Department of Immunology offers excellent laboratory space, access to cutting-edge technologies and a highly collaborative environment. Additional information regarding the department can be found at http://immunology.washington.edu/. This position has a 12-month service period and will remain open until filled. Women, minorities, individuals with disabilities, and veterans are all encouraged to apply. Please submit an application (including a cover letter addressed to Dr. Joan Goverman, Professor and Chair, Department of Immunology and your curriculum vitae, a brief description of proposed research, as well as names and addresses of three references) at: **http://apply.interfolio.com/56686**.

Equal Employment Opportunity Statement
University of Washington is an affirmative action and equal opportunity employer. All qualified applicants will receive consideration for employment without regard to race, color, creed, religion, national origin, sex, sexual orientation, marital status, pregnancy, genetic information, gender identity or expression, age, disability, or protected veteran status.

Commitment to Diversity
The University of Washington is committed to building diversity among its faculty, librarian, staff, and student communities, and articulates that commitment in the UW Diversity Blueprint (http://www.washington.edu/diversity/diversity-blueprint/). Additionally, the University's Faculty Code recognizes faculty efforts in research, teaching and/or service that address diversity and equal opportunity as important contributions to a faculty member's academic profile and responsibilities (https://www.washington.edu/admin/rules/policies/FCG/FCCH24.html#2432).

Western University of Health Sciences
College of Osteopathic Medicine of The Pacific

Associate Professor/Professor and Chair, Department of MEDICAL ANATOMICAL SCIENCES

The College of Osteopathic Medicine of the Pacific (COMP) at Western University of Health Sciences invites applications for a tenure-track Chair position in the joint Department of Medical Anatomical Sciences for the Pomona, CA and Lebanon, OR campuses. Western University is a vibrant and growing graduate health professions university with degree programs in Osteopathic Medicine, Health Sciences, Dentistry, Nursing, Optometry, Pharmacy, Podiatry, and Veterinary Medicine. The department has educational responsibilities in gross anatomy, neuroanatomy, histology, and embryology. Currently, we have 11 primary faculty with backgrounds in clinical practice, classical anatomy, anthropology, and paleontology. Western University seeks applicants (PhD/DO/MD) at the Associate or Full Professor level. The successful applicant will be expected to:

Mentor faculty: Promote faculty development, especially in research, as well as establish research collaboration with other basic and clinical science departments.

Perform effective administrative duties: The Chair is a vital member of the College's leadership team who will participate in setting COMP goals and the assessment of departmental and college outcomes. Good interpersonal skills and administrative experience will be essential.

Maintain an active research program: Competitive salary, start-up support, laboratory space, and access to core facilities will be offered. Preferred applicants will have an established and current research program supported by active and sustained competitive extramural funding.

Participate in health professions education: Limited anatomical teaching to first- and second-year medical students. Applicants must have several years of experience in anatomical teaching and curriculum development. Experience teaching histology, embryology, and/or neuroanatomy is particularly welcome.

The review of applications will begin immediately and continue until the position is filled. Applicants should submit: 1) a cover letter, 2) statement of research activity and current support, 3) teaching philosophy statement, and 4) curriculum vitae with contact information of three references to **http://apptrkr.com/1313828**.

Inquiries can be addressed to the Search Committee Chair: Mathew Wedel, PhD, Department of Medical Anatomical Sciences, College of Osteopathic Medicine of the Pacific, Western University of Health Sciences, 309 E. Second Street, Pomona, CA 91766-1854. Tel #: 909-469-6842. E-mail: mwedel@westernu.edu

Western University of Health Sciences is an Equal Opportunity Employer and actively seeks applications from women and minorities.

AUS | الجامعة الأميركية في الشارقة
American University of Sharjah

American University of Sharjah, one of the Middle East's leading universities, has set itself an ambitious research agenda, pushing the limits of science and engineering and providing exciting opportunities for those seeking to realize the university's research goals.

