

Will a change of seasons
tame COVID-19? p. 1294

Toad hybridization increases
fitness pp. 1304 & 1377

Sustainable chemicals from a
wood biorefinery pp. 1305 & 1385

Science

\$15
20 MARCH 2020

AAAS

SPECIAL ISSUE

ANTARCTICA





It's a matter of expression.

For over 40 years, New England Biolabs has been developing and using recombinant protein technologies in *E. coli* for our own manufacturing processes.

Protein expression can be a very complex, multi-factorial process. Each protein requires a specific environment to correctly and efficiently achieve its secondary and tertiary structures. Proteins may also require post-translational modifications or insertion into a cellular membrane for proper function. Other proteins, once expressed, may be toxic to the host. Thus, no single solution exists for the successful production of all recombinant proteins, and a broad range of expression tools is often required to ensure the successful expression of your target protein.

Our NEBExpress™ portfolio of products include solutions for expression and purification of a wide range of proteins, and is supported by access to scientists with over 40 years of experience in developing and using recombinant protein technologies in *E. coli*. We use these solutions in our own

research and manufacturing processes, and know that quality and performance are critical – all of our products are stringently tested so that you can be sure they will work optimally for your solution, just as we rely on them to work in ours.

Featured products include:

- **Cell-free expression systems** – express analytical amounts of protein in approximately two hours
- ***E. coli* expression and purification kits** – generate and purify high yields of recombinant proteins
- **Competent cells** – express a variety of proteins in *E. coli*, including difficult targets, proteins with multiple disulfide bonds and His-tagged proteins
- **Purification beads, columns and resins** – available for CBD-, MBP- and His-tagged proteins

Access our entire portfolio and request your sample at
www.neb.com/ProteinExpression.

One or more of these products are covered by patents, trademarks and/or copyrights owned or controlled by New England Biolabs, Inc.
For more information, please email us at gbd@neb.com.
The use of these products may require you to obtain additional third party intellectual property rights for certain applications.
© Copyright 2019, New England Biolabs, Inc., all rights reserved.



be INSPIRED
drive DISCOVERY
stay GENUINE

CONTENTS

20 MARCH 2020 • VOLUME 367 • ISSUE 6484

1294

Seasonal changes in humidity, temperature, and other factors may affect the viability of viruses in droplets produced when people sneeze or cough.

NEWS

IN BRIEF

1284 News at a glance

IN DEPTH

1287 Countries test tactics in 'war' against COVID-19

Lockdowns and closings proliferate, but virus testing and contact tracing are lagging
By J. Cohen and K. Kupferschmidt

1289 Coronavirus disruptions reverberate through research

Scientists grapple with halted projects and locked-down labs
By K. Servick, A. Cho, J. Couzin-Frankel, and G. Guglielmi

1290 Oldest fossil of modern birds is a 'turducken'

Ancestral bird skull looks like a duck from the back and a chicken from the front
By G. Vogel

1291 After Brexit, U.K. budget offers boost to science

Unprecedented increase includes money to set up high-risk research agency
By E. Stokstad

1292 Unrest imperils India's census

Opposition to citizenship policies could lead to undercount
By V. Chandrasekhar

1293 Top neuroscientist leaving Mexican university

Alleged incidents of sexual harassment by Raulfo Romo Trujillo began years ago
By I. Gutiérrez and R. Pérez Ortega

FEATURES

1294 Sick time

Dozens of diseases wax and wane with the seasons. Will COVID-19?
By J. Cohen

INSIGHTS

PERSPECTIVES

1298 Challenges of human nutrition research

Facilities to house and feed subjects could increase rigor and advance nutrition science
By K. D. Hall

1300 Is it worth the effort?

Individual variation in dopamine affects the weighting of benefits relative to costs
By A. C. Janes
REPORT p. 1362

1301 Thermal light tunnels its way into electricity

New devices convert low-temperature heat into electricity
By A. P. Raman
RESEARCH ARTICLE p. 1341



1304 & 1377

1303 Neuron-targeted electrical modulation

Engineering neurons to make conductive polymers enables cell type-specific behaviors
By K. J. Otto and C. E. Schmidt
REPORT p. 1372

1304 Shallow ponds prompt fitness-favorable species interbreeding

Dry conditions liberate female toads to drive sexual selection of male mates
By M. Zuk
REPORT p. 1377

1305 Taking on all of the biomass for conversion

A catalytic process converts hardwood into phenol, propylene, and other valuable products
By T. Zhang
REPORT p. 1385

1307 Stanley Cohen (1922–2020)

Award-winning biochemist who shed light on cell growth
By G. Carpenter and R. Coffey
EDITORIAL p. 1283

POLICY FORUM

1308 Time for NIH to lead on data sharing

A draft policy is generally supportive but should start mandating data sharing
By I. Sim et al.

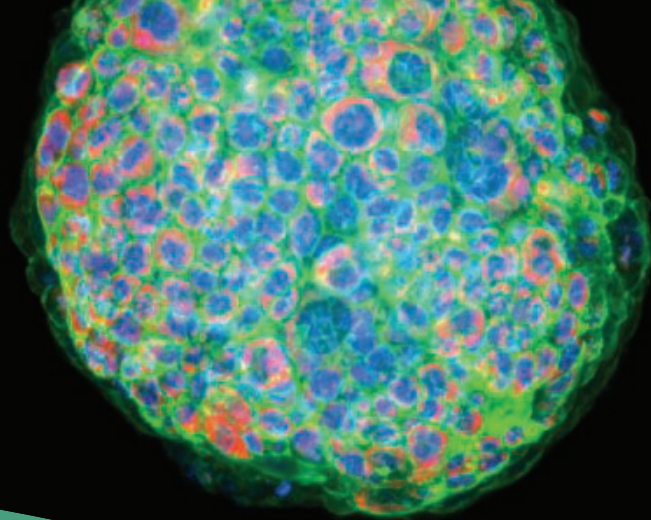
BOOKS ET AL.

1311 Sexism and the stars

A new biography tells the tale of an accomplished astronomer's barrier-breaking life
By J. Carson

1312 Vanquishing false idols, then and now

Francis Bacon's 400-year-old list of scientific foibles holds lessons for modern scientists
By K. P. Weinfurt



CAPTURE the micro moments that matter

NEW Water Immersion Objectives

Gain the sensitivity to capture more phenotypic data at greater depths. Our high-content imagers and expert team are here to help your lab simplify imaging and analysis of complex, 3D assay workflows and fast-track discoveries.

Learn more at
moleculardevices.com/water



SPECIAL SECTION


ANTARCTICA

INTRODUCTION
1316 The frozen continent

NEWS
1318 The lost colony *E. Kintisch*

REVIEWS
1321 History, mass loss, structure, and dynamic behavior of the Antarctic Ice Sheet *R. E. Bell and H. Seroussi*
1326 The Southern Ocean and its interaction with the Antarctic Ice Sheet *D. M. Holland et al.*
1331 The uncertain future of the Antarctic Ice Sheet *F. Pattyn and M. Morlighem*

ON THE COVER



An iceberg from Antarctica drifts in the Southern Ocean. Mass loss from the Antarctic Ice Sheet, whether by iceberg calving or melting, will help determine the future of sea level in our warming climate. How the ice sheet changes over the coming decades to centuries will depend on its structure, dynamics, and interactions with the atmosphere and surrounding ocean. See page 1316. *Photo: Michel Roggo*

LETTERS

1313 Preventing COVID-19 prejudice in academia

By P. Rzymski and M. Nowicki

1313 Humpback dolphins at risk of extinction

By S. Li

1314 A first step for the Yangtze

By Z. Mei et al.

1314 Technical Comment abstracts

RESEARCH

IN BRIEF

1336 From *Science* and other journals

RESEARCH ARTICLES

1339 Human genetics

Insights into human genetic variation and population history from 929 diverse genomes
A. Bergström et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
 DX.DOI.ORG/10.1126/SCIENCE.AAY5012

1340 Cortical genetics

The genetic architecture of the human cerebral cortex
K. L. Grasby et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:
 DX.DOI.ORG/10.1126/SCIENCE.AAY6690

1341 Applied physics

Electrical power generation from moderate-temperature radiative thermal sources

P. S. Davids et al.

PERSPECTIVE p. 1301

1346 Structural biology

Structural basis of G_q and G_i recognition by the human glucagon receptor
A. Qiao et al.

1352 Solar cells

Resolving spatial and energetic distributions of trap states in metal halide perovskite solar cells
Z. Ni et al.

REPORTS

1358 Geomorphology

Latitudinal effect of vegetation on erosion rates identified along western South America
J. Starke et al.

1362 Neuroscience

Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work
A. Westbrook et al.

PERSPECTIVE p. 1300

1366 Cell biology

Golgi-derived PI(4)P-containing vesicles drive late steps of mitochondrial division
S. Nagashima et al.

1372 Biotechnology

Genetically targeted chemical assembly of functional materials in living cells, tissues, and animals
J. Liu et al.

PERSPECTIVE p. 1303

1377 Hybridization

Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates

C. Chen and K. S. Pfennig

PERSPECTIVE p. 1304

1379 Pyroptosis

cFLIP_L protects macrophages from LPS-induced pyroptosis via inhibition of complex II formation
H. I. Muendlein et al.

1385 Chemical engineering

A sustainable wood biorefinery for low-carbon footprint chemicals production
Y. Liao et al.

PERSPECTIVE p. 1305

DEPARTMENTS

1282 EditorialTime to pull together *By H. Holden Thorp***1283 Editorial**Better late than never *By H. Holden Thorp*

PERSPECTIVE p. 1307

1398 Working LifeHow I faced my coronavirus anxiety *By Kai Liu*

New Products.....1394
 Science Careers1395

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2020 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$2148; Foreign postage extra: Air assist delivery: \$98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #125488122. Publications Mail Agreement Number 1069624. Printed in the U.S.A.

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. **Single-copy sales:** \$15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Time to pull together



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

It is now hard to imagine a world that isn't permanently changed by coronavirus disease 2019 (COVID-19). We don't know whether this is an event like 9/11 or the 2008 global financial crisis—where life will mostly go back to the way it was—or whether the institutions and practices of the future will transform in ways that we can't yet imagine. The success of the world's scientists—along with strong political and social leadership—will determine which scenarios unfold, so it is time to focus on what we can all do to help.

On the political front, there is finally some progress as exceptional public servants have emerged as the face of the crisis. Within the United States coronavirus task force, Dr. Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, has appeared as a powerful and uncompromising voice along with Dr. Deborah Birx. Since 2014, Birx has led the highly successful President's Emergency Plan for AIDS Relief, which is one of the greatest international efforts to tackle infectious diseases. That Fauci and Birx are becoming better known to the world, along with their counterparts in other countries, as the steadfast and consistent messengers during this crisis is reassuring for citizens and for the experts who are working tirelessly to find answers.

And indeed, there are very important questions to answer. Will recovery from the first infection confer lasting immunity? Will the first vaccine that proves it works cause side effects that undermine its value? Will the vaccines under development trigger neutralizing antibodies? Do widely used inhibitors of angiotensin-converting enzyme promote or inhibit infection? Will the broad-spectrum antiviral drug remdesivir or viral protease inhibitors thwart the virus?

Then there are also public health and epidemiology questions. Do school closings help or hurt? What happens if hospitals become overwhelmed? If we discover an effective vaccine or drug, can enough be made and delivered to everyone? What are the long-term effects of this crisis on mental health, social well-being, and the economy? What happens when social restrictions, like those in China, are lifted?

We can draw hope from the science at work. I

continue to be inspired by the research papers on COVID-19 submitted to *Science* and appearing on preprint servers around the world. The only way questions will be answered is if scientists can do their work, because scientific knowledge is often the key to knowing what actions to take. So, institutions need to do everything possible to allow these folks to get to the lab safely. Research institutions need to shut down all functions except for clinical care, research on the virus, and public health communication. To support these vital operations, institutions need to provide childcare for scientists and staff whose children are now home from school. And they need to alleviate concerns about the future for these staff by extending tenure clocks, guaranteeing status in graduate school, and extending postdoctoral contracts.

As for the scientific community who are not work-

ing on the virus—we know well that other major problems still exist, such as climate change, inequality, and other diseases. It is understandably very difficult to pause research in other arenas for an indefinite amount of time. This crisis is calling for extraordinary measures, and your supportive responses deserve recognition. Working from home will make it safer for those who must be in buildings and laboratories to do work related to the virus—fewer people in the hallways, lunchrooms, and other public areas will slow the spread

of the virus so that work on COVID-19 can continue. If there is a way for you to assist without slowing these labs, volunteer to do so. If you have colleagues who are working on the virus, an offer of your time to keep an eye on their children or call upon their elderly relatives who are lonely can make a difference.

On so many fronts, this is a battle of a lifetime and a test of our responsibilities for each other and the strength of our compassion. For our part, *Science* will continue to report the news, and make research on COVID-19 freely available as quickly as possible, and we will also continue to support and advocate for the scientists around the world who are leading the charge. Let's maintain social distancing, but pull together, hard. We must. We will.

—H. Holden Thorp

**“This crisis
is calling for
extraordinary
measures...”**

Better late than never

Among the questions that have come up quite often during my first few months as Editor-in-Chief of *Science* are those that concern the Retrospectives that we publish on prominent scientists who have passed away. Does a Retrospective serve as a tribute? Do we run Retrospectives on individuals who made important contributions to science but had other well-known personal drawbacks? Why can't we feature more diverse members of the scientific community in Retrospectives? These have been passionate inquiries that made me wonder about whether we should just stop publishing these pieces to avoid any strife.

After much debate, *Science's* editors decided that it's important to continue publishing Retrospectives. These are an enduring and important part of the scientific record. And a Retrospective is unambiguously a tribute. Generally, we look for someone who can write a personal reflection, relating stories and insights about the individual that stem from the intersections of their lives. It's unreasonable to think that these stories would be anything but positive. So, we're going to remember folks who we feel deserve tribute, and even then, we know there will be disappointments because we simply can't highlight everyone. But we can widen our Retrospective lens. The majority of prominent scientists who are passing away are still mostly white males because they constitute the majority of individuals who were given opportunities to excel in science in those generations. The time will eventually come when more women and people of color will be remembered in Retrospectives. In the meantime, we could seek more diverse voices in these reflections, and we could recognize more diverse individuals who have contributed to the scientific enterprise in ways that are as meaningful as a Nobel Prize.

Which brings me to this week's Retrospective on Stanley Cohen (see page 1307). Certainly, Cohen—who won the Nobel Prize in Physiology or Medicine for discovering nerve growth factor—is deserving of a Retrospective. But Cohen correctly shared the prize with Rita Levi-Montalcini, with whom he worked at Washington University in St. Louis, Missouri. As quoted in today's article, Cohen said of their work together, "On our own we were good and competent. Together we were mar-

velous." Unfortunately, when Dr. Levi-Montalcini died in 2012, *Science* did not run a Retrospective on her. No matter the reason, it was a bad oversight. We missed an important opportunity to praise a remarkable role model for women pursuing excellence in science.

This is important to me because I still hold my faculty title at Washington University, where I have the high honor of being the Rita Levi-Montalcini Distinguished University Professor. I never knew Professor Levi-Montalcini, but I am inspired every day by her extraordinary life story. She went to medical school at the University of Turin and began research in neuroscience immediately after. However, when Mussolini's *Manifesto della Razza* (Manifesto of Race) was issued, Jews could no longer hold academic positions. When Germany invaded Italy, her

family fled to Florence where they survived the holocaust in hiding.

While in hiding, Levi-Montalcini set up a lab in her small living quarters where she examined chicken eggs under a microscope. She wrote that during that time, she would ride trains and read scientific papers while admiring the countryside and smelling the hay. Most scientists can identify with reading scientific papers while on the proverbial hayride of a research career.

On one of those hayrides, she read a paper by the embryologist Viktor Hamburger that changed her life. She went back to her makeshift lab and began the experiments that would lead to the discovery of nerve growth factor.

When the war ended, Hamburger invited her to St. Louis to work with him for a year. She stayed for three decades and forged her collaboration with Cohen.

My colleagues told me many stories about Levi-Montalcini and her graciousness and elegant forcefulness. She was passionate about science and unfettered inquiry. In her Nobel Prize acceptance speech, when describing how Mussolini's fascist directive was signed by 10 scientists, she put "scientists" in quotes. She later became a senator in Italy and died there at the age of 103.

Glad we finally got her tribute into *Science* where it belongs.

Better late than never, Rita.

—H. Holden Thorp



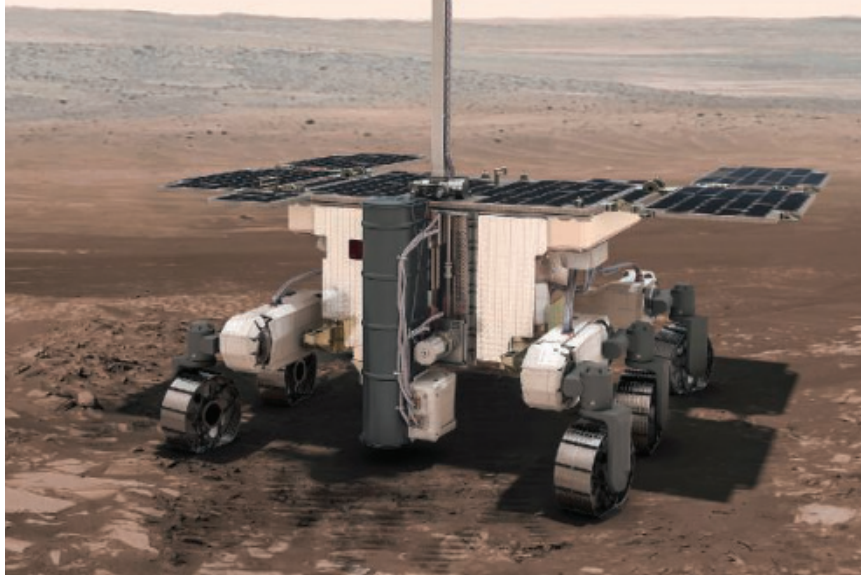
"...we finally got her tribute into *Science* where it belongs."

“I think it's important to destigmatize Huntington's and make it not as scary.”

Neuropsychologist Nancy Wexler, who led the research team that in 1993 identified the gene for Huntington disease, announcing in *The New York Times* that she has the gene and the symptoms.

IN BRIEF

Edited by **Jeffrey Brainard**



The Rosalind Franklin rover will look for buried evidence of martian life.

PLANETARY SCIENCE

Technical woes delay Mars rover

The launch of a rover that would drill beneath the surface of Mars for signs of life has been delayed for 2 years until 2022, the European Space Agency (ESA) and its Russian counterpart, Roscosmos, announced last week. The joint ExoMars mission was threatened by parachute troubles, electrical shorts on its Russian-built lander, and failures of the glue that holds solar panels to Rosalind Franklin, the ESA rover. The coronavirus pandemic further complicated efforts to be ready for a July launch, says instrument team member Francesca Esposito. Although the issues are expected to be fixed in the next few months, mission planners must now wait 26 months for the next favorable alignment of the planets. Instrument team member Valérie Ciarletti is disappointed by the delay, but did not want to risk a repeat of Schiaparelli, a European Mars lander that crashed in 2016. “The scientific payload is impressive,” she says. “It would be a nightmare if it just crashed on Mars.” NASA and China are expected to launch other rovers to Mars this July or August.

Embattled ecologist pushes back

PUBLISHING | A lawyer for behavioral ecologist Jonathan Pruitt, whose colleagues have suggested he fabricated the data behind provocative results on animal personalities and social spiders, has cautioned Pruitt's co-authors and journal editors in recent weeks to hold off retracting his papers. Three have been retracted so far; in a letter the lawyer has requested that no more be withdrawn until misconduct investigations at Pruitt's current and former universities play out. In addition, an online spreadsheet that tracked analyses of the scientist's 160 papers has been taken offline. Its creator—a journal editor who published Pruitt papers—explains that he could not guarantee its accuracy. (He did not reveal whether Pruitt's lawyer warned him.) Pruitt, now at McMaster University, has attributed various data anomalies to inadvertent mistakes. “I just want the process to play itself out, and that will inform the journals about [which] papers are reliable, and which are not,” he said last week.

U.S. expands vaquita protections

CONSERVATION | Last year, researchers found just six vaquitas (*Phocoena sinus*) in the upper Gulf of California, where these small, critically endangered porpoises have been spiraling toward extinction as they become entangled and drown in illegal and legal fishing gear. This month, the Trump administration expanded an existing court-ordered ban on seafood imports from those waters in a bid to protect the vaquita (below). In 2018, the United States prohibited imports of shrimp and curvina, sierra, and chano fish caught in the gulf using gillnets. Beginning 3 April, the ban



IMAGES (TOP TO BOTTOM) ESA; FLIP NICKLIN/MINDEN PICTURES

will extend to almost all fish caught in the gulf using a variety of fishing gear. To be sold in the United States, species named in the ban will require certificates showing they were caught elsewhere in Mexico. But the ban doesn't address the biggest problem facing the vaquita: poachers who gillnet a large fish called the totoaba, sold in China for its perceived medicinal value. Mexico banned gillnet fishing in the gulf in 2017, but enforcement is poor.

Method could turbocharge drugs

CHEMISTRY | For years, drug discovery chemists have struggled to streamline a process that can boost a drug's potency up to 2000-fold: "magic methylation." The reaction sweeps out single hydrogen atoms and replaces them with methyl groups—reshaping the drug molecule to more easily interact with its biological targets. Now, a team of chemists reports it has created a catalyst that performs this delicate exchange with ease on a wide variety of druglike molecules, an advance that could lead to novel treatments for everything from cancer to infectious diseases, they report this week in *Nature*. Physicians may be able to give patients lower doses of higher potency drugs, reducing side effects.

Swine fever vaccine reported

AGRICULTURE | Researchers in China say they have created a vaccine that protects pigs from African swine fever. Since 2018, the contagious viral disease has cut the country's swine population of more than 400 million by about 40%. Scientists at the Harbin Veterinary Research Institute reported this month in *Science China Life Sciences* that they created the attenuated vaccine by deleting seven genes from the virus and found it cannot revert to the virulent strain. "There is justification for guarded optimism," says Dirk Pfeiffer, a veterinary epidemiologist at City University of Hong Kong, although he says more testing of safety and efficacy is needed.

Did Twitter give up on civility?

SOCIAL MEDIA | Some researchers who teamed up with Twitter 2 years ago to make conversations on its platform less toxic say progress has stalled and the social media giant has ignored them, *The Wall Street Journal* (WSJ) reports this week. Other scientists say they struggled to get some of the data promised. "The impression I came away with ... is that they [Twitter] were more sensitive to deflecting criticism than in solving the problem

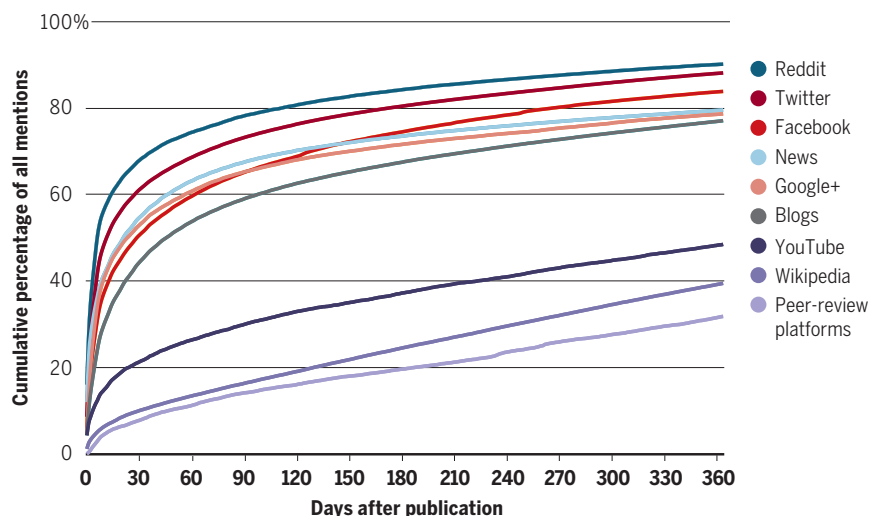
PUBLISHING

'Altmetric' speeds compared

Altmetrics—counts of social media mentions of a journal article—have gained popularity as a promising and faster way to quantify a paper's impact, while adding more texture than its number of citations. Altmetrics provide barometers of interest within days after publication, whereas authors can wait years to see citation counts grow. But the reaction speed of the 12 social media platforms that inform altmetrics scores varies considerably, says a study published this month in *Scientometrics*. Sixty percent of tweets about journal articles are posted during the first 30 days after publication; in contrast, reactions on the Publons and PubPeer commenting platforms appear more slowly. The findings suggest the various social media data used to calculate an altmetric score should not be lumped together, says co-author Zhichao Fang of Leiden University. The study is among the largest examinations of altmetric speed across multiple fields; the authors examined 2.4 million articles in the Web of Science database published from 2012 to 2016. Papers in some disciplines gained social media mentions faster than others; for example, physical science publications outpaced those in computer and social science.

Tweets beat the rest

The study found that reaction speed differs not just by medium (below), but also by discipline and document type; mentions increased faster for journal letters than for research articles.



of harassment," Nathan Matias of Cornell University told WSJ. Twitter blamed the delays on employee turnover and shifting priorities, but vowed to continue the work.

Solar storm risks mapped

GEOMAGNETISM | A map released this week by the U.S. Geological Survey shows where the U.S. electric grid is most vulnerable to space weather. Researchers modeled the effects of a once-in-a-century solar storm, which can hurl charged particles into Earth's magnetic field and cause electric fields to ripple through the ground. Those fields, in turn, trigger electrical surges along power lines that vary in intensity depending on the type of rock deep underground. The geoelectric hazard map shows that power lines in two

U.S. regions are particularly vulnerable: the northern Midwest and east of the Appalachian Mountains.

Michigan removes provost

#METOO | The University of Michigan, Ann Arbor, last week removed toxicologist Martin Philbert as provost as it continues to investigate allegations that he committed sexual misconduct. Michigan placed him on paid administrative leave in January. Before then, at least three people had told the university of instances of harassment or discrimination, and since his suspension, dozens more have accused him of acts of sexual misconduct they say occurred during the past decade, according to the *Detroit Free Press*. Philbert remains a tenured professor but is suspended from teaching.

INFECTIOUS DISEASE

Dispatches from the COVID-19 crisis

The health emergency reverberated as far as the Arctic and into other fields, such as artificial intelligence. Read more at sciencemag.org/tags/coronavirus.

Enforce distancing, experts say

PUBLIC HEALTH | Infectious disease experts across the United States called this week for officials to carry out tougher enforcement of social distancing measures to slow the spread of the virus, including closing schools and nonessential businesses nationwide. Such moves would go beyond the patchwork of closures now in place in some states and major cities. In an open letter circulated on Twitter and by email, more than 370 scientists called for “immediate action on the part of national, state, municipal, and local governments to enforce social distancing in order to truly minimize contact among individuals.”

Travel rules snarl Arctic research

POLAR SCIENCE | Travel restrictions imposed last week by Norway’s Svalbard archipelago to limit the spread of the coronavirus forced the cancellation of research flights in support of the *Polarstern*, the ice-bound German research ship that is the centerpiece of a \$150 million, yearlong scientific mission to the Arctic. The flights, slated to take off from Svalbard in March and April, were to collect key data that would have complemented ground-level observations by ship-based researchers with the project, the Multidisciplinary drifting Observatory for the Study of Arctic Climate. The flights had to be canceled after Svalbard banned entry of nonresidents. The restrictions also threaten a planned rotation in April of 100 researchers onto the ship, which has not reported any cases of COVID-19.

RNA vaccine testing begins

IMMUNOLOGY | A healthy volunteer in Seattle this week became the first person in the United States to receive an experimental vaccine against SARS-CoV-2, the virus that causes COVID-19. Developed by Moderna Therapeutics and the U.S. National Institute of Allergy and Infectious Diseases, the vaccine delivers messenger RNA that codes for a protein found on the surface of SARS-CoV-2. The company plans to enroll 45 healthy people aged 18 to 55 in the phase I study to test the vaccine’s safety and the immune response it triggers. Researchers in China earlier this month announced they had begun to test another experimental vaccine, though they did not detail its composition. A second China vaccine trial was approved to start this week as well, using a harmless adenovirus that makes SARS-CoV-2’s spike protein.

Trove of virus papers debuts

PUBLISHING | Hoping that artificial intelligence can yield new insights into the COVID-19 pandemic, White House science officials unveiled a database this week that pulls together more than 29,000 articles on SARS-CoV-2 and other coronaviruses, including the machine-readable full text of more than 13,000 papers. The project, which involves Microsoft, the Allen Institute for AI, the U.S. National Institutes of Health, and five other organizations, is offering \$1000 prizes to teams that use the data to answer questions such as the virus’ origin.



The ski-equipped German aircraft Polar 6 was to make a series of Arctic research flights over the next month.

THREE Qs

Academic lab aids virus testing

Faculty and staff members, graduate students, and postdocs at the University of Washington (UW), Seattle, School of Medicine got an unusual email last week. The department of laboratory medicine was overwhelmed by demand for processing samples using the COVID-19 test it had developed. It asked staff who could switch from their regular duties or use off time to help, even as the university’s inperson classes were halted. Geoffrey Baird, interim chair of laboratory medicine, told *Science* about a situation that additional academic labs could face as demand for testing escalates.

Q: How much COVID-19 testing are you doing now?

A: The day the [U.S. Food and Drug Administration] allowed us to do testing, we started off at about 100 tests a day and have ramped up to 1600 [as of 11 March]. Our plan is within 3 weeks to be able to do 7000 tests a day, at least.

Q: What prompted this call for help?

A: We’ve been able to optimize our assay to where [it] is no longer the slowest thing [in the process]. The actual stopping block is the number of people who can receive and accept samples. It isn’t highly technical work, but it’s critical to the effort, because we cannot provide a meaningful clinical turnaround time—which has to be less than 12 hours—with the huge backlog of just piles of specimens that would come in together. [The test kit is] a nasopharyngeal swab. We have to make sure it’s appropriately labeled, enter stuff into the computer, and do some minor processing steps to hand off the sample to the [UW] technologists who are licensed to actually do the testing.

Q: Did the call for help work?

A: Within the first day, we got over 130 responses [from UW volunteers], and I believe we’re still getting more. It’s been very, very heartening. We have, still, a need. We’re going to need to keep doing this for weeks and weeks, if not months. And I’m not getting new full-time employees for the lab. We’re at 100 meters and we’re sprinting, but there’s a marathon left.

S [SCIENCEMAG.ORG/NEWS](https://sciencemag.org/news)
Read more news from *Science* online.



IN DEPTH

A woman gets swabbed at a drive-through testing site in West Palm Beach, Florida, on 16 March.

INFECTIOUS DISEASES

Countries test tactics in ‘war’ against COVID-19

Lockdowns and closings proliferate, but virus testing and contact tracing are lagging

By **Jon Cohen** and **Kai Kupferschmidt**

The United States and Europe have stopped hitting the snooze button. After 2 months of mostly waiting and seeing while the coronavirus disease 2019 (COVID-19) alarm sounded ever more loudly, many countries have suddenly implemented strict measures to slow the spread of the disease, which the World Health Organization (WHO) officially declared a pandemic on 13 March. Thousands of events have been canceled; schools, restaurants, bars, and clubs have been closed; and transit systems are at a standstill.

The countries saw little choice. The case numbers exploded, and, in turn, so did the number of deaths. Hospitals in Italy, the hardest hit European country, are overburdened, forcing doctors to make agonizing decisions about whom to treat and on whom to give up. “This is bad,” U.S. President Donald Trump finally acknowledged on 16 March. “This is war,” his counterpart Emmanuel Macron told the French people the same day.

But how to fight that war is still under discussion. The hastily introduced measures vary widely between countries and even within countries. The U.S. government advises against gatherings of more than 10 people, but San Francisco has ordered everyone to stay at home. Italy, France, and Spain have

put their populations on an almost complete lockdown, with police or the military in some places patrolling the streets, but as *Science* went to press, pubs in the United Kingdom remained open. Germany, like many countries, has shut its schools, but they remain open for younger children in Sweden.

The patchwork reflects different phases of the epidemic, as well as differences in resources, cultures, governments, and laws. But there’s also confusion about what works best, and how to balance what is necessary with what is reasonable, especially for an extended period. South Korea, Hong Kong, and Singapore seem to hold important lessons, having turned their epidemics around without the authoritarian tactics used by China. Yet some of the strategies adopted in those countries are missing elsewhere: widespread testing to find cases, tracing their contacts to test or quarantine them, and encouraging—or forcing—infected people to isolate themselves.

No single step will suffice, WHO Director-General Tedros Adhanom Ghebreyesus stressed at a recent press conference. “Not testing alone. Not contact tracing alone. Not quarantine alone. Not social distancing alone. Do it all.”

SOCIAL DISTANCING

There’s little doubt that social distancing—keeping people from getting physically

close—can greatly reduce virus transmission: It was essential to bringing China’s raging epidemic under control in a matter of weeks, according to the report of a joint mission of WHO and the Chinese government released on 28 February (*Science*, 6 March, p. 1061). Other countries are now deciding how far to take that approach.

Many began by banning gatherings of more than 1000 and then successively reduced that number. Some have shut theaters, cinemas, restaurants, and gyms as well as all places of worship. Germany has closed most non-essential stores but extended hours for supermarkets to reduce the number of shoppers at any one time. In some countries, shops are reserving the first hours of the day for older customers at high risk of severe disease.

School closings have sent more than half a billion children home, according to UNESCO. Whether that makes sense is under debate. COVID-19 rarely sickens children, and it’s not clear how often they develop asymptomatic infections and transmit the virus. School closures may have the added benefit of forcing more parents to stay home. On the other hand, some children may end up being looked after by elderly grandparents, and closures may force badly needed health care workers to stay home. Moreover, children could end up missing months of education and many depend on free school lunch programs.

That's why some public health experts say measures should be flexible. Austria and the Netherlands have sent most students home, but schools remain open for children of those working in vital sectors. Singapore has halved class sizes, instituted strict hygiene measures, and staggered break periods to reduce playground contact.

Several countries have now resorted to an extreme measure: forcing almost their entire population to stay home. China led the way in late January, when it penned in more than 50 million people in Hubei province. Some experts argued that Western countries could never enforce such draconian measures—which curtail human rights and cripple economies—but Italy, shocked by the strain on the health care system in the north of the country, followed suit on 9 March. In France, 100,000 police officers began to patrol the streets on 17 March to make sure people stay inside except for essential trips.

TESTING AND ISOLATING

Other countries have beat back the virus without such drastic measures. One example is South Korea, which has seen confirmed infections drop from 909 cases on 29 February to just 74 early this week. “South Korea is a democratic republic; we feel a lockdown is not a reasonable choice,” says Kim Woo-Joo, an infectious disease specialist at Korea University.

Instead, the key to success has been a large, well-organized testing program, combined with extensive efforts to isolate infected people and trace and quarantine their contacts. By 16 March, South Korea had tested more than 270,000 people, many at a network of dozens of drive-through testing stations, a strategy followed elsewhere that eases access to testing and prevents infected people from exposing others in waiting rooms.

But the United States, plagued by an overly bureaucratic system and problems with its test kits, has had a slow start. By 16 March it had done only 74 tests per million inhabitants, compared with 5200 tests per million in South Korea. Only this week did the United States begin to roll out testing on a mass scale. In Europe, Germany is a front-runner, with more than 100,000 tests processed per week, says Christian Drosten, a virologist at the Charité University Hospital in Berlin, who developed the test. But other countries have yet to scale up testing.

The slow rollout has rankled Tedros, who repeats his mantra “testing, testing, test-

ing” almost daily. Countries “cannot fight this pandemic blindfolded,” he said at a 16 March press conference. “They should know where the cases are.” Marcel Salathé, a computational epidemiologist at the Federal Institute of Technology of Lausanne, agrees. “At this point 100% of nations that got it under control did so based on testing and tracing, isolation, quarantining,” he says. What’s required is “a determination to find every single infection and follow up on every potential exposure and break every possible chain of transmission.”

Even if they start to test more widely, some countries may lack the capacity to trace the contacts of those infected. In the United States, the job falls to state and local health departments, which often lack the resources to scale up rapidly. “It’s going to vary immensely by jurisdiction,” says epidemiologist Caitlin Rivers of the Johns Hopkins University Bloomberg School of Public Health.

in 2003, says Mark Woolhouse, an epidemiologist at the University of Edinburgh: “We will be living with this virus indefinitely.” Keeping it at bay might require locking down society for many months, at staggering costs to the economy, social life, and mental health, at least until a vaccine is available. That is inconceivable to Woolhouse and many others.

A few countries are now thinking about gradually letting the population build up immunity by forgoing a complete lockdown and allowing some infections to take place, preferably in low-risk groups such as children or young adults. That’s the strategy Prime Minister Mark Rutte of the Netherlands announced in a televised address in 16 March. “By taking this approach, one in which most people will experience only minor symptoms, we can both build immunity and ensure that our health care system is able to cope,” Rutte said. In a TV interview, epidemiologist Jaap

van Dissel of the Dutch National Institute for Public Health and the Environment explained that the goal was to “titrate” control measures to keep the demand for hospital beds below maximum capacity, a strategy called mitigation. (The U.K. government last week suggested it wanted to build up herd immunity as well, but began to backpedal after receiving pushback.)

A modeling study by researchers at Imperial College London, posted online on 16 March, concluded that even a mitigated epidemic would still overwhelm health care systems and cause at least 250,000 deaths in the United Kingdom and more than 1.1 million in the United States. Suppressing the virus by combining all available measures, including school closings and social distancing of the entire population, is the “only viable strategy

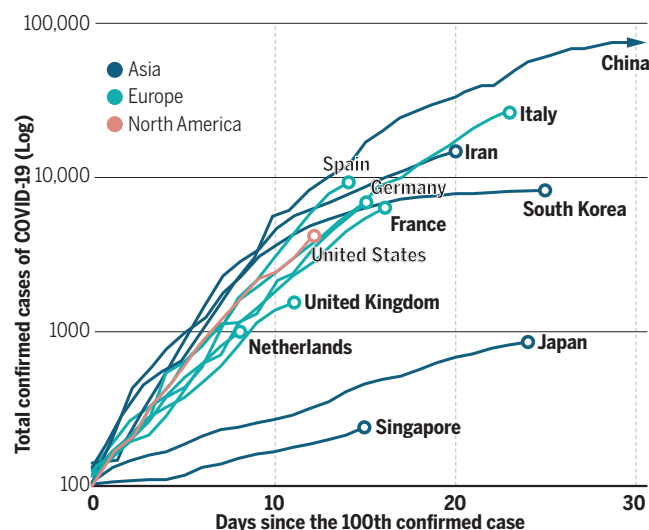
at the current time,” the team wrote.

It did suggest a scheme in which these draconian measures could be relaxed once in a while—a kind of collective “drug holiday”—and then reimposed when case numbers start to climb again. In that scenario, the population would still build up immunity to the virus, but through a series of small outbreaks instead of a massive one. It may not be an attractive scenario, but there may be no other choice. As epidemiologist Seth Berkley, who heads GAVI, the Vaccine Alliance, says, “You cannot say the Earth has to stop for a year or 2 years.” ■

With reporting by Dennis Normile and Martin Enserink.

Exponential growth

China’s outbreak has come to a halt and South Korea has flattened its curve, but COVID-19 case numbers are still rising rapidly in many Western countries.



One way to solve that is to engage the public more actively, says Luciana Borio, the pandemic preparedness point person at the U.S. National Security Council from 2017 to 2019. Borio hopes infected people, if properly educated, will isolate themselves and ask their recent contacts to seek tests, too. “I prefer to empower and educate the population to be able to help them take matters into their own hands,” she says.

NO ENDGAME

For many, the biggest question is: When, and how, will it end? It’s now clear that humanity won’t get rid of COVID-19 as it did with SARS (severe acute respiratory syndrome)



Columbia University, like many other institutions facing local outbreaks, has ramped down research.

RESEARCH COMMUNITY

Coronavirus disruptions reverberate through research

Scientists grapple with halted projects and locked-down labs

By **Kelly Servick, Adrian Cho, Jennifer Couzin-Frankel, and Giorgia Guglielmi**

Bacterial colonies slid into freezers for an indefinite rest. Clinical trial participants told to stay home. Populations of lab mice reduced to the bare minimum. With a surge in cases of coronavirus disease 2019 (COVID-19), research has been upended around the world as access to labs is restricted and travel curtailed to slow the virus' spread.

Many scientists emphasized that disruptions to their work can't compare with the human toll of the pandemic. But the costs to research will be substantial as projects grind to a halt. Institutions facing local outbreaks have made tough decisions about which research to prioritize, often favoring projects addressing the new coronavirus while cutting back others. Many labs have been forced to designate a couple of essential personnel to maintain equipment, cells, and animals, while sending everyone else home. Some are searching for workarounds to keep things ticking, while others resign themselves to losing data and materials.

Cancer geneticist Alberto Bardelli had just 48 hours to close his lab at the University of Turin, after receiving a 7 March email from the institution's scientific director and director general. When he told the

25 lab members about the closure, some cried. "The lab becomes a second family," he says. The researchers rushed to freeze patient samples, but they don't know whether these will be usable once thawed.

At Harvard University, where a 6- to 8-week period of "suspended lab access" began this week, evolutionary biologist Hopi Hoekstra and her team scrambled to plan for their mouse colonies in case the animal care team becomes short-staffed during the outbreak. She thinks she may have to reduce her mouse population by half—hundreds of animals—which could require killing mice. At risk are strains collected in the wild that exhibit unique behaviors, she says. "If something happens, or they're not getting long-term care and we reduce the colony so much that we can't keep them going, [those unique mice are] lost forever."

At the world's biggest atom smasher, the Large Hadron Collider, engineers struggled to complete scheduled upgrades to the collider and the four giant particle detectors it feeds. Only a skeleton crew remained onsite near Geneva to implement work directed by thousands of far-flung researchers. "I've spent probably 3 hours over the past 24 on Skype with somebody, guiding him through the installation of one bit of our electronics," says John Hobbs, a particle physicist at Stony Brook University who works on a detector known as ATLAS.

Shutdown policies remain murky at U.S. national labs run by the Department of Energy (DOE), which offer x-ray sources and other facilities to more than 30,000 visiting researchers each year. Some, including Brookhaven National Laboratory and Argonne National Laboratory, have closed their doors to outside scientists. But operations that could help combat the virus will likely continue. On 12 March, neuroscientist Chris Fall, director of DOE's basic research arm, the Office of Science, wrote to DOE-sponsored researchers urging them to suggest ways they might join the fight. DOE's five particle accelerator-based x-ray sources may help decipher the structures of proteins associated with the virus. And the Office of Science's three supercomputers could help sort through masses of data on existing drugs to see which molecules might bind to the new virus and knock it out of action.

Meanwhile, academic medical centers are considering which of their thousands of clinical trials to maintain. "We felt that with COVID-19, the risk-benefit had shifted," says Deborah Stiles, vice president for research operations and policy at Columbia University. Beyond the risks and benefits of an experimental drug or intervention, managers must now throw in the danger of contracting and spreading the virus during inperson visits. Columbia has paused studies involving human subjects where participants didn't stand to benefit directly. The move allows most trials of therapeutics to continue, although investigators are discouraged from enrolling new participants.

Johns Hopkins University has sorted clinical research into three tiers, reflecting different risk-benefit trade-offs. Studies of potential COVID-19 treatments, such as Gilead's antiviral remdesivir, or those addressing acute, life-threatening conditions, are considered essential. Representing about 10% to 15% of Hopkins's 1500 clinical trials, this tier can continue normally, including enrolling new patients.

Studies in the second tier, including many cancer trials, have halted new enrollment, but can continue as long as they limit face-to-face contact—by mailing drugs, for example. The third tier includes cohort studies that follow volunteers long term. These can proceed only online or by telephone, says Daniel Ford, the university's vice dean for clinical investigation.

Protecting patients and staff from the virus might mean straying from a trial's planned protocol, acknowledges Jonathan Epstein, executive vice dean and chief sci-

entific officer at the University of Pennsylvania. “We’re hopeful that regulatory agencies would understand” this decision, even if it affects the data investigators can gather, he says.

Travel restrictions, meanwhile, are hobbling field research. Epidemiologist Camila González-Beiras and her colleagues in Spain at the Fight AIDS and Infectious Diseases Foundation were unable to carry out the last planned steps in a large trial of treatments for the ulcer-causing skin disease yaws in Papua New Guinea.

The trial, comparing two antibiotic dosing regimens, had collected most of its data, and the team planned to return to the country mainly for ethical reasons. After preliminary evidence showed that multiple doses conferred greater protection, the team wanted to give a second dose to those in the single-dose arm of the study; they also wanted to look for potentially drug-resistant cases. But flying a team from Spain, which has more than 7000 COVID-19 cases, to Papua New Guinea, which had no documented cases as *Science* went to press, “would be very irresponsible,” González-Beiras says.

Unable to continue with hands-on research, many scientists say they’re channeling their energies into drafting manuscripts, analyzing data, and writing grant proposals. Many have just begun to contemplate the consequences of their hibernation—how a drop in productivity could affect funding prospects and tenure decisions, for example.

In hard-hit Italy, at the University of Pavia, structural biologist Federico Forneris ponders changing his research priorities altogether. His lab, which studies synapse formation and collagen production, might have to shift its focus to computational biology, which can be done as telework, if the lockdown goes on too long, he says.

As the pandemic’s impacts sink in, some scientists may find it hard to resist going into the lab in places where that’s still allowed. But others urge them to stay away, for the sake of fighting the pandemic. “We love our science, all of us do ... but there are more important things sometimes,” says evolutionary biologist Richard Lenski of Michigan State University, who last week paused a 32-year experiment that has observed populations of *Escherichia coli* bacteria through more than 73,000 generations. “It’s going to be disruptive to science, but it also reminds us that we have ordinary lives and connections to people that we need to safeguard.” ■

Georgia Guglielmi is a journalist in Cambridge, Massachusetts.

PALEONTOLOGY

Oldest fossil of modern birds is a ‘turducken’

Ancestral bird skull looks like a duck from the back and a chicken from the front

By Gretchen Vogel

Go to a Cajun restaurant in New Orleans, and you might be offered a slice of turducken: a fancy dish of chicken stuffed inside of a duck stuffed into a turkey. Now, paleontologists have their own version: the oldest modern bird skull ever found, which predates the split between the duck lineage and that of both chickens and turkeys—and so has traits of all three.

“This is an incredibly informative specimen,” says Amy Balanoff, a paleontologist at Johns Hopkins University, Baltimore, who wasn’t involved in the work. Whereas the earliest birds, like the 150-million-year-old *Archaeopteryx*, look very different from today’s, the new fossil has clear characteristics of modern land and waterfowl, perhaps offering a glimpse of their common ancestor. Discovered near the Dutch town of Maastricht, in famous fossil beds that formed between 66.8 million and 66.7 million years ago, the turducken lived just before the mass extinction that killed off the dinosaurs. And because at least some of its descendants survived the cataclysm, “it gives us some clues about what characteristics were key in surviving that event,” Balanoff says.

Luck and technology prompted the find, says Daniel Field, a paleontologist at the University of Cambridge, who led the work. John Jagt, a curator at the Maastricht Natural History Museum, had spotted “four very small blocks of rock with broken limb bones poking out” in the museum’s collection, Field says. “It’s hard to imagine a less exciting looking fossil.” Just the same, Field and his postdoctoral fellow Juan Benito put the rock into a computed tomography scanner, hoping the x-rays would reveal the structures inside. When they saw the scan, Field says, their shouts made the technician run back into the room. “She thought we had broken the machine.”



Duck? Chicken? This seagull-size Cretaceous shorebird had features of ducks, chickens, and turkeys.

The scan revealed a complete skull of what looked like a modern bird. The bones in the top and the back of the head closely resemble those of modern ducks, whereas the face and beak have unfused bones, as seen in today’s chickens and turkeys. “You can play this game all day: ‘Oh, it’s a duck! No, it’s a chicken!’” Field says.

Most of the bird’s body is missing, but a piece of leg bone suggests it had long legs for its head size. Combined with the fact that the Maastricht deposits formed in a shallow sea, the fossil’s proportions suggest it was a small shorebird, about the size of a modern seagull.

In a *Nature* paper this week, Field and his colleagues named the bird *Asteriornis maastrichtensis*, for Asteria, the Greek goddess of falling stars who turns herself into a quail. The falling stars nod to the asteroid impact and extinction that struck not long after the bird lived. Some scientists had argued that modern birds evolved in the Southern Hemisphere, because the oldest modern bird fossils found until now came from Antarctica. But the new fossil is likely older than the Antarctic ones, arguing against that assumption.

The ability to look inside the intact rock was crucial to the discovery, Field says. The skull is less than 1 millimeter away from the femur, so “if we had started chipping away, we would have destroyed the skull.” So was the team’s willingness to gamble on an unassuming rock, he adds. “We have to be more hopeful in our collecting.” ■

IMAGE: PHILLIP KRZEMINSKI

After Brexit, U.K. budget offers boost to science

Unprecedented increase includes money to set up high-risk research agency

By Erik Stokstad

For scientists in the United Kingdom, the country's departure from the European Union threatens a major source of grants and collaboration. But now they have been promised help: an unprecedented boost in domestic research funding. Last week, in its budget announcement, the U.K. government said it would increase public R&D by 15% to £13 billion in the fiscal year beginning next month—with even bigger raises to come. “This really is a blockbuster of a budget,” says Graeme Reid, a science policy researcher at University College London. “It will change both the shape and the scale of research and innovation in the U.K.”

Scientists can thank Brexit, which occurred on 31 January. Prime Minister Boris Johnson and his advisers see science as an engine of innovation for the post-Brexit economy. To that end, the new budget includes £800 million to set up a funding agency modeled on the storied U.S. Advanced Research Projects Agency, which spawned economically important inventions such as the internet. The budget also includes £400 million this year for parts of the country that have lagged in capturing research grants, a start to spreading money beyond the “golden triangle” of London, Oxford, and Cambridge. “To me these are quite bold things to try,” says John Womersley, director-general of the European Spallation Source and former head of a U.K. research funding council.

U.K. researchers get a lot of bang for the buck or, rather, power for the pound. Their research is cited more often than that of most of their international peers, even though total R&D spending, public and private, stands at about 1.7% of gross domestic product (GDP)—well below the 2.4% average of economically developed countries. Last year, Johnson and the Conservative Party campaigned on a pledge to reach that average by 2027 and double the public share of R&D funding by 2024. Presenting the new budget outline, finance minister Rishi Sunak said the government would meet that second goal in 2024, with public R&D spending hitting £22 billion, or 0.8% of GDP—more, relative to GDP, than in the United States, Japan, and France.

To reach the overall goal of 2.4%, how-

ever, the private sector needs to increase its R&D spending from £26 billion to £44 billion, despite the economic risks posed by Brexit and now the coronavirus pandemic. To stimulate private investment, the government will put £200 million in a life sciences venture capital fund and spend £900 million on grants to foster business innovation. It is also trying to strengthen connections between academic and industrial researchers, which will require creativity, says Ottoline Leyser, a plant biologist at the University of Cambridge and a member of the prime minister's Council for Science and Technology. “You can't just throw more money,” she says.

By supporting unconventional ideas with distant commercial payoffs, the new agency—modeled on the civilian precursor to the U.S. Defense Advanced Research Projects Agency—could also help. Establishing it with a large pot of money is good, Womersley says, because a high-risk funding agency needs to invest in the long-term and shrug off the inevitable failures. The budget announcement did not reveal when the agency might launch or whether

it will be part of UK Research and Innovation (UKRI), the overarching government funding council created in 2018 to coordinate U.K. research funding. The agency will need leaders who understand the U.K. research landscape, says Anne Glover, a molecular biologist and president of the Royal Society of Edinburgh, Scotland's national academy. “If it's just something like a hobby for politicians, then this is not going to work.”

A more immediate windfall is the extra £400 million to be spent this year on research infrastructure in regions outside the golden triangle, concentrating on basic research and the physical sciences. The budget provides enough funding to support policy experiments, such as giving block grants to local authorities or creating a regional structure for UKRI, says policy analyst Madeleine Gabriel of Nesta, a foundation that studies innovation. “We're only going to build capacity outside the golden triangle by doing something bold to shift the distribution of funding.”

The budget includes another big infrastructure project, a 10-year, £1.4 billion renovation of an animal and plant health research facility in Weybridge. Its focus is protecting U.K. agriculture, but it also studies zoonotic diseases that—like the new coronavirus—spread from animals to humans. James Wilsdon, a science policy expert at the University of Sheffield, says it makes sense to spend money on infrastructure now, to prepare facilities for the ramp-up in research spending.

To effectively absorb more research funding, Womersley says the United Kingdom will need to attract and retain more postdocs and other talent from abroad. The U.K. Home Office in February announced a global talent visa as part of changes to immigration policy. But Brexit has caused lasting reputational damage that may make it harder to recruit from other countries, Glover says. “You can destroy a welcoming environment in an instant and it takes a long time to recover,” she says. Reid, on the other hand, is optimistic about the potential for collaborations, despite Brexit. “You're looking at a country with a strong research base that's just received an unprecedented funding boost. And that makes the U.K. an extraordinary partner of choice.” ■

A post-Brexit bounty

U.K. Prime Minister Boris Johnson wants to double public R&D spending by 2024. The budget overview, announced this week, highlights science projects of different size and duration.

FUNDING (MILLIONS £)	TIME (YEARS)	PURPOSE
1400	10	Upgrade government's animal health science facility
900	?	Business innovation, including space and nuclear fusion technology
80	?	High-risk, high-reward funding agency modeled on ARPA
400	1	Research and infrastructure, particularly in basic and physical sciences
300	5	Math research fellowships to attract global talent
200	?	Health and life science venture capital fund
180	6	New storage and research facility for the Natural History Museum
80	5	Specialist institutions including the Institute of Cancer Research



An enumerator collects data during India's last national census, in 2011.

STATISTICS

Unrest imperils India's census

Opposition to citizenship policies could lead to undercount

By **Vaishnavi Chandrashekhar**

Next month, some 3 million enumerators will fan out across India in a once-a-decade ritual, gathering data for one of Asia's oldest and most respected censuses. But this year could be different. Social scientists in India fear political unrest will disrupt the count, compromising critical demographic data for years to come. In particular, they worry that opposition to controversial new citizenship policies could cause many of India's 1.3 billion people to refuse to participate in the census. And recent attacks on fieldworkers conducting other government surveys have raised concerns about the safety of India's enumerators, who will begin work on 1 April. (The coronavirus outbreak might also disrupt the count.)

"This is a situation we've never faced," says Pronab Sen, India's former chief statistician. "We've always had cooperation in the census' long history."

India has conducted the tally since 1881. "There are very few countries in Asia with such a census," says Perianayagam Arokiasamy, a specialist at the International Institute for Population Sciences. The data play a key role in political and economic decisions. They are used to draw legislative districts, apportion government spending, and inform investment decisions by businesses. Researchers use the data for studies and to design their own surveys. "There is really no other data for understanding the country's demographics," Arokiasamy says.

Three moves by India's government, which is led by the Hindu nationalist Bharatiya Janata Party (BJP), have heightened tensions around collecting population data. BJP leaders have vowed to create the National Register of Citizens, a list of people who can prove, through birth certificates or other ancestry records, that they qualify for citizenship. Last year, the party pushed through a law designed to fast-track citizenship for religious minorities fleeing from persecution in surrounding countries—but pointedly excluded Muslims, a minority group in India. In December 2019, the government also added sensitive new questions to a survey called the National Population Register (NPR), last conducted in 2010. The new NPR, to start in April, will ask respondents when and where their parents were born, for example; such information can be used to determine citizenship. Many observers believe the NPR is intended to help construct the register of citizens.

The moves have sparked protests across India, especially in Muslim communities, where many people fear they do not have the documents needed to prove citizenship, and so will become stateless. When officials in the border state of Assam created a register of citizens last year, some 1.9 million residents, including the family of a former president of India, could not prove citizenship; the state government has begun to build large detention camps, apparently to house noncitizens. In some regions, residents have attacked workers carrying out economic surveys or doing public health work, mistakenly believing they were collecting citizenship information.

Census watchers fear such problems could escalate when the census begins next month, in part because the government will conduct the more controversial NPR at the same time. "My worry is that ... people will confuse the two [surveys] and refuse to give any information," Sen says. Ten of India's 22 states have passed resolutions opposing the NPR.

A faulty census would open a Pandora's box, experts say. Resistance to enumeration could result in undercounts, especially in Muslim communities, leading to reduced funding. "I worry about the implications for the well-being of Muslim communities," says sociologist Sonalde Desai of the University of Maryland, College Park, and a senior fellow at the National Council of Applied Economic Research. "If they don't cooperate with the census, the data for that community will be of poor quality, which could affect services."

Flawed data could also reduce the reliability of other surveys that rely on the census. "Any subsequent survey or household data becomes contaminated," Sen says. And efforts to fill in missing data by statistical tinkering could be technically and politically fraught. "We could do adjustments for undercounts, but how well can we do it?" Desai asks. "What would be the baseline data? Would the adjustments be a political issue?" (It might help, Desai says, if enumerators kept track of people who refuse to be counted.)

Concerns about the census come as experts are already questioning some of India's other national data. Economists have raised questions about the government's gross domestic product calculations, as well as decisions to suppress or discard surveys that reflect poorly on the economy. The Ministry of Statistics and Programme Implementation has set up a panel, chaired by Sen, to examine ways to improve the quality of economic data. "We're still taking stock of data sets on the ground," Sen says. "A lot depends on the census and the economic census."

BJP leaders, meanwhile, have tried to calm protests by promising not to create the national registry of citizens any time soon. And last week, the home minister told Parliament that no one would be marked "D" for "doubtful citizen" in the upcoming population survey. But Sen and some 200 other experts say the government should consider postponing the NPR entirely. Time is running short: The first phase of the census, which identifies households, is set to end in September, and the second phase—which counts individuals—will begin early next year. ■

Vaishnavi Chandrashekhar is a journalist in Mumbai, India.

PHOTO: ANSHUMAN POYREKAR/HINDUSTAN TIMES/GETTY IMAGES

SCIENTIFIC COMMUNITY

Top neuroscientist leaving Mexican university

Alleged incidents of sexual harassment by Ranulfo Romo Trujillo began years ago

By **Inés Gutiérrez**, in Mexico City,
and **Rodrigo Pérez Ortega**

Earlier this month, Mexico's leading university, the National Autonomous University of Mexico (UNAM), announced that renowned neuroscientist Ranulfo Romo Trujillo would leave his position after being disciplined for an unspecified offense.

According to a 4 March press release from UNAM, Romo Trujillo voluntarily asked to be separated from his job at its University City campus in Mexico City. Sources close to the case say he had been temporarily suspended because a female worker made a formal complaint of sexual harassment against him following an incident in January. But current and former UNAM students and staff say that reports of inappropriate behavior by Romo Trujillo had circulated for years before his departure.

Romo Trujillo, who works at UNAM's Institute of Cellular Physiology (IFC), did not respond to repeated requests for comment. He is arguably the most famous neuroscientist in Mexico, studying perception, working memory, and decision-making. He has more than 150 publications, including in top journals such as *Science* and *Nature*; is on the editorial board of *Neuron* and other journals; and is one of 11 Mexican members of the U.S. National Academy of Sciences.

IFC physiologist Marcia Hiriart Urdanivia acknowledged in an email to *Science* that, while director of IFC from 2009 to 2017, she received multiple accounts of sexual harassment or inappropriate conduct by Romo Trujillo. Hiriart Urdanivia says she warned Romo Trujillo that "his career was endangered by such actions." But the women involved did not choose to file official complaints, she says. As a result, "I had no authority to do anything else."

Six former trainees at UNAM told *Science* that they witnessed or experienced misconduct or harassment by Romo Trujillo, including inappropriate and sexually suggestive comments, grabbing women without their consent while he was drunk at the institute, and making unwanted physical advances. The alleged victims and witnesses say they did not file formal complaints because they feared reprisal or thought that his status would protect him from disciplin-

ary action. They asked to remain unnamed.

One woman was a graduate student at IFC in 2006 when, she says, Romo Trujillo harassed her at a graduation party in the institute's cafeteria. According to her, he was drunk when he grabbed her by the arm and pulled her away from the party to take her to his lab to show her the monkeys he works with. "I got really scared," she says. "It was obvious that I did not want to go with him." Many at the party saw the incident, but only one person, a fellow student, helped her get away from Romo Trujillo, she says.

Students and staff say Romo Trujillo's lab now has no female students or technicians. More than 20 members of the IFC commu-



Multiple women alleged sexual harassment by Ranulfo Romo Trujillo of the National Autonomous University of Mexico, University City.

nity told *Science* Romo Trujillo's behavior was well-known, including to those in authority at the institute. "Everybody knows," one says.

Some students say they tried to warn new female students about Romo Trujillo. Others took preventive actions. "When I arrived at the lab, especially on the weekends, and I saw he was there, I used to lock myself up in the lab if I was alone," says one woman, who has been a student there since 2014.

One woman who claims to have been sexually harassed by Romo Trujillo says she did report incidents to Hiriart Urdanivia. But she did not feel she got enough guidance to file a complaint. "It's very difficult for a student without experience or knowledge

to file an official complaint as an individual person," she said.

Hiriart Urdanivia rejects this account and insists she did support students. "If any of the women who spoke to me about Dr. Romo's inappropriate behaviors had wanted to file a formal complaint, they would have had my full support as director," she wrote in an email to *Science*. "I did what I could to make people feel safe."

The January complaint about Romo Trujillo sparked outrage on social media. On 27 February, an anonymous Facebook post claimed that Romo Trujillo's sanction was an 8-day suspension, which the poster argued was far too lenient. The post, which claimed to be from an IFC student, went viral.

Romo Trujillo's departure comes at a time when universities throughout Latin America are only beginning to fight sexual harassment in science (*Science*, 21 February, p. 842). UNAM implemented a gender violence protocol in 2016, and according to officials, 1195 formal complaints have been filed since then. But advocates for women say its responses have often been weak. Starting in September 2019, protests demanding more action from authorities have shut down more than 20 UNAM schools.

On 12 February, UNAM said in a press release that it would change its general statute to view sexual harassment as a serious offense. That action formally recognizes that "violence against women is not allowed nor will it be tolerated," says a UNAM spokesperson. On 6 March, the National Council of Science and Technology, Mexico's main granting agency, similar to the United States's National Science Foundation, announced it would investigate unspecified allegations against Romo Trujillo and another UNAM scholar.

On 9 March, the National College—Mexico's honorary academy of scholars, of which Romo Trujillo is a member—announced that he asked to suspend his membership while he deals with the allegations. Some current students think the discipline and departure of such a prominent researcher will serve as a turning point. "We are tired of this patriarchal system in which the success of a man matters more than the dignity of a woman," says a current Ph.D. student. ■

Inés Gutiérrez is a journalist based in Mexico City.



SICK TIME

Dozens of diseases wax and wane with the seasons. Will COVID-19?

By Jon Cohen

On a December afternoon, 13 days before the winter solstice, six men and women checked into the Surrey Clinical Research Facility, part of the University of Surrey in the United Kingdom. After having their noses swabbed to check for 16 different respiratory viruses, they walked into their own temperature-regulated rooms and, for 24 hours, stayed in a semirecumbent position in dim light. Nurses placed a cannula into a vein of each person's arm, allowing easy sampling of blood that flowed through a tube to portals in the wall. The six subjects could press buzzers for bathroom breaks, where the stool and urine were collected, but otherwise, they were alone in the near-dark.

None of these people were sick. And although the shortest day of the year was approaching, their ritual had nothing to do with pagan rites, Yuletide traditions, or the annual hippie gathering at nearby Stonehenge to celebrate the rebirth of the Sun.

Instead, they were paid volunteers in a study led by infectious disease ecologist Micaela Martinez of Columbia University to investigate a phenomenon recognized 2500 years ago by Hippocrates and Thucydides: Many infectious diseases are more common during specific seasons. "It's a very old question, but it's not very well studied," Martinez says.

It's also a question that has suddenly become more pressing because of the emergence of coronavirus disease 2019 (COVID-19), which has now infected more than 175,000 people around the globe. Some hope the disease might mimic influenza and abate as summer arrives in temperate regions of the Northern Hemisphere, where about half of the world's population lives. U.S. President Donald Trump has expressed that hope repeatedly. "There's a theory that, in April, when it gets warm—historically, that has been able to kill the virus," Trump said on 14 February. But what's known about other diseases doesn't offer much support for the idea that COVID-19 will suddenly disappear over the next few weeks.

Different diseases have different patterns.

People seeking help for pandemic influenza in Brazil in July 2009, when cold weather boosted the spread of the disease.

Some peak in early or late winter, others in spring, summer, or fall. Some diseases have different seasonal peaks depending on latitude. And many have no seasonal cycle at all. So no one knows whether SARS-CoV-2, the virus that causes COVID-19, will change its behavior come spring. "I would caution overinterpreting that hypothesis," Nancy Messonnier, the point person for COVID-19 at the U.S. Centers for Disease Control and Prevention, said at a press conference on 12 February. If the seasons do affect SARS-CoV-2, it could nevertheless defy that pattern in this first year and keep spreading, because humanity has not had a chance to build immunity to it.

Even for well-known seasonal diseases, it's not clear why they wax and wane during the calendar year. "It's an absolute swine of a field," says Andrew Loudon, a chronobiologist at the University of Manchester. Investigating a hypothesis over several sea-

sons can take 2 or 3 years. “Postdocs can only get one experiment done and it can be a career killer,” Loudon says. The field is also plagued by confounding variables. “All kinds of things are seasonal, like Christmas shopping,” says epidemiologist Scott Dowell, who heads vaccine development and surveillance at the Bill & Melinda Gates Foundation and in 2001 wrote a widely cited perspective that inspired Martinez’s current study. And it’s easy to be misled by spurious correlations, Dowell says.

Despite the obstacles, researchers are testing a multitude of theories. Many focus on the relationships between the pathogen, the environment, and human behavior. Influenza, for example, might do better in winter because of factors such as humidity, temperature, people being closer together, or changes in diets and vitamin D levels. Martinez is studying another theory, which Dowell’s paper posited but didn’t test: The human immune system may change with the seasons, becoming more resistant or more susceptible to different infections based on how much light our bodies experience.

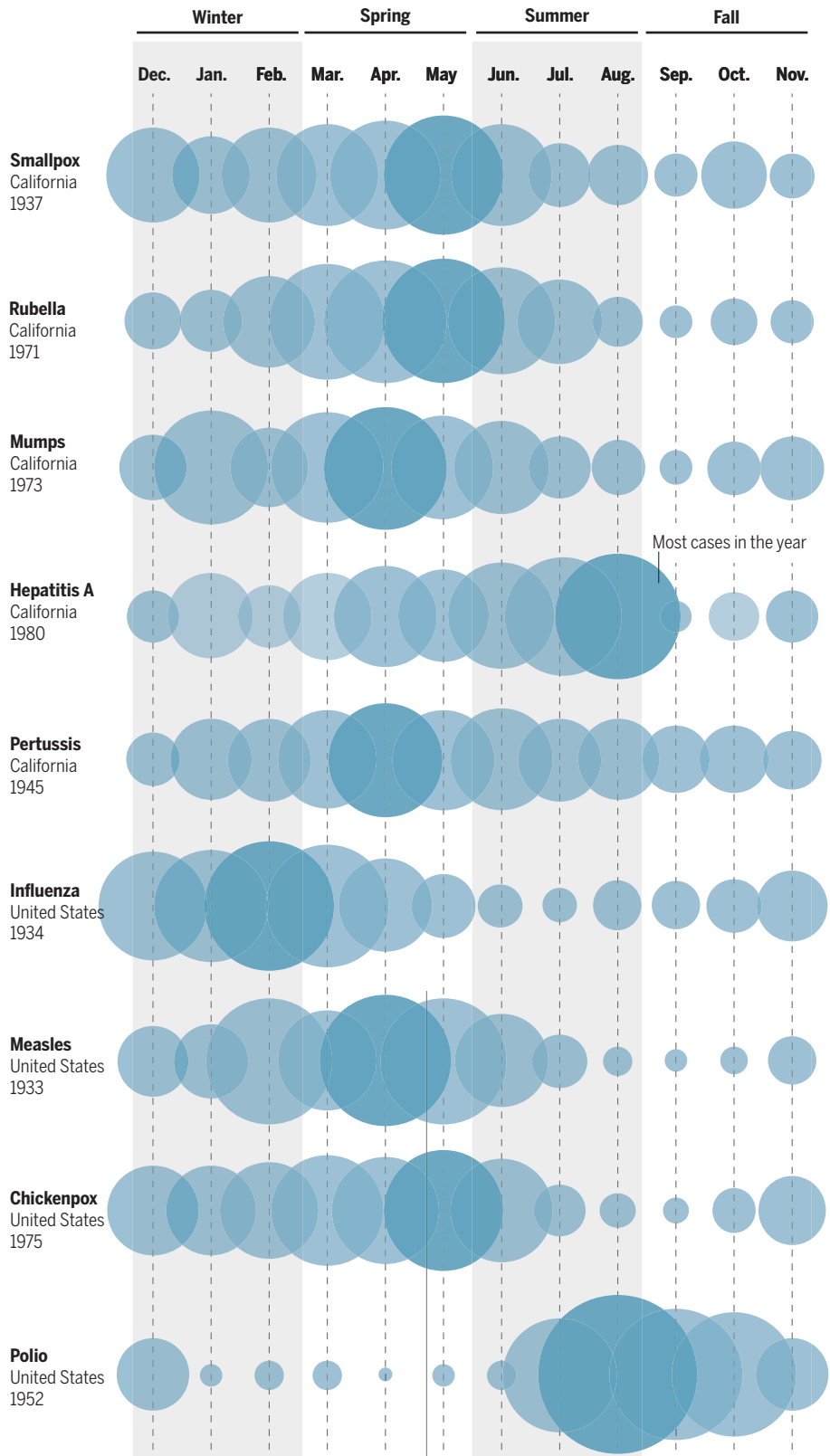
Beyond the urgent question of what to expect with COVID-19, knowing what limits or promotes infectious diseases during particular times of year could inform disease surveillance, predictions, and the timing of vaccination campaigns. It might even point to new ways to prevent or treat them. “If we knew what suppressed influenza to summer-time levels, that would be a lot more effective than any of the flu vaccines we have,” Dowell says.

MARTINEZ BECAME interested in seasonality when, as an undergraduate at the University of Alaska Southeast, she had a job tagging Arctic ringed seals, doing skin biopsies, and tracking their daily and seasonal movements. While working on her Ph.D., her focus on seasonality shifted to polio, a much-feared summer disease before the advent of vaccines. (Outbreaks often led to the closing of swimming pools, which had virtually nothing to do with viral spread.) Polio seasonality in turn made her curious about other diseases. In 2018, she published “The calendar of epidemics” in *PLOS Pathogens*, which included a catalog of 68 diseases and their peculiar cycles.

Except in the equatorial regions, respiratory syncytial virus (RSV) is a winter disease, Martinez wrote, but chickenpox favors the spring. Rotavirus peaks in December or January in the U.S. Southwest, but in April and May in the Northeast. Genital herpes surges all over the country in the spring and summer, whereas tetanus favors mid-summer; gonorrhea takes off in the summer and fall, and pertussis has a higher inci-

The calendar of epidemics

At least 68 infectious diseases are seasonal, according to a 2018 paper by Micaela Martinez of Columbia University. But they’re not in sync, and seasonality varies by location. Here, each bubble represents the percentage of annual cases that occurred in each month. (The data are old because many diseases declined—in some cases to zero—after the introduction of vaccines.)



dence from June through October. Syphilis does well in winter in China, but typhoid fever spikes there in July. Hepatitis C peaks in winter in India but in spring or summer in Egypt, China, and Mexico. Dry seasons are linked to Guinea worm disease and Lassa fever in Nigeria and hepatitis A in Brazil.

Seasonality is easiest to understand for diseases spread by insects that thrive during rainy seasons, such as African sleeping sickness, chikungunya, dengue, and river blindness. For most other infections, there's little rhyme or reason to the timing. "What's really amazing to me is that you can find a virus that peaks in almost every month

of lipids. It interacts with host cells during the infection process and helps dodge immune attacks. Viruses with envelopes are more fragile and vulnerable to adverse conditions, Nathanson says, including, for example, summertime heat and dryness.

A 2018 study in *Scientific Reports* supports the idea. Virologist Sandeep Ramalingam at the University of Edinburgh and his colleagues analyzed the presence and seasonality of nine viruses—some enveloped, some not—in more than 36,000 respiratory samples taken over 6.5 years from people who sought medical care in their region. "Enveloped viruses have a very, very

weather changes. Influenza and RSV both did best when the change in relative humidity over a 24-hour period was lower than the average (a 25% difference). "There's something about the lipid envelope that's more fragile" when the humidity changes sharply, Ramalingam concludes.

Jeffrey Shaman, a climate geophysicist at Columbia, contends that what matters most for the influenza virus is absolute humidity—the total amount of water vapor in a given volume of air—and not relative humidity, which measures how close the air is to saturation. In a 2010 paper in *PLOS Biology*, Shaman and epidemiologist Marc Lipsitch of

the Harvard T.H. Chan School of Public Health reported that drops in absolute humidity better explained the onset of influenza epidemics in the continental United States than relative humidity or temperature. And absolute humidity drops sharply in winter, because cold air holds less water vapor.

Why lower absolute humidity might favor some viruses remains unclear, however. Variables that could affect the viability of the viral membrane could include changes in osmotic pressure, evaporation rates, and pH, Shaman says. "Once you get down to the brass tacks of it, we don't have an answer."

Will SARS-CoV-2, which has an envelope, prove fragile in spring and summer, when absolute and relative humidity climb? The most notorious of the other coronavirus diseases, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), offer no clues. SARS emerged in late 2002 and was driven out

of the human population in the summer of 2003 through intensive containment efforts. MERS sporadically jumps from camels to humans and has caused outbreaks in hospitals, but has never shown widespread human-to-human transmission. Neither virus circulated for long enough, on a wide enough scale, for any seasonal cycle to emerge.

Four human coronaviruses that cause colds and other respiratory diseases are more revealing. Three have "marked winter seasonality," with few or no detections in the summer, molecular biologist Kate Templeton, also at the University of Edinburgh, concluded in a 2010 analysis of 11,661 respiratory samples collected between 2006 and 2009. These three viruses essentially behave like the flu.

That does not mean COVID-19 will as well.



In a study in New York and New Jersey, Micaela Martinez hopes to find out how artificial lighting affects the immune system.

of the year in the same environment in the same location," says Neal Nathanson, an emeritus virologist at the University of Pennsylvania Perelman School of Medicine. "That's really crazy if you think about it." To Nathanson, this variation suggests human activity—such as children returning to school or people huddling indoors in cold weather—doesn't drive seasonality. "Most viruses get transmitted between kids, and under those circumstances, you'd expect most of the viruses to be in sync," he says.

Nathanson suspects that, at least for viruses, their viability outside the human body is more important. The genetic material of some viruses is packaged not only in a capsid protein, but also in a membrane called an envelope, which is typically made

definite seasonality," Ramalingam says.

RSV and human metapneumovirus both have an envelope, like the flu, and peak during the winter months. None of the three is present for more than one-third of the year. Rhinoviruses, the best-known cause of the common cold, lack an envelope and—ironically—have no particular affinity for cold weather: The study found them in respiratory samples on 84.7% of the days of the year and showed that they peak when children return to school from summer and spring holidays. Adenoviruses, another set of cold viruses, also lack an envelope and had a similar, nonseasonal pattern, circulating over half the year.

Ramalingam's team also studied the relationship between viral abundance and daily

The virus can clearly transmit in warm, humid climates: Singapore already has more than 240 cases. Two new papers published on preprint servers last week come to opposite conclusions. One, co-authored by Lipsitch, looked at COVID-19 spread in 19 provinces across China, which ranged from cold and dry to tropical, and found sustained transmission everywhere. The second study concludes that transmission appears to occur only in specific bands of the globe that have average temperatures between 5°C and 11°C and 47% to 70% relative humidity.

The other coronaviruses may be more susceptible to seasonal changes in the environment simply because they've been infecting people for much longer. Once a high percentage of the population develops immunity, an unfavorable environment can provide the extra push needed to temporarily exile those viruses. But that's not the situation with COVID-19. "Even though there might be a big seasonal decline, if enough susceptible people are around, it can counter that and continue for a long time," Martinez says. Lipsitch doesn't think the virus will go poof in April either. Any slowdown "is expected to be modest, and not enough to stop transmission on its own," he wrote in a recent blog post.

IN SURREY, MARTINEZ is investigating a different factor that might eventually affect COVID-19 incidence. Her subjects have returned to the clinic repeatedly—at the winter and summer solstices and again at the spring and fall equinoxes—so the researchers can evaluate how their immune system and other physiology change over the course of the day and from season to season.

She doesn't expect to show that our immunity is, say, weaker in the winter and stronger in the summer. But by counting different immune system cells, assessing metabolites and cytokines in the blood, deciphering the fecal microbiome, and measuring hormones, Martinez's team hopes to learn whether the seasons "restructure" the immune system, making some types of cells more abundant in certain parts of the body, and others less, in ways that influence our susceptibility to pathogens.

Animal studies support the idea that immunity changes with the seasons. Ornithologist Barbara Helm from the University of Groningen and her colleagues, for example, studied European stonechats, small songbirds that they caught and then bred in captivity. By taking multiple blood samples over the course of 1 year, they found that the birds ramp up their immune systems in the summer, but then tamp them down in the fall, the time they migrate, presumably because migration is a big drain on their energy.

Melatonin, a hormone primarily secreted at night by the pineal gland, is a major driver of such changes. The hormone keeps track of the time of day but is also a "biological calendar" for the seasons, says Randy Nelson, an endocrinologist at West Virginia University who specializes in circadian rhythms. When nights are long, more melatonin is released. "The cells say, 'Oh, I'm seeing quite a bit of melatonin, I know, it's a winter night.'" In studies of Siberian hamsters—which, like humans, are diurnal—Nelson and his co-workers have shown that administering melatonin or altering light patterns can change immune responses by up to 40%.

The human immune system, too, seems to have an innate circadian rhythm. For instance, a vaccine trial in 276 adults by researchers at the University of Birmingham randomly assigned half to receive an influenza vaccine in the morning and the other half in the afternoon. Participants in

"If we knew what suppressed influenza to summertime levels, that would be a lot more effective than any of the flu vaccines we have."

Scott Dowell, Bill & Melinda Gates Foundation

the morning group had significantly higher antibody responses to two of the three flu strains in the vaccine, the researchers reported in 2016.

There's evidence of seasonal variation in the actions of human immune genes as well. In a massive analysis of blood and tissue samples from more than 10,000 people in Europe, the United States, Gambia, and Australia, researchers at the University of Cambridge found some 4,000 genes related to immune function that had "seasonal expression profiles." In one German cohort, expression in white blood cells of nearly one in four genes in the entire genome differed by the seasons. Genes in the Northern Hemisphere tended to switch on when they were switched off south of the equator, and vice versa.

Just how these massive changes might affect the body's ability to fight pathogens is unclear, however, as immunologist Xaquín Castro Dopico and colleagues explain in a 2015 paper describing the findings. And some changes could be the result of an infection, instead of the cause. The team tried to eliminate people who had acute infections, but "of course a seasonal infectious burden likely plays a part," says Dopico,

who is now at the Karolinska Institute. And seasonal immunity changes could not explain all the complex variation in seasonality that diseases show. "They're all out of sync with each other," Nathanson points out. He's also skeptical that a seasonal immune system change could be large enough to make a difference. "It would have to be pretty markedly different."

Martinez, however, says she has found intriguing hints. Early analyses from her Surrey study, which collected its final data in December 2019, don't reveal anything about seasonality yet, but they do show that specific subsets of white blood cells that play central roles in immune system memory and response are elevated at certain times of day. She hopes to firm up the finding by launching a similar but larger study next year.

Martinez cautions that artificial light may play havoc with natural circadian rhythms, with unpredictable effects on disease susceptibility. To explore possible impacts, she has a separate study underway, with Helm, in both urban and rural parts of New York and New Jersey. They have installed light sensors on trees and poles and outfitted participants with devices that monitor light exposure and body temperature. "The fact that people really are just kind of washing out the rhythms in light exposure can be problematic," she says.

"EXPERIMENTS OF NATURE" could also offer insights into the factors affecting disease seasonality, Dowell suggested in his 2001 paper. People from the Southern and Northern hemispheres who have adapted to different seasons regularly mix on cruise ships or at conventions, where they are confronted by the same pathogens—witness the massive COVID-19 outbreak on the *Diamond Princess*, which was docked and quarantined in Yokohama, Japan, for 2 weeks last month. Researchers could potentially analyze whether they were infected at different rates.

Whatever the answers, they might eventually bring important public health benefits, Martinez says. For example, "If we know how best to administer vaccines, in terms of what time of year and the best time of day to take advantage of our immune systems, then we can get a lot more bang for our buck," she says.

The global COVID-19 emergency may bring more attention to the research and help speed discoveries, she says. But for now, no one knows whether rising humidity, longer days, or some as-yet-unsuspected seasonal effect will come to the rescue—or whether humanity must confront the pandemic without any help from the seasons.

Time will tell. ■

INSIGHTS

PERSPECTIVES

PHYSIOLOGY

Challenges of human nutrition research

Facilities to house and feed subjects could increase rigor and advance nutrition science

By **Kevin D. Hall**

Nutrition is fundamentally important for human health (1), but there is widespread public confusion about what constitutes a healthy diet. Flip-flopping headlines report conflicting information about whether individual foods (e.g., butter, eggs, meat), nutrients (e.g., saturated fat, cholesterol, sodium), or eating patterns (e.g., Mediterranean versus ketogenic diets) result in improved, worsened, or unchanged health. However, public confusion about nutrition belies expert consensus

regarding important aspects of healthy diets. For example, it is widely agreed that Western diets high in ultra-processed food are deleterious and that considerable health improvements would likely result from shifting the population toward eating mostly minimally processed foods (2). But expert consensus erodes when discussing detailed questions of optimal human nutrition or the physiological mechanisms underlying the body's response to diet changes. Rigorous controlled feeding studies would help to address such questions and advance human nutrition science, a field whose overall veracity has recently been questioned (3, 4).

Much of the criticism of nutrition science has been directed at nutritional epidemiology, a field that investigates associations

between diet and health outcomes in large numbers of people. Although nutritional epidemiology has ardent defenders (5, 6), its critics suggest that it is plagued by measurement error, reverse causality, selection bias, weak effects, analytical flexibility, and unmeasured or residual confounders that can result in spurious relationships between diet variables and health outcomes (7). Increased funding for large, long-term randomized diet intervention trials has been suggested as a way to mitigate reliance on nutritional epidemiology and improve causal inference about the effects of diet on human health (8). However, such trials have their own challenges, including the impracticality of randomizing large numbers of people to eat different diets for months or

National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD 20892, USA.
Email: kevinh@niddk.nih.gov

In 1945, a domiciled feeding study carried out at the University of Minnesota involved participants being fed a semistarvation diet.

years while ensuring high levels of adherence throughout.

Indeed, most randomized diet intervention trials do not actually study the effects of different diets; rather, they investigate the effects of differing diet advice. In other words, subjects are randomized to receive education and support to consume diets that are assigned by the investigators. Although diet-advice trials assess real-world effectiveness, their results conflate adherence to a given diet with the effects of that diet.

Knowledge about the effects of diet per se is required for advancement of fundamental nutrition science. However, studies in free-living people have a limited ability to provide such knowledge because it is not currently possible to accurately and objectively quantify their food intake. Indeed, most human nutrition studies rely on self-reported diet measures that are known to have systematic biases, such as underestimation of energy intake. Furthermore, errors in self-reported diet measurements may be associated with other variables (e.g., socioeconomic status) or health outcomes (e.g., obesity) that can result in biased associations (9).

Rather than relying on self-reported diet assessments, some diet intervention trials provide food to their free-living subjects, but these studies seldom verify whether all the food is eaten. Even when subjects are instructed to eat only the food provided by the study, substantial quantities of off-study food may be consumed amounting to several hundred kilocalories per day that can confound study results (10, 11). To understand how these challenges impede the progress of human nutrition science, imagine trying to develop a new drug without being confident that researchers could administer known quantities of the drug or measure its pharmacokinetics, pharmacodynamics, or dose response. Successful pharmaceutical development requires such studies because they investigate benefits and risks of the drug under highly controlled conditions where questions of patient adherence are minimized because the researchers administer the drug. The inability to conduct such trials would severely impede the drug development process. Why should human nutrition science be expected to advance without the benefit of well-controlled diet efficacy studies?

Therefore, it is important to conduct human nutrition studies where subjects can comfortably reside at a research facility, thereby allowing investigators to control and objectively measure their food intake. Subjects enrolled in such domiciled feeding

studies are required to stay at the research facility for periods of days, weeks, or months without leaving to ensure that they consume the provided food under observation while avoiding exposure to off-study food.

Domiciled feeding studies have a long history of yielding important discoveries about human nutrition and metabolism. For example, many of the physiological responses to starvation and nutritional rehabilitation were revealed in a controlled feeding study of 32 male volunteers who simultaneously resided at the University of Minnesota for a continuous 48-week period during the Second World War (12) (see the photo). The subjects were fed a baseline diet for 12 weeks followed by a 24-week semistarvation diet, after which they were fed several rehabilitation diets for the final 12 weeks. The resulting detailed physiological and psychological measurements in response to known diets would have been impossible had the subjects not been domiciled during this classic study.

Unfortunately, domiciled feeding studies have become prohibitively expensive in the United States since the National Institutes of Health ceased directly funding Clinical Research Centers (13). Very few centers around the world currently conduct domiciled feeding studies, and their study populations often comprise students, staff, and faculty, which limits their generalizability. Furthermore, the few facilities conducting domiciled feeding studies are typically limited to housing and feeding only a handful of subjects at a time, which restricts their power and duration.

Such limitations are surmountable. Investment in research facilities for domiciled feeding studies could provide the infrastructure and staff required to simultaneously house and feed dozens of subjects comfortably and safely. One possibility would be to create centralized domiciled feeding facilities that could enable teams of researchers from around the world to recruit a wide range of subjects and efficiently conduct rigorous human nutrition studies that currently can only be performed on a much smaller scale in a handful of existing facilities.

Well-designed domiciled feeding studies can increase the rigor of human nutrition science and elucidate the fundamental mechanisms by which diet affects human physiology. For example, such studies can investigate complex interactions among changes in diet, the microbiota, and its role in modulating host physiology. The effects of meal timing and circadian biology could be advanced by enabling precisely controlled periods for eating and sleeping. Personalized nutrition and nutrient-genomic interaction studies could be facilitated by reducing the usual noise of unknown diet variability to focus on indi-

vidual physiological variability in response to controlled diets. Nutrient requirements and their dependence on overall dietary and physical activity patterns could be assessed in a variety of populations of men and women of different ethnicities and ages. The effects of diet on physical and cognitive performance could also be carefully evaluated. Comprehensive assessment of the effects of diet interventions on common health conditions such as obesity, metabolic syndrome, and type 2 diabetes, as well as rare diseases such as those that result from inborn errors of metabolism, could also be rigorously determined in domiciled subjects.

Although domiciled feeding studies can provide important mechanistic insights, their artificial environment may limit generalizability and application to free-living populations. Furthermore, domiciled feeding studies alone are insufficient for determining what constitutes a healthy diet because it is impossible to continuously house for several years the large numbers of subjects that would be required to objectively measure both food intake and clinical endpoints, such as cardiovascular events or diabetes progression. Therefore, long-term nutrition studies in free-living people will always be required.

Nonetheless, domiciled feeding studies can help to improve long-term human nutrition studies. For example, the development and validation of objective diet assessment technologies requires domiciled feeding studies because the only way to objectively know what people eat is to house them continuously in a research facility and directly measure their food intake. Advancement of objective diet assessment technologies has been identified as a top priority for human nutrition science (14) and promising new technologies are emerging, such as sensors and cameras that detect food intake. Biomarkers of diet are also being developed, such as plasma concentrations of vitamin C and carotenoids as indicators of fruit and vegetable intake. Domiciled feeding studies can validate objective diet assessment technologies and biomarkers in diverse subject groups consuming a variety of known diets. These validated technologies and standardized biomarkers can then be deployed in large, long-term nutrition studies to monitor diet adherence and improve understanding of the relationships between diet and disease, and diet and health.

Domiciled feeding studies can also help researchers to design and interpret large, long-term nutrition studies. For example, surrogate biomarkers of disease risk often change rapidly in response to controlled diet interventions. When surrogate markers are causally related to disease risk, then it may be possible to cautiously extrapolate the re-

sults of domiciled feeding studies, especially those that test dose responses, and to estimate the effects of diet changes on long-term disease risk. Such information can be useful for planning long-term randomized diet trials by helping to avoid underpowered studies whose null statistical results might be misinterpreted to conclude that the diet had no real effect when even a small undetected effect might be important, especially on the population scale.

For example, prior to devoting many millions of dollars to a large, long-term randomized trial of a Westernized Mediterranean diet intended to prevent cardiovascular disease, domiciled feeding studies could be used to help develop and validate biomarkers of varying degrees of adherence to the dietary pattern while also evaluating surrogate markers of disease risk in response to known diet changes. For a relatively small fraction of the overall investment, data from such a domiciled feeding study could be used to help plan and interpret the results of the large, long-term randomized trial.

The advancement of human nutrition science has enormous benefits for health and the economy (15). Knowledge of nutrition requires triangulation of evidence from a variety of study designs, including observational studies and randomized trials in free-living people. Facilitating more domiciled feeding studies will lead to fundamental new discoveries about the mechanistic physiological responses to diet and will improve human nutrition research in all its forms. ■

REFERENCES AND NOTES

1. GBD 2017 Diet Collaborators, *Lancet* **393**, 1958 (2019).
2. D. L. Katz, S. Meller, *Annu. Rev. Public Health* **35**, 83 (2014).
3. J. P. Ioannidis, *BMJ* **347**, f6698 (2013).
4. S. E. Nissen, *Ann. Intern. Med.* **164**, 558 (2016).
5. D. Mozaffarian, N. G. Forouhi, *BMJ* **360**, k822 (2018).
6. A. Satija *et al.*, *Adv. Nutr.* **6**, 5 (2015).
7. J. P. A. Ioannidis, *JAMA* **320**, 969 (2018).
8. J. P. Ioannidis, *Am. J. Clin. Nutr.* **103**, 1385 (2016).
9. M. Cainzos-Achirica *et al.*, *Curr. Cardiovasc. Risk Rep.* **12**, 4 (2018).
10. S. K. Das *et al.*, *Am. J. Clin. Nutr.* **85**, 1023 (2007).
11. K. D. Hall *et al.*, *Int. J. Obes.* **43**, 2350 (2019).
12. A. Keys *et al.*, *The Biology of Human Starvation* (Univ. of Minnesota Press, 1950).
13. A. Vella *et al.*, *Diabetes* **65**, 2821 (2016).
14. Interagency Committee on Human Nutrition Research, *National Nutrition Research Roadmap 2016–2021: Advancing Nutrition Research to Improve and Sustain Health* (2016).
15. A. A. Toole, F. Kuchler, *Improving Health Through Nutrition Research: An Overview of the U.S. Nutrition Research System* (Economic Research Service, U.S. Department of Agriculture, 2015).

ACKNOWLEDGMENTS

Thanks to N. K. Fukagawa, M. B. Katan, K. C. Klatt, P. Ohukainen, M. L. Reitman, and E. J. Weiss for insightful comments. Supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases.

10.1126/science.aba3807

NEUROSCIENCE

Is it worth the effort?

Individual variation in dopamine affects the weighting of benefits relative to costs

By Amy C. Janes

Before undertaking any task, humans implicitly determine whether reaching the goal is worth the effort. Weighing costs and benefits is a fundamental brain function that often occurs unconsciously, allowing for the adaptive use of resources to attain goals. The neurotransmitter dopamine is a key player in this process (1). On page 1362 of this issue, Westbrook *et al.* (2) clarify the role of dopamine by showing that increasing an otherwise weak dopamine signal shifts attention toward the rewarding outcome, resulting in greater readiness to perform cognitive effort to reach the goal. As such, increasing dopamine appears to be beneficial specifically for those individuals with relatively lower dopamine function. This finding may explain the efficacy of dopamine-enhancing medications such as Ritalin (methylphenidate), which is prescribed to treat attention deficit hyperactivity disorder (ADHD) and has been used without a prescription by students as a “study enhancing drug.”

Dopamine is found throughout the brain in several neurobiological pathways that mediate processes including movement, reward, and cognitive functions such as learning and working memory (3). Given the range of functions influenced by dopamine, there is a need to better understand how dopamine within distinct brain re-

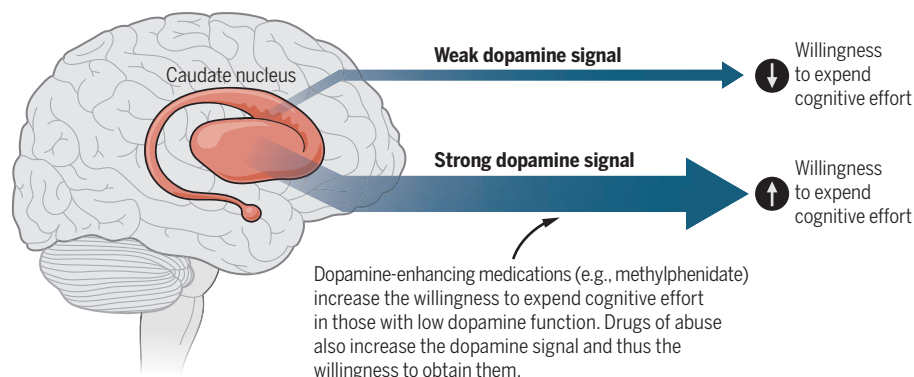
gions affects nuanced elements of cognition and behavior. For example, Westbrook *et al.* expand on the finding that a blunted dopamine signal can result in cognitive dysfunction (4). Specifically, they show that the willingness to expend cognitive effort is diminished in those with lower dopamine function in the caudate nucleus, a portion of the brain involved in goal-motivated behavior (5). This finding blends two known roles of dopamine—motivation and cognition—by indicating that goal-related attention drives the motivation to engage cognitive resources.

Westbrook *et al.* also show that a blunted willingness to expend cognitive effort can be increased by pharmacologically enhancing the dopamine signal using the dopamine agonist methylphenidate. This is consistent with prior findings that dopamine enhancement leads to increased willingness to expend effort in patient populations who have disorders with an underlying dopamine deficit, such as ADHD and Parkinson's disease (6, 7). Thus, dopamine-enhancing medications may not improve cognitive ability per se, but drive the willingness to expend cognitive effort (8).

More precisely, this greater willingness to expend effort occurs because dopamine-enhancing medications raise the salience of, and attention to, goal-related stimuli that would otherwise evoke a response too weak to warrant the expenditure of cognitive effort (8). Methylphenidate and similar drugs

Caudate dopamine affects the weight of benefits

Individual differences in dopamine function in the caudate nucleus relate to one's willingness to expend cognitive effort, which can be influenced both by medications and drugs of abuse that enhance the dopamine signal.





Drugs of abuse increase caudate dopamine and reduce the perception of effort to acquire them.

therefore enhance this aspect of cognition specifically for individuals with relatively lower dopamine function in the caudate nucleus. This indicates that individual variation in dopamine function matters, both for explaining differences in cognitive function across healthy individuals, and also for treatment efficacy. Such reduced dopamine function may render healthy individuals more prone to illicitly use methylphenidate as a study drug. Similarly, those with ADHD who have lower caudate dopamine function are likely to have enhanced efficacy from this treatment (see the figure).

Unfortunately, the effects of dopamine on cost-benefit weighting do not discriminate on the basis of the longer-term repercussions of reaching the goal. For example, methylphenidate also increases the dopamine signal in response to drug-related stimuli in those with a substance use disorder (9), which in the context of the findings of Westbrook *et al.* will enhance the willingness to expend effort to attain the drug-related goal. Even without methylphenidate's facilitatory effects, dopamine plays a key role in substance use disorders, as all abused substances enhance dopamine release (10). After repeated drug use, environmental stimuli associated with drug abuse also evoke dopamine release in regions that include the caudate nucleus (11), the same brain region shown by Westbrook *et al.* to be linked with the mo-

tivation to expend cognitive effort. These drug-associated stimuli can then drive the strong desire to acquire the drug of abuse, contributing to relapse even when individuals are trying to abstain (12). Variation in dopamine function plays a role in this example as well. Specifically, drug-associated stimuli evoke more activation in reward-related brain regions such as the striatum and prefrontal cortex, in individuals with a genetic variation in the gene encoding the dopamine transporter (DAT) that results in slower clearance of dopamine and an amplification of the dopamine signal (13).

Determining whether a goal is "worth the effort" is just one aspect of goal-related decision-making. Such decisions are made through the integration of information processed by the striatum, including the caudate nucleus, and other brain regions like the frontal cortex, which allows for learned experiences and potential future consequences to also shape goal-related behavior (3). The influence of dopamine within these frontal brain regions is not straightforward but depends on the relative balance of D1- and D2-like dopamine receptors (14). These receptors act like a lock into which dopamine fits, and their activation results in the propagation of the dopamine signal. This points out that the "dopamine signal" is not only driven by dopamine synthesis, as was measured by Westbrook *et al.*, but also involves the activation of specific receptor subtypes. Next steps should integrate these findings to explain how dopamine synthesis and receptor function across brain regions work together to affect cognition and behavior. Such findings will create a clearer understanding of cognitive function and psychiatric diseases such as ADHD and substance use disorders. ■

REFERENCES AND NOTES

1. M. T. Treadway *et al.*, *J. Neurosci.* **32**, 6170 (2012).
2. A. Westbrook *et al.*, *Science* **367**, 1362 (2020).
3. S. N. Haber, *Dialogues Clin. Neurosci.* **18**, 7 (2016).
4. A. Nieoullon, *Prog. Neurobiol.* **67**, 53 (2002).
5. E. M. Tricomi, M. R. Delgado, J. A. Fiez, *Neuron* **41**, 281 (2004).
6. M. A. Addicott *et al.*, *Pharmacol. Biochem. Behav.* **183**, 14 (2019).
7. T. T.-J. Chong *et al.*, *Cortex* **69**, 40 (2015).
8. N. D. Volkow *et al.*, *Biol. Psychiatry* **57**, 1410 (2005).
9. N. D. Volkow *et al.*, *Neuroimage* **39**, 1266 (2008).
10. R. C. Pierce, V. Kumaresan, *Neurosci. Biobehav. Rev.* **30**, 215 (2006).
11. N. D. Volkow *et al.*, *J. Neurosci.* **26**, 6583 (2006).
12. A. C. Janes *et al.*, *Biol. Psychiatry* **67**, 722 (2010).
13. T. R. Franklin *et al.*, *Neuropsychopharmacology* **34**, 717 (2009).
14. S. B. Floresco, *Front. Neurosci.* **7**, 62 (2013).

ACKNOWLEDGMENTS

A.C.J. is supported by NIDA grant K02 DA042987. She provides consulting services for Axial Biotherapeutics.

Functional Integration of Addiction Research Lab,
Department of Psychiatry, Harvard Medical School,
McLean Hospital, Belmont, MA 02478, USA.
Email: ajanes@mclean.harvard.edu

APPLIED PHYSICS

Thermal light tunnels its way into electricity

New devices convert low-temperature heat into electricity

By Aaswath P. Raman

In 1824, a 28-year-old French engineer named Sadi Carnot published a now-celebrated treatise that sought to understand how efficiently heat could be converted to work (1). While Carnot laid the foundations for our modern understanding of thermodynamics, he also implicitly identified a pathway for energy recovery and harvesting that may prove to be an essential component of the 21st-century response to climate change. Carnot's eponymous cycle revealed that in any energy conversion process, some heat will always be rejected to a cold reservoir—typically the ambient environment. In the United States alone, 61% of the energy we consume is rejected in this manner as "wasted" heat (2). However, this waste heat need not actually be lost. It can be recovered, in principle, by driving another energy conversion device to generate electricity. On page 1341 of this issue, Davids *et al.* propose and implement a way to directly convert lower-temperature heat into electricity (3).

Recovering heat rejected in transportation, power generation, and industrial processes motivates a broad range of contemporary research in solid-state materials and devices. A large fraction of waste heat is lost radiatively, emitted as incoherent, broadband electromagnetic waves at long-wave and mid-infrared (IR) wavelengths. Converting this thermal radiation to electricity has been an exciting, although challenging, subset of this line of inquiry. In particular, thermophotovoltaic devices have shown considerable promise at being able to convert emitted thermal radiation to electricity through a photovoltaic mechanism. Substantial progress has been made in thermophotovoltaics by using photonic approaches (4) as well as near-

Department of Materials Science and Engineering,
University of California, Los Angeles, CA 90049, USA.
Email: aaswath@ucla.edu

10.1126/science.abb0265

field effects (5); however, these systems have typically focused on high-temperature sources ($>1000^{\circ}\text{C}$), whereas more than 95% of wasted heat in the United States—and 85% of the associated work potential—is below 400°C (2). The direct conversion of longer-wavelength thermal radiation poses numerous challenges, such as lower incident photon flux and the limited availability of efficient low-bandgap semiconductors. An alternative approach is to use a rectifying antenna, which finds wide application at lower-energy microwave frequencies. In these devices, incident oscillating electromagnetic waves are funneled by an antenna-like structure and drive a direct electrical current through a fast diode (6). At higher frequencies, ultrafast direct tunnel diodes in a metal-insulator-metal con-

all process shares superficial similarities to a photovoltaic system because it uses a pn junction, the current instead is generated from photon-assisted tunneling between two metal-oxide semiconductor diodes rather than a depletion region.

The authors demonstrate a peak power density of $61\text{ }\mu\text{W}/\text{cm}^2$ for a 350°C radiative source. This appears relatively modest at first glance, as the achieved power corresponds to a considerably lower efficiency than the Carnot limit. However, these results are orders of magnitude better than previous unipolar metal-oxide semiconductor tunnel junction diodes (8, 9). The use of multiple interdigitated bipolar junctions—which yields a periodic well structure where charge is stored that can be additionally pumped from p+ to metal and from metal

authors leverage a phonon resonance of silicon dioxide in the $8\text{-}\mu\text{m}$ range, which overlaps well with the blackbody spectrum of thermal emitters in the 200° to 400°C range. This specific resonance yields wavelength ranges where silicon dioxide has a near-zero permittivity (10) and enables funneling of incident thermal radiation to improve tunneling efficiency. Although this initial performance is compelling, it leaves a large fraction of incident thermal radiation uncoupled at wavelengths away from silicon dioxide's phonon resonance. Intriguing opportunities and challenges thus lie in combining photonic microstructures and exploiting the intrinsic dispersion of different materials to make this photonic concentration more broadband. In doing so, a larger fraction of incident thermal radiation could be harnessed.

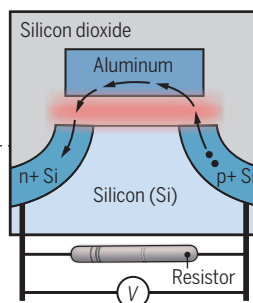
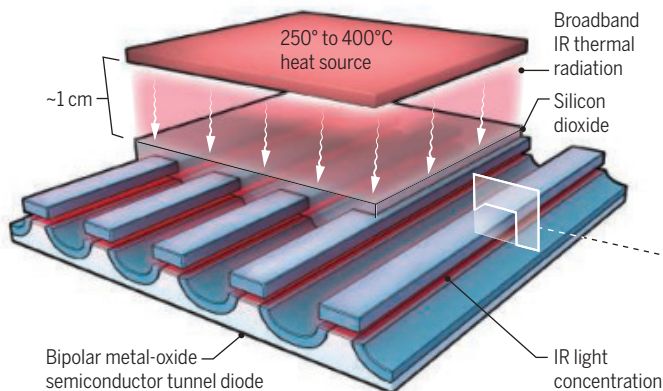
The authors' use of the globally standard complementary metal-oxide semiconductor (CMOS) platform allows for the scalability that will eventually be needed for this technology. However, a substantial amount of wasted heat still lies at temperatures below 250°C (2). Conversion of low-temperature radiated heat to electricity, all the way down to ambient (0° to 50°C) heat rejected as thermal radiation to the sky (11, 12), represents an intriguing frontier for energy efficiency but must ultimately be able to deliver meaningful performance at an attractive cost point. Indeed, the laws of thermodynamics that Carnot glimpsed in his prescient experiments two centuries ago place fundamental constraints on our ability to convert energy. Nonetheless, considerable headroom still remains to better control and harness our collective thermal energy footprint at the very largest scales. ■

Converting heat into power

Infrared (IR) radiation from lower-temperature heat sources is usually wasted, but certain specialized devices can convert this radiation into electricity. A bipolar semiconductor tunnel diode accomplishes this through a combination of light funneling and light-facilitated electron tunneling.

A device with many junctions

By using many tunnel diodes (see inset), a larger voltage (V) is generated from the IR radiation of the heat source. The light is funneled into a very thin silica barrier between doped silicon and an aluminum grating.



Bipolar semiconducting tunnel diode

The concentrated light (red) drives charge carriers to tunnel from the p-type to the n-type silicon and thus generates a rectifying current.

figuration have shown promise (7), but the large asymmetry needed in tunnel diodes at the small voltages associated with IR thermal radiation has remained a roadblock.

Dauids *et al.* tackle this challenge and demonstrate a bipolar metal-oxide semiconductor tunnel junction diode that converts incident photons in the long-wave IR part of the electromagnetic spectrum (7 to $14\text{ }\mu\text{m}$) to electricity. Incident electromagnetic waves are coupled by a grating to an electromagnetic mode that confines light strongly in a 3- to 4-nm silicon dioxide barrier between the metal grating and a base layer of doped silicon (see the figure). The strong electromagnetic field concentration drives photon-assisted tunneling of electrons from the doped p-type silicon into the metal and to the n-type silicon part. Although the over-

all process shares superficial similarities to a photovoltaic system because it uses a pn junction, the current instead is generated from photon-assisted tunneling between two metal-oxide semiconductor diodes rather than a depletion region.

Converting thermal radiation to electricity from low-temperature sources (or higher-temperature ones such as the Sun) can be more effectively harnessed through photonic approaches that concentrate light or enhance light-matter interaction. These approaches are essential to enable improved performance for thermophotovoltaic devices (4, 5). In a similar vein, the authors cleverly exploit both photonic design and materials properties to enhance the photon-assisted tunneling effect. The

REFERENCES AND NOTES

1. S. Carnot, *Réflexions sur la puissance motrice du feu* (1824).
2. U.S. Department of Energy, Advanced Research Projects Agency–Energy, *Request for Information (RFI) DE-FOA-0001607 on Lower Grade Waste Heat Recovery* (2016); <https://arpa-e-foa.energy.gov/FileContent.aspx?FileID=80197dc7-ad62-4d58-a8c4-4bdbc-c0a52c6>.
3. P. S. Dauids *et al.*, *Science* **367**, 1341 (2020).
4. A. Lenert *et al.*, *Nat. Nanotechnol.* **9**, 126 (2014).
5. A. Fiorino *et al.*, *Nat. Nanotechnol.* **13**, 806 (2018).
6. J. O. McSpadden, T. Yoo, K. Chang, *IEEE Trans. Microw. Theory Tech.* **40**, 2359 (1992).
7. S. Grover, O. Dmitriyeva, M. J. Estes, G. Model, *IEEE Trans. Nanotechnol.* **9**, 716 (2010).
8. P. S. Dauids *et al.*, *Nat. Nanotechnol.* **10**, 1033 (2015).
9. J. Shank *et al.*, *Phys. Rev. Appl.* **9**, 054040 (2018).
10. A. Shahsafi *et al.*, *Phys. Rev. Appl.* **10**, 034019 (2018).
11. P. Santhanam, S. Fan, *Phys. Rev. B* **93**, 161410 (2016).
12. A. P. Raman, W. Li, S. Fan, *Joule* **3**, 2679 (2019).

ACKNOWLEDGMENTS

Supported by a Sloan Research Fellowship in Physics (Alfred P. Sloan Foundation).

Neuron-targeted electrical modulation

Engineering neurons to make conductive polymers enables cell type–specific behaviors

By **Kevin J. Otto** and **Christine E. Schmidt**

Conductive polymers have been widely studied and used for biomedical applications—including as biosensors, neural prostheses, and bioactuators—and for drug delivery and tissue engineering (1). Conductive polymers are organic chains of alternating single and double bonds, which endow the polymers with metal-like semiconductive properties. Exogenous application of electrical stimulation to these polymers can promote cellular activities such as proliferation, adhesion, migration, differentiation, and protein secretion. Because many cells and tissues, particularly neurons, are responsive to electrical fields, conductive polymers are attractive for biological and medical applications. On page 1372 of this issue, Liu *et al.* (2) report a genetically targeted approach to assemble conductive polymers in neurons. This in turn remodels membrane electrical properties and enables cell type–specific cellular and behavioral modulation, such as control of neuronal firing, as demonstrated in cultures of rat hippocampal neurons, mouse brain slices, human cortical spheroids, and in living *Caenorhabditis elegans* worms.

Commonly studied conductive polymers include poly(3,4-ethylenedioxythiophene (PEDOT), polyaniline (PANI), and polypyrrole. Conductive polymers have traditionally been synthesized as standalone biomaterials that are used in cultured cells or for implantation in vivo (1). Integrating conductive polymers into tissues is critical for localized delivery of electric fields. There have been attempts to polymerize electroactive materials directly into tissues to provide a more seamless interface between the conductive substrate and cells. The first reports of successful in vivo polymerization of PEDOT in the brains of rats demonstrated that it did not negatively affect behavior (3, 4). Although these studies show some local specificity, the polymers are ubiquitous throughout the neural space and thus do not provide cell-type specificity. Alternative injectable neural interfaces are being developed (5); however, these are also not targeted to specific cells.

Liu *et al.* demonstrate in vivo polymerization of PANI conductive polymers that are manufactured according to specific cell types and modify the electrical properties of the cell membrane. They genetically engineered neurons, using adeno-associated virus (AAV) vectors, to express peroxidase enzyme on the outside of cell membranes. Peroxidases catalyze polymerization of aniline monomer and dimer precursors when infused into cells or tissues. This approach could provide more cell-specific targeting of electric fields (see the figure).

Bioelectronic medicine, which aims to electrically modulate neural elements for

target- and organ-specific effects, promises improved specificity and efficacy over traditional pharmaceutical medicine (6). The approach has been heralded as the frontier of medicine (7), and a roadmap has guided numerous funding opportunities (8). Intended therapeutic benefits of bioelectronic medicines are often contingent on activating a predominant effect (such as excitatory or inhibitory) and/or conveying unidirectional information (efferent or afferent) in nerves. Side effects arise from the activation of off-target cells or tissues and can result in undesired outcomes, such as stimulation of the cough reflex in vagal nerve stimulation and seizure episodes in deep brain stimulation. Thus, targeting a subpopulation of neurons is attractive. This is challenging in current systems because of spatial proximity of different neural elements, such as excitatory neurons located near inhibitory elements or efferent axons close to afferent axons. Genetically targeting cell-specific expression of conductive polymers could overcome this challenge.

Autonomic neural modulation (such as vagus nerve stimulation) is a subset of bioelectronic medicine in which cell-type specificity is particularly desirable. Autonomic neural modulation involves electrical stimulation of the autonomic nervous system—for example, to decrease sympathetic activity (“fight or flight”) and increase parasympathetic activity (“calm and composed”) as a therapeutic strategy for the treatment of diseases such as epilepsy and depression. The first in-human evidence of autonomic neural modulation—vagal stimulation for epilepsy (9)—paved the path for other clinical uses. Yet, there is room for clinical improvement. Autonomic nerves often contain motor and sensory axons as well as sympathetic and parasympathetic information. Parasympathetic motor axons can drive hormonal release in off-target organs; sensory axons could miscommunicate organ state information to the brain. As greater understanding of the innervation of end organs is discovered, it is likely that cell type–specific activation will result in desirable organ effects without the undesirable side effects.

Another approach of bioelectronic medicine uses neuromodulation for prostheses. Stimulation of peripheral nerves has been enabled by advances in materials,

Modes of neural modulation

Altering nerve activity is an important therapeutic strategy. Pharmacological modulation results in systemic and unwanted side effects. Bioelectronic approaches increase targeting of defective neural pathways. However, ensuring that specific cells are modulated could allow behavioral modification with minimal side effects.

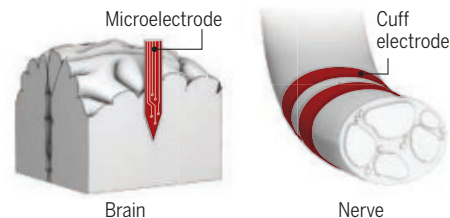
Pharmacology

Undesired side effects
Therapeutic effects



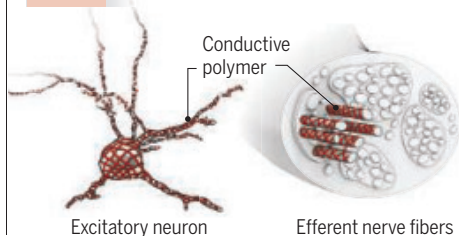
Bioelectronic medicine

Undesired side effects
Therapeutic effects



Cell-specific stimulation

Undesired side effects
Therapeutic effects



J. Crayton Pruitt Family Department of Biomedical Engineering, Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL, USA. Email: kevin.otto@bme.ufl.edu; schmidt@bme.ufl.edu

electronics, and devices (10, 11) and has emerged as a successful approach for amputees by conveying sensory information from limb prostheses directly to the users' nervous system (12–14). Cell-type specific modulation could improve these applications; for example, sensory fibers related to a prosthesis's physical interaction with the environment (somatosensation and proprioception) need to be excited while avoiding undesirable side effects if motor axons were simultaneously excited.

Integrating cell type-specific conducting polymers into tissues could also enhance the regeneration of tissues after injury or disease. The body contains electric fields that play roles during embryogenesis and wound healing by guiding cell migration and eliciting protein secretion. Exogenous application of electrical fields affects similar processes. Conductive polymers have been explored as electroactive substrates for regenerative medicine to stimulate tissue repair (15).

There remain considerable hurdles to the translation of cell type-specific neural modulation approaches. Liu *et al.* use AAV transfection, which is still being developed for applications in humans because of the regulatory challenges of gene therapy. It also remains to be elucidated whether this approach will be viable in higher-order species, especially vertebrates. Last, it is unclear whether the biologically assembled conductive polymers are transmembrane or extracellular, which could affect cellular modulation. ■

REFERENCES AND NOTES

1. N. K. Guimard, N. Gomez, C. E. Schmidt, *Prog. Polym. Sci.* **32**, 876 (2007).
2. J. Liu *et al.*, *Science* **367**, 1372 (2020).
3. S. J. Wilks, A. J. Woolley, L. Ouyang, D. C. Martin, K. J. Otto, *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2011**, 5412 (2011).
4. L. Ouyang, C. L. Shaw, C. C. Kuo, A. L. Griffin, D. C. Martin, *J. Neural Eng.* **11**, 026005 (2014).
5. J. K. Trevathan *et al.*, *Adv. Healthc. Mater.* **8**, 1970090 (2019).
6. K. J. Tracey, *Bioelectron. Med.* **1**, 1 (2014).
7. K. Famm, B. Litt, K. J. Tracey, E. S. Boyden, M. Slaoui, *Nature* **496**, 159 (2013).
8. K. Birmingham *et al.*, *Nat. Rev. Drug Discov.* **13**, 399 (2014).
9. J. K. Penny, J. C. Dean, *Epilepsia* **31** (Suppl. 2), S40 (1990).
10. C. J. Bettinger, *Bioelectron. Med.* **4**, 6 (2018).
11. R. A. Green, N. H. Lovell, G. G. Wallace, L. A. Poole-Warren, *Biomaterials* **29**, 3393 (2008).
12. E. L. Graczyk, B. P. Delhay, M. A. Schiefer, S. J. Bensmaia, D. J. Tyler, *J. Neural Eng.* **15**, 046002 (2018).
13. E. L. Graczyk *et al.*, *Sci. Transl. Med.* **8**, 362ra142 (2016).
14. S. Wendelken *et al.*, *J. Neuroeng. Rehabil.* **14**, 121 (2017).
15. B. Guo, P. X. Ma, *Biomacromolecules* **19**, 1764 (2018).

ACKNOWLEDGMENTS

The authors are supported by National Institutes of Health, National Science Foundation, and Defense Advanced Research Projects Agency. We thank M. Urdaneta, B. Ibarra, E. W. Dirr, and J. Murbach for helpful discussions.

10.1126/science.abb0216



EVOLUTIONARY BIOLOGY

Shallow ponds prompt fitness-favorable species interbreeding

Dry conditions liberate female toads to drive sexual selection of male mates

By Marlene Zuk

If there is one thing that biologists used to agree on, it was that mating between two animals of different species yields unfavorable results. (Plants have a dodgy reputation for insouciant polyploidy and hence have often been ignored.) Basic biology classes teach that different species usually cannot interbreed successfully and that rarely produced cross-bred offspring (hybrids) are often infertile or of lower fitness. Now, on page 1377 of this issue, Chen and Pfennig (7) contradict conventional wisdom about the disadvantages of hybridization and provide a connection between species diversification, sexual selection, and, ultimately, the context dependence of behavioral evolution.

In general, scientists expect mate selection to favor precise species-recognition mechanisms, and indeed, many organisms display such finely tuned abilities (2, 3). Architects of the Modern Evolutionary Synthesis (which merges Mendelian genetics with Darwinian evolution) were so convinced of the importance of species recognition that they virtually dismissed

sexual selection—that is, selection due to reproductive competition—as a force in evolution. Indeed, in 1942, evolutionary biologist Ernst Mayr proclaimed (4), “it is now recognized that many phenomena previously thought to promote intraspecific sexual selection are actually specific recognition marks.”

Hybrids now have a more respectable reputation. Spurred partly by advancements made in genomics over the past few decades, biologists have come to acknowledge more and more the importance of hybrids in generating biodiversity at a macroevolutionary scale (5). For example, whole-genome analysis facilitated the assessment of introgression (gene flow from one species into the gene pool of another) and revealed a history of hybridization in numerous taxa (including humans) without having to rely on observable morphological differences between groups (5). Scientists also now understand that sexual selection in the form of mate choice and mate competition account for differences in male and female behavior, appearance, and life history that go far beyond recognition of conspecifics (6).

Chen and Pfennig studied hybridization between two species of spadefoot toads that live in ephemeral ponds of varying

Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, USA. Email: mzung@umn.edu

A. S. bombifrons female (left) will mate with an *S. multiplicata* male (right) to increase their offspring's fitness under harsh environmental conditions.

depth in the United States and Mexico (see the photo). In shallow ponds, but not deep ones, female Plains spadefoot toads, *Spea bombifrons*, prefer to mate with a male of another species, the Mexican spadefoot toad, *S. multiplicata*. The resulting hybrid tadpoles develop faster than nonhybrids, which means they stand a greater chance of becoming adults before harsh conditions cause the ponds to dry up. But the *S. bombifrons* females do not mate with just any heterospecific; they prefer males with particularly attractive mating calls.

Female *S. bombifrons* that follow these rules yield offspring with greater fitness than if they mated with male *S. bombifrons*. This means that females exert sexual selection pressure on males of a different species, a phenomenon that has not previously been reported.

Two of the many intriguing aspects of this finding are the demonstration that hybridization can be adaptive and that females drive the process. It has long been known that animals do not always mate with the "correct" species and sometimes even mate with inanimate objects. Australian jewel beetles, for example, famously attempt to copulate with beer bottles, but the actors in such cases are virtually always males (7). The sex difference in such so-called mistakes is usually attributed to the higher cost of misidentification for females, who generally make the larger investment in offspring (8). Mating with the "wrong" species could, therefore, mean loss of reproductive success for an entire breeding season. Conversely, a male that mates with a heterospecific species or an object risks only the time and effort expended in the mating. It might even be beneficial to have a broad filter for acceptable mates, because missing a mating opportunity is likely more costly than wasting sperm. Although the bottle-infatuated beetles do experience reduced survival, perhaps scientists have been too hasty in dismissing all male-initiated hybridization attempts as mistakes.

The study by Chen and Pfennig suggests many avenues for future research, starting with the question of just how widespread sexual selection across species might be. Opportunities could be afforded in the many groups of animals that occur in species "flocks" or "swarms," including cichlid fishes, Hawaiian crickets, and drosophilid flies. Plants are another unexplored avenue for deciphering the cross-species mating phenomenon. Could incompatibility

mechanisms be similarly flexible depending on environmental circumstances?

At a more proximate level, it would be worth investigating the cue used by the female toads to determine what constitutes a shallow enough pond to trigger the attraction to heterospecific males. The females spend up to several hours swimming and submerged in the ponds before they select a mate, but how the water level is assessed and how that evaluation translates into a preference for a particular mating-call frequency remain unknown. Females of many species, from butterflies to birds, choose egg-laying or nesting sites (9), but such habitat selection is not generally linked to mate choice, as it is in the toads.

The toads might also illustrate another phenomenon gaining attention: evolutionary rescue. In this case, an adaptive change allows persistence of a population that would otherwise face extinction (10). The classic case of evolutionary rescue entails a large population that, when subjected to a rapid ecological change, evolves its way out of the new pressure because genetic variation enables the population to adapt. Here, the behavioral plasticity of female *S. bombifrons*, rather than their genetic variability, is key to their survival. The females exert a preference for particular males, but only when it serves as an adaptive response to impending death of the tadpoles in shallow ponds. Whether such behavioral responses account for other instances of evolutionary rescue remains to be seen.

The new work underscores the complex interaction of genes, environment, and behavior in evolution. Although the importance of the first two are widely acknowledged, behavior is sometimes dismissed as ephemeral and less likely to contribute to evolutionary changes. The interspecific hybridization that simultaneously rescues *S. bombifrons* in dry environments and exerts selective pressure on the mating calls of *S. multiplicata* occurs only because of a finely tuned sensory response to both ecological and behavioral signals. ■

REFERENCES AND NOTES

1. C. Chen, K. S. Pfennig, *Science* **367**, 1377 (2020).
2. M. J. Ryan, A. S. Rand, *Evolution* **47**, 647 (1993).
3. T. C. Mendelson, K. L. Shaw, *Trends Ecol. Evol.* **27**, 421 (2012).
4. E. Mayr, *Systematics and the Origin of Species from the Viewpoint of a Zoologist* (Columbia Univ. Press, 1942).
5. S. A. Taylor, E. L. Larson, *Nat. Ecol. Evol.* **3**, 170 (2019).
6. M. Zuk, L. W. Simmons, *Sexual Selection: A Very Short Introduction* (Oxford Univ. Press, 2018).
7. D. T. Gwynne, D. C. F. Rentz, *Aust. J. Entomol.* **22**, 79 (1983).
8. M. A. Peterson *et al.*, *Evolution* **59**, 2639 (2005).
9. J. N. Thompson, *Entomol. Exp. Appl.* **47**, 3 (1988).
10. A. McDermott, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12116 (2019).

10.1126/science.abb2398

CHEMICAL ENGINEERING

Taking on all of the biomass for conversion

A catalytic process converts hardwood into phenol, propylene, and other valuable products

By Tao Zhang

The old saying that "you can make everything from biomass except money" sums up the huge challenges of converting lignocellulose into fuels and chemicals in a cost-competitive manner. Most efforts separate its more readily processed cellulose and hemicellulose components from lignin. However, upgrading the methoxylated phenylpropanoid lignin biopolymer (a process called valorization) has become the bottleneck (1). On page 1385 of this issue, Liao *et al.* (2) address this key issue in a more holistic way. Rather than separate the components, they converted woody biomass into phenol and propylene, two bulk chemicals widely used in the polymer industry, and coproduced valuable phenolic oligomers and a carbohydrate pulp amenable to bioethanol production. The total conversion efficiency is impressive, up to 78% by weight (wt %) based on the initial mass input. Moreover, both technoeconomic analysis and life-cycle assessment reveal that such an integrated process can be profitable and sustainable.

Intensive research efforts have been invested toward valorization of lignin (3), but most studies have focused on the conversion of lignin-related model compounds; few succeeded with real lignocellulose. In contrast to those chemistry-centered studies, Liao *et al.* is more engineering-guided. In particular, the efficient valorization of lignin involves reductive catalytic fractionation (RCF), extraction of phenolic monomers, and catalytic funneling (see the figure). On a lignin basis, this process provided phenol in 20 wt % yield with coproduction of 9 wt % propylene. Most state-

State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China. Email: taozhang@dicp.ac.cn

of-the-art lignin conversion strategies have produced a mixture of phenols, arenes, and their oligomers (4, 5) that cannot be directly used as chemicals without further expensive and energy-intensive separation operations in downstream utilization.

In their first RCF step, hardwood reacted with added hydrogen (H_2) over a metal catalyst (ruthenium on a carbon support, Ru/C) to yield lignin oil containing phenolic monomers in near-theoretical yields (~50 wt%) and a solid carbohydrate pulp. This “lignin-first” strategy overcame the condensation of the reactive intermediates and achieves near-complete delignification with little carbohydrate degradation. The monomers in the lignin oil were easily extracted with a less than sixfold added mass of *n*-hexane in a cost-effective way.

Given the excellent yields of phenolic monomers and the high extraction efficiency, Liao *et al.* sought an effective catalytic process to convert the lignin monomers into phenol, which is the key step in demonstrating the concept of lignin-to-phenol biorefinery because the phenolic monomers contain both methoxy groups and other versatile substitute groups on different positions of the benzene ring (1). Removal of these functional groups without destroying the benzene ring and the phenolic hydroxy represents a huge challenge in catalysis, given their similar bond dissociation energy.

To reach this target, a stepwise catalytic process was needed. First, removal of methoxy groups was explored. To make the process economically viable, a solvent- and sulfur-free continuous catalytic gas-phase hydroprocessing was performed over a non-noble metal catalyst, nickel on a silica support (Ni/SiO₂). Model compound study indicates that the well-dispersed Ni/SiO₂ catalyst showed remarkable versatility on different methoxylated alkylphenols through either direct demethoxylation or tandem demethylation-dehydroxylation. Hydroprocessing of the lignin monomers mixture provides high selectivity (75 to 85%) toward *n*-propylphenols and ethylphenols at nearly complete conversion.

The major byproducts of methoxy cleavage were water and methane, and the reaction proceeded without formation of carbon monoxide or carbon dioxide. Water was beneficial to maintain robust catalytic activity in the next dealkylation step. Methane, together with small molecules (such as excess H_2 , ethylene, methanol, and methyl acetate) generated in other steps,

could be incinerated to provide heating, cooling, and electricity.

In the subsequent dealkylation step, the researchers developed a hierarchical catalyst based on ZSM-5, which they named Z140-H, that had a balanced network of micro- and mesopores for the dealkylation of the crude alkylphenol condensates. Near-quantitative transformation through a carbenium mechanism led to a combined yield for phenol and olefins of 82%. Zeolites with smaller or larger pore sizes did not work, because of either pore restriction or a lack of the pore confinement needed for shape-selective conversion. These findings underline the importance of zeolite hierar-

A technoeconomic analysis indicates that self-energy supply is enough to operate the integrated biorefinery. For a factory with capacity production of 100 kilotons of biophenol per year, an internal rate of return of 23.33% and a payout time of ~4 years for a plant with a lifetime of 20 years are anticipated. Moreover, life-cycle assessment suggested a much lower carbon footprint for phenol and propylene compared to their fossil-based counterparts. Taking into account the manufacturing investment, production efficiency, and product price, as well as the lower carbon footprint, this process shows great potential for practical application.

The study by Liao *et al.* not only offers an attractive approach toward the production of phenol and propylene from lignin, it also realizes full utilization of all components in lignocellulose. However, some key issues must be addressed for large-scale commercialization. For instance, nonhardwood feedstocks such as pinewood, cornstalk, and bagasse should be studied to improve versatility. In the RCF step, a large amount of Ru/C catalyst (10% based on wood weight) was used, and a non-noble metal catalyst would be less costly and more sustainable (6). Further, unlike fuels, the value of chemicals depends on purity, especially for phenol and propylene as monomers for the polymer industry. At present, how to obtain these two products in sufficiently high purity is not clear.

For biorefineries to play a greater role in the production of liquid fuels and chemicals, feedstock conversion must be maximized. Robust catalysts combined with high-efficiency process engineering are still desirable. In fundamental studies, some elegant progress has been achieved recently, including a solar energy-driven lignin-first approach (7), a formaldehyde stabilization strategy (8), and mild redox-neutral depolymerization (9). ■

REFERENCES AND NOTES

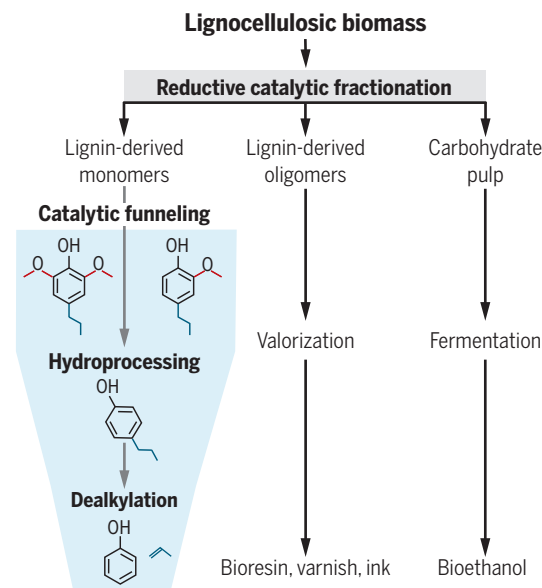
1. A. J. Ragauskas *et al.*, *Science* **344**, 1246843 (2014).
2. Y. Liao *et al.*, *Science* **367**, 1385 (2020).
3. C. Li, X. Zhao, A. Wang, G. W. Huber, T. Zhang, *Chem. Rev.* **115**, 11559 (2015).
4. R. Rinaldi *et al.*, *Angew. Chem. Int. Ed.* **55**, 8164 (2016).
5. W. Schutyser *et al.*, *Chem. Soc. Rev.* **47**, 852 (2018).
6. H. Guo *et al.*, *ChemSusChem* **9**, 3220 (2016).
7. X. J. Wu *et al.*, *Nat. Catal.* **1**, 772 (2018).
8. L. Shuai *et al.*, *Science* **354**, 329 (2016).
9. Y. X. Liu *et al.*, *ACS Catal.* **9**, 4441 (2019).

ACKNOWLEDGMENTS

Support from the National Natural Science Foundation of China (21690080, 21721004) is gratefully acknowledged.

Conquer and divide

Liao *et al.* processed hardwood pulp through reductive catalytic fractionation to produce three feedstocks for further upgrading. The diverse mix of lignin-derived monomers undergoes catalytic funneling to produce mainly phenol and propylene.



chization in the dealkylation reaction and the need for customized catalysts.

To make the technology sustainable and economically feasible, full valorization and utilization of all of the components in the woody biomass are essential. After extraction of monophenols from lignin crude oil, Liao *et al.* disclosed a new application of the residue oligomers to replace fossil-based *p*-nonylphenol by making high-quality printing ink, which provides an unexplored market potential for phenolic oligomers, a major fraction in almost all the lignin depolymerization processes. As for the carbohydrate pulp generated in the RCF step, it can be readily converted to ethanol after a near-simultaneous saccharification-fermentation process.

RETROSPECTIVE

Stanley Cohen (1922–2020)

Award-winning biochemist who shed light on cell growth

By **Graham Carpenter**¹ and **Robert Coffey**²

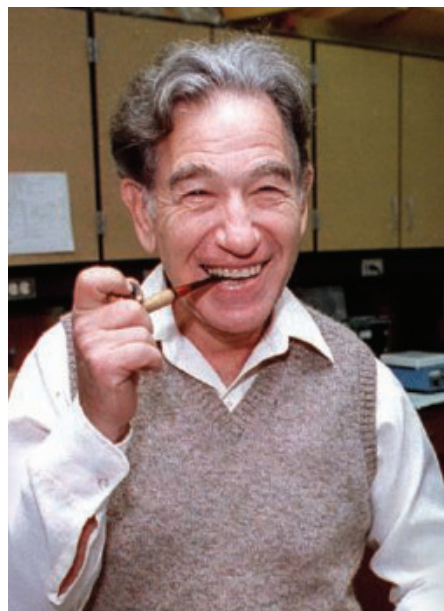
Stanley Cohen, biochemist and Nobel laureate, died on 5 February at age 97 in Nashville, Tennessee, where he had served on the faculty of Vanderbilt University since 1959. With neurobiologist Rita Levi-Montalcini, Cohen discovered the first growth factor, a hormone-like protein that regulates cell responses such as proliferation and differentiation.

Born to Russian immigrants in Brooklyn, New York, Cohen graduated from Brooklyn College in 1945 with majors in chemistry and zoology. He then received a master's degree in zoology from Oberlin College in Ohio and a Ph.D. in biochemistry from the University of Michigan in 1948. After a postdoctoral fellowship, he became a faculty member at Washington University in St. Louis in 1953. There, his collaboration with Levi-Montalcini resulted in the discovery of nerve growth factor (NGF). Levi-Montalcini had made the crucial observation that NGF's biological activity was due to a diffusible substance. Cohen then proved that a single protein was the responsible agent—a critical advance that made NGF a scientifically recognized agent. Both Levi-Montalcini and Cohen were creative thinkers and determined experimentalists. As Cohen once remarked, "On our own we were good and competent. Together we were marvelous."

In 1959, Cohen moved to Vanderbilt University as an assistant professor of biochemistry, where he discovered and purified epidermal growth factor (EGF) from mice. His bioassay for EGF was based on its capacity to induce precocious eyelid opening in newborn mice. It is difficult to imagine the time, coordination, patience, skill, and luck required for this to succeed. Years later, another group reported that in this assay, newborn female mice also exhibited precocious opening of the vagina. Stan's shy response was that he had also noticed this but was too embarrassed to report it.

In the following decades, Cohen relentlessly and productively pursued the mechanism of action underlying EGF. The

identification of the EGF cell surface receptor (EGFR) provided a starting point to elucidate the mechanism by which EGF provoked biological responses, such as cell proliferation, in target cells. By showing that EGF binding to the EGFR in membrane preparations activated a biochemical response, Cohen paved the way for the first analysis of a hormone's mechanism at the biochemical level. When investigators in the lab of biochemist Edward Krebs in Seattle called Cohen to say that they had success-



fully repeated his experiments, Cohen was enormously proud. He always took great care to delay publication of data in his lab until repeatability was assured.

Finally, Cohen's group demonstrated that the EGF:EGFR complex was trafficked from the cell surface to intracellular lysosomes. These studies upended the accepted view that bioactive molecules, bound to receptors, simply associated and dissociated at the cell surface. These seminal investigations have proven to be true for other bioactive proteins, including hormones, growth factors, viruses, and toxins.

Cohen's contributions served as a model that has led to the identification of nearly 100 other growth factors and at least 50 receptors, all with important biological activities in humans. His trailblazing leadership led to numerous scientific awards, culmi-

nating in the 1986 Nobel Prize in Medicine or Physiology. Cohen's EGFR studies later provided the basis for the development of clinically effective drugs used to combat cancer. This application was unanticipated when the Nobel Committee made its prescient decision, a decade before the advent of precision medicine.

Stan's productivity belied that fact that his lab group seldom included more than three other people: two dedicated research assistants and one postdoctoral fellow (students were very rare). With the exception of one 10-year period, the Cohen lab was continuously funded by one rather modest National Institutes of Health grant. Stan's relentless focus and thinking was the key to the group's success. Puffing on a corn cob pipe, he spent hours walking the hallways at Vanderbilt either devising the ideal experiment or stopping to test his ideas with colleagues (including the two of us). Inevitably, the pipe was left on a nearby refrigerator or stuffed in his pocket where it would burn a hole. This routine was well known by everyone and the mislaid pipes would be returned by professors, technicians, or janitorial staff. This intense experimental focus also made him an easy target for April Fools' jokes perpetrated by lab personnel.

Stan approached life with optimism and a smile. Always humble, he attributed his professional success to paying careful attention to the data and to luck. Colleagues were always impressed at how Stan logically identified the next question and then went to his lab to do the critical experiment himself. Stan had a driving need to be as close to the experiment as possible, which included the validity of reagents—he never accepted experiments conducted with "kits."

In addition to working intensely, Stan enjoyed listening to music, playing the clarinet (with local amateur musicians), whitewater canoeing, and hacking away at a tennis ball. Shortly after retirement, Stan moved for a few years to Arizona, where he discovered a passion for off-road driving. He also devoted much of his time to volunteering in a public junior high school as a science adviser.

Stan is survived by his devoted wife Jan, three sons, and two granddaughters. His stellar career serves as a reminder that basic observations in the lab, pursued doggedly, can have long-lasting, beneficial consequences for humankind. Today, Stan's discoveries help countless patients with advanced lung, colorectal, and head and neck cancer, for which pharmacological blockade of the EGFR is a mainstay of treatment. ■

¹Department of Biochemistry, Vanderbilt University, Nashville, TN 37212, USA. ²Vanderbilt University Medical Center, Nashville, TN 37232, USA.
Email: graham.carpenter@vanderbilt.edu

POLICY FORUM

DATA

Time for NIH to lead on data sharing

A draft policy is generally supportive but should start mandating data sharing

By **Ida Sim**^{1,2}, **Michael Stebbins**³, **Barbara E. Brier**^{4,5}, **Atul J. Butte**¹, **Jeffrey Drazen**⁶, **Victor Dzau**⁷, **Adrian F. Hernandez**⁸, **Harlan M. Krumholz**⁹, **Bernard Lo**¹⁰, **Bernard Munos**¹¹, **Eric Perakslis**⁸, **Frank Rockhold**⁸, **Joseph S. Ross**⁹, **Sharon F. Terry**¹², **Keith R. Yamamoto**¹, **Deborah A. Zarin**⁴, **Rebecca Li**^{2,13}

The U.S. National Institutes of Health (NIH), the largest global funder of biomedical research, is in the midst of digesting public comments toward finalizing a data sharing policy. Although the draft policy is generally supportive of data sharing (1), it needs strengthening if we are to collectively achieve a long-standing vision of open science built on the principles of findable, accessible, interoperable, and reusable (FAIR) (2) data sharing. Relying on investigators to voluntarily share data has not, thus far, led to widespread open science practices (3); thus, we suggest steps that NIH could take to lead on scientific data sharing, with an initial focus on clinical trial data sharing.

In 2013, the White House directed all U.S. federal research funding agencies with more than \$100 million in annual research and development expenditures to develop programs to ensure access to the results of publicly funded research, including peer-reviewed publications and digital data (4). The directive was explicit: “[D]igitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.” In 2015, the NIH issued its plans for responding to this directive, asserting its explicit intent “to make public access to digital scientific data the standard for all NIH-funded research” and to “[e]nsure that

data management plans include clear plans for sharing research data” (5). In November 2019, the NIH assured the U.S. Government Accountability Office that “it is in the process of developing an agency-wide data management and sharing policy, including compliance mechanisms, to fully implement its public access plan” (6).

Under the draft policy, NIH would require researchers to submit a plan describing the “rationale for decisions about which scientific data will be preserved and shared.” However, the draft policy does not specify a minimum standard or time frame for data sharing and, most importantly, stops short of a definitive mandate for sharing. In the absence of an explicitly stated requirement,

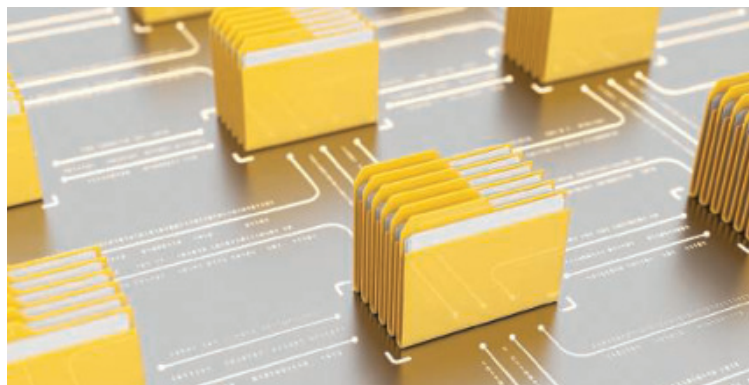
To be sure, there are challenges to implementing FAIR data sharing. For some types of data, sharing may be legitimately delayed or restricted to protect confidential commercial information or for reasons of national or personal security. Privacy considerations are paramount when sharing individual participant-level data from human studies, which legitimize additional protections.

Although it would advance the entire research enterprise, mandatory data sharing would have perhaps its broadest and most immediate impact on clinical trials, where sharing of participant-level data will not only accelerate discovery but would also meet the ethical imperative to honor trial participants’ assumption of personal risk by

maximizing the potential scientific value of the data. Substantial advances have been made in recent years in the technology, infrastructure, and governance of participant-level clinical trial data sharing. Several repositories have established successful models of sharing and have demonstrated assurance of patient privacy and security and feasibility of use and are experiencing accelerating user uptake (7–9). Concern over where and how to share

clinical trial data is no longer a viable rationale for delay, even as we acknowledge that more needs to be done to ease researcher use of these repositories.

One argument for delaying mandated sharing is the desire to introduce the requirement only when all support infrastructure is in place. Although additional standards, policy, and support infrastructure are needed, NIH should not let the perfect be the enemy of good progress. A phased program beginning with mandatory sharing of clinical trial data, with expansion to other types of data as standards and best practices around data stewardship emerge, seems wise. NIH al-



we are concerned that researchers will be able to comply with this policy even if their plan was to effectively withhold data from public access (for example, some current plans amount to little more than “email me”). As written, the draft policy may not practically result in data being shared by default for NIH-sponsored research. Moreover, the draft policy requires submission of data sharing plans only after the proposal review process has concluded, which researchers could interpret as meaning that data sharing plans are not a core part of good scientific practice, unlike trial recruitment or statistical analysis plans.

¹University of California San Francisco, San Francisco, CA, USA. ²Vivli, Cambridge, MA, USA. ³Science Advisors, Washington, DC, USA. ⁴Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University, Cambridge, MA, USA. ⁵Department of Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA. ⁶Pulmonary and Communications Divisions, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. ⁷National Academy of Medicine, Washington, DC, USA. ⁸Duke University, Durham, NC, USA. ⁹Yale University, New Haven, CT, USA. ¹⁰Greenwall Foundation, New York, NY, USA. ¹¹FasterCures (Milken Institute), Washington, DC, USA. ¹²Genetic Alliance, Washington, DC, USA. ¹³Center for Bioethics, Harvard Medical School, Boston, MA, USA. Email: ida.sim@ucsf.edu 361667513

ready has the requisite technical (such as ClinicalTrials.gov) and policy infrastructure (10) that can be extended to support a mandate for participant-level clinical trial data sharing. Mandates may also drive additional public and private investments to bolster a wider data sharing ecosystem. For clinical trials, for example, substantial private investment in data repositories has helped to expand data sharing capacity for both industry and academic trials (11).

Another concern about a sharing requirement is the imposition of rigid standards. Certainly, investigators should have some flexibility—for example, when to share data and which platform to use—but should be held accountable for meeting minimum standards of data stewardship and availability. Flexibility in how to share, not whether to share, should be the policy framework. For clinical trials in particular, refusing to share breaks trust with trial participants' strong desire to share (12). Past experience has shown that science can flourish because of, not in spite of, mandated data sharing: NIH's 2014 Genomic Data Sharing Policy implemented explicit data sharing requirements with sanctions for noncompliance. After some early resistance from the genetics investigator community, genomic data sharing is now widely accepted by researchers and research institutions and was a major enabler of the extraordinary scientific and economic value gained from the Human Genome Project.

Specific, practical, and implementable NIH policies can help transform academic culture and practice toward routine data sharing. Building on the current draft plan, we recommend that NIH establish mandated sharing of participant-level data from interventional clinical research, for which the ethical arguments for sharing are most compelling. Below, we recommend new enforceable policies for implementing such a mandate. These recommendations should apply to all prospective human subjects research and could subsequently be adapted for other biomedical research data sharing.

NIH should establish standards for clinical research data sharing repositories, maintain a list of approved repositories, and promote awareness and use of these repositories. NIH should take the lead in facilitating interoperability among approved repositories and fostering close coordination with ClinicalTrials.gov to minimize burdens to investigators and ensure that the data are "findable" and can be understood in the context of the full range of studies on a particular topic.

NIH should require all clinical research proposals to include a data sharing plan that commits to sharing participant-level data in an approved repository. This data sharing plan should be explicitly scored in the

grant review process; scores should affect the overall funding decision. Once funded, researchers should be required to post the data sharing plan, selected repository, and anticipated date of data availability to ClinicalTrials.gov before enrollment of the first participant to provide public accountability. Subsequent-year funding (for the duration of the study) should be contingent on meeting these new ClinicalTrials.gov reporting requirements. Applications should include the methods and appropriate budget in the main grant proposal to ensure appropriate data stewardship so that data will be findable and sharable in approved repositories at the conclusion of the study.

In addition to the current requirement that human studies report summary results to ClinicalTrials.gov within the time frames established under the law (generally 1 year, with some exceptions) (13), NIH should also require reporting of participant-level data to an approved repository within a reasonable time period. Although there is disagreement about when participant-level data should be reported, the U.S. National Academy of Medicine has deemed 18 months after trial completion to be a reasonable embargo period (14).

NIH should establish mechanisms for applicants to report (such as in biosketches) whether and how they executed on data sharing plans from previous grants. Consequences for failing to report and share results should include loss of eligibility for future funding. Conversely, exemplary past data sharing practices should be recognized and rewarded within the grant review process. Because technical and policy issues remain, NIH should continue to support efforts to address challenges to effective FAIR data sharing, such as data management, alignment for reuse, and sharing infrastructures.

Of course, academic institutions must be partners in this effort. Academia must train and support investigators to meet data sharing objectives; partner with approved data sharing platforms; recognize data contributorship in hiring, promotions, and tenure decisions (15); and train and reward investigators for reusing data as a valuable complement to generating data through primary studies. As a major source of funding for academic medical centers, the Clinical and Translational Science Award program and other major NIH networks should include institutional data sharing practices in its evaluation and funding criteria. Academia, NIH, scientific societies, and other stakeholders should work together to achieve clear and consistent rewards for and enforcement of all data sharing requirements, including sanctions for noncompliance. Fair and robust oversight is essential to ensure

that the fruits of federally funded research are put to maximal use.

We suggest that limited data sharing arises not from culture but from policy. Researcher reluctance to share is a rational response to existing incentive systems that measure and reward individual achievement partly on the basis of the accumulation and use of closely held data sets. NIH has outsized influence to adjust these incentives by mandating and making funding contingent on data sharing, realigning researcher behavior with core values of scholarship. Once data sharing becomes the norm, researchers and the general public will benefit, and in turn, sharing will itself become an incentive. ■

REFERENCES AND NOTES

1. NIH, "Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance" (notice no. NOT-OD-20-013, NIH, 2019); <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-013.html>.
2. M. D. Wilkinson et al., *Sci. Data* **3**, 160018 (2016).
3. A. Bergeris, T. Tse, D. A. Zarin, *JAMA* **319**, 406 (2018).
4. J. P. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research" (Office of Science and Technology Policy, 2013); https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
5. NIH, "Plan for Increasing Access to Scientific Publications" (NIH, 2015); <https://grants.nih.gov/grants/nih-public-access-plan.pdf>.
6. U.S. Government Accountability Office (GAO), "Federal Research: Additional Actions Needed to Improve Public Access to Research Results" (report no. GAO-20-81, GAO, 2019); <https://www.gao.gov/assets/710/702847.pdf>.
7. F. Rockhold, P. Nisen, A. Freeman, *N. Engl. J. Med.* **375**, 1115 (2016).
8. R. Liet al., *NAM Perspect.* **10**, 31478/201811b (2018).
9. J. S. Ross et al., *Sci. Data* **5**, 180268 (2018).
10. NIH, "NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information" (NIH, 2017); <https://grants.nih.gov/policy/clinical-trials/reporting/understanding/nih-policy.htm>.
11. R. Banzi et al., *Trials* **20**, 169 (2019).
12. M. M. Mello, V. Lieou, S. N. Goodman, *N. Engl. J. Med.* **378**, 2202 (2018).
13. D. A. Zarin, T. Tse, R. J. Williams, S. Carr, *N. Engl. J. Med.* **375**, 1998 (2016).
14. Institute of Medicine, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* (National Academies Press, 2015).
15. H. H. Pierce, A. Dev, E. Statham, B. E. Bierer, *Nature* **570**, 30 (2019).

ACKNOWLEDGMENTS

I.S., R.L., and B.E.B. are cofounders of Vivli, a nonprofit platform for clinical trial data sharing, and currently serve on the Vivli Board of Directors. R.L. is an employee of Vivli, and I.S. is a paid consultant of Vivli. M.S. serves on the Vivli Board of Directors, and B.L. and J.D. serve on Vivli's External Advisory Committee. F.R. serves as a Vivli Senior Advisor and Principal Investigator of Supporting Open Access for Researchers, Duke Clinical Research Institute's data sharing platform partially funded by Bristol-Myers Squibb. Vivli receives funding support from the Arnold Ventures, Doris Duke Charitable Trust, the Helmsley Charitable Trust, the pharmaceutical industry, and academic institutional partners. J.S.R. and H.M.K. are co-founders of the Yale Open Data Access Project, a Yale University platform for clinical trial sharing, which shares trials for and receives support from Johnson and Johnson. A.J.B. receives research support from the National Institute of Allergy and Infectious Diseases as principal investigator of ImmPort, an open-access data repository, including clinical trials data. Additional author disclosures are provided as supplementary materials.

SUPPLEMENTARY MATERIALS

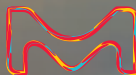
science.sciencemag.org/content/367/6484/1308/suppl/DC1

10.1126/science.aba4456

Unleash the impossible

Whatever projects you're working on, from a simple test to life-sustaining medical breakthroughs, you can count on the unparalleled Sigma-Aldrich® portfolio along the way. With proven, reliable, cutting-edge lab and production materials, we are always expanding and perfecting our global portfolio. So we'll be with you from day one to the day your impossible dream suddenly doesn't seem so impossible.

To find out more, visit:
SigmaAldrich.com/UnleashTheImpossible



MilliporeSigma, the vibrant M and Sigma-Aldrich are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources.

© 2020 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved.

The Life Science Business of Merck KGaA, Darmstadt, Germany operates as MilliporeSigma in the US and Canada.

Sigma-Aldrich.
Lab & Production Materials

NEW!

Low-Noise Ultra-Fast Digital Patch Clamp Amplifier System

- Integrated data acquisition interface
- High bandwidth for fastest signal characterization
- Single-channel and whole-cell patch clamp recordings
- Digital compensation circuitry for precision and signal fidelity
- Quick and easy setup
- Bundled SutterPatch® software built on Igor Pro platform



dPatch®

The next generation Digital Patch Clamp Amplifier System. Combining high-speed, high-resolution digital processing, precision A/D circuitry, integrated data acquisition and bundled SutterPatch® software, the dPatch system provides capabilities previously out of reach for the electrophysiologist. Available in either a single- or double-headstage configuration, the dPatch meets the requirements of today's experiments and anticipates the demands of tomorrow's.



SUTTER INSTRUMENT

PHONE: +1.415.883.0128
FAX: +1.415.883.0572
EMAIL: INFO@SUTTER.COM
WWW.SUTTER.COM

AAAS Travels

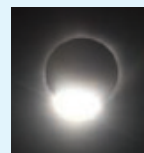
Members and friends of AAAS are invited to join AAAS Travels on fascinating trips to all 7 continents!

Argentina Eclipse December 8-19, 2020



Total Solar Eclipse - Dec. 14, 2020

Explore the magnificence of Argentina, from the premiere city of Buenos Aires and historic La Plata to San Martín de los Andes, and the thundering wonder of Iguazú Falls. View the Total Solar Eclipse just north of San Martín de los Andes. \$5,995 pp + air.



Diamond Ring photo courtesy David Morrison

Easter Island



December 2-9, 2020

There are about one thousand 10 to 30-ft tall moais found on Easter Island, carved out of the volcanic rock. Our adventure in Easter Island will give you a great look at the heritage of this remarkable location with an excellent local guide. \$2,995 pp share + air

For a detailed brochure, please call (800) 252-4910

All prices are per person twin share + air



BETCHART EXPEDITIONS Inc.
17050 Montebello Rd
Cupertino, California 95014
Email: AAASInfo@betchartexpeditions.com
www.betchartexpeditions.com

**What Stars Are Made Of:
The Life of Cecilia
Payne-Gaposchkin**
Donovan Moore
Harvard University Press,
2020. 320 pp.



class physicists.”) Payne-Gaposchkin was also taught by Niels Bohr, whose quantum theory of atomic structure would enable her to come to her own revolutionary conclusions.

By the end of her time at Cambridge, it had become clear to Payne-Gaposchkin that she would never be employed as an astronomer in England. So she secured a fellowship at the Harvard Observatory and moved to America. Here, she was granted research opportunities, but the discrimination she had experienced at home continued.

One of the most egregious perpetrators of this discrimination was Harvard’s president, Abbott Lawrence Lowell, who declared that Payne-Gaposchkin would never be named a professor as long as he was alive. “Lowell had tried to limit Jewish enrollment at Harvard to 15 percent, and he tried to ban black students from living in the freshman dorms. In both instances, the Harvard Board of Overseers overruled him,” writes Moore. “The board did not overrule him, however, when he decreed in 1928 that women should not receive teaching appointments from the Harvard Corporation.” Payne-Gaposchkin was devastated: “‘I had no official status,’ Cecilia recalled. ‘I was paid so little that I was ashamed to admit it to my relations in England.’”

Despite these and other hardships, Payne-Gaposchkin’s accomplishments were remarkable. She wrote several books and more than 270 journal articles, was elected to both the Royal and American Astronomical Societies and the American Philosophical Society, earned an honorary doctorate from Smith College, and was the first woman to receive the American Astronomical Society’s lifetime achievement award. In 1956, after Lowell’s death, she was named the first female professor at Harvard. She died just before the election that would have admitted her to the National Academy of Sciences.

To Moore, Payne-Gaposchkin was the classic driven scientist. “She endured everything from laboratory slights to classroom derision because there was no choice. She was driven to understand, which meant that nothing in the way would stop her.” This is a view Payne-Gaposchkin echoed in her own memoir: “Do not undertake a scientific career in quest of fame or money. There are easier and better ways to reach them. Undertake it only if nothing else will satisfy you; for nothing else is probably what you will receive.” ■

10.1126/science.aba9179



BOOKS *et al.*

SCIENCE LIVES

Sexism and the stars

A new biography tells the tale of an accomplished astronomer’s barrier-breaking life

By **Jennifer Carson**

In the early 20th century, astronomers believed in a uniformity principle that held that all objects in the universe were made of the same elements, in approximately the same amounts. In 1925, however, Cecilia Payne, a Ph.D. student at Harvard, discovered that stars are composed of a million times more hydrogen than was previously assumed. But because she was young and female, the scientific community rejected her findings. It would take several decades before Payne-Gaposchkin received the recognition she was due. In *What Stars Are Made Of*, using compact and skillful prose, Donovan Moore charts Payne-Gaposchkin’s scientific life from grade school standout to world-class astronomer.

Moore includes impressive context on the misogyny that existed within British academic society at the turn of the 20th century. He notes, for example, the Cambridge University tradition wherein, when women entered a lecture hall, the roomful of men would stomp along in time with the women’s footsteps. It was a culture of discrimination that at times turned violent, as it did on 24 October 1921, when the university announced its decision to grant women “titular degrees,” and a furious mob of Cambridge men stormed the women’s college, bashing in its iron gates with a coal cart.

The reader meets physics luminaries and future Nobel Prize winners in the Cavendish laboratory, where Payne-Gaposchkin trained as an undergraduate. The lab was directed by J. J. Thomson when she arrived and Ernest Rutherford when she left. (Thomson, incidentally, believed that women “simply did not have the intellectual capacity to be world-

The reviewer is a physics professor and freelance writer in Los Angeles, CA, USA. Email: jenniferecarson@gmail.com

CLASSICS REVISITED

Vanquishing false idols, then and now

Francis Bacon's 400-year-old list of scientific foibles holds lessons for modern scientists

By Kevin P. Weinfurt

In the early 17th century, the English philosopher and statesman Francis Bacon envisioned a bold, multiphase program to accumulate knowledge of the natural world. A critical part of this plan was *Novum Organum*, which celebrates its 400th anniversary this year. In this work, Bacon attempted to undo the centuries-old dominance of Aristotelian forms of inquiry, encouraging readers to instead apply inductive reasoning to carefully curated observations of the natural world.

"Book One" of *Novum Organum* addressed why so little progress had thus far been made in understanding nature. Here, Bacon cautioned against "idols and false notions" that can interfere with the quest for scientific knowledge, providing the first and possibly the most comprehensive catalog of human foibles that can threaten the integrity of science.

Bacon referred to these shortcomings as "Idols" of "the Tribe," "the Cave," "the Marketplace," and "the Theater." The "Idols of the Tribe" are the tendencies of the mind to leap to incorrect conclusions, for example, our inclination "to suppose the existence of more order and regularity in the world"

than is actually there. To combat a prominent Idol of the Tribe—the tendency to seek and be moved by confirmatory evidence more so than by disconfirmatory evidence (what we now call confirmation bias)—Bacon directed the scientist to construct a "Table of Deviations, or of Absence in Proximity" that documents observations that are similar to an affirmative but for which an association does not hold (e.g., although heat accompanies the Sun's rays, heat does not accompany the rays of the Moon).

"Idols of the Cave" refers to how people occupying different "caves," or groups, differ in their scientific beliefs and practices. Here, Bacon described how scientists can become attached to ideas or practices "either because they fancy themselves the authors and inventors thereof, or because they have bestowed the greatest pains upon them and become most habituated to them."

"Idols of the Marketplace" refers to the intellectual risks entailed in the use of language to conduct science. Bacon regarded this category of idol as the "most troublesome of all," perhaps because one cannot escape using language and because the negative effects of this necessary practice can be so insidious. Included in this category is the tendency to assign names to things that do not exist, which encourages one to believe that they do exist. This happens whenever researchers use

a single label (e.g., "breast cancer") to refer to a collection of phenomena ("breast cancer" actually refers to many different pathophysiologic conditions). "Idols of the Theater," or "Idols of the System," refers to people's tendency to cling to dogmatic systems of belief that portray a tidy and/or entertaining but ultimately inaccurate picture of nature.

Bacon, a contemporary of Galileo and Shakespeare, wrote *Novum Organum* at a time when many still believed that truths about the world were handed down by monarchs and ministers. He spoke for the burgeoning empirical sciences, encouraging readers to use the inductive method to throw off the shackles of authority. But if we are to realize his vision for a practice of science that frees people from the shackles of both authority and their own minds, it might be a good idea to update the idols to reflect the modern challenges threatening the scientific enterprise today.

The Idols of the Tribe, for example, need to expand to include the social psychological tendencies of group dynamics. The Idols of the Cave, which originally alluded to the conflicting interests of the scientist, must now emphasize the ubiquity and variety of financial and nonfinancial conflicts of interest attendant to the massive institutional and bureaucratic systems that have risen up around scientific activities. To the Idols of the Marketplace, meanwhile, we might add that the pressures within academic science to publish and win grants have exacerbated the use of trendy yet ill-defined terms. And finally, the Idols of the Theater might be updated to include the uncritical adherence to systems of ritualized rules intended to automate the inductive activities of scientists.

One year after the publication of *Novum Organum*, Bacon, who was seriously in debt, was accused of corruption, briefly jailed in the Tower of London, and barred from Parliament for life. It took several decades before his work began to receive wide praise, and in 1660 it inspired the creation of the Royal Society. A modern reader might be similarly inspired by *Novum Organum's* subtlety of thought, commitment to understanding nature as it is, and excitement about the potential for science to be a liberating force for humankind. ■

The reviewer is at the Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, USA. Email: kevin.weinfurt@duke.edu



A 17th-century lecturer provides an empirical demonstration of an optical principle.

10.1126/science.aba5802



U.S. college students consider next steps after an announcement that classes will be held online for 3 weeks to mitigate the spread of COVID-19.

Edited by Jennifer Sills

Preventing COVID-19 prejudice in academia

The coronavirus disease 2019 (COVID-19) outbreak has triggered global panic (1). Because the disease emerged in China and most of the affected individuals so far have been of Asian origin, a backlash of Asia-phobic reactions has been reported in various regions of the world (2–4). As it moves through Europe, Europeans may also be at risk of discrimination. Given that xenophobia during outbreaks is not uncommon (5), universities should proactively develop policies that support students, faculty, and staff affected by discriminatory behavior both now and in the future. To study or work abroad, members of academia often must travel far from home, adapt to another culture, and overcome a language barrier. Facing prejudice, including discrimination related to COVID-19, may add to feelings of isolation (6, 7) and affect career development, especially for students.

Preventive measures by universities to lessen prejudice should include transparency about the disease status, data gathering, and direction about appropriate behavior. University administrators should release reassuring statements about the local COVID-19 situation that supplement the information released by health authorities. Academic administrators should survey students and staff of Asian origin (as well as others if appropriate) to determine whether they have experienced any prejudice related to COVID-19 and whether

they expect university authorities to take any additional action. The administrators should also release statements that explain that in Asia, people wear masks for a variety of reasons, such as to filter polluted air, make fashion or political statements, or provide social indicators that they want to be left alone in public spaces (8). Typical surgical face masks do not necessarily indicate someone is sick, and as many students are likely aware, they do little to prevent catching viral infections (9).

Universities can also launch social media campaigns that support Asian and Asian-American students (and other targets) in the form of infographics or videos. Both university administrators and department heads should issue a notice that COVID-19-related prejudice or xenophobic reactions from academic staff and other students will not be tolerated and will be treated in accordance with anti-discrimination laws. Finally, university leaders at all levels should encourage students and academic staff to provide extra support and kindness to Asian and other international students during the ongoing outbreak.

Piotr Rzymski^{1*} and Michał Nowicki²

¹Department of Environmental Medicine, Poznań University of Medical Sciences, 0-781 Poznań, Poland. ²Department of Histology and Embryology, Poznań University of Medical Sciences, 60-806 Poznań, Poland.

*Corresponding author.

Email: rzymskipiotr@ump.edu.pl

REFERENCES AND NOTES

1. *Lancet* **395**, 537 (2020).
2. K. Shimzu, *Lancet* **395**, 685 (2020).
3. J. Wen et al., *Anatolia*, 10.1080/13032917.2020.1730621 (2020).
4. L. Asmelash, "UC Berkeley faces backlash after stating 'xenophobia' is 'common' or 'normal' reaction to coronavirus," *CNN* (2020).

5. B. A. O'Shea et al., *Soc. Psychol. Pers. Sci.*, 10.1177/1948550619862319 (2019).
6. K. Yan, D. C. Berliner, *Asia Pacific Educ. Rev.* **12**, 173 (2011).
7. T. Hamamura, P. G. Laird, *J. Multicult. Couns. Dev.* **42**, 205 (2014).
8. A. Burgess, M. Horii, *Soc. Health Illness* **34**, 1184 (2012).
9. M. Loeb et al., *JAMA* **302**, 1871 (2009).

10.1126/science.abb4870

Humpback dolphins at risk of extinction

The humpback dolphin genus (*Sousa* spp.) has recently been revised to contain four species: *S. teuszii*, *S. plumbea*, *S. chinensis*, and *S. sahulensis* (1). All four species are listed on Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (2). *S. chinensis* and *S. sahulensis* are classified as Vulnerable on the International Union for Conservation (IUCN) Red List, *S. plumbea* is Endangered, and *S. teuszii* is Critically Endangered (3, 4). The humpback dolphins and their habitats—shallow, coastal waters of the eastern Atlantic, Indian, and western Pacific oceans—are threatened by fishing, vessel traffic, habitat degradation and destruction, environmental contaminants, and prey depletion, putting the humpback dolphins at risk of extinction (4–7). Interventions at the national and international levels are urgently needed.

Environmental pollution and other anthropogenic activities affect humpback dolphins throughout their territories. High levels of organochlorines have been found in the blubber of humpback dolphins in China and South Africa (8, 9). Large-scale dredging, drilling, land reclamation,

construction blasting, boat traffic, resource extraction, and other coastal development projects are concentrated within humpback dolphin habitat and threaten their survival in Africa, India, southeastern China, and northern Australia (4, 7, 9, 10).

Saving the humpback dolphins, a charismatic megafauna and valuable genetic resource, is important to nearshore marine biodiversity (4). Yet dolphin populations in poor, developing countries receive little study and management attention (4). Even in the territories of relatively wealthy and well-developed nations, such as Australia, China, and South Africa, the state and local governments have failed to provide adequate protection for the humpback dolphins to prevent population declines (4, 7, 10).

To ensure the safety of humpback dolphins, scientists and conservationists must thoroughly evaluate and recognize the conservation status and risk factors of the various species, subspecies, and populations. Governments of countries that are home to humpback dolphins should also take urgent action to support extensive and substantive national and international collaboration to put in place a set of conservation actions, such as helping coastal fishermen to broaden their income sources and decrease their reliance on nearshore fishing, thereby reducing the threat to humpback dolphins and their habitats.

Songhai Li

Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, 572000, China. Email: lish@idsse.ac.cn

REFERENCES AND NOTES

1. T. A. Jefferson, H. R. Rosenbaum, *Mar. Mamm. Sci.* **30**, 1494 (2014).
2. CITES, Appendices I, II, and III (2019); <https://cites.org/eng/app/appendices.php>.
3. "Humpback dolphin" search, The IUCN Red List of Threatened Species (2017); www.iucnredlist.org/search?query=humpback%20dolphin&searchType=species.
4. T. A. Jefferson, B. E. Curry, *Adv. Mar. Biol.* **72**, 17 (2015).
5. M. Liu *et al.*, *Anim. Conserv.* **20**, 182 (2017).
6. S. Li *et al.*, *Integr. Zool.* **13**, 495 (2018).
7. L. Karczmarski *et al.*, *Adv. Mar. Biol.* **73**, 27 (2016).
8. D. Gui *et al.*, *Chemosphere* **114**, 106 (2014).
9. T. Gill *et al.*, *Adv. Mar. Biol.* **72**, 119 (2015).
10. G. J. Parra, D. Cagnazzi, *Adv. Mar. Biol.* **73**, 157 (2016).

10.1126/science.abb5744

A first step for the Yangtze

As China has rapidly developed over the past 40 years, the Yangtze River has suffered a severe loss of biodiversity. The baiji (Yangtze river dolphin) is likely extinct (1), the Yangtze finless porpoise is critically endangered and declining (2), and more than 30% of Yangtze fish species are at the brink of extinction (3). This year, a Chinese research team reported on the extinction of the Chinese paddlefish (4). The decline of

vertebrate biodiversity in the Yangtze River is primarily attributable to illegal fisheries and excessive fishing (5). Chinese policy to discourage overfishing has improved, but more must be done to prevent the extinction of many of these species.

In 2002, China initiated an annual 3-month closure of all commercial fisheries on the Yangtze River. In 2016, the closure was increased to 4 months. Two years later, fisheries were banned in all of China's aquatic biological reserves (2). Then, in January 2020, China instituted a 10-year ban on all commercial fishing in the Yangtze River and its tributaries, including the adjoining lakes (6). The new regulations were enacted in an effort to save endangered Yangtze River aquatic life and prevent further extinctions. The initial 10-year period of the fishery ban was chosen because it will extend through two to three generations for most of the fish species in the Yangtze River (7). Although implementation of the fishing ban will face many logistic difficulties, this is a momentous first step, which, together with a shift in public support for conservation (2), gives us great hope for the protection of aquatic life in the Yangtze River.

Despite these positive steps, many species remain at risk of extinction. It will be important to quantify the extinction risk for species in the Yangtze Basin in the coming years. We must also anticipate the disproportionate effects on biodiversity that could result from the loss of groups with a longer evolutionary history, fewer system branches, and fewer species (8, 9). Because of its unique ancient evolutionary characteristics, the baiji was given conservation priority among the 4510 mammals in the world (9), and the Chinese paddlefish was the basal group in all ray-finned fish (10); now both are likely extinct. To prevent further extinctions, in addition to adhering to the fishing ban, we must invest more in research to identify those (potentially less iconic) species that are at greatest risk of extinction and which would result in the greatest loss of global biodiversity.

Zhigang Mei^{1*}, Peilin Cheng^{2*}, Kexiong Wang¹, Qiwei Wei², Jay Barlow³, Ding Wang¹

¹Key Laboratory of Aquatic Biodiversity and Conservation of the Chinese Academy of Sciences, Institute of Hydrobiology, Wuhan 430072, China. ²Key Laboratory of Freshwater Biodiversity Conservation, Ministry of Agriculture of China, Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan 430223, China. ³National Oceanic and Atmospheric Administration, Southwest Fisheries Science Center, La Jolla, CA 92037, USA. *Corresponding author. Email: meizhigang@ihb.ac.cn (Z.M.); chengpeilin@yfi.ac.cn (P.C.)

REFERENCES AND NOTES

1. S. T. Turvey *et al.*, *Biol. Lett.* **3**, 537 (2007).
2. J. Huang *et al.*, *Biol. Conserv.* **241**, 108315 (2019).
3. L. Cao *et al.*, *Biodivers. Sci.* **24**, 598 (2016).

4. H. Zhang *et al.*, *Sci. Total Environ.* **710**, 136242 (2019).
5. D. Dudgeon, in *River Conservation: Challenges and Opportunities*, S. Sabater, A. Elsosegi, Eds. (Fundación BBVA, Bilbao, Spain, 2013), pp. 129–167.
6. Notice of the Ministry of Agriculture and Rural Affairs on the scope and time of arrests in key waters of the Yangtze River Basin (www.cjyzbgs.moa.gov.cn/tzgg/201912/t20191227_6334009.htm) [in Chinese].
7. F. Liu *et al.*, *Acta Hydr. Sinica* **43**, 144 (2019).
8. D. P. Faith, *Cladistics* **8**, 361 (1992).
9. N. J. Isaac *et al.*, *PLOS One* **2**, e296 (2007).
10. L. C. Hughes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6249 (2018).

10.1126/science.abb5537

TECHNICAL COMMENT ABSTRACTS

Comment on "Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity"

Max R. Lambert, Molly C. Womack, Allison Q. Byrne, Obed Hernández-Gómez, Clay F. Noss, Andrew P. Rothstein, David C. Blackburn, James P. Collins, Martha L. Crump, Michelle S. Koo, Priya Nanjappa, Louise Rollins-Smith, Vance T. Vredenburg, Erica B. Rosenblum Scheele *et al.* (Reports, 29 March 2019, p. 1459) bring needed attention to the effects of amphibian infectious disease. However, the data and methods implicating the disease chytridiomycosis in 501 amphibian species declines are deficient. Which species are affected, and how many, remain critical unanswered questions. Amphibians are imperiled; protective actions require public support and robust science. Full text: [dx.doi.org/10.1126/science.aay1838](https://doi.org/10.1126/science.aay1838)

Response to Comment on "Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity"

Ben C. Scheele, Frank Pasmans, Lee F. Skerratt, Lee Berger, An Martel, Wouter Beukema, Aldemar A. Acevedo, Patricia A. Burrows, Tamile Carvalhal, Alessandro Catenazzi, Ignacio De la Riva, Matthew C. Fisher, Sandra V. Flechas, Claire N. Foster, Patricia Frías-Álvarez, Trenton W. J. Garner, Brian Gratwicke, Juan M. Guayasamin, Mareike Hirschfeld, Jonathan E. Kolby, Tiffany A. Kosch, Enrique La Marca, David B. Lindenmayer, Karen R. Lips, Ana V. Longo, Raúl Maneyro, Cait A. McDonald, Joseph Mendelson III, Pablo Palacios-Rodríguez, Gabriela Parra-Olea, Corinne L. Richards-Zawacki, Mark-Oliver Rödel, Sean M. Rovito, Claudio Soto-Azat, Luis Felipe Toledo, Jamie Voyles, Ché Weldon, Steven M. Whitfield, Mark Wilkinson, Kelly R. Zamudio, Stefano Canessa

Lambert *et al.* question our retrospective and holistic epidemiological assessment of the role of chytridiomycosis in amphibian declines. Their alternative assessment is narrow and provides an incomplete evaluation of evidence. Adopting this approach limits understanding of infectious disease impacts and hampers conservation efforts. We reaffirm that our study provides unambiguous evidence that chytridiomycosis has affected at least 501 amphibian species. Full text: [dx.doi.org/10.1126/science.aay2905](https://doi.org/10.1126/science.aay2905)

Fundación BBVA

winning

13th Edition BBVA Foundation Frontiers of Knowledge Awards

With the collaboration of



The **BBVA Foundation Frontiers of Knowledge Awards** recognize and reward world-class research and artistic creation, prizing contributions of singular impact for their originality and significance. The name of the scheme is intended to denote not only research work that substantially enlarges the scope of our current knowledge – pushing forward the frontiers of the known world – but also the meeting and overlap of different disciplinary areas and the emergence of new fields.

The Frontiers of Knowledge Awards honor fundamental disciplinary or interdisciplinary advances across a broad expanse of the knowledge map of the 21st century.

The BBVA Foundation is assisted in the award process by the **Spanish National Research Council (CSIC)**.

Categories

- 1/ **Basic Sciences (Physics, Chemistry, Mathematics)**
- 2/ **Biology and Biomedicine**
- 3/ **Information and Communication Technologies**
- 4/ **Ecology and Conservation Biology**
- 5/ **Climate Change**
- 6/ **Economics, Finance and Management**
- 7/ **Humanities and Social Sciences**
- 8/ **Music and Opera**

In Humanities and Social Sciences, the award will alternate annually between these two disciplinary domains, with this thirteenth edition dedicated to the Humanities.

Nomination

Nominations are invited from scientific or artistic societies and organizations, public or private R&D centers, university and hospital departments, schools of music, orchestras, and organizations working on or around the issue of climate change, as well as other institutions specified in the call conditions.

Entry submission

The nomination period concludes at **23:00 GMT on June 30, 2020**.

www.frontiersofknowledgeawards-fbbva.es

Fundación BBVA

Plaza de San Nicolás, 4 · 48005 Bilbao
Paseo de Recoletos, 10 · 28001 Madrid
awards-info@fbbva.es



THE FROZEN

By Jesse Smith

The earliest recorded sighting of Antarctica was made 200 years ago in 1820, when the first Russian Antarctic expedition, led by Captain Fabian Gottlieb von Bellingshausen, encountered what is now called the Fimbul Ice Shelf. In commemoration of that historic event, we present this special issue on Antarctica. The defining aspect of Antarctica is its massive ice sheet. The Antarctic Ice Sheet is Earth's largest

reservoir of fresh water, with the capacity to raise sea level by more than 60 meters if it were to completely melt into the ocean. Additionally, Antarctica is one of the great archives of past climate: The EPICA Dome C ice core, for example,

provides a continuous record of the past 800,000 years and represents one of the crowning achievements of modern climate science.

This special issue describes the formation of the Antarctic Ice Sheet and the geological processes controlling its existence; the ice sheet's evolution, as affected by its interaction with the surrounding ocean; and how the continent's ice is expected to change in our warming future. We also examine a more biological facet of the region—the mysterious recent die-off of penguins on a remote Antarctic island. Together, the contents of this special issue provide a window into the physical aspects and fauna of this remarkable part of the world.

Full moon over
sea ice formations
during polar
night at Atka Bay,
Antarctica

PHOTO: STEFAN CHRISTMANN/NPL/MINDEN PICTURES

NEWS

The lost colony p. 1318

REVIEWS

History, mass loss, structure, and
dynamic behavior of the Antarctic
Ice Sheet p. 1321

The Southern Ocean and its interaction
with the Antarctic Ice Sheet p. 1326

The uncertain future of the Antarctic
Ice Sheet p. 1331

CONTINENT

THE LOST COLONY

Some 900,000 king penguins vanished without a trace. Why?

By **Eli Kintisch**

Where on Earth, wondered Henri Weimerskirch, were all the penguins? It was early 2017. Colleagues had sent the seabird ecologist aerial photos of Île aux Cochons, a barren volcanic island halfway between Madagascar and Antarctica that humans rarely visit. The images revealed vast areas of bare rock that, just a few decades before, had been crowded with some 500,000 pairs of nesting king penguins and their chicks. It appeared that the colony—the world’s largest king penguin aggregation and the second biggest colony of any of the 18 penguin species—had shrunk by 90%. Nearly 900,000 of the regal, meter-high, black, white, and orange birds had disappeared without a trace. “It was really incredible, completely unexpected,” recalls Weimerskirch, who works at the French national research agency CNRS.

Soon, he and other scientists were planning an expedition to the island—the first in 37 years, and only the third ever—to search for explanations. “We had to go see for ourselves,” says CNRS ecologist Charles Bost.

As the researchers prepared for the journey, they had to grapple with the logistical, political, and scientific challenges that have long bedeviled biologists trying to understand Antarctica’s remote ecosystems. The vast distances, rough weather, and rugged terrain make travel difficult and expensive. They needed a ship—and a helicopter, because frigid seas and rocky shores make for perilous boat landings on Antarctic islands. Complying with the tough permitting and biosecurity rules governing the French-controlled island—

meant to prevent researchers from disturbing fragile ecosystems—required careful planning and paperwork that took months to complete. And once they arrived, they would have precious little time: just 5 days to investigate a multitude of suspects in the disappearance, including disease, predators, and a warming Southern Ocean.

In all likelihood, they would never be able to return. “We knew this was going to be a one-shot expedition,” recalls conservation biologist Adrien Chaigne, an expedition organizer who works for the National Nature Reserve of the French Southern and Antarctic Territories, which manages the island. “It was a real kind of pressure.”

CONSTRAINTS LIKE THAT have long faced biologists seeking to understand life at the remote bottom of the planet. Two centuries ago, researchers wanting to visit the region had to tag along with explorers, whalers, or seal hunters. The Adélie penguin, for example, was first identified by a naturalist who joined an 1837 expedition to southeastern Antarctica led by the French explorer Jules Dumont d’Urville, who named the place Terre Adélie after his wife. The harrowing voyages rewarded them with surprises: In 1895, botanists certain no plant could survive the frigid Antarctic were shocked to discover lichens on Possession Island, near Île aux Cochons.

Today, modern research budgets and a network of polar research stations have made Antarctica more accessible. Biologists have flocked to the region to tackle an array of fundamental questions, including how animals evolved to survive subzero temperatures and how ecosystems are organized in the vast, productive Southern Ocean. Cli-

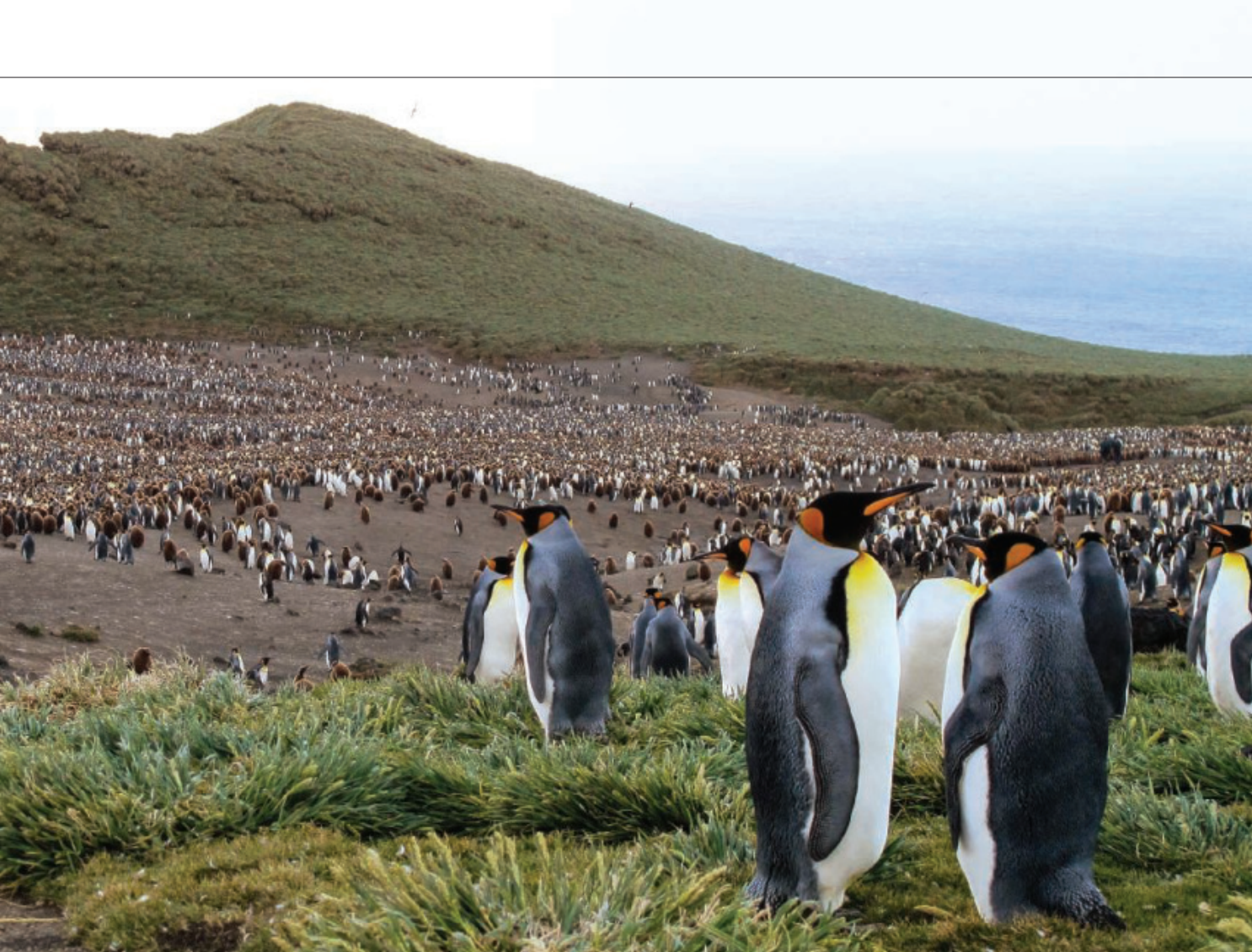


The king penguin throngs on Île aux Cochons in November 2019 were a fraction of their past size.

mate change, which has made the Antarctic one of the fastest changing places on Earth, has inspired studies of shifting ice and acidifying seas. The potential for discovery makes the region addictive, says marine biologist Deneb Karentz of the University of San Francisco. “Once you go as a scientist you always want to go back.”

But even today, Antarctic research is challenging. “If it takes you 2 hours to collect samples back home, it could take 10 in Antarctica,” Karentz says. Holes drilled in sea ice to collect samples, for example, often need poking to remain open. The harsh conditions can claim valuable gear. In 1987, shifting sea ice swept away a plexiglass frame Karentz was using to study microorganisms beneath the surface. She scrambled to replace it with materials scrounged from a nearby research station. In Antarctica, she says, “You have to be resourceful.”

SUCH LESSONS weren’t lost on Weimerskirch and Bost, both veterans of Antarctic research, when a helicopter from the *Marion Dufresne*, a French research vessel, delivered the pen-



guin researchers and their 700 kilograms of gear to Île aux Cochons in November 2019. It was the middle of king penguin nesting season, and they were greeted by the raucous honking and chirping of tens of thousands of chicks. They also saw vast empty swaths of bedrock, where previous generations of birds had scraped away the soil. The scientists estimate penguins once stood flipper to flipper on some 67 square kilometers of ground that is now abandoned.

The researchers were eager to find out what had caused those losses. King penguins, numbering 3.2 million birds in the Antarctic region, aren't in immediate danger; in fact their numbers are mostly rebounding from centuries of human hunting. By contrast, half the world's penguin species are threatened (see graphic, p. 1320), and several have recently experienced disturbing die-offs. But big losses of even relatively healthy species could point to broader threats, which is why the calamity on Île aux Cochons triggered alarm bells.

King penguins should be relatively easy to study. Unlike their ice-bound cousins,

such as emperor penguins, king penguins live on islands dotting the subantarctic region. That means they can be reliably and repeatedly counted in satellite images over time, and scientists can camp alongside their breeding colonies to observe them. (Other ice-dependent species, like emperor penguins, are more peripatetic.) During

the lengthy breeding season, the parents trade off tasks, with one incubating eggs or rearing fluffy brown chicks while the other heads to sea to catch fish and other sea creatures. These foraging round-trips can cover 500 kilometers or more, electronic tags attached to the birds have shown.

The researchers' first priority was to attach such tags to 10 penguins, to see whether foraging changes might have contributed to the losses. It wasn't easy.

The team's permits stipulated they utilize just one well-trodden trail and operate only on the edge of the colony. Breaks in the rain allowed the scientists to glue transmitters on the birds' feathers.

Other researchers, meanwhile, set up traps, cameras, and night-vision optics to look for cats and mice, which were introduced by whalers or sealers long ago and are known to eat seabird eggs and chicks. The scientists also took samples of penguin blood, to be screened later for diseases and other data. And they collected feathers and dug up penguin bones that



could yield further ecological clues, including dietary changes.

"The first 2 days were intense," Chaigne says. "We knew it was possible that bad weather could end the expedition any day." Luckily, they avoided serious storms, and by the end of their fifth day the researchers had tagged the penguins and gathered the samples they sought.

REAMS OF DATA remain to be digested. But the researchers have already ruled out some possible explanations for the massive penguin decline. Land predators, for instance, don't seem to have played a role. Examinations of chicks and adult penguins, as well as excavated bones, revealed no signs of cat or mouse bites, and the team's cameras recorded no attacks. (Rabbits, seen on previous expeditions, were curiously missing.)

Nor, it seems, had the penguins simply moved somewhere nearby. A second smaller colony on the island, a natural site for relocation, had just an estimated 17,000 pairs, not enough to explain the massive drop-off in the main group. And Bost says there's no obvious indication—in satellite images, for instance—that the colony relocated to some other island.

That leaves one main explanation, Bost says: "If the penguins are not here, they died." But what killed them?

Not disease, apparently. The team is waiting on final blood analyses, but they saw few ailing birds or fresh corpses. "We thought we'd see carrion, individuals in bad condition," Chaigne says. But the birds looked healthy.

Instead, he and his colleagues suspect that changes in the surrounding ocean forced the penguins to swim farther to find food. Studies of other king penguin colonies suggest foraging birds from Île aux Cochons normally swim toward an oceanic feature hundreds of kilometers to the south known as the polar front or Antarctic convergence. The front marks the northern extent of the colder Antarctic waters. The penguins are attracted by the many sea creatures that gather at such thermal edges—especially the bird's main prey, lanternfish, which form huge schools some 100 meters or more below the surface.

The polar front doesn't stay in the same place every year. During some years, climate anomalies known as the El Niño-Southern

Oscillation and the Subtropical Indian Ocean Dipole cause ocean waters in the region to warm, and the polar front shifts south, closer to the pole and farther from Île aux Cochons. During the longer foraging trips, hunger might force the parent left back at the colony to leave the nest to feed—leaving chicks vulnerable to predators or starvation. The longer swims might also make the adult penguins more vulnerable to deadly stress and predation. And those anomalous years offer a preview of how the Southern Ocean is expected to warm in the coming decades, steadily shifting the polar front farther south.

Evidence that a warming ocean could threaten the penguins comes from a 2015 study that Bost and his colleagues did at a smaller king penguin colony, on Possession Island, some 160 kilometers west of Île aux

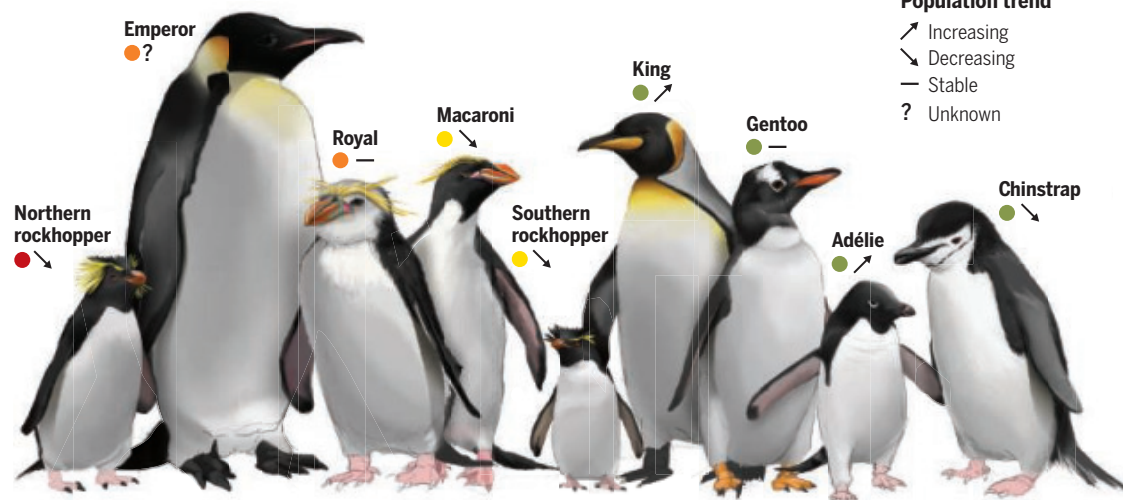
changes could cut king penguin numbers by half by the end of the century.

Whether that scenario explains the Île aux Cochons crash may never be entirely clear. (Another possibility is that the colony just grew unusually large during some bountiful decades, then fell back when conditions became more typical.) But the tags the researchers placed on the 10 penguins during the expedition could offer some new clues. Five are still transmitting and could continue to provide data into early 2021.

Already, the tags have offered some surprises: They show that a few of the penguins headed north—not south—from the island to forage. That could mean the birds are hunting at a different thermal edge, known as the sub-Antarctic front. "It's a small sample size of course," Weimerskirch says, "but it's very interesting." The tag data

Antarctica's birds in plight

"Penguins are in trouble," researchers warned last year in *Conservation Biology*. Populations of more than half of the world's 18 species are declining, they noted, including several of the nine species that live in Antarctica and the sub-Antarctic (below). Threats include habitat loss, drowning in fishing gear, and climate change.



Cochons. The island hosts France's Alfred Faure research station, and less strict biosecurity rules allow researchers to continually monitor the colony and climate and oceanographic conditions. The study, published in *Nature Communications*, analyzed 124 foraging routes taken by 120 tagged birds over 16 years. It found that in years when the polar front moved south, the penguins had to travel hundreds of kilometers farther. During "these very unfavorable environmental conditions," the researchers wrote, "the penguin breeding population experienced a 34% decline."

Building on that study, a 2018 paper published in *Nature Climate Change* forecast that warming seas and other environmental

might also reveal a trend toward longer foraging trips, which could suggest the worrisome forecasts about the impact of climate change are accurate.

The unexpected calamity on Île aux Cochons could be a harbinger of that dire future, researchers fear, and perhaps of declines at other penguin colonies as well. But after their frenzied 5 days on the island, the scientists are resigned to monitoring its birds from afar, knowing that the authorities aren't likely to approve another expedition any time soon. The only glimpses of the penguins' fate will come from occasional helicopter flights over the island and, when clouds cooperate, images snapped by satellites orbiting far above. ■



Transantarctic Mountains along the edge of the Ross Ice Shelf from a ROSETTA-ice imaging flight

REVIEW

History, mass loss, structure, and dynamic behavior of the Antarctic Ice Sheet

Robin E. Bell^{1,*} and Helene Seroussi²

Antarctica contains most of Earth's fresh water stored in two large ice sheets. The more stable East Antarctic Ice Sheet is larger and older, rests on higher topography, and hides entire mountain ranges and ancient lakes. The less stable West Antarctic Ice Sheet is smaller and younger and was formed on what was once a shallow sea. Recent observations made with several independent satellite measurements demonstrate that several regions of Antarctica are losing mass, flowing faster, and retreating where ice is exposed to warm ocean waters. The Antarctic contribution to sea level rise has reached ~8 millimeters since 1992. In the future, if warming ocean waters and increased surface meltwater trigger faster ice flow, sea level rise will accelerate.

Two hundred years ago, an expedition led by Fabian Gottlieb von Bellingshausen and Mikhail Lazarev discovered mainland Antarctica, the most remote and inhospitable continent. Today, Antarctic is an ice-covered continent where change is emblematic of the impacts humans have on the global climate. Hidden beneath the ice sheets are a rich diversity of terrains and hydrologic systems of mountains, lakes, and dynamic subglacial water networks (1, 2). The changes we are now witnessing (3–5) are concentrated in the low-elevation regions as well as the Antarctic Peninsula, the furthest north part of the continent. Evidence of change comes from satellite measurements of ice mass, velocity (4), and elevations (3). Large floating ice shelves have disintegrated (6), and the location where the ice goes afloat is moving inland (7). Future vulnerabilities arise from interactions with the

warming ocean, melting of the ice surface, and the disappearance of the ice shelves. Looking forward, coastal communities around the globe need to know how much sea level will rise, and how much Antarctica will change is one of the greatest unknowns. Both improved models and observations are essential to improve the scientific community's response to the question of how much sea level will rise over the coming decades and centuries.

Origin and history

Over the past 100 million years or more, Antarctica shifted from a green tree-covered continent (8) to a continent encased in ice as it became tectonically isolated while the global climate cooled (9). The tectonic isolation of the continent began more than 200 million years ago, when it was at the center of the Gondwana supercontinent with a climate similar to that of modern New Zealand. The supercontinent breakup occurred slowly, first with Africa (170 million years ago), then India (145 million years ago), and last, Australia (90 million years ago), shifting away from what today is Antarctica

as the Southern Ocean began to form. The final step occurred 34 million years ago as the Drake Passage (10), between South America and the Antarctic Peninsula, and the Tasmanian Gateway, south of Australia, opened. The global oceans thereby were effectively linked. The Antarctic Circumpolar Current, the strongest ocean current on the planet, began to circulate around the continent, and Antarctica was isolated. As this tectonic isolation occurred, global temperatures and the concentration of atmospheric carbon dioxide (CO₂) began dropping steadily (Fig. 1) (11). A period of gradual climate cooling was marked by a sudden cooling of the ocean temperatures together with a decline in CO₂ 34 million years ago (at the Eocene-Oligocene boundary). The dropping temperatures and CO₂, together with the tectonic isolation, induced the first ice formation on the continent's high elevations.

The morphology of the Antarctic continent determined how it became glaciated. The continent consists of the cratonic shield of East Antarctica, the thin low-lying West Antarctic rift system, and the Antarctic Peninsula, an elongated ridge produced by convergent tectonics (12). East Antarctica is dominated by two major mountain ranges. In the center of the craton at 80°S, the Gamburtsev Mountains rest at an elevation of 3000 m. The Transantarctic Mountains, with elevations reaching 4500 m above sea level, separate the elevated shield of East Antarctica from the thinned crust of the West Antarctic rift system (13). The early ice in East Antarctica formed on the Gamburtsev and Transantarctic Mountains and in the high elevations of Dronning Maud Land. This early ice cover quickly grew from mountain glaciers into an ice cap (2), slowly eroding the Gamburtsev Mountains. For the next 20 million years, East Antarctica remained glaciated, but little ice persisted in West Antarctica.

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964-8000, USA. ²Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive MS 300-323, Pasadena, CA 91109-8099, USA.

*Corresponding author. Email: robinb@ldeo.columbia.edu

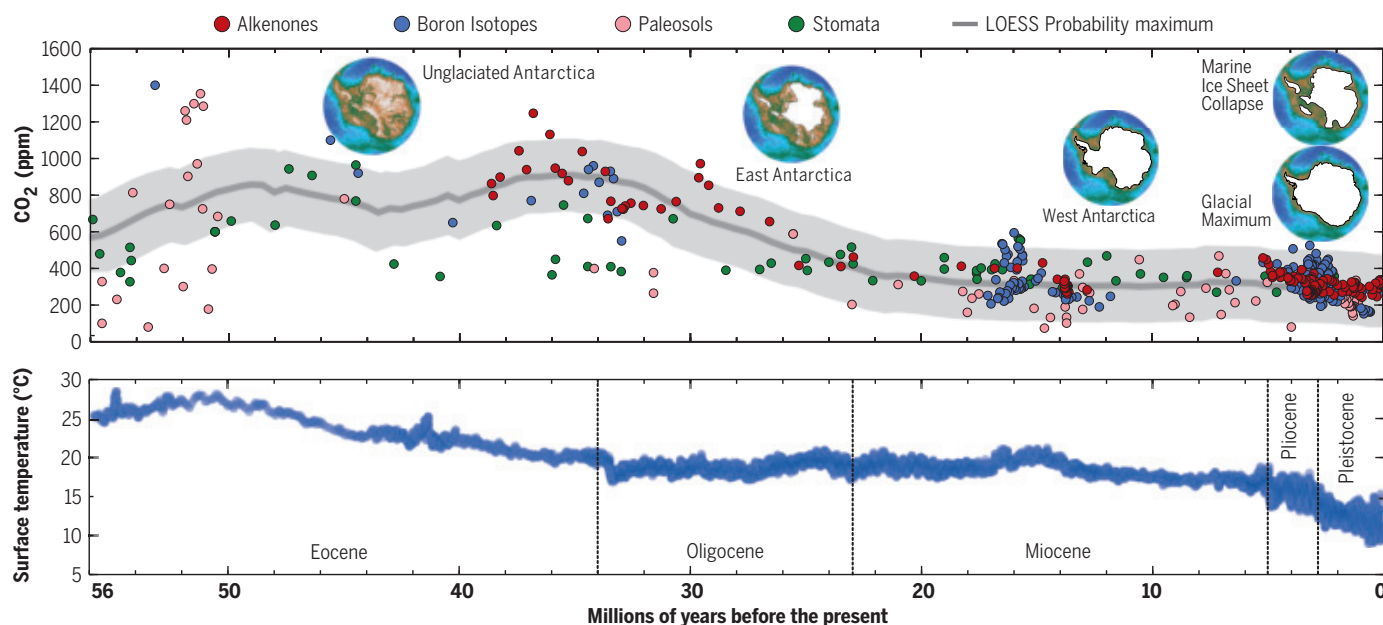


Fig. 1. Development of Antarctic ice together with global CO₂ and ocean surface temperature. (Top) Global CO₂. (Bottom) Ocean surface temperature. Onset of East Antarctic ice occurred 34 million years ago as ocean temperatures and CO₂ dropped (53, 54). Development of West Antarctica marine ice sheet at 14 million years ago began with the next major drop in global temperature.

Two extreme modes of Antarctic ice have occurred since the onset of glaciation in West Antarctica was first covered with ice: ice extending all the way to the continental shelf during cold periods, such as the Last Glacial Maximum 25,000 years ago, and retreat beyond the present extent, with partial collapse of marine portions of Antarctica during some past warm periods (16).

Given that much of West Antarctica is below sea level, establishing a large ice sheet required substantial cooling. When the global climate began to cool again ~14 million years ago, ice started to cover West Antarctica. Again, tectonics may have contributed to the encasing of West Antarctica in ice as uplift of Marie Byrd Land (12) established elevated terrain on which an ice sheet could grow. By 10 million years ago, the West Antarctic Ice Sheet filled the low-lying rift system.

Once fully established, two insolation cycles paced the ice growth and decay: the obliquity (or wobble) of the orbit, with a ~40,000-year period, and the eccentricity, with a 100,000-year period. The background temperature appears to control which of these frequencies is dominant. Currently, the ice core records indicate a 100,000-year frequency to the ice sheet expansion and contraction (14). Before 800,000 years ago, the 40,000-year period dominated ice advance and retreat. During cool periods, the ice in both East and West Antarctica expanded, reaching the edge of the continental shelf margin (Fig. 1). At the ice sheet's maximum extent, sediments transported by the ice were deposited where the ice went afloat, expanding the shallow reaches of the continental shelf each time the ice sheets grew. These low-lying, marine portions of the continent were slower to become ice-covered and are the ice sheet regions most susceptible to collapse when the climate warms. During some past warm periods, the West Antarctic Ice Sheet retreated beyond its current extent and in some instances completely collapsed (15). Low-lying portions of

East Antarctica are likely to have collapsed as well during past warm periods (16). Evidence for the collapse comes from sediments in the Ross Sea and global sea level records where the past sea level cannot be replicated without loss of ice from both East and West Antarctica (Fig. 1).

Structure

There are three major components of an ice sheet system: grounded slow-moving ice, fast-flowing ice streams or outlet glaciers and floating ice shelves (Fig. 2). The grounded slow-moving ice contains the vast majority of the ice (Fig. 2) and is in contact with the underlying rocks. This ice moves very slowly by means of internal deformation at rates on the order of 1 m/year (4) and is up to 4775 m thick (17). The fast flowing ice streams and outlet glaciers are conveyor belts that move the ice toward the ocean, are up to 100 km across, and slide rapidly over the underlying topography at rates of up to 4 km/year (4). Although the surface of glaciers and ice streams are fractured by crevasses, water and till (water-saturated sediments) at their base reduce the basal friction and enable their fast flow (18). The grounded Antarctic holds enough ice to raise sea level rise by 58 m (19).

Ice shelves are expansive, flat, floating bodies of ice attached to the fast flowing ice streams that cover an area of >1.5 million km² all around the continent (20). They do not affect sea level rise directly but slow the flow of the glaciers and the ice streams that feed them by providing some backward stress. The location where the ice in glaciers and ice streams

starts floating over the ocean is called the grounding line. If more ice crosses the grounding line, global sea level will rise. Inland retreat of the grounding line is indicative of a shrinking ice sheet. The large ice shelves Ross, Filchner-Ronne, and Amery provide substantial back-stress and are on average 300 m thick. Snow accumulation on these large ice shelves' surfaces and ice discharge from the grounded ice balance ice loss from ocean melting at their bases and large icebergs calving off their fronts (20).

East Antarctica is the largest ice sheet on the planet, with thicknesses greater than 4600 m (19). In some areas, the bedrock underlying the ice is above sea level, but extensive portions are below sea level. The top of the ice sheet, Dome A, is at 4200 m over the Gamburtsev Mountains (Figs. 2 and 3), whereas the deepest point, carved by erosion during successive advance and retreat of an ice stream, is located more than 3500 m below sea level under Denman Glacier (19). Beneath the thick ice are large lakes—Vostok (27), 90°E, and Sovetskaya (22)—with up to 1000 m of water (Figs. 2 and 3). These systems have been sealed from the atmosphere for ~34 million years since the onset of Antarctic glaciation. The iconic ice core records of temperature and CO₂ come from the deep cores at Vostok [400,000 years (14)], Dome C [800,000 years (23)], and Dome F [720,000 years (24)].

West Antarctica is classified as a marine ice sheet because the topography beneath the ice sheet is largely below sea level. The ice is underlain by marine sediments deposited by glacial processes, rifted sedimentary basins,

crystalline bedrock, and active volcanic terrains (12, 13). The low-viscosity warm lithosphere beneath the region is reflected in the ongoing vertical movement of up to 4 cm/year measured with GPS (25) and caused by ice sheet thinning. The basal hydrology beneath West Antarctica is complex, with hundreds of pockets of water that fill and drain on a decadal time scale producing meter-scale surface elevation changes (1). Two sides of the West Antarctic Ice Sheet are buttressed by the large Ross and Ronne-Filchner ice shelves, whereas the Amundsen Sea sector drains into small ice shelves (Fig. 4) (26). The ice sampled at the base of West Antarctica is 68,000 years (Fig. 2) (27).

The Antarctic Peninsula is the northernmost and warmest region of Antarctica. More than 500 glaciers drain ice from the central plateau. Along the eastern side, facing the Weddell Sea, the Peninsula glaciers feed the Larsen ice shelves (Fig. 4).

Evidence for change

The evidence for recent changes of Antarctic ice is quantified by three independent measurements primarily derived from satellite and airborne systems: decreasing mass from gravity, dropping surface elevation, and increased surface velocities (Fig. 4).

Changes in ice mass are measured from space with the pair of GRACE (Gravity Recovery and Climate Experiment) satellites, which capture changes in the gravity field experienced by each spacecraft as they orbit Earth together (28). The original pair of satellites resolved monthly changes in Antarctic ice mass from 2002 to 2017 (Fig. 4), and a new pair of satellites, GRACE Follow-On, was launched in 2018 to continue the record of the Earth and Antarctic mass changes. The observed changes must be corrected for modeled changes in motion of the solid Earth to include the crust and the mantle rebound owing to past and ongoing ice mass changes. Determining both the Earth structure and the history of past ice sheet changes are the greatest challenges in separating the observed mass changes into the solid Earth component and the changes in ice mass. The GRACE data (28) show mass loss in West Antarctica, focused in the Amundsen and Bellingshausen Sea sectors, and mass gain in some regions of East Antarctica and along Kamb Ice Stream (Fig. 4).

Lowering of the surface elevation has been measured with altimeters from space and aircraft in the same regions where mass loss is observed. Both radar (Cryosat and European Remote Sensing satellites) and laser altimeters [Ice, Cloud, and land Elevation Satellite 1 (ICESat1) and ICESat2 satellites, Operation IceBridge airborne] are used to measure ice surface elevation. Laser observations are impeded by cloud cover, whereas radar measurements penetrate into the upper portions of the snowpack, introducing some ambiguity. Dropping elevation over the

25-year altimetry period is pronounced in the Amundsen and Bellingshausen Sea sectors of West Antarctica and Wilkes Land in East Antarctica (3). Pine Island, Thwaites, and Smith-Pope-Kohler Glaciers experienced the greatest elevation drop over this period, with changes of up to 9 m/year (Fig. 4). Supporting the observed GRACE mass gain, some margins of East Antarctica are increasing in elevation because of increased snowfall (Fig. 4) (29). Along the Siple Coast, the interior of Kamb Ice Stream has thickened at a rate of ~0.65 m/year over the past 20 years because of its stagnation (Fig. 4) (3). Velocity measurements based on interferometric synthetic aperture radar (SAR) measurements and speckle tracking (4) allow the flow of large regions to be observed with accuracies of several tens of meters per year. The changes in ice sheet velocity are striking in the Peninsula, where a substantial acceleration of glaciers feeding the Larsen ice shelves was observed after their collapse, as well as in

the Amundsen Sea sector. In this region, Pine Island Glacier's velocity doubled from the 1990s to the 2010s (Fig. 4), while its grounding line position, accurately estimated by differential interferometric SAR, retreated by more than 30 km (7). The velocity observations are used to calculate the flux of ice discharge into the ocean and, combined with modeled surface accumulation, to estimate the ice mass gain or loss for the different catchment basins. Through an international collaboration, the scientific community has demonstrated the robust agreement between these three different methods and highlighted the ongoing changes of the Antarctic Ice Sheet (5).

Between 1950 and 2000, the average air temperature in the Peninsula increased by 4°C (30). During this warming period, the Larsen A and B ice shelves collapsed in 1995 and 2002, respectively (Fig. 4). The glaciers feeding the Larsen B Ice Shelf sped up after the loss of the backward stress or buttressing (6, 31). Before

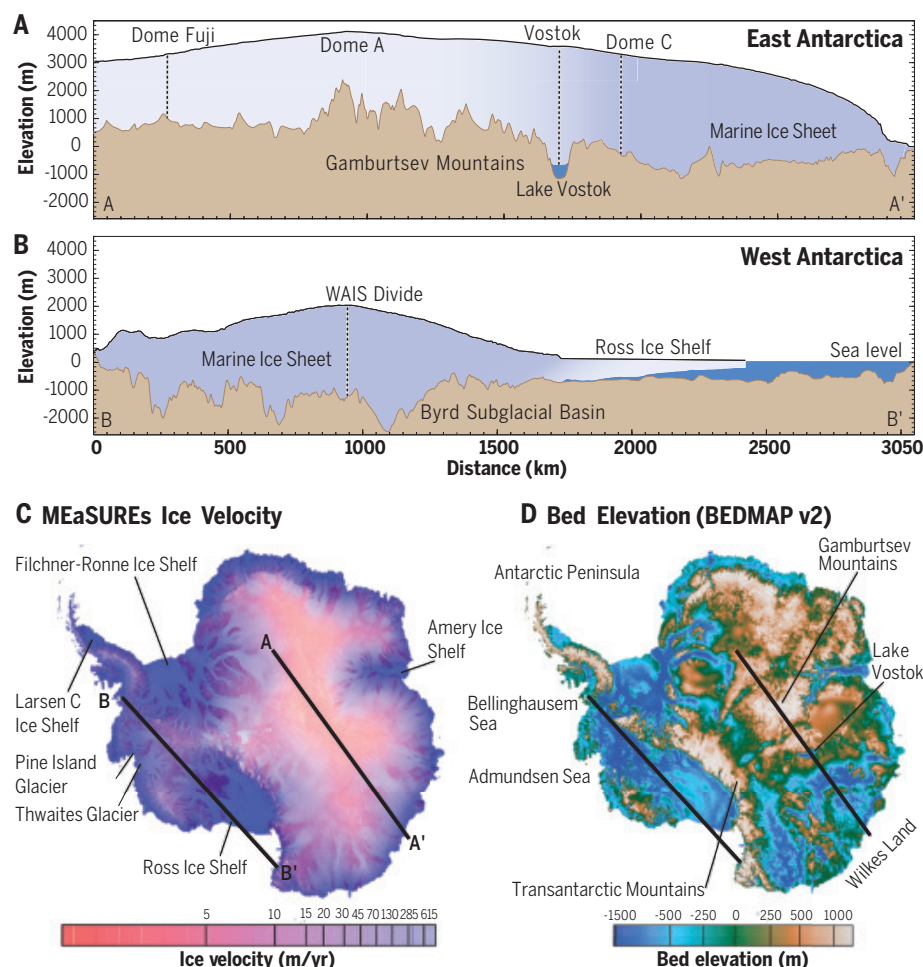


Fig. 2. Structure of the Antarctic Ice Sheet. (A) East and (B) West Antarctic cross section profiles. (C and D) Location of profiles are shown on (C) surface velocity and (D) subglacial bed topography (17). West Antarctica maximum elevation of 2200 m is nestled in the deep Byrd Subglacial Basin, with depths 2500 m below sea level. The East Antarctic Ice Sheet nucleated on the high Gamburtsev Mountains, with the maximum elevation at Dome A reaching 4200 m, covers deep subglacial lakes such as Lake Vostok and has portions that are marine. The portions of the ice sheets with bedrock elevation below sea level are shaded blue.

the Larsen B collapse, the surface of the ice shelf surface was covered by lakes, indicating that warming air temperatures and surface meltwater can destabilize ice shelves, leading to faster flow of Antarctic ice into the global oceans and highlighting the protecting role of ice shelves (32).

Vulnerabilities

These remote sensing observations allow scientists to observe ice sheet changes and decipher the causes of such changes. Both the ocean surrounding Antarctica and the atmosphere, especially in the Peninsula region, have warmed over the 25-year observational record of ice change (33, 34). Antarctica is losing most of its mass through increased ice flow of the outlet glaciers and ice streams. This contrasts with the Greenland Ice Sheet, where half of the loss is due to faster ice flow and half is due to increased melting of the ice sheet surface (35). Surface melt is not yet a major contributor to ice loss in Antarctica, and global climate models suggest that an increase in snowfall in East Antarctica could partially offset the dynamic

mass loss (36). Although these changes have been ongoing for the past three decades, more rapid and dramatic mass loss cannot be excluded. The marine portions of the ice sheet with subglacial topography that deepens inland and glaciers with thick marine terminating fronts are prone to instabilities (37, 38).

Although the surface waters surrounding Antarctica are cold, the underlying waters of the Circumpolar Deep Water are warmer and can influence the ice sheet when they reach the ice shelves and grounding lines, where the ice becomes afloat. The concentration of changes in West Antarctica points to the dominant role the warming ocean plays in recently observed change (39, 40). At the base of the outlet glaciers in the Amundsen Sea, the topography beneath the ice either rises inland or drops. A bed topography dropping inland with ice getting thicker is referred to as a reverse slope. This reverse slope for a marine ice sheet has long been at the core of a concept called the marine ice sheet instability (37). As a glacier retreats across a reverse slope, glacier retreat means thicker ice at the grounding line, and therefore, more ice

is leaving the ice sheet, while the region that accumulates snow is reduced. The ice sheet is out of balance. The greater flux of ice results in thinning and additional retreat until a region with an inland rising slope is encountered to stabilize the grounding line (38). In addition to this, ice dynamics imbalance on a reverse slope; the thicker ice at the grounding line means more ice is exposed to warming ocean waters (40). A perturbation can nudge an outlet glacier off a stable point into a region with a reverse slope and have consequences for decades (39, 41). An extreme El Nino event in the 1940s appears to have triggered the grounding line retreat still ongoing in the Pine Island catchment in West Antarctica (42).

Another trigger for rapid and sustained increased ice flux is the collapse of buttressing ice shelves. This concept was widely debated in the science community until the acceleration of the ice flow in the glaciers feeding the Larsen B ice shelf after its collapse in 2003 was observed (6, 31). Shortly before the collapse, this ice shelf surface was covered with lakes, leading to the hypothesis that hydrofracture and loading from lakes can damage an ice shelf sufficiently to induce a catastrophic collapse (32). This mechanism has been incorporated into some ice sheet models (41) but assumes that meltwater is stationary and that little water is transported across an ice shelf. It is now clear that surface water can flow from the grounded ice onto ice shelves (43) and coalesce into rivers atop the ice surface that end as waterfalls at their front (44). Hydrology could therefore have a stabilizing impact on ice sheet mass balance as the distribution of meltwater increases.

How large and thick marine ice sheets behave after the complete removal of ice shelf buttressing has not yet been witnessed. The question of how quickly ice shelves can collapse and how the glaciers feeding them respond remains open (45).

Unknowns and future directions

Increasingly, communities around the globe are asking how much the sea level will rise in the coming decades. Although we now know that the answer for each community must incorporate knowledge of local processes, such as isostatic uplift from unloading of ice at the end of the last glacial period and subsidence owing to sediment compaction, changing ocean volume from Antarctic mass loss remains one of the largest contributors to communities' unknown future. Gaps in our fundamental knowledge of the bathymetry close to the ice sheet and in regions covered by sea ice and ice shelves, the temperature of the deep water masses, the fate of surface meltwater, and the basal conditions beneath the ice sheets introduce limits into our ability to project the future. It is essential that we refine our projections through

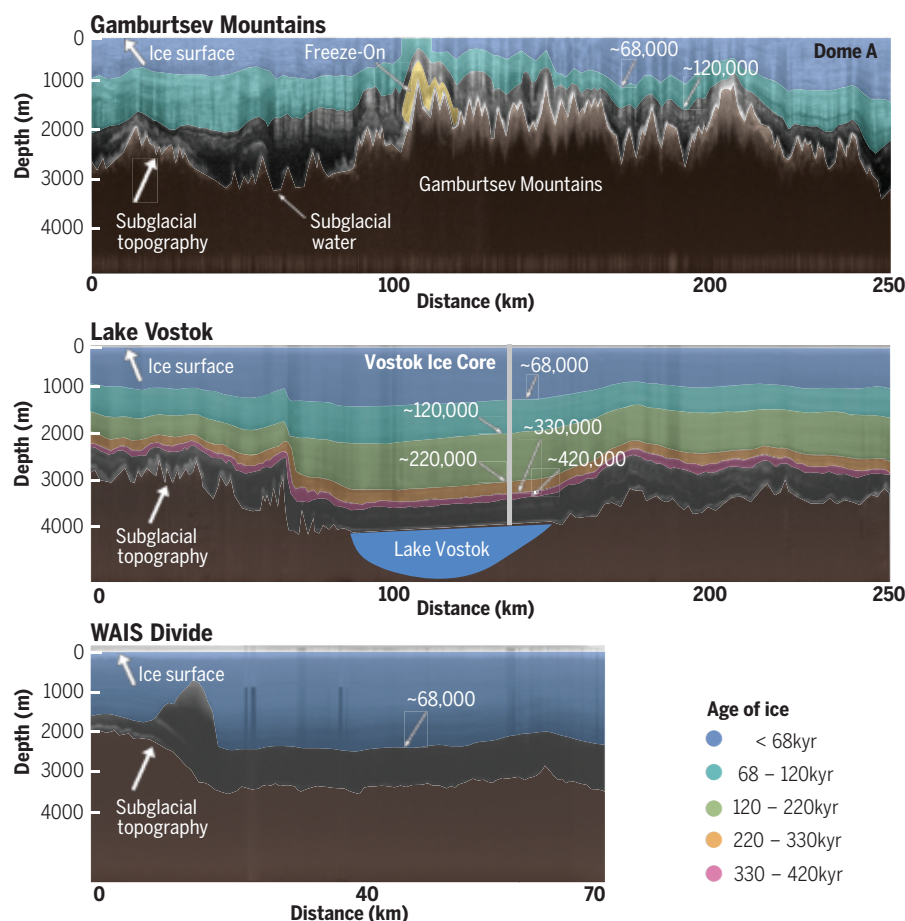


Fig. 3. Radar cross sections over the Gamburtsev Mountains, Lake Vostok, and West Antarctica.

(Top) Gamburtsev Mountains. (Middle) Lake Vostok. (Bottom) West Antarctica. Location of profiles are approximately along the profiles shown in Fig. 2. Radar layers indicate ice stratigraphy. Ice is deformed as it flows over mountains, but layers remain flat as the ice flows over Lake Vostok, the location of the first deep ice core. The color shading highlights the age of the ice sheets. Basal freeze-on is observed in the Gamburtsev Mountain profile.

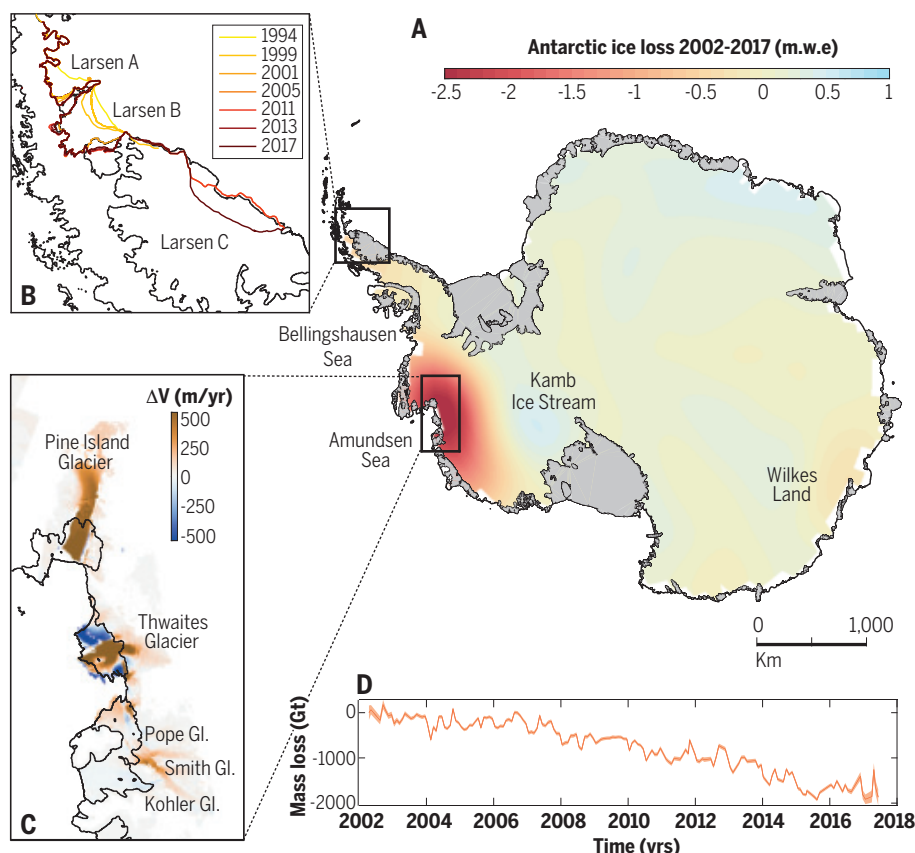


Fig. 4. Evolution of the Antarctic Ice Sheet over the past two decades. (A) Spatial ice mass loss (m.w.e.) estimated from GRACE-collected data over the 2002–2017 period (28). Gray areas show the extension of the floating ice shelves that do not contribute directly to sea level rise. (B) Ice front retreat in the Antarctic Peninsula for the Larsen A, B, and C ice shelves between 1995 and 2017 (6, 31). (C) Change in ice velocity between 2005 and 2017 (meters per year) for glaciers in the Amundsen Sea Sector (4). Black lines represent the ice front and grounding lines. (D) Time series of mass loss (gigatons) and associated uncertainties estimated from GRACE-collected data (28).

expanded observational efforts and improved ice sheet models.

Protecting individual cities with walls and barriers only protects those living behind the protection. An ice sheet–based solution might be more equitable. Possible approaches to slow the flow of ice and reduce the future sea level rise are being considered. Early proposals include snowblowers depositing snow in the interior of Antarctica, building protective berms or curtains to isolate the ice from warming ocean waters or cooling the margins of fast flowing glaciers (46–48). However, serious concerns about the efficacy and cost-benefit of such solutions remain. Looking at global solutions to the changing ice is essential for humanity's evolving coastlines but should be thoroughly investigated because unintended consequences and unknown feedbacks are likely (49).

Conclusions

Over the 200 years since Antarctica was first spotted, our knowledge of the continent has shifted from the notion of a stagnant piece of ice to a constantly evolving continent interacting with the ocean around, the atmosphere

above, and the solid Earth under it and affected by human activities. Advancing our knowledge of the basic history and fundamental processes that control the ice sheet evolution is crucial to future generations. This knowledge will improve predictive capabilities of Antarctica's evolution and help better inform coastal communities worldwide.

REFERENCES AND NOTES

- H. A. Fricker, T. Scambos, R. Bindshadler, L. Padman, *Science* **315**, 1544–1548 (2007).
- T. T. Creyts et al., *Geophys. Res. Lett.* **41**, 8114–8122 (2014).
- H. D. Pritchard, R. J. Arthern, D. G. Vaughan, L. A. Edwards, *Nature* **461**, 971–975 (2009).
- J. Mougnot, E. Rignot, B. Scheuchl, R. Millan, *Remote Sens.* **9**, 364 (2017).
- IMBIE team, *Nature* **558**, 219–222 (2018).
- J. Wuite et al., *Cryosphere* **9**, 957–969 (2015).
- E. Rignot, J. Mougnot, M. Morlighem, H. Seroussi, B. Scheuchl, *Geophys. Res. Lett.* **41**, 3502–3509 (2014).
- J. Francis et al., 100 million years of Antarctic climate evolution: Evidence from fossil plants, in *Antarctica: A Keystone in a Changing World* (National Academies Press, 2008), pp. 19–28.
- R. M. DeConto, D. Pollard, *Nature* **421**, 245–249 (2003).
- R. Livermore, C. D. Hillenbrand, M. Meredith, G. Eagles, *Geochim. Geophys. Geosyst.* **8**, Q01005 (2007).
- C. Escutia et al., *Oceanography* **32**, 32–46 (2019).
- T. A. Jordan, T. R. Riley, C. S. Siddoway, *Nat. Rev. Earth Environ.* **1**, 117–132 (2020).
- K. Tinto et al., *Nat. Geosci.* **12**, 441–449 (2019).
- J.-R. Petit et al., *Nature* **399**, 429–436 (1999).

- T. Naish et al., *Nature* **458**, 322–328 (2009).
- R. H. Levy et al., *Nat. Geosci.* **12**, 132–137 (2019).
- P. Fretwell et al., *Cryosphere* **7**, 375–393 (2012).
- R. B. Alley, D. D. Blankenship, C. R. Bentley, S. T. Rooney, *Nature* **322**, 57–59 (1986).
- M. Morlighem et al., *Nat. Geosci.* **13**, 132–137 (2019).
- E. Rignot, S. Jacobs, J. Mougnot, B. Scheuchl, *Science* **341**, 266–270 (2013).
- A. P. Kapitsa, J. K. Ridley, G. de Q. Robin, M. J. Siegert, I. A. Zotikov, *Nature* **381**, 684–686 (1996).
- R. E. Bell, M. Studinger, M. A. Fahnestock, C. A. Shuman, *Geophys. Res. Lett.* **33**, L02504 (2006).
- J. Jouzel et al., *Science* **317**, 793–796 (2007).
- K. Kawamura et al., *Nature* **448**, 912–916 (2007).
- V. R. Barletta et al., *Science* **360**, 1335–1339 (2018).
- E. Rignot, J. Mougnot, B. Scheuchl, *Geophys. Res. Lett.* **38**, L10504 (2011).
- WAIS Divide Project Members, *Nature* **520**, 661–665 (2015).
- M. M. Watkins, D. N. Wiese, D. N. Yuan, C. Boening, F. W. Landerer, *J. Geophys. Res. Solid Earth* **120**, 2648–2671 (2015).
- C. Boening, M. Lebeck, F. Landerer, G. Stephens, *Geophys. Res. Lett.* **39**, 21 (2012).
- J. Turner et al., *Nature* **535**, 411–415 (2016).
- H. Rott, F. Müller, T. Nagler, D. Floricioiu, *Cryosphere* **5**, 125–134 (2011).
- A. F. Banwell, D. R. MacAyeal, O. V. Sergienko, *Geophys. Res. Lett.* **40**, 5872–5876 (2013).
- S. S. Jacobs, A. Jenkins, C. F. Giulivi, P. Dutrieux, *Nat. Geosci.* **4**, 519–523 (2011).
- L. D. Trusel et al., *Nat. Geosci.* **8**, 927–932 (2015).
- J. Mougnot et al., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9239–9244 (2019).
- H. Seroussi et al., *The Cryosphere* **2020**, 1–54 (2020).
- J. Weertman, *J. Glaciol.* **13**, 3–11 (1974).
- C. Schoof, *J. Geophys. Res.* **112**, F03S28 (2007).
- H. Seroussi et al., *Geophys. Res. Lett.* **44**, 6191–6199 (2017).
- P. Dutrieux et al., *Science* **343**, 174–178 (2014).
- R. M. DeConto, D. Pollard, *Nature* **531**, 591–597 (2016).
- J. A. Smith et al., *Nature* **541**, 77–80 (2017).
- J. Kingslake, J. C. Ely, I. Das, R. E. Bell, *Nature* **544**, 349–352 (2017).
- R. E. Bell et al., *Nature* **544**, 344–348 (2017).
- A. A. Robel, A. F. Banwell, *Geophys. Res. Lett.* **46**, 12092–12100 (2019).
- M. J. Wolovick, J. C. Moore, *Cryosphere* **12**, 2955–2967 (2018).
- J. Feldmann, A. Levermann, M. Mengel, *Sci. Adv.* **5**, eaaw4132 (2019).
- K. Frieler, M. Mengel, A. Levermann, *Earth Syst. Dynam.* **7**, 203–210 (2016).
- O. Gürses, V. Kolatschek, Q. Wang, C. Rodehacke, *Cryosphere* **13**, 2317–2324 (2019).
- D. N. Wiese, D.-N. Yuan, C. Boening, F. W. Landerer, M. M. Watkins, Antarctica mass variability time series version 1 from JPL GRACE Mascon CRI Filtered, version 1 (Physical Oceanography Distributed Active Archive Center, 2017); <https://doi.org/10.5067/TEMSC-ANTS1>.
- J. Mougnot, B. Scheuchl, E. Rignot, MEASURES annual Antarctic ice velocity maps 2005–2017, version 1 (NASA National Snow and Ice Data Center Distributed Active Archive Center, 2017); <https://doi.org/10.5067/9T4EPQXTJYW9>.
- H. Rott et al., *Cryosphere* **12**, 1273–1291 (2018).
- G. L. Foster, D. L. Royer, D. J. Lunt, *Nat. Commun.* **8**, 14845 (2017).
- J. Hansen, M. Sato, G. Russell, P. Kharecha, *Philos. Trans. A Math. Phys. Eng. Sci.* **371**, 20120294 (2013).

ACKNOWLEDGMENTS

A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. GRACE data comes from (50), accessed 28 January 2020. Antarctic velocity data comes from (51), accessed 28 January 2020. Ice front positions for Larsen A and B comes from (6, 52). Cryo portal Enveo snow, glaciers, and ice sheet products and services dataset was accessed 6 February 2020. The radar data in Fig. 3 is from Gamburtsev Mountains–International Polar Year AGAP project Line 560; http://pods.ideo.columbia.edu/1986/legacyData/AGAP/Data_Level_1/RADAR/DecimatedSAR_images/L560_WholeLineEchogram.jpg; Lake Vostok NASA Icebridge Flight 20131127, data available at <https://data.cresis.ku.edu/data/rds/>; and CRESIS data 20091224_frames 24, 25, and 26. Data are available at the same web server. B. Hönisch provided guidance on the paleoclimate data in Fig. 1. I. Cordero and C. Dieck Locke assisted with the figure production. K. Tinto, R. Constantino, B. Keisling, and C. Siddoway provided important feedback on the manuscript. Funding from Lamont Doherty Earth Observatory of Columbia University and the Old York Foundation supported this work. **Competing interests:** None declared.

10.1126/science.aaz5489



Cliffs of the Ross Ice Shelf, Antarctica

REVIEW

The Southern Ocean and its interaction with the Antarctic Ice Sheet

David M. Holland^{1,2*}, Keith W. Nicholls³, Aurora Basinski^{1,2}

The Southern Ocean exerts a major influence on the mass balance of the Antarctic Ice Sheet, either indirectly, by its influence on air temperatures and winds, or directly, mostly through its effects on ice shelves. How much melting the ocean causes depends on the temperature of the water, which in turn is controlled by the combination of the thermal structure of the surrounding ocean and local ocean circulation, which in turn is determined largely by winds and bathymetry. As climate warms and atmospheric circulation changes, there will be follow-on changes in the ocean circulation and temperature. These consequences will affect the pace of mass loss of the Antarctic Ice Sheet.

Ice shelves surround much of the continent of Antarctica. They begin where the Antarctic Ice Sheet separates from the underlying ocean floor in the southernmost reaches of the Southern Ocean. Ice shelves affect the ice sheet in a number of ways, one important one being that they can impede the flow of inland ice into the ocean and thereby slow sea level rise. In places, cavities beneath ice shelves hold relatively cold waters, helping to keep such shelves intact; other cavities contain warmer waters, threatening the existence of those shelves. Source waters for ice shelf cavities originate in the Southern Ocean, and their distribution is largely controlled by circumpolar wind patterns. Winds, therefore, dominate the interaction between the Southern Ocean and the Antarctic Ice Sheet. Future changes in wind patterns will be the principal driver of ice sheet contribution to sea level change. Understand-

ing how winds and ocean circulation work to control the interaction between ice shelves and the surrounding waters is essential for predicting the future of the Antarctic Ice Sheet.

Ocean–ice-sheet interaction

The direct interaction of the Southern Ocean (SO) with the Antarctic Ice Sheet (AIS) takes place largely through ice shelves. Ice shelves are the floating part of the AIS; they are formed along the coast in many locations, fed by the inland ice flowing under the action of gravity. Water is a unique substance in that its solid phase is less dense than its liquid, allowing the ice shelves to float over the surface of the ocean. On the underside of an ice shelf, interaction between the waters of the SO and the AIS leads to both melting and formation of ice and modification of the water masses in contact with the ice base. The melting itself is in part driven by another unique physical aspect of water in its solid phase: As pressure increases, the melting point drops, which means that the deeper the ice base, the greater the potential for it to be melted by ocean waters. Although the eastern portion of the AIS is

largely grounded on a bed that is above current-day sea level, the opposite is the case for much of West Antarctica, meaning that a thinning in the ice could result in it going afloat, thereby leaving it vulnerable to changes in ocean temperature (1). The study of the interaction of the SO with the AIS is important, as the future of the AIS and global sea level are intimately tied up with the fate of the ice shelves, because the shelves serve to hold back and buttress vast amounts of inland ice (3), controlling its flow into the ocean [see Bell and Seroussi (3) in this issue].

The SO, which surrounds the AIS, is connected to the World Ocean through the Global Conveyor Belt (4), a planetary-scale circulation that imports warm water to the SO, modifies it in part by interaction with ice, and exports a deeper, cooler, and fresher water mass back to the World Ocean (5). This interaction occurs through the SO's interface with sea ice, icebergs, and ice shelves. The SO is affected by dominant westerly winds that drive an eastward ocean current that encircles the continent: the Antarctic Circumpolar Current. This, the largest current on the planet, is guided by ocean bathymetry and coastal landmass outline, and flows close to the continent in some locations and much further offshore in others. The current remains well offshore, particularly in places where there are large-scale ocean gyres (Fig. 1). Important to the interaction of the SO with the AIS is the existence of a warmer layer of water at depth beneath the surface waters. This Circumpolar Deep Water (CDW), ubiquitous at a depth centered at ~500 m, is denser than the overlying surface water despite being warmer owing to its greater salt content. Perhaps surprising, these waters largely originate at the other end of the planet, in the North Atlantic where Arctic-bound Gulf Stream waters are transformed into a water mass known as North Atlantic Deep Water, which is subsequently

¹Department of Mathematics, New York University, New York, NY 10012, USA. ²New York University Abu Dhabi Institute, Abu Dhabi, United Arab Emirates. ³British Antarctic Survey, Cambridge CB3 0ET, UK.

*Corresponding author. Email: david.holland@nyu.edu

transported at a subsurface depth by the Global Conveyor Belt to the SO where it appears as CDW (4, 5). The CDW is further modified by a seasonal cycle of sea-ice advance and retreat, which by surface area is Earth's largest such cycle. The sea-ice cycle creates a "salt-pump" that raises the salinity of the CDW (6). The surface waters and the CDW are vertically separated by a "thermocline," a region of strong vertical change in temperature.

Flanked by the grounded inland ice of the AIS and the deep offshore CDW of the SO lies the continental shelf, which in Antarctica is overdeepened, having an average seafloor depth of about 600 m, compared with an average of around 100 m elsewhere in the global ocean (7). A second distinct feature of the continental shelf is that it has a retrograde (negative) slope heading inland to the AIS in most places (8), in contrast to most of the rest of the world, where the slope is prograde. This reverse slope means that once an ice sheet grounded below sea level (i.e., a marine ice sheet) begins to advance over such a retrograde slope, a positive feedback begins, leading to further ice sheet growth (9). The opposite occurs in the case of an initial

retreat [see "marine ice sheet instability" discussed by Bell and Seroussi (3) in this issue]. One potential trigger for an initial retreat that can lead to this instability is for more of the warm water in the SO (i.e., CDW) to come into contact with the ice shelves. Ice shelves reside over the southern reaches of the continental shelf, and for CDW to reach the sub-ice-shelf ocean cavities, it must first cross the continental shelf break, the location at which the continental shelf drops away to the deep ocean. To be able to flow onto the continental shelf, the CDW must reside in the water column at or above the depth of the shelf break. Winds and local bathymetry ultimately dictate the degree to which CDW can penetrate onto the continental shelf.

The access that CDW has to the continental shelf is greatly influenced by the winds far offshore, away from the continental shelf, that modify the depth at which CDW flows over the deep ocean. The higher in the water column the CDW sits, the greater the chance that it will cross onto the continental shelf. Over the continental shelf itself, local winds drive the cold surface waters onshore or offshore. To maintain the same water column thickness over the continental shelf, the onshore or offshore motion of the overlying surface waters forces the underlying CDW in the opposite direction. Further complicating this picture is the formation of very salty (and hence dense) shelf waters when sea ice is formed in winter by cold, southerly winds. The sea ice once formed subsequently can be blown far offshore by such winds, allowing further sea-ice formation, and thus further increasing the salinity of the water column. The dense continental shelf waters

that form can block the lighter offshore CDW from getting onto the continental shelf.

In addition to wind influencing the process of whether the CDW crosses the shelf break, the bathymetry of the sea floor also affects the exchange. Because of Earth's rotation, ocean currents generally flow perpendicular to strong gradients in the seafloor depth, such as found along a continental shelf break, and they thus flow along isobaths at the shelf break. Nonetheless, the flow can cross the shelf break in particular locations; for example, at sea floor troughs cut into the continental shelf by past ice sheet advances to the shelf break (10). There is also a correlation between the curvature of the shelf break and the ocean flow crossing the shelf break—that is, places where the shelf break isobaths curve in front of the ocean flow. Momentum advection can force the flow across the shelf break in such locations (11). All these features and processes of wind and bathymetry, taken cumulatively, control the presence or absence of CDW on the continental shelf (12, 13). The present-day distribution of where CDW is, and is not, on the continental shelf is shown in Fig. 1.

The properties and fluxes of the ocean waters entering a sub-ice-shelf cavity are controlled by the processes occurring over the continental shelf. Once within a cavity, the waters continue to flow along the retrograde seafloor toward the grounding zone, the region where the inland grounded ice sheet first goes afloat. When the waters reach the grounding zone, they cause melting, water mass modification, and a return flow out of the cavity. Common to all cavities is that the heat exchange between the ocean and ice is controlled by a complex boundary layer at the interface, which determines the basal melt rate of the ice shelf. The nature of the boundary layer itself depends on the stratification of the ambient water column, the level of turbulence, the strength of tidal currents, and the roughness and slope of the ice base, among other factors. The degree of melting that occurs in the various ice shelf cavities fringing the AIS can be broadly classified into two major types: low-melting in cold-water cavities (Fig. 2A) and high-melting in warm-water cavities (Fig. 2B), with the latter having modified CDW (i.e., CDW that is slightly cooler because it has mixed with another water mass) directly in

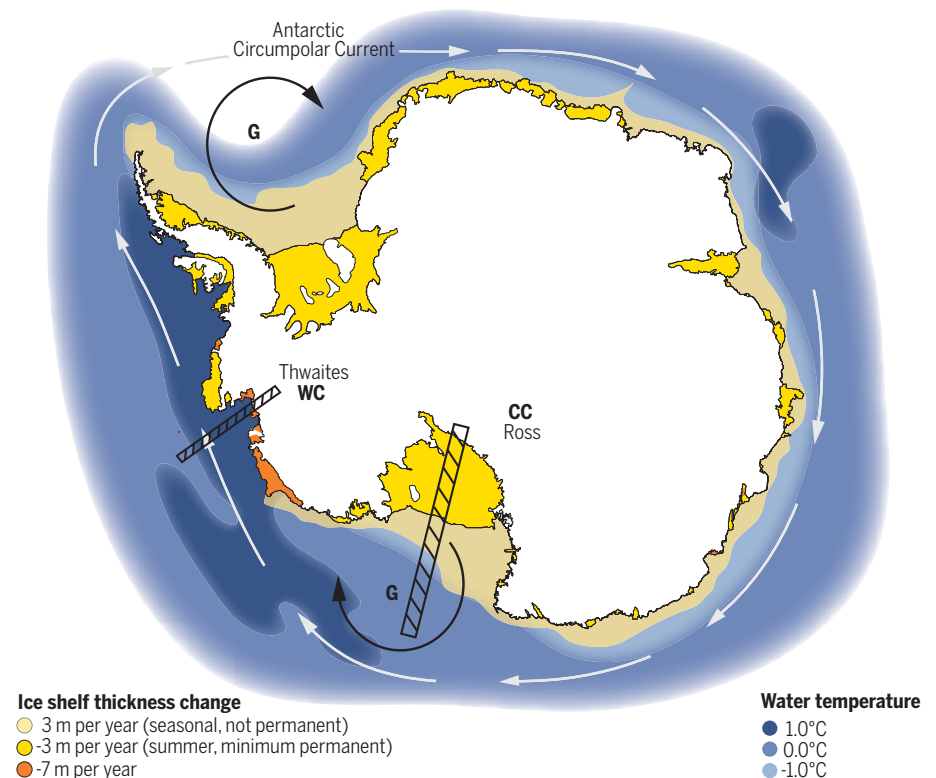


Fig. 1. Oceanographic context. Schematic illustrating the Antarctic Circumpolar Current of the Southern Ocean (arrowed white lines) encircling the Antarctic Ice Sheet. Major ocean gyres are indicated by "G". The Antarctic continent is shown as solid white, bathymetry of the continental shelf and shelf break as light yellow shading, temperature of surrounding ocean waters at 500-m depth as light blue (cold) and dark blue (warm), and ice shelves around the perimeter of the continent as dark yellow (slow basal melting) and orange (fast basal melting). Warm-water cavities occur when CDW comes onto the continental shelf; an example is Thwaites Glacier located along the transect "WC". Cold-water cavities occur where CDW does not come onto the shelf; an example is the Ross Ice Shelf located along the transect "CC".

contact with the boundary layer at the ice base and the former not. Warm-water cavities are found only where the Antarctic Circumpolar Current, which carries the CDW around the SO, is located close to the continental shelf break where CDW can potentially move from offshore onto the continental shelf. By contrast, cold-water cavities are to a great extent protected from CDW by the coastal landmass outline, ocean gyres, and strong off-ice-shelf winds.

For a cold-water cavity, persistent off-ice-shelf wintertime winds cause sea ice to be formed over the continental shelf and transported away from the coast, thereby transforming the continental shelf waters into High Salinity Shelf Water (HSSW). This cold (near surface freezing point) and salty water mass is denser than the CDW that is found offshore beyond the continental shelf break. Additionally, a dynamic feature forms at the shelf break—the Antarctic Slope Front—a geophysical-fluid dynamics consequence of the presence of contrasting water masses (HSSW and CDW) adjacent to one another on either side of a strong change in bathymetry at the continental shelf break. Consequently, the HSSW effectively blocks offshore CDW from getting onto the continental shelf. The dense HSSW, which has a temperature close to the surface freezing point, floods the ice shelf cavity along the retrograde slope from the open continental shelf inland to the grounding zone. As increasing pressure lowers the melting point of ice, the HSSW is above the melting point when it encounters the ice base and therefore has the capacity to cause melting. The water mass that results from the chilling and freshening of

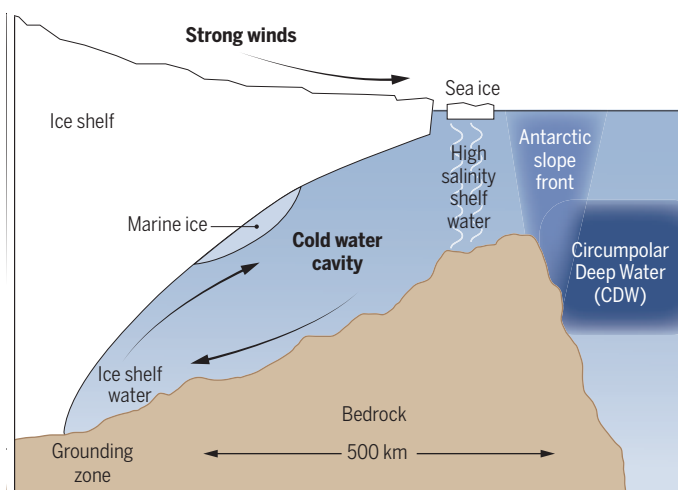
the HSSW, known as Ice Shelf Water (ISW), has a temperature that is below the surface freezing point as a result of its interaction with ice at pressure. The added meltwater renders the ISW overall positively buoyant, and it rises along the ice shelf base, flowing back toward the ice shelf front. At some point, as it rises and the pressure decreases, the in situ freezing point increases above the temperature of the ISW, and so ice forms in the water column, accreting at the ice base to create marine ice. This melt of ice at the grounding zone and redeposition further up along the ice shelf base, and the associated movement of the water in the cavity, is known as an “ice pump circulation” (14).

In the case of a warm-water cavity, the absence of well-organized, off-ice-shelf winds in such a location reduces the production and off-coast transport of sea ice, and no dense HSSW is produced over the continental shelf. This in turn leads to the absence of an Antarctic Slope Front, allowing the offshore CDW to flow onto the continental shelf, forming a thermocline at the interface with the colder surface waters. The Coriolis force in the Southern Hemisphere causes moving fluid to curve to the left. The broad easterlies that blow along the coast of Antarctica therefore induce a southward transport of surface waters toward the coast, which increases the depth of the thermocline, reducing the thickness of CDW on the continental shelf. Once on the continental shelf, the CDW flows down to the grounding zone, primarily along deep sea floor troughs, and comes into contact with the ice base, causing intense melting (15). Despite the resultant meltwater being

cooler than CDW, it is also relatively fresh and thus positively buoyant, and rises along the ice shelf base. This density-driven circulation contributes to the melting, as it results in an overall more vigorous melt-driven circulation with higher turbulence, increasing the transport of heat toward the ice base. In this setting, there is no marine ice formed at the base of the ice shelf, as the waters in the circulation are above the in situ freezing point at all depths. The inflowing CDW has far greater heat content than can be extracted by the basal melting, resulting in the vast majority of the heat content imported to the cavity being re-exported.

Currently, many ice shelves with warm-water cavities are observed from remote sensing to be undergoing rapid change (Fig. 1). Numerical models are the only predictive tool for studying the fate of such ice shelves. However, the present generation of those models demonstrate considerable uncertainty in the future behavior of the ice shelves, suggesting that the rate of retreat can vary greatly depending on the details of how melt occurs in the grounding zone (16). To improve numerical models, there is a pressing need for field observations to study important physical processes occurring in this critical zone, as sketched in Fig. 3 and outlined in Box 1. The change in friction when ice transitions from being grounded to floating is one such process. On the inland side of the region, where the ice is grounded, the ice experiences basal friction with the underlying bed, whereas on the other side of the region, where the ice is freely floating over the ocean cavity, the ice experiences effectively no basal friction. This transition partially regulates the volume flux of ice across

A Cold water cavity



B Warm water cavity

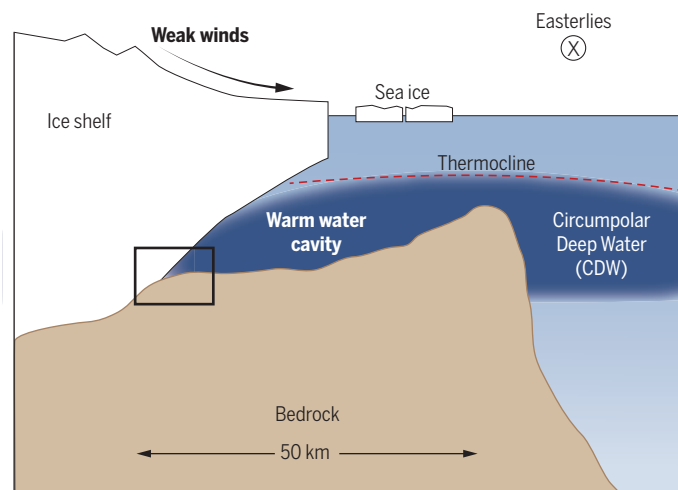


Fig. 2. Interaction of water masses with cold- and warm-water cavities.

(A) A vertical slice illustrating the water masses interacting with a cold-water cavity (see transect CC in Fig. 1). The schematic shows a weak connectivity from (right to left) of offshore warm, circumpolar deep water (CDW) to the cold, salty water residing over the continental shelf, to the water in the ice shelf cavity, to that at the grounding zone, where the ice shelf first goes

afloat. (B) A vertical slice illustrating the water masses interacting with a warm-water cavity (see transect WC in Fig. 1). The schematic shows the CDW on the continental shelf and entering the sub-ice-shelf cavity. Owing to the increased melt rates, the ice shelf itself (and hence the cavity) tends to be an order-of-magnitude shorter than the cold-water case shown in (A). The boxed area is described in Fig. 3.

ILLUSTRATION: N. CARY/SCIENCE

the grounding zone. Additionally, beneath the grounded ice, a subglacial inflow of fresh water into the ocean cavity at the grounding zone can occur, adding positive buoyancy and influencing ocean circulation in the cavity. The floating ice can also be affected by ocean tides that can lift and flex the ice shelf, temporarily moving the position of the grounding zone. Tidal currents also enhance the flow of waters in the cavity, affecting the oceanic exchange of heat at the ice shelf base.

Changes in the SO and its interaction with the AIS

Whereas ice shelf basal melt is controlled by ocean water properties, the latter are dictated by atmospheric processes. Of course, sea ice has a mediating influence on the ocean, but it is the atmosphere that dominates. On the large scale, the general circulation of the atmosphere, when viewed as an east-west average, consists of a sequence of north-south overturning cells, one of which is the Polar Cell of the Southern Hemisphere. The Polar Cell consists of descending air over the AIS but rising air over the SO (see Fig. 4A). The ascending air creates a trough of low pressure over the SO, with westerlies to the north and easterlies to the south, each driven by the pressure trough and modified by the Coriolis force.

The dominant mode of atmospheric variability for these winds in the Southern Hemisphere is the Southern Annular Mode (SAM). It is represented by a normalized index defined as the zonal pressure difference between 40°S and 65°S (17). SAM is then a measure of the strength of the westerlies and is known to have increased, meaning that the westerlies have been moving south and strengthening (18). There is decadal variability in SAM that comes from a teleconnection with the tropics (19–21). Additionally, SAM is strongly positively affected by ozone depletion (22) and greenhouse gases (23), implying an anthropogenic influence.

SAM is a zonally averaged index. However, if we examine the systems surrounding Antarctica without east-west averaging, we can see embedded highs and lows (see Fig. 4A). These highs and lows also exhibit strong decadal variability, largely driven by oceanic and atmospheric changes, originating in the tropics (24–26). These winds, whether viewed as a zonal average (as in the SAM) or with detailed east-west regional variations, influence how much and where CDW comes close to the Antarctic coast (27–29). Where the lows persist, the coastal easterlies are stronger than the average around the coast, forcing more surface water toward the coast by the Coriolis force. The increase in the flow of surface water to the coast decreases the amount of CDW below the surface water, and this has a tendency to lessen the amount of CDW on the continental shelf. Thus, a stronger low results in less CDW on the shelf break than a weaker low. The posi-

tion of the lows also influences the ultimate amount of CDW on the shelf.

The ongoing retreat of the Thwaites Glacier system has directed much interest toward better quantifying the amount of CDW on the Amundsen Sea continental shelf. The Amundsen Sea Low, adjacent to Thwaites Glacier in West Antarctica (Fig. 1), has deepened in recent decades (30) and has influenced the amount of CDW beneath the Thwaites Ice Shelf (Fig. 2B). Ice cores from the West Antarctic Ice Sheet suggest that higher surface temperatures in the middle of the last century were a result of influence from the tropics (31). Additionally, sediment cores taken from the ocean floor in the general area suggest that a retreat has been ongoing in this area since the middle of the last century, further suggesting a tropical tele-

connection (32–35). An analysis based on ensemble simulations from a computer model spanning the past century and looking forward through the next shows that although the winds adjacent to the Thwaites Glacier are dominated by internal climate system variability primarily from the tropics (24, 26), there is a suggestion that a trend due to anthropogenic forcing is emerging (Fig. 4B) (36). These changing winds could have a substantial impact on accelerating the retreat of Thwaites Glacier. The Thwaites Glacier system serves as a specific and important example of how the SO interacts with the AIS and can affect its mass balance, in this case via an ice shelf with a warm-water cavity.

More broadly, the mass balance of the entire AIS is a competition between mass gain through

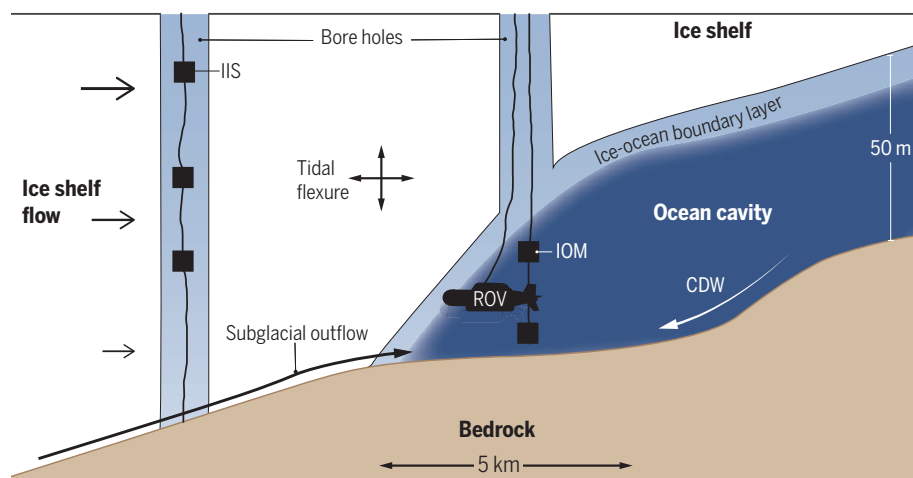


Fig. 3. Schematic showing a zoomed-in vertical slice through a grounding zone (see box at grounding zone in Fig. 2B for spatial context). The grounding zone describes the transition region where the inland ice, under the action of gravity, flowing toward the ocean, goes afloat to form an ice shelf extending out over the ocean, creating a sub-ice-shelf ocean cavity in the process.

Box 1. Observing grounding-zone processes in a warm-ocean cavity.

Physical processes at the grounding zone are not well understood, yet are critical to modeling grounding-zone change. One approach to understanding these physical processes is outlined in Fig. 3. Direct observations can be made via hot-water-drilled boreholes, both on the grounded and floating portions. On the grounded side, a down-hole instrumented ice string (IIS) can include a fiber-optic-based distributed temperature sensing (DTS) system to observe the vertical temperature profile and embedded gauges to measure shear in ice flow. On the floating side, a borehole into the ocean cavity can first allow a brief spatial exploration of the cavity, right up to the exact spot where the ice first ungrounds, using a fiber-optic tethered remotely operated vehicle (ROV) having a suite of instruments such as cameras, sensors for temperature and salinity, current meters, and side-scan sonar. Subsequently, to gain temporal information, an instrumented ocean mooring (IOM) can be placed permanently in the ocean cavity and include both a string of temperature and salinity sensors as well as turbulence gauges in the ice-ocean boundary layer. Innovations in ice and ocean instrumentation have been moving forward at a rapid pace over the past decade (45), allowing these types of integrated observations to be made for the first time. Although instrumentation technology has advanced, making these observations is challenged by the remoteness of the locations and the harshness of the environment (often heavily crevassed at the surface from where field camps need to be established), resulting in the planning and execution of field campaigns taking years to achieve. Such observations are presently being collected in a warm-ocean cavity by the International Thwaites Glacier Collaboration (43), and more will be needed both at Thwaites and elsewhere to better understand grounding-zone change.

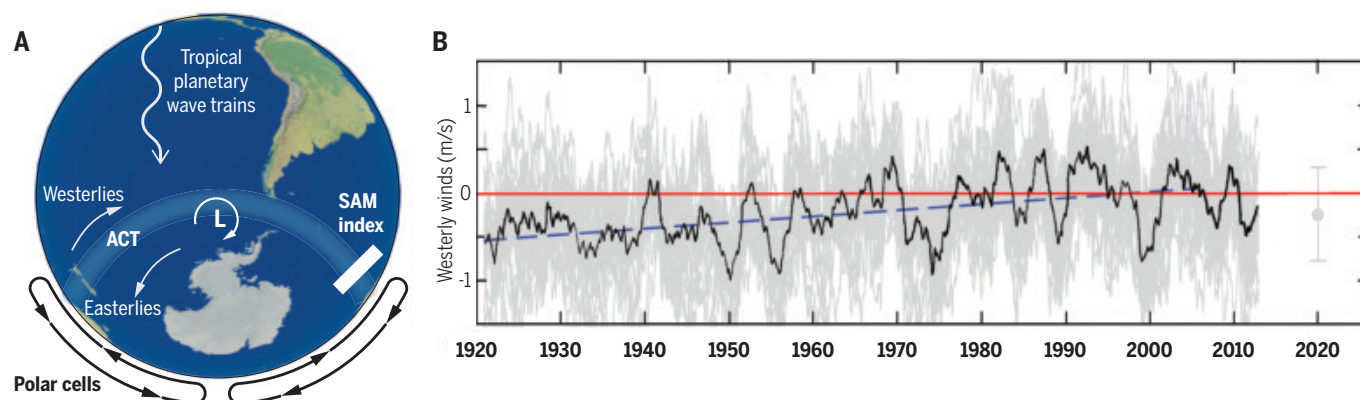


Fig. 4. Atmospheric context wind variability. (A) The approximately east-west oriented Antarctic Circumpolar Trough (ACT) supports a band of westerly winds to the north of the trough, and easterlies to the south. Coriolis force causes the westerlies and easterlies to drive the upper ocean waters to the north and south, respectively, creating a divergence of surface waters all along the ACT. This creates an upwelling of the subsurface CDW, which can cause it to rise above the depth of the continental shelf break in places and allow the possibility of it crossing onto the shelf and interacting with the AIS. Whether or not the CDW gets onto the continental shelf is determined by processes sketched in Fig. 2, A and B. (B) A time series of westerly winds in the Amundsen

Sea adjacent to Thwaites Glacier (36). Positive values indicating westerly winds favor a greater transport of warm CDW into the cavity below Thwaites Ice Shelf, thereby increasing melting and consequently grounding-zone retreat. Negative values indicating easterly winds represent the opposite behavior. The winds were created using a global climate model that simulates large, natural, decadal variability in the climate system, principally originating in the tropics. Superimposed on this variability is a centennial-scale trend in positive wind values that is likely being forced by anthropogenic sources. The study suggests that this upward trend will lead to an increase in warm waters at Thwaites Ice Shelf, and hence increased melting and retreat.

snowfall over the continent and mass loss, the latter approximately equally split between ice-berg calving and ice shelf basal melting (37). Under the current situation of a warming global atmosphere, AIS precipitation is likely to increase, as a warmer atmosphere can hold more moisture and therefore produce greater precipitation (38). At the same time, a warming atmosphere would increase surface melt of the ice sheet, leading to mass loss. Additionally, a warming atmosphere will lead to changes in wind patterns, with unknown outcomes with respect to increasingly warm ocean waters reaching ice shelf cavities and accelerating ice loss. The warm-water cavities are sensitive (39), as they currently have a largely unconstrained flow of warm water onto the continental shelf and into their sub-ice-shelf cavities. The cold-water cavities are currently protected by dense and saline continental shelf waters that block transport of CDW to the ice shelf cavities (Fig. 2A), although this situation could change under a changing climate (40). It should also be noted that the anticipated mass loss processes of surface and basal melt have an inherently faster time scale than that of the atmospheric-driven gain through precipitation. This implies that the mass loss effects may occur faster than they can be offset by precipitation.

Reductions in the extent and thickness of ice shelves and the retreat of their grounding zones associated with the ice-ocean processes occurring in cavities would have major consequences for the inland ice, causing it to flow faster to the grounding zone as the buttressing force of the ice shelves is weakened or lost (41). There is some evidence that the resulting reduction in buttressing produces a near-instantaneous

response in the speedup of the inland ice (42). As the inland ice flows faster toward the ocean and is melted, the ice feeding the ice shelf becomes thinner, leading to overall decreased mass of the AIS. Of all the processes governing this mass and implied sea-level change, perhaps the most important ones are those in the grounding zone, which are also arguably the least well understood at present. This is a key motivation for the ongoing integrated field, remote-sensing, and modeling campaign called the International Thwaites Glacier Collaboration (43). Studies such as this into the retreat of grounding zones are needed to feed into fully coupled global climate models, so that climate models can accurately represent such processes to project future sea level with confidence [see Pattyn (44) in this issue].

REFERENCES AND NOTES

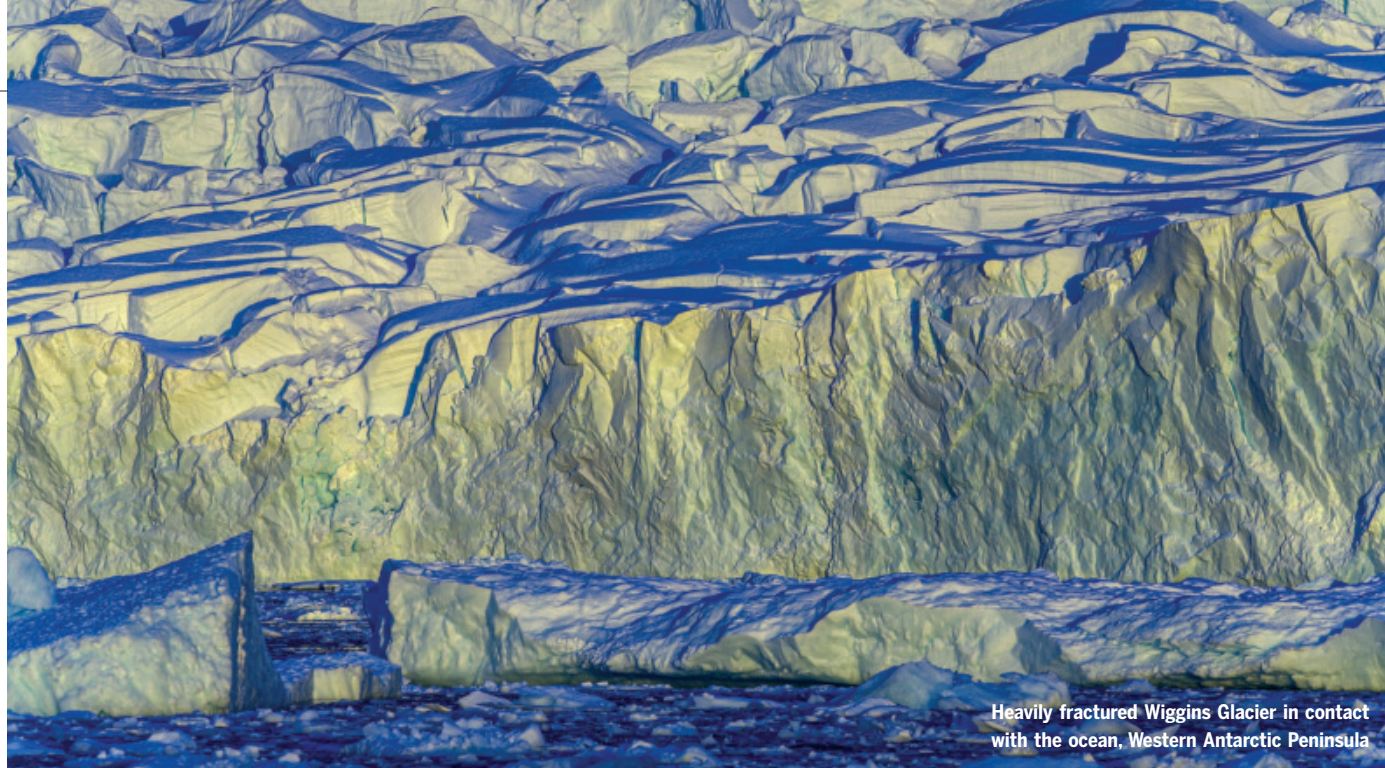
1. E. Rignot, S. S. Jacobs, *Science* **296**, 2020–2023 (2002).
2. R. Thomas *et al.*, *Science* **306**, 255–258 (2004).
3. R. E. Bell, H. Seroussi, *Science* **367**, 1321–1325 (2020).
4. W. Broecker, *Oceanography* **4**, 79–89 (1991).
5. D. Iudicone, S. Speich, G. Madec, B. Blanke, *J. Phys. Oceanogr.* **38**, 1401–1425 (2008).
6. W. S. Broecker, T.-H. Peng, *Global Biogeochem. Cycles* **1**, 251–259 (1987).
7. U. S. ten Brink, W. P. Rogers, R. M. Kirkham, in *Sixth International Symposium on Antarctic Earth Sciences* (Terra Scientific Publishing Company, Ranzan, JP, 1992), pp. 763–771.
8. M. Morlighem *et al.*, *Nat. Geosci.* **13**, 132–137 (2020).
9. J. Weertman, *J. Glaciol.* **13**, 3–11 (1974).
10. F. O. Nitsche, S. S. Jacobs, R. D. Larer, K. Gohl, *Geochem. Geophys. Geosyst.* **8**, Q10009 (2007).
11. M. S. Dinniman, J. M. Klinck, W. O. Smith Jr., *Deep. Res. Part II Top. Stud. Oceanogr.* **50**, 3103–3120 (2003).
12. S. S. Jacobs, H. H. Hellmer, A. Jenkins, *Geophys. Res. Lett.* **23**, 957–960 (1996).
13. A. Jenkins *et al.*, *Oceanography* **29**, 106–117 (2016).
14. E. L. Lewis, R. G. Perkin, *J. Geophys. Res.* **91**, 11756 (1986).
15. A. Khazendar *et al.*, *Nat. Commun.* **7**, 13243 (2016).
16. R. J. Arthern, C. R. Williams, *Geophys. Res. Lett.* **44**, 2352–2359 (2017).

17. G. J. Marshall, *J. Clim.* **16**, 4134–4143 (2003).
18. D. W. J. Thompson, S. Solomon, *Science* **296**, 895–899 (2002).
19. R. L. Fogt, D. H. Bromwich, *J. Clim.* **19**, 979–997 (2006).
20. Q. Ding, E. J. Steig, D. S. Battisti, J. M. Wallace, *J. Clim.* **25**, 6330–6348 (2012).
21. M. L. L'Heureux, D. W. J. Thompson, *J. Clim.* **19**, 276–287 (2006).
22. R. L. Fogt *et al.*, *J. Clim.* **22**, 5346–5365 (2009).
23. J. M. Arblaster, G. A. Meehl, *J. Clim.* **19**, 2896–2905 (2006).
24. T. Lachlan-Cope, W. Connolley, *J. Geophys. Res. Atmos.* **111**, (2006).
25. Q. Ding, E. J. Steig, D. S. Battisti, M. Küttel, *Nat. Geosci.* **4**, 398–403 (2011).
26. X. Li, D. M. Holland, E. P. Gerber, C. Yoo, *Nature* **505**, 538–542 (2014).
27. S. Jacobs, *Phys. Eng. Sci.* **364**, 1657–1681 (2006).
28. P. Dutrieux *et al.*, *Science* **343**, 174–178 (2014).
29. M. Thoma, A. Jenkins, D. Holland, S. Jacobs, *Geophys. Res. Lett.* **35**, L18602 (2008).
30. M. N. Raphael *et al.*, *Bull. Am. Meteorol. Soc.* **97**, 111–121 (2016).
31. D. P. Schneider, E. J. Steig, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12154–12158 (2008).
32. P. Dutrieux *et al.*, *Science* **343**, 174–178 (2014).
33. A. Jenkins *et al.*, *Nat. Geosci.* **11**, 733–738 (2018).
34. F. S. Paolo *et al.*, *Nat. Geosci.* **11**, 121–126 (2018).
35. J. A. Smith *et al.*, *Nature* **541**, 77–80 (2017).
36. P. R. Holland, T. J. Bracegirdle, P. Dutrieux, A. Jenkins, E. J. Steig, *Nat. Geosci.* **12**, 718–724 (2019).
37. E. Rignot, S. Jacobs, J. Mouginot, B. Scheuchl, *Science* **341**, 266–270 (2013).
38. B. Medley, E. R. Thomas, *Nat. Clim. Chang.* **9**, 34–39 (2019).
39. I. Joughin, B. E. Smith, B. Medley, *Science* **344**, 735–738 (2014).
40. H. H. Hellmer, F. Kauker, R. Timmermann, J. Determann, J. Rae, *Nature* **485**, 225–228 (2012).
41. R. B. Alley *et al.*, *Annu. Rev. Earth Planet. Sci.* **43**, 207–231 (2015).
42. G. H. Gudmundsson, F. S. Paolo, S. Adusumilli, H. A. Fricker, *Geophys. Res. Lett.* **46**, 13903–13909 (2019).
43. ITGC Thwaites Glacier; <https://thwaitesglacier.org/>.
44. F. Pattyn, M. Morlighem, *Science* **367**, 1331–1335 (2020).
45. SCAR, “The Southern Ocean Observing System (SOOS) 2012 Report” (2013), (available at <https://www.scar.org/library/policy/antarctic-treaty/atcm-xxvii-and-cep-xvi-2013/2813-atcm36-ip005/file/>).
46. H. D. Pritchard *et al.*, *Nature* **484**, 502–505 (2012).

ACKNOWLEDGMENTS

Funding: D.M.H. and A.B. were supported by NSF grant PLR-1739003 and NYU Abu Dhabi grant no. G1204. K.W.N. is supported by NERC grant NE/S006656/1. **Competing interests:** The authors declare no competing interests. This is ITGC contribution no. 012.

10.1126/science.aaz5491



Heavily fractured Wiggins Glacier in contact with the ocean, Western Antarctic Peninsula

REVIEW

The uncertain future of the Antarctic Ice Sheet

Frank Pattyn^{1*} and Mathieu Morlighem²

The Antarctic Ice Sheet is losing mass at an accelerating pace, and ice loss will likely continue over the coming decades and centuries. Some regions of the ice sheet may reach a tipping point, potentially leading to rates of sea level rise at least an order of magnitude larger than those observed now, owing to strong positive feedbacks in the ice-climate system. How fast and how much Antarctica will contribute to sea level remains uncertain, but multimeter sea level rise is likely for a mean global temperature increase of around 2°C above preindustrial levels on multicentennial time scales, or sooner for unmitigated scenarios.

Major uncertainties in predicting and projecting future sea level rise are due to the contribution of the Antarctic Ice Sheet (1). These uncertainties essentially stem from the fact that some regions of the ice sheet may reach tipping points, defined as (regionally) irreversible mass loss, with a warming climate. The exact timing of when these tipping points might occur remains difficult to assess, allowing for a large divergence in timing of onset and mass loss in model projections. The instability mechanisms responsible for these tipping points are closely related to the shape of the bed under the ice sheet (Fig. 1). The West Antarctic Ice Sheet (WAIS), which has the potential to raise sea level by 5.3 m (2), has its current base grounded well below sea level, and the bed deepens from the periphery of the ice sheet toward the interior (a so-called retrograde bed slope). Marine basins are also present in certain areas of the East Antarctic

Ice Sheet (EAIS) (Fig. 1), which has a far greater sea level contribution potential of 52.2 m (2). Marine ice sheets are in direct contact with the ocean under floating ice shelves around the coast, and changes in ocean circulation or heat content may lead to rapid ice loss on time scales of decades to centuries. The uncertainty in the timing and extent of potential tipping points also stems from our poor knowledge of both drivers of change and mechanisms that operate in the dynamics of marine ice sheets. Despite these shortcomings, multimodel comparisons like Ice Sheet Modeling Intercomparison Project 6 (ISMIP6) allow for a more standardized approach that enables outliers to be more clearly identified. Hence, uncertainties in future projections have since been reduced, and more robust projections of sea level contributions from the Antarctic Ice Sheet are to be expected.

Observations and drivers of dynamical mass change

Recent satellite observations indicate that the contribution of the Antarctic Ice Sheet to sea level rise has considerably increased in recent

years (3). Antarctica has been contributing, on average, 0.15 to 0.46 mm/year to sea level between 1992 and 2017, accelerating to 0.49 to 0.73 mm/year between 2012 and 2017 (4). Most ice loss is concentrated in West Antarctica, where the thinning of floating ice shelves is causing glacier flow to accelerate and grounding lines (the contact between the grounded ice sheet and the ice shelf floating on the ocean) to retreat.

The ice flow acceleration and thinning of Pine Island Glacier, Thwaites Glacier, and nearby glaciers that drain into the Amundsen Sea (Fig. 2), which dominate the mass loss from the WAIS, result from ice shelf thinning and shrinkage and associated grounding-line retreat. This is thought to be a response to a wind-driven increase in the circulation of warm Circumpolar Deep Water (CDW) onto the continental shelf reaching ice shelf cavities and grounding lines (5). The strengthening of the regional westerly winds that have forced warmer waters to the grounding zones is attributed primarily to remote changes occurring in the tropics (6). However, changes in larger-scale circulation owing to the recent stratospheric cooling resulting from ozone depletion and increased concentration of greenhouse gases have also been identified as potential drivers (7). Thwaites Glacier is today undergoing the largest changes of any ice-ocean system in Antarctica (8). This ongoing mass loss will be modulated, but likely not reversed, by variability in the ocean (9).

The EAIS is closer to a balanced state, but this remains poorly constrained in terms of surface mass balance (essentially precipitation-evaporation) and glacial isostatic adjustment (GIA) in response to volume change stemming from the last glacial-interglacial period. Recent

¹Laboratoire de Glaciologie, Université Libre de Bruxelles, Brussels, Belgium. ²Department of Earth System Science, University of California, Irvine, CA, USA.

*Corresponding author. Email: fpattyn@ulb.ac.be

studies reveal that some ice shelves in East Antarctica, once thought to be stable, are also exposed to ocean heat and are experiencing high rates of basal melt (10); hence, the discharge of the EAIS may increase if the atmospheric and oceanic conditions change.

Antarctic surface mass balance derived from reconstructions of ice core records show large but opposing trends across West Antarctica, especially for recent decades, whereas precipitation changes are less pronounced in East Antarctica (11). A key attribute of precipitation

events is the penetration of warm, moist air masses over the ice sheet, which may dominate the annual total precipitation and make such events primarily responsible for most interannual variations in precipitation (12).

Dynamics of the marine ice sheet

The mass balance of the Antarctic Ice Sheet, and therefore its contribution to sea level, is determined by the balance between mass gain and mass loss. The ice sheet gains mass from snowfall on its surface and loses mass primarily

by ocean-induced melting beneath its floating ice shelves along the coast and by calving icebergs that drift away and melt in the ocean. Although the surface mass balance has been relatively stable over the past decades, ice flow in several sectors of the ice sheet has accelerated, thereby increasing ice discharge. The dominant process triggering these large, rapid changes is the loss of ice shelf buttressing. This is initiated by changes in ocean circulation and, to a lesser extent, atmospheric drivers that control summer surface-melt rates (13, 14). In particular, the warmer waters of the CDW move toward the ice fronts and ice shelf grounding zones along troughs in the bathymetry, causing increased melting at the ice-ocean interface. This process thins the ice shelves, reducing drag along their sides and at local pinning points on seafloor highs, which in turn reduces the buttressing, that is, the resistive stress that the ice shelves exert on the grounded ice (8). Thinning ice shelves lead to faster grounded-ice flow, which in turn leads to further thinning, causing previously grounded ice to float as the grounding zone retreats farther inland. This process can be particularly fast and unstable along retrograde slopes (i.e., the bed deepens inland), because more ice crossing the grounding zone and a smaller accumulation area (15, 16) create a positive-feedback process known as the marine ice sheet instability (MISI; Fig. 3). The process may halt when the bedrock rises upward—that is, when a prograde bed slope or pronounced ridge at the bed is encountered—or when ice shelves exert enough buttressing to stop further grounding-line retreat.

The retreat until 2010 of Pine Island Glacier has been attributed to enhanced ocean-induced melt, although its recent slowdown may be due to a combination of reduced forcing and a concomitant increase in glacier buttressing (17). It is possible that some glaciers, such as Pine Island Glacier and Thwaites Glacier, may already be undergoing MISI (9). Thwaites Glacier is now in a less-buttressed state because its ice shelf is mostly unconfined, and several simulations using state-of-the-art ice sheet models indicate continued mass loss and possibly MISI or MISI-like behavior, even under present climatic conditions (18–20).

More recently, the hypothesis of marine ice cliff instability (MICI) has emerged (14, 21), postulating that ice cliffs become unstable and collapse if higher than ~90 m above sea level, facilitating the rapid retreat of ice sheets. This process may have been important in Antarctica during past warm periods (14) by enhancing MISI (Fig. 3). During Pliocene warm periods, sea level was 10 to 20 m higher than it is now (22), requiring extensive retreat or collapse of the Greenland, West Antarctic, and marine-based sectors of the East Antarctic ice sheets. The MICI mechanism allows for increasing the model sensitivity such that the high sea level

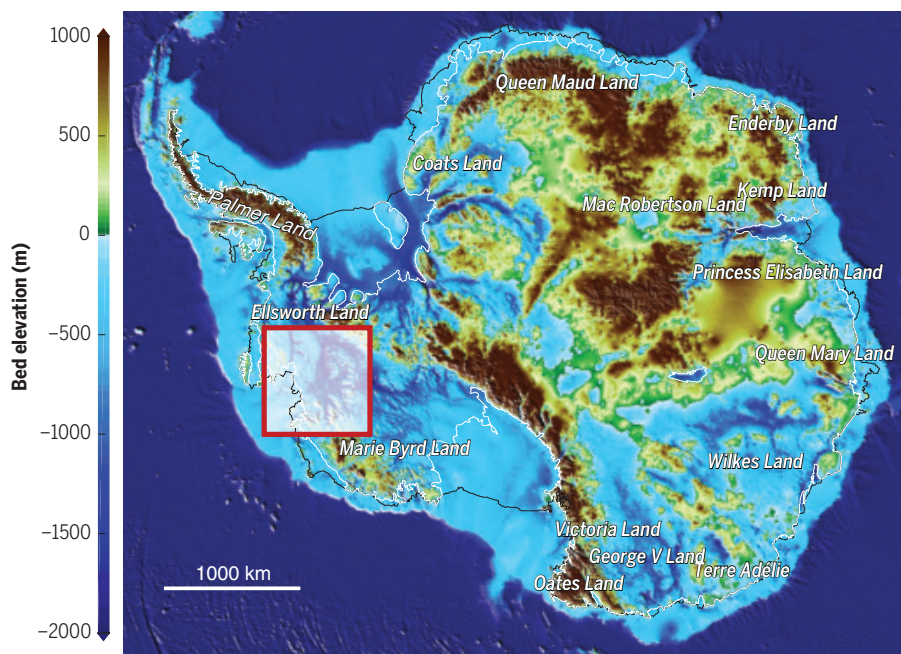


Fig. 1. Bed topography (bathymetry) of Antarctica. Blue areas are marine based (below sea level). The ice sheet grounding line is plotted in white, and the ice front is plotted in black. The area enclosed by the red square indicates the Amundsen Sea Embayment, shown in Fig. 2. Image modified from (2).

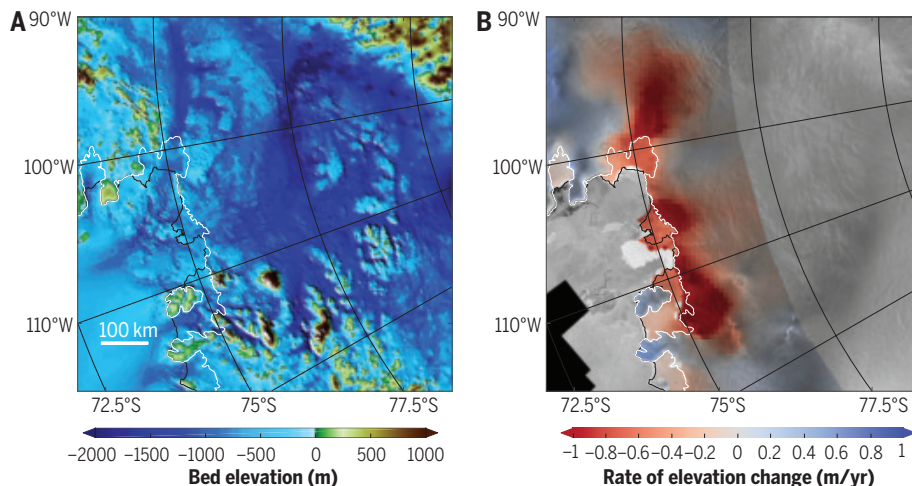


Fig. 2. Bed topography and ice elevation change in the Amundsen Sea Embayment. (A) Bed topography (bathymetry) of the Amundsen Sea Embayment. Image modified from (2). (B) Rate of ice sheet elevation change (from 2003 to 2009) from Ice, Cloud and land Elevation Satellite (ICESat) Geoscience Laser Altimeter System (GLAS) laser altimetry (45).

stands present during that period can be reached (14). However, contrary to the MISI hypothesis, MICI is not supported by a formal linear stability analysis (16), which hampers an adequate representation in marine ice sheet models. Furthermore, MICI has not been observed at such a scale in Antarctica, and so it remains unclear how rapidly an ice cliff would retreat as a function of its height (23). So far, models including MICI parameterized the rate of retreat based on the observed retreat rate of Jakobshavn Isbræ in West Greenland, which reached 3 km/year when its ice shelf collapsed in the early 2000s.

Cliff instability requires an a priori collapse of ice shelves and is favored by, among others, hydrofracturing through the increase of water pressure in surface crevasses, which widens and deepens them (21, 24, 25). Contrary to MISI, MICI could also occur on prograde bed slopes. Evidence from the Larsen B collapse, and rapid front retreat of Jakobshavn Isbræ, suggests that hydrofracturing could lead to the rapid collapse of ice shelves and potentially produce high, mechanically unsustainable ice cliffs (21, 24). However, its current impact is limited, because only a few Antarctic ice shelves have collapsed as of now. Moreover, recent work shows that the critical cliff height increases with time scale (i.e., the longer the time scale, the taller the cliff needs to be before collapse is possible), and therefore, ice shelf buttressing must be removed on time scales of less than 1 day to produce rapid brittle fracturing of a subaerial ice cliff at heights

attainable in ice sheets (23). Compelling evidence from observations at the Ross Sea shows that there has been no immediate grounding-line retreat after cliff collapse in the past (26). More research into the dynamics of ice cliffs is needed, and the existence of MICI remains controversial today.

Projecting the future of the Antarctic Ice Sheet

A major factor that limits reliable projections of the future Antarctic Ice Sheet response is how global warming relates to ocean dynamics that bring CDW onto and across the continental shelf, potentially increasing subshef melt. Because of this uncertainty, several studies apply linear extrapolations of present-day observed melt rates or simple parameterizations of ice-ocean melting rates, mostly focusing on unmitigated climate scenarios, such as Representative Concentration Pathway (RCP) 8.5. Numerous large-scale modeling studies conducted in the past decade have simulated future collapse of the WAIS under various climate-warming scenarios (13, 14, 27–30). These studies found that future grounding-zone retreat into the central WAIS region is expected on time scales of a few centuries to a millennium, contributing several meters to global mean sea level rise. However, although the time of onset of collapse is quite different across models and scenarios, all models produce WAIS collapse under unmitigated emission scenarios on multicentennial time scales.

Whole Antarctic simulations for unmitigated emission scenarios (RCP8.5) show a large scatter on centennial and multicentennial time scales (Fig. 4). However, the introduction of MICI in one ice sheet model (14) results in future sea level rise estimates almost one order of magnitude larger than those of other studies (Fig. 4). Although projected contributions of the Antarctic Ice Sheet to sea level rise by the end of this century for recent studies hover between 0 and 0.45 m (5 to 95% probability range), the MICI model occupies a range of 0.2 to 1.7 m (Fig. 4). The discrepancy is even more pronounced for 2300, at which point the MICI results and other model estimates no longer agree within uncertainty bounds. Given the uncertainty range on Pliocene sea level stands, MICI is not necessarily required to lead to rapid multimeter sea level rise (31), and other mechanisms related to basal conditions may well be able to accelerate mass loss on shorter time scales (30, 32).

Not all feedbacks in marine ice sheets enhance ice loss and collapse. Several mechanisms may slow down rapid ice retreat. For instance, as glaciers thin, the pressure that they exert on Earth's crust decreases, and so the bed rises in response to the reduction in ice mass. The lithosphere is a viscoelastic material, and the rate of uplift has two distinct response times: The elastic response is instantaneous but limited in magnitude, whereas the viscous response is slow but larger in magnitude. A low-viscosity asthenosphere and a thin lithosphere (known as a weak Earth structure), as observed under WAIS, will produce a faster and more localized viscoelastic response of solid Earth on decadal rather than millennial time scales (33). When the bedrock rises, the grounding-line retreat may slow down as the height above hydrostatic equilibrium increases inland. Simulations that account for this negative feedback show that bedrock uplift delays the collapse of the WAIS, leading to slower mass loss (34) compared with models that keep a fixed bedrock geometry. Although this mechanism has a strong impact on model simulations on multicentennial to millennial time scales, it is not yet clear whether it is important on the scale of decades.

Sea level commitment and tipping points

On multicentennial to multimillennial time scales, feedbacks with the atmosphere and ocean increase in importance. When subjected to perturbed climatic forcing over these time scales, ice sheets manifest large changes in their volume and distribution. These changes typically occur with a considerable lag in response to the forcing applied, which leads to the concept of sea level commitment, that is, ice mass losses that will occur in the long-term future are committed to that loss at a much earlier stage. Ice sheets are subject to threshold behaviors in their stability, because a change in

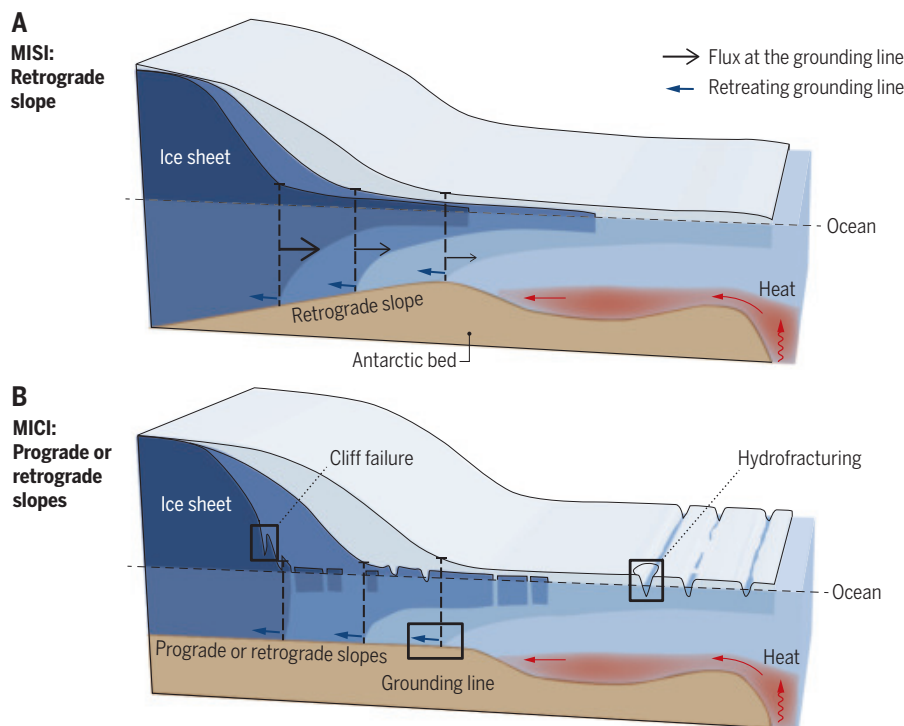


Fig. 3. Schematics of the marine ice sheet instability and marine ice cliff instability. (A) MISI, invoking unstable grounding line retreat on retrograde bed slopes due to reduced ice shelf buttressing. (B) MICI, where grounded ice cliffs may rapidly collapse after ice shelf breakup. Images modified from (1).

boundary conditions such as climate forcing can cause the current ice sheet configuration to become unstable through, for instance, MISI. Crossing these tipping points leads the system to equilibrate to a qualitatively different state (a complete collapse of the WAIS, for example). The existence of a tipping point implies that ice sheet changes are potentially irreversible. In other words, returning to a preindustrial climate may not necessarily stabilize the ice sheet once the tipping point has been crossed. Reversibility, however, may be possible over large climate cycles, such as a glacial-interglacial cycle.

The projected long-term sea level rise contribution of the Antarctic Ice Sheet for warming levels associated with the high-mitigation RCP2.6 scenario is limited to well below 1 m, although with a probability distribution that is not Gaussian but skewed with a long tail toward high values owing to potential MICI (7). However, substantial future retreat in some basins (such as Thwaites Glacier) cannot be ruled out, because grounding-line retreat may continue even with no additional forcing (18–20, 32). The long-term sea level rise contribution of the Antarctic Ice Sheet therefore crucially depends on the behavior of individual ice shelves and outlet glacier systems and whether they enter MISI for a given level of warming. Under sustained warming, a threshold for the survival of Antarctic ice shelves, and thus the stability of the ice sheet, seems to lie between 1.5° and 2°C above the present mean annual air temperature (28). Crossing these thresholds implies commitment to large ice sheet changes and sea level rise that may take thousands of years to be fully realized and may be irreversible on longer time scales (7).

Understanding key physical processes

Considerable progress has been made over the past decade with respect to understanding fundamental processes at the interface between ice sheets, atmosphere, and ocean and mechanisms of ice sheet instability. However, along with missing knowledge on the drivers of change, some key physical processes inherent to the dynamics of retreating marine ice sheets are still poorly understood. These processes include (i) ice-ocean interface processes responsible for subshelf melt, (ii) calving and (hydro)fracture processes, (iii) ice sheet basal sliding and subglacial sediment deformation, and (iv) GIA. This missing knowledge reduces our capability to accurately predict the timing and magnitude of the onset of enhanced mass loss or define potential tipping points of the Antarctic Ice Sheet.

As discussed above, increased subshelf melting (i) has triggered the observed acceleration of large Antarctic outlet glaciers in the Amundsen Sea sector during the past decade (3, 4, 8), and it is therefore critical that numerical ice sheet models represent the processes governing subshelf melt accurately. Subshelf melting is either parameterized or computed through coupling with an ocean model. Parameterizations typically relate subshelf melting to ocean temperature and/or ice shelf depth, in either a linear or a quadratic fashion, which leads to higher melting close to the grounding line (35). Other parameterizations relate subshelf melting to the distance to the grounding line, to the ice shelf and cavity depths, or, more recently, by using melt rates from a plume model that are extended spatially using physically motivated scalings that depend on local slope and ice draft (35). More accurate representations of subshelf melting can be achieved through coupling to an ocean model, which should lead to considerable improvements compared with simple parameterizations, because it accounts

for the transfer of heat, freshwater, and momentum between the two bodies.

Iceberg calving (ii) is responsible for the other part of the ice mass loss at the margins of the Antarctic Ice Sheet. Calving occurs when ice chunks break off from the edge of floating ice shelves in Antarctica. The rate at which icebergs detach from the ice shelf, or calving rate, determines the dynamics of the ice front. When the ice front is stationary, the calving rate is equal to the flow velocity of the ice. The calving rate therefore modulates buttressing induced by ice shelves and hence indirectly controls upstream grounded ice speed and subsequent sea level rise contribution. The large amount of ice lost through calving is common for Antarctica, but its representation and quantification in models are hampered by the difficult access to field sites, a high variability in time and space, and its inherent discontinuous nature, as opposed to the continuum approach used in most models. Until recently, calving rates were essentially either assumed to be equal to ice velocity (i.e., by keeping the ice

front fixed in space) or based on empirical relationships that are not well constrained by observations. Recent studies apply continuum damage mechanics to simulate crevasse formation. This approach represents initial ice microfractures and their vertical development as crevasses, which in turn weakens the ice through damage and decreases ice viscosity and which can be advected with the ice flow (36). Hydrofracturing, based on the surface meltwater widening and deepening crevasses, is also ubiquitously parameterized in ice sheet models and forms the precursor for MICI (21, 24). Calving remains one of the grand challenges of ice sheet modeling, and no general calving law exists yet, which profoundly limits our ability to model catastrophic calving events.

Basal conditions (iii) and GIA (iv) both have an impact on how ice sheets respond to forcing. Although the physics of GIA is well understood, the upper mantle viscosity under the Antarctic Ice Sheet is poorly constrained. Similarly, the mechanics of basal friction and how it varies spatially remain largely unknown. Models typically rely on simple friction laws that depend on the basal velocity linearly or nonlinearly (37), which is generally a good approximation for a hard bedrock. Many Antarctic ice streams, however, are known to be lying on soft beds that have a layer of deformable till. Recent studies and laboratory experiments suggest that the rheology of the till is plastic at large strain, and new

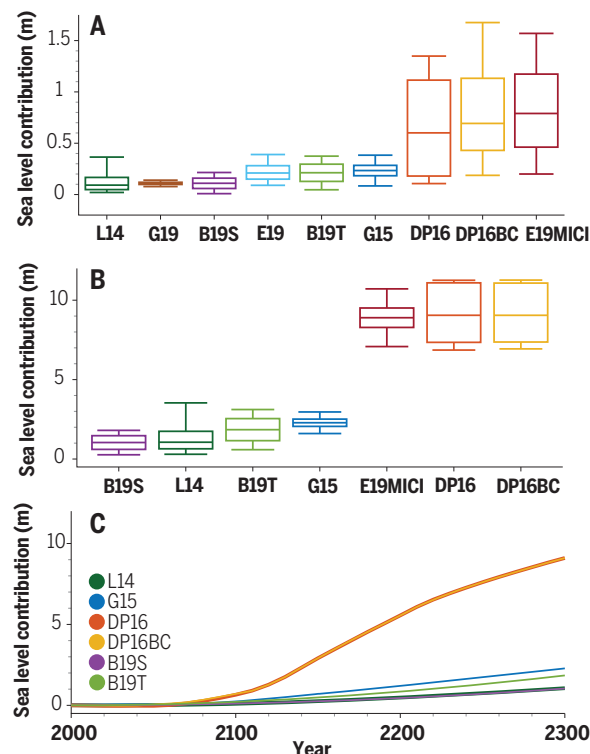


Fig. 4. Projections of Antarctic sea level contribution.

(A and B) Projections of Antarctic sea level contribution in 2100 (A) and 2300 (B) for different studies performed under RCP8.5. Boxes and whiskers show the 5th, 25th, 50th, 75th, and 95th percentiles. (C) Median projections of Antarctic sea level contribution from 2000 to 2300 (RCP8.5). The studies plotted are L14 (46); G15 (28); DP16 (14); DP16BC, bias-corrected simulations (14); B19S, simulations with Schoof's parameterization (30); B19T, simulations with Tsai's parameterization (30); E19, simulations without MICI (31); E19MICI, simulations with MICI (31); and G19 (32). Figure modified with permission from Elsevier (47).

parameterizations are being developed to account for both soft and hard beds (37). The development and validation of these new friction laws are critical to further improve the predictive skills of numerical models.

Challenges to reduce uncertainties

Besides understanding key physical processes, their representation in ice sheet models is also crucial. One way to assess the accuracy in the representation of physical processes in current ice sheet models is to organize large, international intercomparison projects. For example, the Marine Ice Sheet Model Intercomparison Project for planview models (MISMIP3d) greatly improved the representation of grounding-line migration by conforming models to known analytical solutions (38). These numerical experiments demonstrated that to resolve grounding-line migration in marine ice sheet models, a sufficiently high spatial resolution needs to be adopted, because membrane stresses need to be resolved across the grounding line to guarantee mechanical coupling, unless parameterizations are used (14) based on analytical solutions (16). Therefore, a series of ice sheet models have implemented subelement parameterizations or a spatial grid refinement, which also favors accurate data assimilation (27). In transient simulations, the adaptive mesh approach enables the finest grid to follow the grounding-line migration (27). These higher spatial resolutions on the order of hundreds of meters in the vicinity of grounding lines also pose new challenges about data management for modeling purposes and demand precise bathymetry to resolve the grounding zone (2). Nevertheless, recent theoretical developments with respect to grounding-line stability in response to buttressing (39), basal drag (40), and external forcing (41) demonstrate that further efforts are required in the verification and validation of numerical ice sheet models.

Intercomparisons are also essential for improving coupled ocean–sub-ice-shelf cavity–ice sheet models within a global system context (42). To better understand the influence of model initialization, an initial state intercomparison exercise (initMIP) has been developed (43). initMIP is the first set of experiments of the Ice Sheet Model Intercomparison Project for CMIP6 (ISMIP6), which is the primary Coupled Model Intercomparison Project Phase 6 (CMIP6) activity focusing on the Greenland and Antarctic Ice Sheets (42).

Besides multimodel ensembles, such as ISMIP6, uncertainty quantification within the model parameter space is a powerful tool to characterize and investigate uncertainty in projections (29, 30) and to improve projections of future sea level rise. One of the advantages of uncertainty quantification is that it can quantify the uncertainty in the projections associated to different input parameters, related either

to external forcing or to physical properties of the ice sheet (e.g., initial conditions, coefficients in parameterizations). It therefore makes it possible to show where progress should be made to reduce the uncertainty in projections of sea level rise most efficiently.

Model initialization remains another important factor, which relies on two distinct, but often combined, approaches: spin up versus data assimilation. The first approach spins up the model over glacial-interglacial periods, which ensures that the internal properties of the ice sheet are consistent with each other but may provide an inaccurate representation of the present-day ice sheet geometry and flow speed, which may introduce considerable biases on short-term (i.e., decadal to centennial) projections. The alternative is the assimilation of data, such as satellite-derived surface flow speeds, thinning and thickening rates, and so on. These two approaches lead to large differences in the initial conditions from which projections are made and therefore create a substantial spread in projected contributions to future sea level rise (43). Although data assimilation techniques cannot ensure consistent internal properties of the ice sheet, they are improving for centennial projections with the increasing access to high-resolution satellite products, which even allow for characterizing the subglacial conditions to a far better degree (44). They also enable the improvement of ice-thickness and bedrock datasets at a high resolution for the Antarctic Ice Sheet (2). One of the challenges for the coming years is that the volume of data available is increasing exponentially, but ice sheet models are not equipped to ingest large amounts of data from different sensors at different resolutions and acquired at different times. Some progress has been made by relying on tools such as automatic differentiation, but these methods have not yet been applied to large-scale systems such as the entire Antarctic Ice Sheet.

Eventually, the full coupling between ice, ocean, and atmosphere must be considered, which is currently the subject of ongoing research but remains limited to decadal or multidecadal time scales owing to the high computational cost of coupled models. Full ice–ocean coupling on the Thwaites drainage basin revealed a continued mass loss over the coming decades at a sustained rate and shows that uncoupled simulations greatly overestimate the rate of grounding-line retreat compared with the coupled model (20). Whole Antarctic semi-coupled simulations, on the other hand, show that meltwater from Antarctica will trap warm water below the sea surface, creating a positive feedback that increases Antarctic ice loss (32).

The increase in computational efficiency enabling high-spatial resolution modeling, the availability of high-resolution datasets of bed topography and of high-resolution satellite-based

ice surface velocity and changes in ice velocity, longer time series on ice sheet changes, and the improved initialization of ice sheet models are now allowing the ice sheet modeling community to produce increasingly robust projections on the future behavior of the Antarctic Ice Sheet. Closing knowledge gaps in drivers, forcing, and processes and an improved understanding of feedbacks between the different systems will be necessary to more accurately comprehend when and how future tipping points of the ice sheet are reached, because they have a profound impact on global sea level rise around the planet.

REFERENCES AND NOTES

1. F. Pattyn et al., *Nat. Clim. Chang.* **8**, 1053–1061 (2018).
2. M. Morlighem et al., *Nat. Geosci.* **13**, 132–137 (2020).
3. E. Rignot et al., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1095–1103 (2019).
4. A. Shepherd, H. A. Fricker, S. L. Farrell, *Nature* **558**, 223–232 (2018).
5. S. Schmidtke, K. J. Heywood, A. F. Thompson, S. Aoki, *Science* **346**, 1227–1231 (2014).
6. P. Dutrieux et al., *Science* **343**, 174–178 (2014).
7. D. W. J. Thompson et al., *Nat. Geosci.* **4**, 741–749 (2011).
8. F. S. Paolo, H. A. Fricker, L. Padman, *Science* **348**, 327–331 (2015).
9. K. Christianson et al., *Geophys. Res. Lett.* **43**, 10817–10825 (2016).
10. S. R. Rintoul et al., *Sci. Adv.* **2**, e1601610 (2016).
11. B. Medley, E. R. Thomas, B. Medley, *Science* **344**, 735–738 (2014).
12. J. Turner et al., *Geophys. Res. Lett.* **46**, 3502–3511 (2019).
13. J. Feldmann, A. Levermann, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14191–14196 (2015).
14. R. M. DeConto, D. Pollard, *Nature* **531**, 591–597 (2016).
15. J. Weertman, *J. Glaciol.* **13**, 3–11 (1974).
16. C. Schoof, *J. Geophys. Res. Earth Surf.* **112**, F03S28 (2007).
17. L. Favier et al., *Nat. Clim. Chang.* **4**, 117–121 (2014).
18. I. Joughin, B. E. Smith, B. Medley, *Science* **344**, 735–738 (2014).
19. I. J. Nias, S. L. Cornford, A. J. Payne, *J. Glaciol.* **62**, 552–562 (2016).
20. H. Seroussi et al., *Geophys. Res. Lett.* **44**, 6191–6199 (2017).
21. D. Pollard, R. M. DeConto, R. B. Alley, *Earth Planet. Sci. Lett.* **412**, 112–121 (2015).
22. G. R. Grant et al., *Nature* **574**, 237–241 (2019).
23. F. Clerc, B. M. Minchew, M. D. Behn, *Geophys. Res. Lett.* **46**, 12108–12116 (2019).
24. J. N. Bassis, C. C. Walker, *Proc. Royal Soc., Math. Phys. Eng. Sci.* **468**, 913–931 (2012).
25. A. A. Robel, A. F. Banwell, *Geophys. Res. Lett.* **46**, 12092–12100 (2019).
26. P. J. Bart, M. DeCesare, B. E. Rosenheim, W. Majewski, A. McGlannan, *Sci. Rep.* **8**, 12392 (2018).
27. S. L. Cornford et al., *Cryosphere* **9**, 1579–1600 (2015).
28. N. R. Golledge et al., *Nature* **526**, 421–425 (2015).
29. C. Ritz et al., *Nature* **528**, 115–118 (2015).
30. K. Bultuis, M. Arnst, S. Sun, F. Pattyn, *Cryosphere* **13**, 1349–1380 (2019).
31. T. L. Edwards et al., *Nature* **566**, 58–64 (2019).
32. N. R. Golledge et al., *Nature* **566**, 65–72 (2019).
33. V. R. Barletta et al., *Science* **360**, 1335–1339 (2018).
34. E. Larour et al., *Science* **364**, eaav7908 (2019).
35. L. Favier et al., *Geosci. Model Dev.* **12**, 2255–2283 (2019).
36. S. Sun, S. L. Cornford, J. C. Moore, R. Gladstone, L. Zhao, *Cryosphere* **11**, 2543–2554 (2017).
37. J. Brondex, F. Gillet-Chaulet, O. Gagliardini, *Cryosphere* **13**, 177–195 (2019).
38. F. Pattyn et al., *J. Glaciol.* **59**, 410–422 (2013).
39. M. Haseloff, O. V. Sergienko, *J. Glaciol.* **64**, 417–431 (2018).
40. O. V. Sergienko, D. J. Wingham, *J. Glaciol.* **65**, 833–849 (2019).
41. G. H. Gudmundsson, F. S. Paolo, S. Adusumilli, H. A. Fricker, *Geophys. Res. Lett.* **46**, 13903–13909 (2019).
42. S. M. J. Nowicki et al., *Geosci. Model Dev.* **9**, 4521–4545 (2016).
43. H. Seroussi et al., *Cryosphere* **13**, 1441–1471 (2019).
44. F. Gillet-Chaulet et al., *Geophys. Res. Lett.* **43**, 10,311–10,321 (2016).
45. H. D. Pritchard, R. J. Arthern, D. G. Vaughan, L. A. Edwards, *Nature* **461**, 971–975 (2009).
46. A. Levermann et al., *Earth Syst. Dyn.* **5**, 271–293 (2014).
47. E. Hanna et al., *Earth Syst. Res.* **201**, 102976 (2020).

ACKNOWLEDGMENTS

We thank K. Bultuis for drafting Fig. 3; N. Golledge; and two anonymous reviewers, as well as the editor, for their insightful comments. **Competing interests:** None declared.

10.1126/science.aaz5487

RESEARCH

IN SCIENCE JOURNALS

Edited by Michael Funk

BIOTECHNOLOGY

From genetics to material to behavior

Introducing new genes into an organism can endow new biochemical functions or change the patterns of existing functions, but extending these manipulations to structure at the tissue level is challenging. Combining genetic engineering and polymer chemistry, Liu *et al.* directly leveraged complex cellular architectures of living organisms to synthesize, fabricate, and assemble bioelectronic materials (see the Perspective by Otto and Schmidt). An engineered enzyme expressed in genetically targeted neurons synthesized conductive polymers in tissues of freely moving animals. These polymers enabled modulation of membrane properties in specific neuron populations and manipulation of behavior in living animals. —SYM

Science, this issue p. 1372; see also p. 1303

Artist's conception of neurons, some of which are genetically engineered to produce a membrane-associated conducting polymer

APPLIED PHYSICS

Electricity from thermal sources

It is desirable to harvest as much energy as possible from processes that produce useful amounts of heat and convert it from waste into electrical power. Thermoelectrics and thermophotovoltaics can harness and convert heat waste but tend to operate at high temperatures. Davids *et al.* designed and fabricated a complementary metal-oxide semiconductor infrared photonic device that can harvest and recover energy from low-temperature thermal sources (see the Perspective by Raman). Using a new conversion mechanism, they experimentally demonstrate large thermal-to-electrical power generation in a bipolar grating-coupled tunneling device, rivaling the best thermoelectric devices. The device design could be used for

energy harvesting of waste heat and the development of compact thermal batteries. —ISO

Science, this issue p. 1341; see also p. 1301

CELL BIOLOGY

PI(4)P regulates mitochondrial fission

Mitochondria are dynamic intracellular organelles, the shape and number of which are regulated by various cell-signaling pathways. Mitochondrial division is driven by the recruitment of a constricting guanosine triphosphatase protein at sites of contact with the endoplasmic reticulum, but other factors, including lysosomes, are also involved. Nagashima *et al.* now document an essential role for Golgi-derived vesicles bearing a specific lipid—phosphatidylinositol 4-phosphate, or PI(4)P—in the final steps of mitochondrial

division. Disruption of PI(4)P production results in mitochondrial morphological defects indicative of an inability to complete fission. —SMH

Science, this issue p. 1366

COMPUTER VISION

Dodgeball for drones

Uncrewed aerial vehicles, or drones, use frame-based cameras for object perception. However, these cameras have

exposure times on the order of milliseconds, which limits their responsiveness. Falanga *et al.* instead used event cameras with reaction times of microseconds. For sense-and-avoid maneuvers, these bioinspired neuromorphic cameras are more responsive than conventional cameras because they measure changes of brightness in an image (events), thus effectively sensing motion. With the addition of event-based detection



A drone dodges objects using event-based cameras and onboard processing.

algorithms, the authors have designed a drone that can dodge multiple objects. —MML

Sci. Robot. **5**, eaaz9712 (2020).

CHEMICAL ENGINEERING

Every twig and splinter used

Plant-based production of commodity chemicals faces steep competition from fossil resources, which are often cheaper and easier to partition. Sustainable use of renewable resources requires strategies for converting complex and recalcitrant biomolecules into streams of chemicals with extraordinary efficiency. Liao *et al.* developed a biorefinery concept in which wood is eventually fully converted into useful chemicals: phenol, propylene, pulp amenable to ethanol production, and phenolic oligomers that can be incorporated into ink production (see the Perspective by Zhang). A life-cycle assessment and techno-economic analysis highlight the efficiency of the process and reveal the potential for such biorefinery strategies to contribute to sustainable chemicals markets. —MAF

Science, this issue p. 1385; see also p. 1305

ANTIBODIES

Antibody assembly in lampreys

For B lymphocytes in jawless vertebrates to produce antibodies, a combination of gene cassettes must be stitched together to create a functional antibody gene. Circumstantial evidence based on gene expression data previously implicated the cytidine deaminase CDA2 in this process. Morimoto *et al.* used CRISPR-Cas9-mediated mutagenesis to show that loss-of-function mutations in the CDA2 gene result in the loss of antibody gene assembly without disrupting the formation of functional genes encoding lamprey T cell receptors. These methods establish lampreys as

a genetically tractable model system. —IW

Sci. Immunol. **5**, eaba0925 (2020).

STRUCTURAL BIOLOGY

Choosing a partner that fits

G protein-coupled receptors (GPCRs) are responsible for transducing diverse signals from outside to inside cells. This process requires specificity both in ligand binding to GPCRs and in coupling between GPCRs and their intracellular partners, G proteins. Qiao *et al.* determined the structure of the human glucagon receptor (GCGR), a type B GPCR, bound to glucagon and one of two heterotrimeric G proteins, G_s or G_{i1}. GCGR signals mainly through G_s, and the structures provide a basis for this specificity. Conformational changes in GCGR, relative to the inactive state, create a binding cavity for the G proteins. The pocket is opened sufficiently to accommodate a bulky binding motif in G_s. G_{i1} can still bind but the pocket does not close around it, so there is a smaller interaction interface. —VV

Science, this issue p. 1346

GEOMORPHOLOGY

Erosion-vegetation interactions

The impact of vegetation on erosion rates is hard to gauge. Although vegetation can hold soils in place mechanically, root systems can also loosen soils or even help to fracture rock. These processes can increase erosion, especially because areas of heavy vegetation tend to be in areas with high precipitation rates. Starke *et al.* tackled this issue using a large set of observations that span 3500 km of the Andes mountain range. They found a complex set of interactions where increasing vegetation decreases erosion in more arid regions but can accelerate erosion in vegetation dense regions. —BG

Science, this issue p. 1358

IN OTHER JOURNALS

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

COGNITION

Avian statisticians

We humans, even those not well versed in the study of statistics, make statistical inferences regularly when we decide to, say, choose from a bowl with a high ratio of chocolate chips to nuts. In the animal world, this kind of inference, and the ability to broadly apply it when making choices, has also been found in chimpanzees, but whether it exists outside of this lineage has been debated. Bastos and Taylor looked for statistical ability in parrots, which are increasingly recognized as having high-level cognitive functions. When trained to understand that certain tokens conferred a reward, keas (scavenging parrots native to New Zealand) consistently judged their chances of acquiring one under different circumstances. They were as successful as humans at avoiding samples offered from containers with the fewest reward tokens and consistently chose samples offered from containers with the most. They also spotted when the experimenters introduced biases. —SNV

Nat. Commun. **11**, 828 (2020).



Parrots, like this New Zealand kea, are increasingly recognized for having high-level cognitive abilities.

BRAIN MAPPING

Connecting the human amygdala

The amygdala is a brain structure that is affected in many different psychiatric disorders. We still have an inadequate understanding of its role within the organization of the human brain. Sylvester *et al.* used repeated sampling and precision mapping to define three amygdala subdivisions in 10

individuals based on connectivity patterns with the cortex. These subdivisions occupied similar locations in different subjects and similar network connectivity. One subdivision has preferential functional connectivity to the default mode network, which engages when an individual is focused on a specific task; a second, medially located subdivision preferentially connects to the dorsal attention network; and a third connects to

SEISMOLOGY

The red planet quakes

The first unsuccessful attempt to detect seismic activity on Mars was in 1975 on the Viking landers. More than 40 years later, Giardini *et al.* finally detected marsquakes with the seismometer on the InSight mission that landed on Mars in 2018. Most of the detected marsquakes have been small, but there were a few that could be as large as a magnitude 4. Although most of the 174 events were likely due to seismic activity, some may have been caused by meteorite impact or other sources. The catalog forms a basis for further investigation into the rock properties in the martian interior. —BG

Nat. Geosci. **13**, 205 (2020).



Seismic activity on Mars (shown) has been detected by NASA's InSight lander.

a ventrally located amygdala subdivision, but does not show any functional network preferences. These data may help to develop biologically plausible biomarkers and targets for intervention in psychiatric patients. —PRS

Proc. Natl. Acad. Sci. U.S.A. **117**, 3808 (2020).

CANCER

Cancer therapy in good order

Treatment of cancer patients with two or more drugs acting through different mechanisms is a strategy that has prolonged many lives. Whether the drugs within these combination therapies are delivered concurrently or sequentially can have a major impact on efficacy. A new study illustrates this principle for drugs that inhibit cell cycle kinases CDK4 and CDK6 (CDK4/6 inhibitors), which

have attracted great interest because of their clinical efficacy in breast cancer. Studying mouse models of pancreatic cancer, Salvador-Barbero *et al.* found that sequential treatment with Taxol (which inhibits mitosis) followed by a CDK4/6 inhibitor (which prevents cell cycle entry) offered substantially more therapeutic benefit than concurrent treatment with the drugs. Mechanistically, this is because the CDK4/6 inhibitor prevents cancer cells from repairing the chromosomal damage caused by Taxol. —PAK

Cancer Cell **10.1016/j.ccell.2020.01.007** (2020).

PSYCHOLOGY

The psychology of the alt-right

The political movement known as the "alt-right" has increased

in popularity in the United States over the past several years. However, empirical descriptive research on the psychological characteristics that unite members of the movement is needed. Forscher and Kteily conducted detailed survey work to determine the popularity of the movement and its psychological profile. On the basis of their findings, the authors estimate that 6% of the U.S. population, and 10% of people who voted for Trump in the 2016 election, identify as being part of the alt-right. Alt-right members do not indicate feelings of economic anxiety, but rather exhibit preferences for social group-based hierarchies favoring whites. These results have implications for understanding the role of inter-group relations and conflict in U.S. electoral politics. —TSR

Perspect. Psychol. Sci. **15**, 90 (2020).

CHEMINFORMATICS

Machine learning for natural extracts

Natural products and their derivatives continue to be an important source of drug candidates because of their structural diversity and wide-ranging biological activities, which are unmatched by synthetic compounds. Natural products are generally complex mixtures with chemical constituents that are not well characterized. Reher *et al.* report a nuclear magnetic resonance-based machine-learning tool, SMART, for rapid structural analysis of major constituents from crude natural extracts and for the discovery of new natural products. For example, SMART automatically characterized a cyanobacterial extract mixture and isolated a new chimeric macrolide, symplocolide A; it also dereplicated several known natural products. The proposed cheminformatic tool paves the way for new computer-aided approaches to natural product drug discovery. —YS

J. Am. Chem. Soc. **142**, 4114 (2020).

SIGNALING

A decoy insulin receptor in worms?

Insulin signaling in the worm *Caenorhabditis elegans* appears to be regulated by expression of a truncated, alternatively spliced form of the receptor that lacks the intracellular signaling domain of the receptor. Expression of the spliced form of the receptor, DAF-2B, is regulated in the worm and serves to modulate the effects of insulin-like peptides. Expression of DAF-2B alters sensitivity to insulin. Martinez *et al.* suggest that the modified receptor might alter insulin signaling by sequestering insulin peptides on the inactive receptor, although such binding was not shown. Interaction with the full-length receptor is also a possibility. The results raise the intriguing possibility that a spliced receptor might function similarly in mammals and contribute to the control of insulin signaling. —LBR

eLife **9**, e49917 (2020).

CREDIT: NASA/JPL

RESEARCH ARTICLE SUMMARY

HUMAN GENETICS

Insights into human genetic variation and population history from 929 diverse genomes

Anders Bergström*, Shane A. McCarthy†, Ruoyun Hui†, Mohamed A. Almarri†, Qasim Ayub, Petr Danecek, Yuan Chen, Sabine Felkel, Pille Hallast, Jack Kamm, Hélène Blanché, Jean-François Deleuze, Howard Cann†, Swapan Mallick, David Reich, Manjinder S. Sandhu, Pontus Skoglund, Aylwyn Scally, Yali Xue§, Richard Durbin§, Chris Tyler-Smith§*

INTRODUCTION: Large-scale human genome-sequencing studies to date have been limited to large, metropolitan populations or to small numbers of genomes from each group. Much remains to be understood about the extent and structure of genetic variation in our species and how it was shaped by past population separations, admixture, adaptation, size changes, and gene flow from archaic human groups. Larger numbers of genome sequences from more diverse populations are needed to answer these questions.

RATIONALE: We sequenced 929 genomes from 54 geographically, linguistically, and culturally diverse human populations to an average of 35× coverage and analyzed the variation among them. We also physically resolved the haplotype phase of 26 of these genomes using linked-read sequencing.

RESULTS: We identified 67.3 million single-nucleotide polymorphisms, 8.8 million small insertions or deletions (indels), and 40,736 copy number variants. This includes hundreds of thousands of variants that had not been

discovered by previous sequencing efforts, but which are common in one or more population. We demonstrate benefits to the study of population relationships of genome sequences over ascertained array genotypes, particularly when involving African populations.

Populations in central and southern Africa, the Americas, and Oceania each harbor tens to hundreds of thousands of private, common genetic variants. Most of these variants arose as new mutations rather than through archaic introgression, except in Oceanian populations, where many private variants derive from Denisovan admixture. Although some reach high frequencies, no variants are fixed between major geographical regions.

We estimate that the genetic separation between present-day human populations occurred mostly within the past 250,000 years. However, these early separations were gradual in nature and shaped by protracted gene flow. All populations thus still had some genetic contact more recently than this, but there is also evidence that a small fraction of present-day structure might be hundreds of thousands of years older. Most populations expanded in size over

the past 10,000 years, but hunter-gatherer groups did not.

The low diversity among the Neanderthal haplotypes segregating in present-day populations indicates that, while more than one Neanderthal individual must have contributed genetic material to modern humans, there was

likely only one major episode of admixture. By contrast, Denisovan haplotype diversity reflects a more complex history involving more than one episode of admixture.

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aay5012>

We found small amounts of Neanderthal ancestry in West African genomes, most likely reflecting Eurasian admixture. Despite their very low levels or absence of archaic ancestry, African populations share many Neanderthal and Denisovan variants that are absent from Eurasia, reflecting how a larger proportion of the ancestral human variation has been maintained in Africa.

CONCLUSION: The discovery of substantial amounts of common genetic variation that was previously undocumented and is geographically restricted highlights the continued value of anthropologically informed study designs for understanding human diversity. The genome sequences presented here are a freely available resource with relevance to population history, medical genetics, anthropology, and linguistics. ■

The list of author affiliations is available in the full article online.

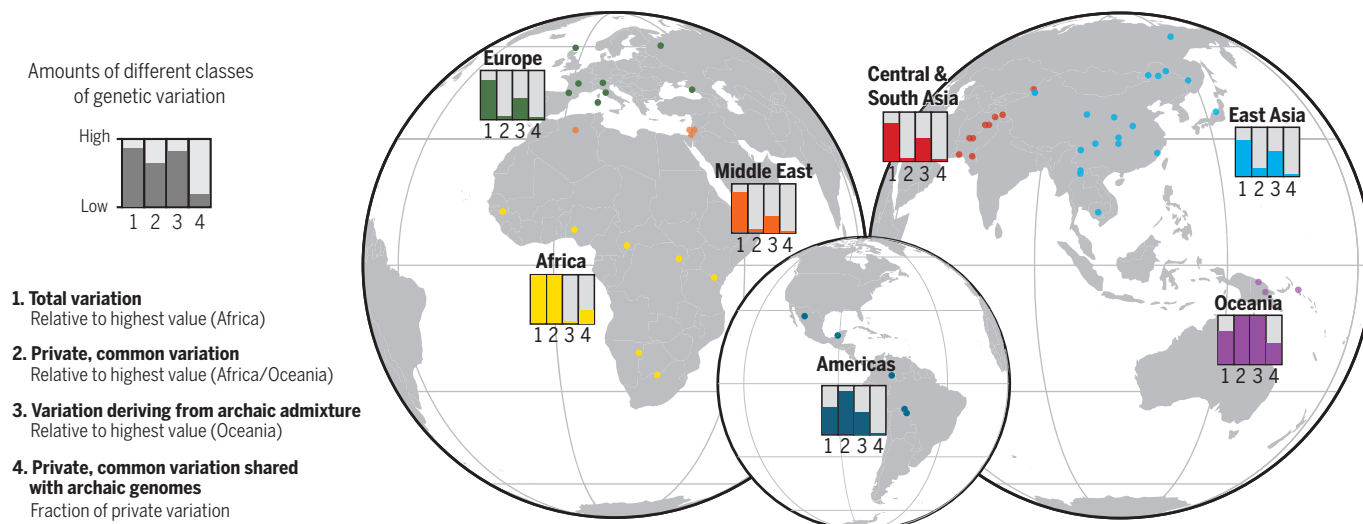
*Corresponding author. Email: ab34@sanger.ac.uk (A.B.); cts@sanger.ac.uk (C.T.-S.)

†These authors contributed equally to this work.

‡Deceased.

§These authors contributed equally to this work.

Cite this article as A. Bergström *et al.*, *Science* 367, eaay5012 (2020). DOI: 10.1126/science.aay5012



Structure of genetic variation across worldwide human populations. Shown is a schematic illustration of the approximate amounts of four different classes of genetic variation found in different geographical regions. The origins of the populations included in the study are indicated by dots.

RESEARCH ARTICLE SUMMARY

CORTICAL GENETICS

The genetic architecture of the human cerebral cortex

Katrina L. Grasby*† and Neda Jahanshad*† et al.

INTRODUCTION: The cerebral cortex underlies our complex cognitive capabilities. Variations in human cortical surface area and thickness are associated with neurological, psychological, and behavioral traits and can be measured in vivo by magnetic resonance imaging (MRI). Studies in model organisms have identified genes that influence cortical structure, but little is known about common genetic variants that affect human cortical structure.

RATIONALE: To identify genetic variants associated with human cortical structure at both global and regional levels, we conducted a genome-wide association meta-analysis of brain MRI data from 51,665 individuals across 60 cohorts. We analyzed the surface area and

average thickness of the whole cortex and 34 cortical regions with known functional specializations.

RESULTS: We identified 306 nominally genome-wide significant loci ($P < 5 \times 10^{-8}$) associated with cortical structure in a discovery sample of 33,992 participants of European ancestry. Of the 299 loci for which replication data were available, 241 loci influencing surface area and 14 influencing thickness remained significant after replication, with 199 loci passing multiple testing correction ($P < 8.3 \times 10^{-10}$; 187 influencing surface area and 12 influencing thickness).

Common genetic variants explained 34% (SE = 3%) of the variation in total surface area

and 26% (SE = 2%) in average thickness; surface area and thickness showed a negative genetic correlation ($r_G = -0.32$, SE = 0.05, $P = 6.5 \times 10^{-12}$), which suggests that genetic influences have opposing effects on surface area and thickness. Bioinformatic analyses showed that total surface area is influenced by genetic variants that alter gene regulatory activity in neural progenitor cells during fetal development.

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aay6690>

By contrast, average thickness is influenced by active regulatory elements in adult brain samples, which may reflect processes that occur after mid-fetal development, such as myelination, branching, or pruning.

When considered together, these results support the radial unit hypothesis that different developmental mechanisms promote surface area expansion and increases in thickness.

To identify specific genetic influences on individual cortical regions, we controlled for global measures (total surface area or average thickness) in the regional analyses. After multiple testing correction, we identified 175 loci that influence regional surface area and 10 that influence regional thickness. Loci that affect regional surface area cluster near genes involved in the Wnt signaling pathway, which is known to influence areal identity.

We observed significant positive genetic correlations and evidence of bidirectional causation of total surface area with both general cognitive functioning and educational attainment. We found additional positive genetic correlations between total surface area and Parkinson's disease but did not find evidence of causation. Negative genetic correlations were evident between total surface area and insomnia, attention deficit hyperactivity disorder, depressive symptoms, major depressive disorder, and neuroticism.

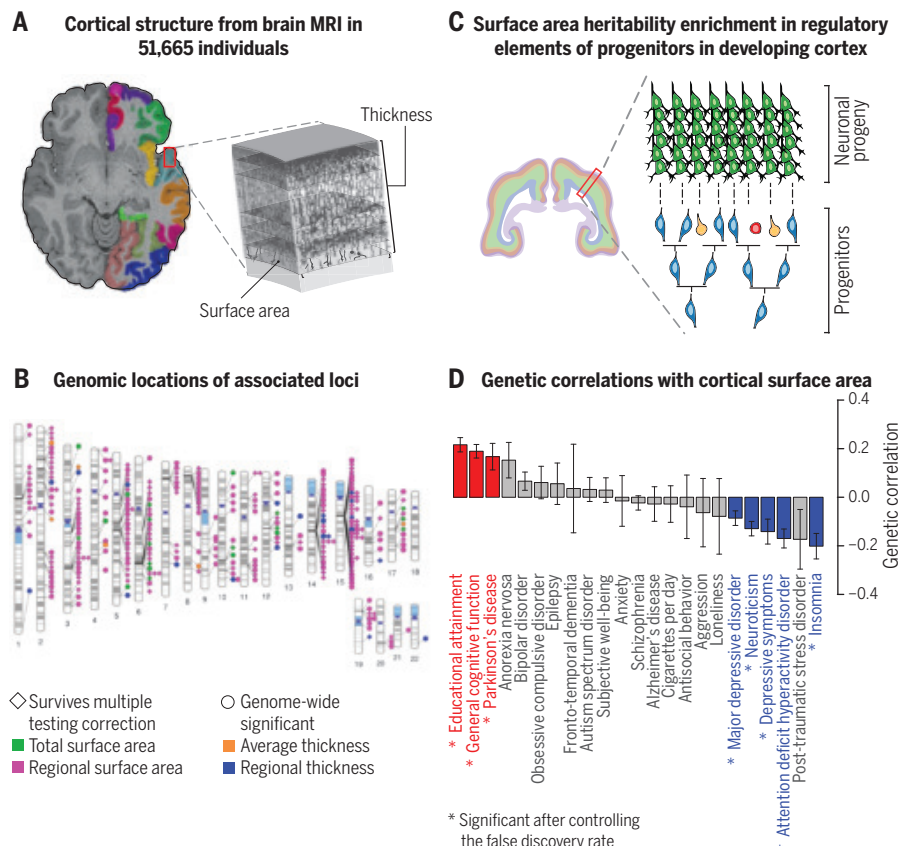
CONCLUSION: This large-scale collaborative work enhances our understanding of the genetic architecture of the human cerebral cortex and its regional patterning. The highly polygenic architecture of the cortex suggests that distinct genes are involved in the development of specific cortical areas. Moreover, we find evidence that brain structure is a key phenotype along the causal pathway that leads from genetic variation to differences in general cognitive function. ■

The complete list of authors and affiliations is available in the full article online.

*Corresponding authors: Katrina L. Grasby (katrina.grasby@qimrberghofer.edu.au); Neda Jahanshad (njahansh@usc.edu); Jason L. Stein (jason_stein@med.unc.edu); Paul M. Thompson (pthomp@usc.edu); Sarah E. Medland (sarah.medland@qimrberghofer.edu.au)

†These authors contributed equally to this work.

Cite this article as K. L. Grasby et al., *Science* 367, eaay6690 (2020). DOI: 10.1126/science.aay6690



Identifying genetic influences on human cortical structure. (A) Measurement of cortical surface area and thickness from MRI. (B) Genomic locations of common genetic variants that influence global and regional cortical structure. (C) Our results support the radial unit hypothesis that the expansion of cortical surface area is driven by proliferating neural progenitor cells. (D) Cortical surface area shows genetic correlation with psychiatric and cognitive traits. Error bars indicate SE.

RESEARCH ARTICLES

APPLIED PHYSICS

Electrical power generation from moderate-temperature radiative thermal sources

Paul S. Davids*, Jared Kirsch, Andrew Starbuck†, Robert Jarecki†, Joshua Shank, David Peters

Moderate-temperature thermal sources (100° to 400°C) that radiate waste heat are often the by-product of mechanical work, chemical or nuclear reactions, or information processing. We demonstrate conversion of thermal radiation into electrical power using a bipolar grating-coupled complementary metal-oxide-silicon (CMOS) tunnel diode. A two-step photon-assisted tunneling charge pumping mechanism results in separation of charge carriers in pn-junction wells leading to a large open-circuit voltage developed across a load. Electrical power generation from a broadband blackbody thermal source has been experimentally demonstrated with converted power densities of 27 to 61 microwatts per square centimeter for thermal sources between 250° and 400°C. Scalable, efficient conversion of radiated waste heat into electrical power can be used to reduce energy consumption or to power electronics and sensors.

All objects at finite temperature radiate owing to thermal fluctuations of their atomic constituents in a characteristic spectrum that depends on the object's surface temperature and spectral emissivity. Radiative heat transfer from the Sun (6050°C effective solar blackbody spectrum; air mass 1.5) is the dominant radiative energy resource available on Earth. Photovoltaic power generation is an effective and rapidly growing technology for converting this incident radiation into electrical power. However, radiative processes from cooler terrestrial sources or man-made waste heat can give rise to considerable net energy exchange that is also readily available as an electrical power source, provided that it can be efficiently converted.

Thermophotovoltaic (TPV) devices that convert radiation from broadband thermal sources into electrical power are promising technologies for solar energy conversion and waste heat recovery. These devices work by heating a secondary thermal source that acts as a selective emitter. The emission spectrum is filtered and matched to a small bandgap semiconductor device (1). The semiconductor device is typically a p- and n-type (pn) junction designed such that the absorption takes place in the depletion width of the device. The absorption of a photon in the depletion region of the semiconductor creates an electron hole pair that is separated by the internal field, resulting in the separation of charge and an open-circuit voltage induced across the device. These devices typically work in the range of temperatures from 1000 to 2000 K or wavelengths λ of 1.4 to 3.0 μm , which corresponds

to bandgap energies E_g of 0.43 to 0.86 eV. Wien's law requires that, as the temperature of the blackbody source is decreased, the wavelength at peak power increases such that, for source temperatures between 100° and 400°C, the spectral range is in the thermal infrared (7 to 12 μm). Narrow bandgap semiconductors at room temperature matched to this infrared band have considerable photon generation from thermal fluctuations and thermal generation across the bandgap resulting in appreciable noise. For this reason, infrared detectors are typically cooled below room temperature to reduce background noise driven by these thermal fluctuations in the narrow-gap semiconductor (2). Furthermore, the power density in the working bandwidth of the device is decreased because of the exponential nature of Planck's radiation law, thus making TPV conversion from a moderate-temperature source very challenging. Improvements in TPV conversion efficiency for moderate-temperature sources using vacuum near-field enhancement have been proposed (3–5). These near-field TPV devices show great promise, but practical limits of vacuum gaps and planarity from micrometer to nanometer length scales remain a challenge for larger-scale power generation.

Approaches for energy conversion

Alternative approaches for thermal-to-electrical conversion on the basis of direct rectification of infrared radiation using ultrafast tunneling have been proposed (6–8). This approach is not based on square-law absorption and generation of carriers in a semiconductor, but rather it relies on high-speed direct tunneling within an asymmetric tunnel diode to separate charge. In the microwave part of the spectrum, antenna-coupled diode rectifiers are used as efficient converters from high-speed gigahertz

signals to direct current (dc) (9–11). These devices use standard antennas to channel microwave radiation into a stripline guided mode, which is loaded with a fast diode to rectify the signal. Impedance-matching techniques are used to minimize reflectance from the antenna and effectively maximize power transfer to the load. This is inherently a narrow-band conversion process, differing from that in infrared and optical devices. These microwave rectifying antenna (rectenna) devices can have conversion efficiencies >85% (10), and attempts to scale to infrared and optical frequencies (30 to 200 THz) have been explored (12–15). A key component in these scaled devices is the ultrafast direct tunnel diode. Metal-insulator-metal (MIM) or double insulator tunnel (MIIM) diodes have been examined for this purpose (16–18). However, large diode asymmetry at small voltages is required for rectification of infrared radiation, and further progress is needed in this area. A promising approach for making highly asymmetric tunnel diodes on the basis of advanced complementary metal-oxide-silicon (CMOS) fabrication has recently been developed, which brings new design tools and scalable processing to large-area antenna-coupled tunnel diode devices (13, 19).

Electrical power generation from a direct grating-coupled tunnel MOS diode rectifier illuminated by a thermal source has recently been experimentally observed (20). Peak power densities of 1 to 8 nW/cm² have been measured for thermal sources between 400° and 450°C in a large area n+ MOS grating-coupled tunnel diode. The peak power is seen to increase as the source temperature is increased, and impedance matching into the load resistance is as predicted by the simple rectifier model. These antenna-coupled tunnel diodes strongly interact photonic resonances with longitudinal optical (LO) phonon resonances in the polar oxide tunnel barrier (13, 19). The confined and enhanced transverse optical field in the tunnel barrier occurs because of the epsilon-near-zero (ENZ) dispersion in the polar gate oxide, and the resonant antenna or grating structure is necessary to couple into the LO phonon mode (21–24). The transverse field in the tunnel barrier is spatially varying and enhanced relative to the incident field amplitude and can be orders of magnitude larger (13, 19, 20). Moreover, the enhanced transverse field confined in the tunnel barrier occurs over a finite bandwidth near the ENZ material resonance and is a result of the avoided crossing of the photonic and LO phonon resonances. A mechanism for the generation of a direct photocurrent from the confined transverse field in the barrier is based on a model of photon-assisted tunneling (PAT) in a uniform barrier (25). PAT is obtained as a perturbation expansion in the strength of the electromagnetic interaction in the confined tunnel barrier and leads to a multiphoton

Sandia National Laboratory, P.O. Box 5800, Albuquerque, NM 87185-1082, USA.

*Corresponding author. Email: pdavids@sandia.gov

†These authors contributed equally to this work.

tunneling process (25, 26). In this model, tunneling is a multistep process that results from the absorption of a photon in an occupied state near the Fermi energy of the metal followed by subsequent field-enhanced tunneling into an unoccupied state of the silicon leading to a small direct photocurrent. A similar time-reversed process occurs in the semiconductor, resulting in a backflow current into the metal. The resulting direct current is due to the difference in these currents, which arises from

the difference in effective mass for electrons in the metal and the semiconductor. This simple picture accurately predicts the temperature transitions and current-voltage (IV) dependence on temperature for measured rectenna power generation. The full perturbative dynamic current model predicts a nonlinear expansion of the PAT current in terms of nonlinear conductance in analogy with nonlinear susceptibilities in conventional nonlinear optics.

Photon-mediated charge pumping

PAT and the spatially varying confined optical field in the tunnel barrier suggest a newly devised means for thermal photovoltaic conversion from a low-grade thermal source in the temperature range of 100° to 400°C (Fig. 1A). An interdigitated bipolar pn junction array under the tunneling gate electrode acts as a charge pump moving electrons from p-type region to n-type well in a half period of the optical field. Figure 1B shows a schematic of multiple periods of a bipolar grating-coupled tunnel diode. The underlying band structure and associated charge wells are also shown. An instantaneous electron particle current, J_T , shown in Fig. 1B leads to charge separation into the n and p wells. An optical micrograph of the fabricated three-terminal bipolar device is shown in Fig. 1C. The simulated instantaneous enhanced transverse field at the grating resonance is shown in Fig. 1D, which can be translated into a spatially varying voltage across the grating-metal width. At maximum current, the metal-n+ and metal-p+ tunnel diodes are both forward-biased tunnel junctions. After a half cycle of the optical field, a minimal backflow current occurs owing to the reverse biasing of the two tunnel junctions. This is analogous to photovoltaic conversion in a pn junction, except that the device current does not arise from square-law generation in the depletion region but from ultrafast PAT in the two metal-oxide-semiconductor tunnel diodes and subsequent charge pumping. The interdigitated p and n regions are separately contacted, and power generation is measured across a variable external load resistance, R , that shorts the pn junction. As will be demonstrated, the advantage of using multiple pn junctions is that the open-circuit voltage, V_{oc} , is not limited by the induced alternating current (ac) infrared voltage on the tunnel diode. This effective diode voltage multiplier circuit results in orders-of-magnitude improvement in electrical power generation compared with direct rectification approaches (20).

Device modeling

A device model for thermal photovoltaic conversion in an ideal bipolar antenna-coupled tunnel diode rectifier consists of a buried symmetric pn junction under an equilibrium MOS metal gate, as shown in Fig. 2A. The model development proceeds by considering the confined and enhanced electric field in the tunnel barrier, which gives rise to a spatially varying voltage across the metal gate. The device is illuminated by a thermal source modeled as a blackbody broadband emitter. The enhanced and confined transverse field in the gap is obtained from the incident field, $E_{\max}(\omega) = \gamma(\omega)E_0(\omega)$, where γ is the ENZ enhancement factor and $E_0(\omega)$ is the incident field. The power per unit area emitted from a

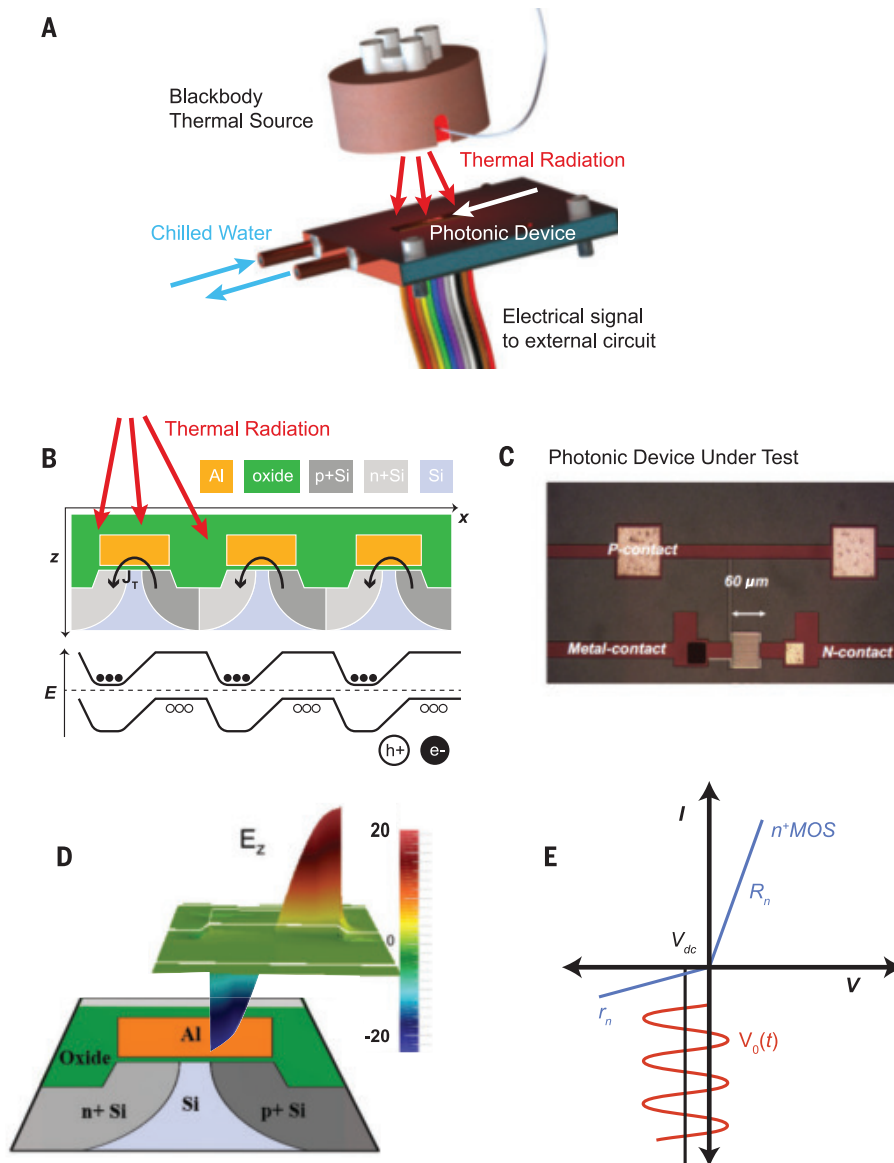


Fig. 1. Bipolar grating-coupled tunnel diode thermal photovoltaic device. (A) Illustration of thermal illumination of a bipolar thermal photovoltaic device in a vacuum radiometry setup. Device is packaged and mounted on a chilled stage with temperature stabilized at 20°C. (B) Schematic of multiple periods of bipolar thermal photovoltaic device, illustrating the charge-pumping mechanism. E , energy; h+, holes; e-, electron. (C) Image of actual bipolar grating-coupled tunnel diode at resonance with front-side contacting scheme. (Grating area is 60 μm by 60 μm.) (D) The modeled transverse spatial field profile in a thin tunnel barrier at peak field confinement. This confined field leads to the driven photon-assisted tunneling. (E) The model IV tunneling characteristic for the n+ MOS tunnel diode. R_n and r_n are the diode resistances in forward and reverse bias and the rectification of the tunneling current. (Model p+ MOS leads to similar IV characteristics.)

blackbody source at temperature T_s in a narrow bandwidth, $\Delta\nu$, is $S = dP/dA$, where $S = M_0(\nu, T_s)\Delta\nu$ with $M_0(\nu, T_s)$ being the spectral exitance of a blackbody in free space (eq. 15 in the supplementary materials). The incident free space electric field amplitude is obtained from the Poynting flux, $S = |E_0(\omega)|^2/2Z_0$, where Z_0 is the permittivity of free space.

The resulting instantaneous voltage at any point along the metal width is

$$V(x, t) = V_{dc} - V_m e^{i\omega t} \cos\left(\frac{\pi}{w}\left(x + \frac{w}{2}\right)\right) \quad (1)$$

the complex voltage profile across the device (Fig. 2B). A small dc bias (V_{dc}) is added to the harmonic time-varying term to ac-

count for self-bias of the device. Here, $V_m = d\gamma\sqrt{2Z_0M_0(\nu, T_s)\Delta\nu}$ is the ENZ enhanced blackbody induced voltage, and d is the tunnel oxide thickness. Estimates for the incident voltage amplitude in a fixed bandwidth from the blackbody source are shown in Fig. 2C. The node of the voltage in the model oscillates around the center of the metal. The voltage node, $x_c(t)$, is a function of the time, but we can use the two end points for the limits of integration for the spatially averaged current shown in Fig. 2B. Explicitly, we find that

$$x_c^\pm = \pm\left(-\frac{w}{2} + \frac{w}{\pi} \cos^{-1}\left(\frac{V_{dc}}{V_m}\right)\right) \quad (2)$$

which splits the current into two regions that, in general, vary with time. We can define two instantaneous current amplitudes as

$$I_n(t) = \int_{-w/2}^{x_c^+(t)} \frac{dx}{w} \frac{V(x, t)}{Z_n(t)} \quad (3)$$

$$I_p(t) = \int_{x_c^-(t)}^{w/2} \frac{dx}{w} \frac{V(x, t)}{Z_p(t)} \quad (4)$$

where Z_n and Z_p are the time-dependent diode impedances for each section, respectively. The half-cycle time average of the current amplitudes is of interest. Because the nodal position of the voltage and the impedance of the n and p sections vary with time, we need to make some assumptions to simplify the integration. To do so, we split the integration into two quarter cycles; cycle 1, where $0 \leq t < T/4$, and cycle 2, where $T/4 \leq t < T/2$. For cycle 1, the nodal position is assumed to be x_c^+ for the whole cycle, and impedance is constant and given by $Z_n = R_n$ and $Z_p = R_p$ in this time interval (R_n and R_p are the forward bias tunnel resistances). Both tunnel diodes are forward-biased, and a large tunnel current flows from the metal to the n+ region, I_{fn} , and from p+ to the metal, I_{fp} . Likewise, for cycle 2, the nodal position is assumed to be x_c^- , and impedance is constant and given by $Z_n = r_n$ and $Z_p = r_p$ in this time interval (r_n and r_p are the reverse bias resistances for the n and p tunnel diodes, respectively; see Fig. 2, D and E, for the resistances). The two tunnel diodes are reverse-biased, and small backflow current I_{mn} and I_{mp} is observed. The magnitude and direction of the tunnel currents are indicated by arrows in Fig. 2B. The real parts of the half-cycle averaged currents are

$$I_n = \left[\frac{V_{dc}}{4} \left(\frac{1}{R_n} + \frac{1}{r_n} \right) + \frac{x_c}{w} \frac{V_{dc}}{2} \left(\frac{1}{R_n} - \frac{1}{r_n} \right) - \frac{\sqrt{V_m^2 - V_{dc}^2}}{\pi^2} \left(\frac{1}{R_n} - \frac{1}{r_n} \right) \right] \quad (5)$$

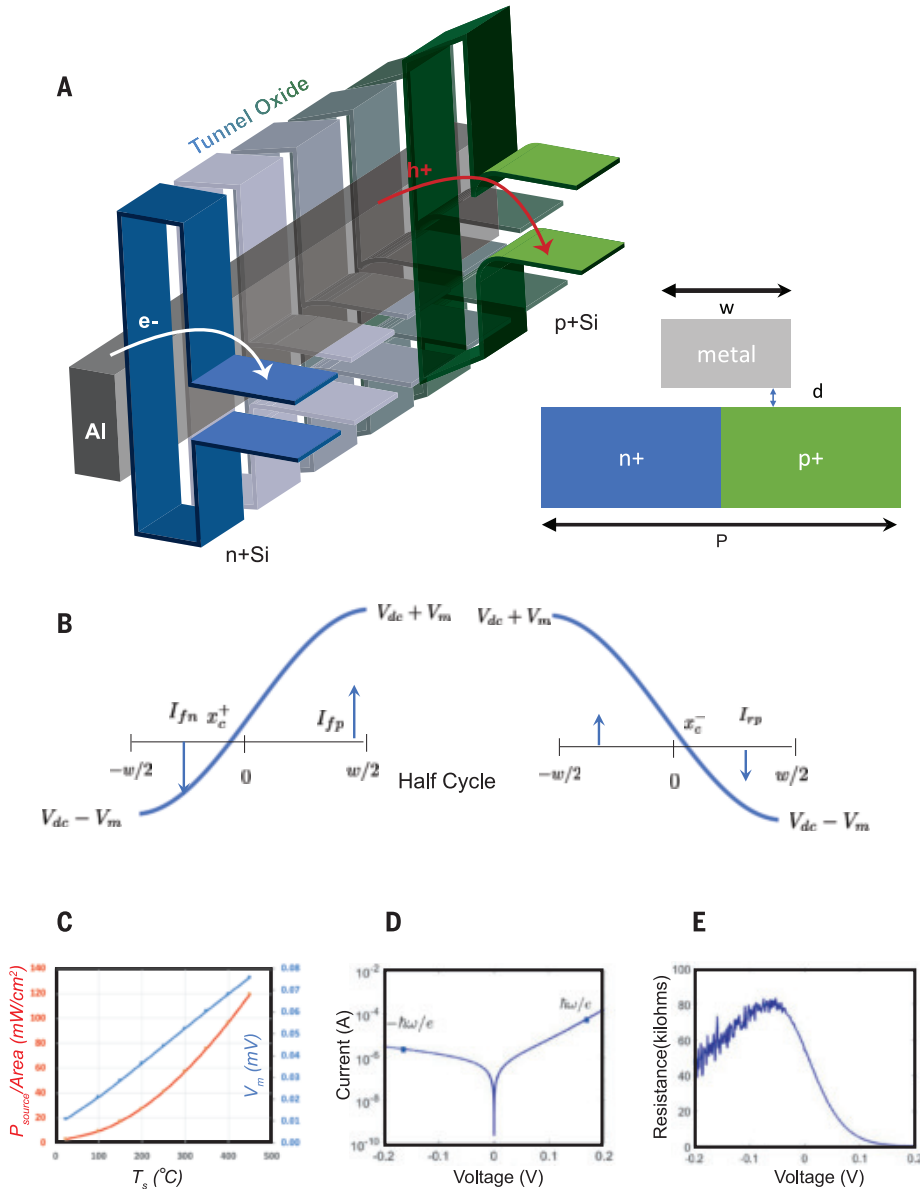


Fig. 2. Bipolar grating-coupled tunnel diode model. (A) Equilibrium band diagram of the bipolar device under a metal gate showing electron and hole particle currents. Inset shows unit cell geometry. The period of the grating (P) is $3 \mu\text{m}$, the metal width (w) is $1.8 \mu\text{m}$, and the tunnel oxide thickness (d) is 3 to 4.5 nm . (B) The instantaneous voltage profile in the device at $t = 0$ and at $t = T/2$. The spatially varying currents occur in both n and p+ regions, and the voltage node shifts to a negative x position. The half-period instantaneous voltage profile and currents across the device. The voltage node shifts to a positive x position. (C) Integrated blackbody source power per unit area (red curve) for bandwidth between $c/8.0 \mu\text{m}$ and $c/7.0 \mu\text{m}$ with field enhancement $\gamma = 20$ and $d = 4 \text{ nm}$. The blue curve is the associated ac voltage amplitude V_m . (D) Measured tunnel diode characteristic for typical n+ MOS tunnel diode with resonant PAT single-photon voltage marked. (E) Extracted resistance from n+ MOS tunnel diode. $R_n \approx 200 \text{ ohms}$ and $r_n \approx 50,000 \text{ ohms}$ at the indicated photovoltages (Fig. 1E).

$$I_p = \left[\frac{V_{dc}}{4} \left(\frac{1}{R_p} + \frac{1}{r_p} \right) - \frac{x_c V_{dc}}{w} \left(\frac{1}{R_p} - \frac{1}{r_p} \right) + \frac{\sqrt{V_m^2 - V_{dc}^2}}{\pi^2} \left(\frac{1}{R_p} - \frac{1}{r_p} \right) \right] \quad (6)$$

and the imaginary parts of half-cycle averaged currents are

$$i_n = i \frac{\sqrt{V_m^2 - V_{dc}^2}}{\pi^2} \left(\frac{1}{R_n} - \frac{1}{r_n} \right) \quad (7)$$

$$i_p = -i \frac{\sqrt{V_m^2 - V_{dc}^2}}{\pi^2} \left(\frac{1}{R_p} - \frac{1}{r_p} \right) \quad (8)$$

The complex current is simply a result of the complex admittance of the tunnel diode $I = (G + iB)V$, where G is the conductance, B is the capacitive part, and $V = V_{dc} + V_0(x)e^{i\omega t}$. The self-consistent dc voltage can be estimated by the current-matching condition, which requires that the half-cycle currents be equal and in opposite directions, such that $I_n = I_p$ at V_{dc} . For matched tunnel diodes, $R_n = R_p$ and $r_n = r_p$, the self-consistent dc bias is $V_{dc} = V_m$. However, in general, these equations lead to a transcendental equation for the self-consistent dc bias voltage.

A key feature of the bipolar conversion device is the periodic buried pn junction under the metal. The interdigitated nature of the structure leads to essentially a periodic well structure for storage of charge that can be pumped by the combined action of the two forward-biased tunnel junctions, metal to n+ and p+ to metal, respectively. We can extend the analysis of grating-coupled diode array as a simple (ideal) pn junction photovoltaic by considering the total device current across the pn junction as $I = I_0(\exp(\beta eV) - 1) - I_{ph}$, where the photon-assisted tunnel current, I_{ph} , acts as a photocurrent in analogy with light absorption in the depletion region of the pn junction. Here, I_{ph} is given by Eq. 5 or 6, which are equal at the self-consistent bias point. The resulting open-circuit voltage across the pn junction is

$$V_{oc} = \frac{1}{\beta e} \ln \left(1 + \frac{I_{ph}}{I_0} \right) \quad (9)$$

Similarly, the short-circuit current is $I_{sc} = -I_{ph}$. The open-circuit voltage across the pn junction can far exceed the confined voltage amplitude in the tunnel barrier, because $V_{th} = kT/e \gg V_m$ (Fig. 2C). The larger the open-circuit voltage, the greater the power generation in the bipolar device. This model implies that the RC time constant of the pn junction is such that it cannot respond to the infrared (35-THz) input signal, and the device charges across a nearly static depletion capacitance.

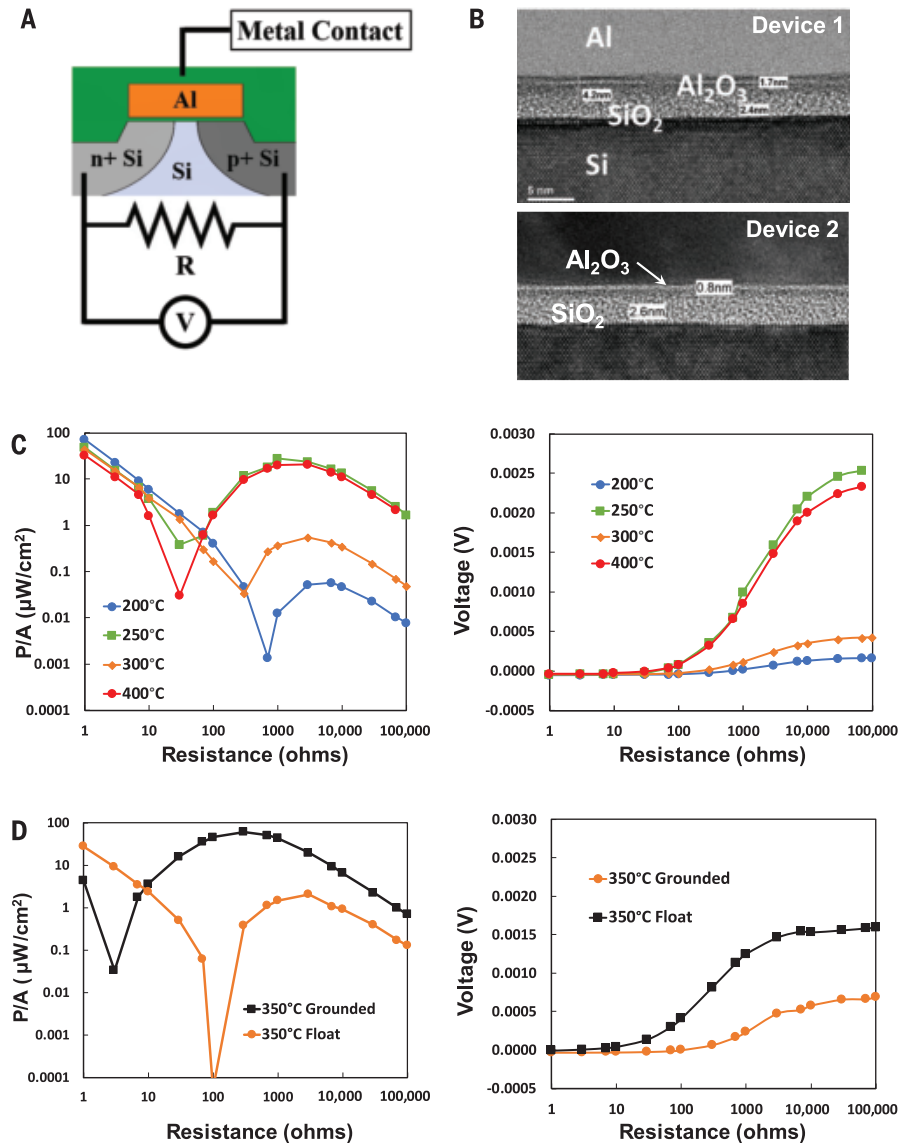


Fig. 3. Bipolar device power generation. (A) Circuit schematic of device contacts for power generation. (B) TEM cross sections through a nominal 4-nm gate oxide stack (device 1) and through a nominal 3.5-nm gate oxide stack (device 2). (C) Measured power density for device 1 as a function of the load resistance for various source temperatures and measured voltage across pn junction shorted by a load resistor versus load resistance for various source temperatures. (D) Measured power density for device 2 as a function of the load resistance at fixed source temperature for grounded and floating metal gates and measured voltage across pn junction shorted by a load resistor versus load resistance.

Experimental results

Electrical power generation from a moderate-temperature source is measured using a vacuum TPV setup (Fig. 1A) with an ~2-mm gap between sample and heat source, which has been previously described (20). The bipolar grating-coupled device active area is ~60 μm by 60 μm, and the grating pitch is 3.0 μm with a metal width of 1.8 μm (Fig. 2A). The device is a three-terminal device with interdigitated n and p regions with separate n, p, and metal contacts. The induced voltage from the thermal source at fixed temperature is measured

by shorting the pn junctions with a variable load resistor and measuring the induced voltage as a function of the load resistance with a nanovoltmeter (Fig. 3A).

The process and device parameters play a key role in the performance of the bipolar device. The oxide thickness and composition affect the tunneling resistance and the ENZ field concentration, respectively. Moreover, the implant conditions and thermal annealing cycles greatly affect the pn junction characteristics, such as resistance and depletion width from dopant diffusion under the metal

gate. Figure 3B shows two transmission electron microscope (TEM) cross-sectional images of two different device fabrication runs taken under the gate metal. Device 1 has a nominal targeted gate oxide of 4 nm, and device 2 has a nominal gate oxide thickness of 3.5 nm. For device 1, a nominal oxide thickness was targeted at 4 nm (4.2 nm measured), and we see a thin alumina layer formed at the SiO_2/Al interface. This alumina layer arises from contact silicide formation during thermal processing. Its makeup as alumina has been confirmed by energy-dispersive x-ray spectroscopy (EDS). Figure 3C shows the measured power density and the measured voltage as a function of load resistance at various source temperatures. The peak power density of $\sim 27 \mu\text{W}/\text{cm}^2$ occurs at source temperatures of 250° and 400°C , with an open-circuit voltage in excess of 2 mV for the two high-power cases. The peak power generation trend is not monotonic with increasing source temperature, and this can be attributed to the formation of an additional alumina layer. The impact of the thick alumina layer shifts the peak power density to lower source temperatures because the alumina LO phonon resonance occurs at roughly 200°C or $\lambda_{\text{LO}} = 10.1 \mu\text{m}$ —compare this with SiO_2 LO phonon mode at 400°C or $\lambda_{\text{LO}} = 8.1 \mu\text{m}$. There is therefore a complex interaction of the gate oxide LO phonon resonance and the device design parameters that determines the output power for this newly realized form of TPV conversion. This presents an opportunity to adjust the operational temperature of the device by tuning the LO phonon resonance.

Device 2 shown in Fig. 3B, from a different process run, targets an oxide thickness of 3.5 nm nominal (3.4 nm pictured). Figure 3D shows the measured power density generated and the measured voltage as a function of the load resistance measured across the pn junction for the nominal 3.5-nm device. The three-terminal nature of the device allows for measurement in two different wiring configurations. The default configuration has been to leave the metal contact floating. This was the condition in the previous measured device and in the current 3.5-nm device shown in orange. Alternatively, we can ground the metal contact by removing any buildup of charge on the gate, and the measured power density increases by a factor of $10\times$, shown in black (Fig. 3D). The peak power density for the 350°C source is $61 \mu\text{W}/\text{cm}^2$ and occurs at a load resistance of ~ 250 ohms. The peak measured open-circuit voltage for the grounded case is 1.6 mV compared with 0.5 mV for the ungrounded case. In both the 3.5- and 4-nm device cases, the open-circuit voltage is obtained at large load resistance and is shown to greatly exceed the predicted ac voltage amplitude shown in Fig. 2C. By comparison to a direct unipolar rec-

tenna, the open-circuit voltage cannot exceed the peak ac voltage amplitude. Thus, these bipolar devices far exceed the limit of direct rectification, suggesting that PAT and charge separation can be further improved through device and process optimization.

Outlook

Efficient conversion of moderate-temperature radiative thermal sources represents a largely untapped resource for energy harvesting. Microsystem-based radiative thermal-to-electrical energy converters built on our bipolar grating-coupled tunneling device represent a scalable compact energy-harvesting technology. These devices can be used as stand-alone energy converters or in conjunction with thermoelectric power generators, where they need only view the thermal source.

We have presented an alternative mechanism for thermal photovoltaic conversion that does not rely on square-law absorption but uses PAT to pump charge into n- and p-type wells in a bipolar grating-coupled device. A spatially and temporally varying enhanced transverse infrared electric field is confined in the tunnel barrier that results from resonant broadband coupling into the oxide LO phonon resonance. A simple device model for the gate-shunted photocurrent produced by PAT is developed, and the open-circuit voltage measured across the pn junction is seen to exceed the simulated peak ac voltage across the device. The charge separation by PAT in the bipolar device is akin to the Seebeck effect in a thermoelectric couple. The confined transverse infrared field drives PAT in a similar fashion as the temperature gradient in each leg of a thermoelectric couple creates the charge separation due to the internal electron and hole currents. In general, the ideal diode picture for the pn junction does not represent the observed IV characteristics across the pn junction owing to the nature of the diffused pn junction under the metal electrode and the abrupt junction in the field. Electrical power density improvements of several orders of magnitude ($\times 10^4$ to 10^5) are seen experimentally in the three-terminal bipolar device relative to recent direct unipolar rectenna devices (20). The best-performing device (device 2) had a measured electrical power density of $61 \mu\text{W}/\text{cm}^2$ from a 350°C thermal source. This results in an estimated conversion efficiency for a single conversion bandwidth of 1 THz of 0.4%, which is approaching TPV conversion efficiencies but at substantially cooler source temperatures (I) (see supplementary materials). Further improvements in power generation can be achieved using structured emitters on the thermal source with matched polarization-insensitive antenna designs and alternative gate dielectric materials in the antenna-coupled tunnel diode for ENZ reso-

nance matching to the thermal source. The full power of the modified CMOS device design and process optimization requires further exploration.

REFERENCES AND NOTES

- W. R. Chan et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5309–5314 (2013).
- E. L. Dereniak, G. D. Boreman, *Infrared Detectors and Systems* (Wiley Series in Pure and Applied Optics, John Wiley & Sons, 1996).
- A. Fiorino et al., *Nat. Nanotechnol.* **13**, 806–811 (2018).
- M. Lim et al., *J. Quant. Spectrosc. Radiat. Transf.* **210**, 35–43 (2018).
- B. Zhao et al., *Nano Energy* **41** (suppl. C), 344–350 (2017).
- S. P. Kwok, G. I. Haddad, G. Lobov, *J. Appl. Phys.* **42**, 554–563 (1971).
- A. Sanchez, C. F. Davis Jr., K. C. Liu, A. Javan, *J. Appl. Phys.* **49**, 5270–5277 (1978).
- A. Sanchez, S. K. Singh, A. Javan, *Appl. Phys. Lett.* **21**, 240–243 (1972).
- J. O. McSpadden, L. Fan, K. Chang, *IEEE Trans. Microw. Theory Tech.* **46**, 2053–2060 (1998).
- J. O. McSpadden, T. Yoo, K. Chang, *IEEE Trans. Microw. Theory Tech.* **40**, 2359–2366 (1992).
- Y.-H. Suh, K. Chang, *IEEE Trans. Microw. Theory Tech.* **50**, 1784–1789 (2002).
- M. Dagenais, K. Choi, F. Yesilkoy, A. N. Chryssis, M. C. Peckerar, *Proc. SPIE Int. Soc. Opt. Eng.* **7605**, 76050E (2010).
- P. S. Davids et al., *Nat. Nanotechnol.* **10**, 1033–1038 (2015).
- A. Sharma, V. Singh, T. L. Bougher, B. A. Cola, *Nat. Nanotechnol.* **10**, 1027–1032 (2015).
- D. R. Ward, F. Hüser, F. Pauly, J. C. Cuevas, D. Natelson, *Nat. Nanotechnol.* **5**, 732–736 (2010).
- S. Grover, O. Dmitriyeva, M. J. Estes, G. Moddel, *IEEE Trans. Nano Technol.* **9**, 716–722 (2010).
- P. Maraghechi, A. Foroughi-Abari, K. Cadien, A. Y. Elezzabi, *Appl. Phys. Lett.* **99**, 253503 (2011).
- P. Maraghechi, A. Foroughi-Abari, K. Cadien, A. Y. Elezzabi, *Appl. Phys. Lett.* **100**, 113503 (2012).
- E. A. Kadlec, R. L. Jarecki, A. Starbuck, D. W. Peters, P. S. Davids, *Phys. Rev. Appl.* **6**, 064019 (2016).
- J. Shank et al., *Phys. Rev. Appl.* **9**, 054040 (2018).
- M. Z. Alam, I. De Leon, R. W. Boyd, *Science* **352**, 795–797 (2016).
- A. Ali, M. G. Silveirinha, A. Salandrino, N. Engheta, *Phys. Rev. B* **75**, 155410 (2007).
- D. W. Berreman, *Phys. Rev.* **130**, 2193–2198 (1963).
- S. Vassant, J.-P. Hugonin, F. Marquier, J.-J. Greffet, *Opt. Express* **20**, 23971–23977 (2012).
- P. S. Davids, J. Shank, *Phys. Rev. B* **97**, 075411 (2018).
- P. K. Tien, J. P. Gordon, *Phys. Rev.* **129**, 647–651 (1963).

ACKNOWLEDGMENTS

P.S.D. thanks D. B. Burckel and R. Sanchez of Sandia for many useful discussions of this work. **Funding:** This work was funded by Sandia's Laboratory Directed Research and Development (LDRD) program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, a wholly owned subsidiary of Honeywell International Inc., for the United States Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States government. **Author contributions:** P.S.D. developed the device concept and design. A.S. and R.J. developed the process flow and fabricated the devices. J.K. and J.S. performed all experiments. D.P., J.S., and P.S.D. developed theory and simulated device performance. All authors contributed to the preparation of the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1341/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (27–30)

13 November 2019; accepted 6 February 2020
Published online 20 February 2020
10.1126/science.aba2089

STRUCTURAL BIOLOGY

Structural basis of G_s and G_i recognition by the human glucagon receptor

Anna Qiao^{1,2,3*}, Shuo Han^{1,2*}, Xinmei Li^{3,4*}, Zhixin Li^{5*}, Peishen Zhao^{6*}, Antao Dai^{1,7}, Rulve Chang⁵, Linhua Tai^{3,4}, Qiuxiang Tan^{1,2}, Xiaojing Chu^{1,2}, Limin Ma^{1,2}, Thor Seneca Thorsen⁸, Steffen Reedtz-Runge⁸, Dehua Yang^{1,7}, Ming-Wei Wang^{1,3,5,7,9}, Patrick M. Sexton^{5,6}, Denise Wootten^{5,6,†}, Fei Sun^{3,4,10,†}, Qiang Zhao^{2,3,11,†}, Beili Wu^{1,3,9,11,†}

Class B G protein-coupled receptors, an important class of therapeutic targets, signal mainly through the G_s class of heterotrimeric G proteins, although they do display some promiscuity in G protein binding. Using cryo-electron microscopy, we determined the structures of the human glucagon receptor (GCGR) bound to glucagon and distinct classes of heterotrimeric G proteins, G_s or G_{i1} . These two structures adopt a similar open binding cavity to accommodate G_s and G_{i1} . The G_s binding selectivity of GCGR is explained by a larger interaction interface, but there are specific interactions that affect G_i more than G_s binding. Conformational differences in the receptor intracellular loops were found to be key selectivity determinants. These distinctions in transducer engagement were supported by mutagenesis and functional studies.

Upon binding to extracellular agonists, G protein-coupled receptors (GPCRs) stimulate various signaling pathways by recruiting different heterotrimeric G proteins ($G\alpha\beta\gamma$) to mediate a wide variety of physiological functions (1). The selective coupling between a GPCR and specific G proteins is critical for the physiological action of the receptor in response to its endogenous ligands and therapeutic agents. However, the molecular details that define how an individual GPCR recognizes different G protein subtypes remain elusive. Class B GPCRs canonically exert their physiological actions by producing cyclic adenosine monophosphate (cAMP) through G_s signaling; however, they also couple to other G proteins such as $G_{i/o}$ and $G_{q/11}$, leading to diverse cellular responses (2–7). Recently, structures of four class B GPCRs bound to G_s were determined by single-particle cryo-electron microscopy (cryo-EM) (8–12), but the lack of structures with other G proteins limits our understanding of molecular mechanisms driving pleiotropic coupling and biased agonism that are important considerations for drug development.

The human glucagon receptor (GCGR), a member of the class B GPCR family, is critical to glucose homeostasis by triggering the release of glucose from the liver (13). Previous studies in native tissues and recombinant cell lines using different assays demonstrated that glucagon, in addition to promoting cAMP for-

mation, activates other downstream effectors that are pertussis toxin-sensitive or phospholipase C-dependent, revealing $G_{i/o}$ and $G_{q/11}$ signaling of GCGR (14–18). Selective activation of G_i in mouse hepatocytes in vivo was also reported to cause a pronounced increase in glucose production and severely impaired glucose homeostasis (19), which suggest that other subtypes of hepatic G proteins contribute to glucose regulation. There is interest in GCGR as a therapeutic target for type 2 diabetes and obesity (20). However, glucagon biology is complex; it can increase energy expenditure but high levels are diabetogenic, and drugs that selectively target GCGR are not currently available for treatment of diabetes and obesity. To better elucidate the molecular mechanisms underlying the G protein selectivity of GCGR, we determined the cryo-EM structures of GCGR in complex with its cognate ligand glucagon and heterotrimeric G_s or G_{i1} protein. These structures, combined with pharmacological data, provide important insights into GCGR activation, pleiotropic coupling, and G protein specificity.

Overall structures of G_s - and G_{i1} -bound GCGR

To obtain the GCGR- G_s complex, we replaced the native signal peptide of GCGR with that of hemagglutinin and removed 45 residues at the receptor C terminus (construct 1). Functional assays show that these modifications had little effect on glucagon binding and G_s and G_i acti-

vation of the receptor (fig. S1, A to C). To solve the GCGR- G_i structure, we further introduced three mutations—E126R, T200^{2,73b}W, and A366^{6,57b}M (construct 2)—to increase glucagon binding affinity and glucagon potency in G protein activation (fig. S1, A to C). [Superscripts refer to the Wootten numbering system for class B GPCRs, a modified form of the Ballesteros-Weinstein system for class A GPCRs (21).] These mutations may stabilize the receptor in a conformation favorable for G_i coupling and thus improve the stability and yield of the glucagon-GCGR- G_{i1} complex (fig. S1D). The glucagon-GCGR- G_s and glucagon-GCGR- G_{i1} structures were determined by cryo-EM single-particle analysis with an overall resolution of 3.7 Å and 3.9 Å, respectively (Fig. 1, figs. S2 and S3, and table S1) (22).

In the glucagon-GCGR- G_s and glucagon-GCGR- G_{i1} complexes, the glucagon binds at a site similar to that of the peptide in a structure of GCGR bound to the partial agonist NNC1702 (23) (fig. S4A). Structural differences in the peptide binding site between these structures occur in the region of the second and third extracellular loops (ECL2 and ECL3) and their connected transmembrane helices IV, V, VI, and VII (fig. S5, A to C). These conformational rearrangements may initiate the conformational changes of the receptor transmembrane helical bundle on the extracellular side that accompany receptor activation and transducer coupling in both the G_s - and G_{i1} -bound complexes (22).

A common G protein-binding pocket for G_s and G_{i1}

The intracellular half of the receptor in the glucagon-GCGR-G protein structures exhibits conformational changes relative to the inactive GCGR structure. The intracellular tip of helix VI moves away from the central axis of the helical bundle by ~19 Å (fig. S5D). Furthermore, to create a binding cavity for the G proteins, the intracellular ends of helices V and VII move outward by 8 Å and 2 Å, respectively. These conformational transitions are conserved regardless of the class of G protein that is coupled, generating a common binding pocket for both G_s and G_i (Fig. 2A).

In contrast to the common binding pocket of GCGR for G_s and G_{i1} , structures of class A GPCR-G protein complexes revealed differential positioning of helix VI (Fig. 2B), leading to proposals that the positional difference of helix VI is a major determinant for the coupling

¹CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China. ²State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China. ³University of Chinese Academy of Sciences, Beijing 100049, China. ⁴National Laboratory of Biomacromolecules, National Center of Protein Science–Beijing, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

⁵School of Pharmacy, Fudan University, Shanghai 201203, China. ⁶Drug Discovery Biology and Department of Pharmacology, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria 3052, Australia. ⁷National Center for Drug Screening, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China. ⁸Novo Nordisk A/S, Måløv 2760, Denmark. ⁹School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China. ¹⁰Center for Biological Imaging, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ¹¹CAS Center for Excellence in Biomacromolecules, Chinese Academy of Sciences, Beijing 100101, China.

*These authors contributed equally to this work.

†Corresponding author. Email: denise.wootten@monash.edu (D.W.); feisun@ibp.ac.cn (F.S.); zhaoq@sim.ac.cn (Q.Z.); beiliwu@sim.ac.cn (B.W.)

specificity of G_s and $G_{i/o}$ in GPCRs (24–30). The displacement of the intracellular tip of helix VI in the glucagon-GCGR-G protein structures is larger than in any class A GCGR-G protein structures but is more similar to the G_s -bound structures (Fig. 2D and fig. S6) (22). The C-terminal $\alpha 5$ helix of $G\alpha$ subunits plays a key role in coupling selectivity (31, 32). The amino acid sequence of the $\alpha 5$ helix in $G\alpha_s$ and $G\alpha_i$ differs at positions G.H5.23 and G.H5.24 [common $G\alpha$ numbering system (33)] ($G\alpha_s$, Y^{G.H5.23} and E^{G.H5.24}; $G\alpha_{i/o}$, C^{G.H5.23} and G^{G.H5.24}). The bulkier residues in $G\alpha_s$ require a larger pocket than $G\alpha_i$ to accommodate packing of their side chains (Fig. 2, E and F). Accordingly, it was hypothesized that G_s and $G_{i/o}$ binding to GPCRs requires a different opening size of the intracellular binding cavity (28). This hypothesis was supported by molecular dynamics simulation studies on the β_2 adrenergic receptor (β_2 AR) in complex with C-terminal peptides derived from $G\alpha_s$ or $G\alpha_i$ (34).

Contrary to this hypothesis, the two GCGR-G protein structures as well as other G_s -bound class B GPCR structures (8–12) display a similar outward shift of helix VI, forming a common binding cavity for recognition of both G_s and G_i , where the backbone conformations overlay for both the receptor and the far C terminus of the $G\alpha$ $\alpha 5$ helix (Fig. 2, A and C, and fig. S4B). However, although GCGR couples to both G proteins through this common pocket, it does so with differing efficiencies (fig. S1, A and B, and fig. S7). The measured interaction interface formed between the $\alpha 5$ C terminus (residues G.H5.16 to G.H5.26) and GCGR is larger for $G\alpha_s$ (802 Å²) than for $G\alpha_i$ (551 Å²). Therefore, preferential coupling to G_s can be explained by the open G protein-binding pocket (relative to the class A GPCR- $G_{i/o}$ structures) that is required to accommodate canonical binding to the bulkier $\alpha 5$ helix in G_s , but may still allow interaction with the less bulky G_i $\alpha 5$ helix (Fig. 2, G and H, and fig. S7). This concept likely extends to other GPCRs where the size of the G protein-binding pocket in the receptor core may reflect the receptor's ability to couple to multiple G proteins—a theory that is consistent with recent studies where receptors that canonically couple to G_s (and $G_{q,11,12,13}$) are generally more promiscuous than those that are classified as G_i -coupled (31).

Intracellular loops mediate G protein recognition and specificity

Although the $\alpha 5$ helix of $G\alpha$ proteins is a key contributor to G protein selectivity, interactions with additional domains of the G protein also contribute to specificity (31). The interaction surface between GCGR and the $\alpha 5$ C terminus is larger for $G\alpha_s$ than for $G\alpha_i$; however, this surface only forms 60% of the interaction surface for $G\alpha_s$ (total interface 1276 Å², 1418 Å² including $G\beta$ interactions), whereas it contrib-