LEADING A NEW ERA IN RESEARCH

AUS is a preeminent and top-ranked university in the region, currently embarking on a massive new era in research, founding seven new research entities:

Biosciences and Bioengineering Research Institute (BBRI)
Materials Science and Engineering Research Institute (MSERI)
Gulf Environments Research Institute (GERI)
Smart Cities Research Institute (SCRI)
High-Performance Computing Center (HPCC)
Geospatial Analysis Center (GAC)
Genomics Research Laboratory (GRL)

To learn more about the AUS research agenda, and the opportunities available, visit: www.aus.edu/research

By **Richard C. Larson**

What are you waiting for?

When I was hired as an assistant professor in 1969, mandatory retirement at age 65 was the law of the land for tenured faculty members. I was 26 years old at the time, so that seemed impossibly far away. But by the time I was 50, two amendments to federal law had removed all age limits. I could stay in my tenured position forever! That's how, in 2011, I found myself still an active professor at the Massachusetts Institute of Technology (MIT) in Cambridge at age 68. I might still be in my tenured job today, if not for a meeting that year with the official who administered my federal research funding.

The official asked me to investigate how eliminating mandatory retirement had affected the availability of positions for new assistant professors. The question struck me as important but not personally relevant—until my colleagues and I got our results.

Our initial intuition was that there would be no substantial long-term effect. We expected to find that the number of open positions dipped just after the law's two changes. After all, the number of available tenure-track faculty slots is essentially fixed—at MIT, there are approximately 1000. To create room for a new faculty member, an existing one has to leave. But after a brief dip, we thought, retirements should return to normal, creating room for new recruits.

"Professors older than 65 were blocking the way of many young scholars."

One word for such intuition: wrong! Through modeling, we discovered that eliminating the retirement age had reduced the number of new slots for MIT assistant professors by 19%, from 57 to 46 per year. Put simply, without a mandatory retirement age, senior faculty members are much slower to leave. When our paper was published, I viewed it as just another finding. But eventually, I had serious reflections about what the results really meant.

Around that time, I had hired a postdoc named Navid Ghaffarzadegan. He was a superlative young scholar. Yet he worried that, like many postdocs, he might not be able to get the tenure-track position he sought. There are simply too many applicants seeking too few positions. And I began to realize that I and other professors older than 65 were blocking the way of many young scholars who seek academic careers. I started to wonder whether it was time for me to step aside, but the idea of leaving the job I had been tied to for so long was hard to swallow. (In the meantime, Navid secured that tenure-track position and is now an associate professor with tenure.)

Then, the dean of the engineering school heard about our paper and asked me to go over the details with him. It must have resonated with him, because he briefed the department heads about the need for a flexible after-tenure option that would vacate a position and open the way for a new hire. They soon invented "professor, post-tenure," tossing out an earlier option with the horrendous name "professor without tenure, retired," or PWOTR, pronounced "pee-water."

Once "professor, post-tenure" was announced in 2016, I found it increasingly attractive. It wasn't the same as "emeritus"—not full retirement. I could retain my office, teach and supervise students, and be a principal investigator on research grants—all with great flexibility. I would get to choose which projects I wanted to do and be paid accordingly, up to 49% of my previous salary. I could also access retirement and pension funds. My wife and I would be able to spend more time together and with our children and grandchildren. Decision made!

I submitted my tenure resignation in 2017. I've enjoyed every minute since, busy as ever but only on activities I select—such as MIT BLOSSOMS, a project I co-founded to create interactive video lessons for high school math and science classes. I feel lucky to have this option. Too few institutions offer these types of transitional positions to ease the challenge for us senior professors. At 74, I in essence removed 9 years from someone else's career. I should have stepped aside sooner. ■

Richard C. Larson is a post-tenure professor in the Institute for Data, Systems, and Society at the Massachusetts Institute of Technology in Cambridge. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER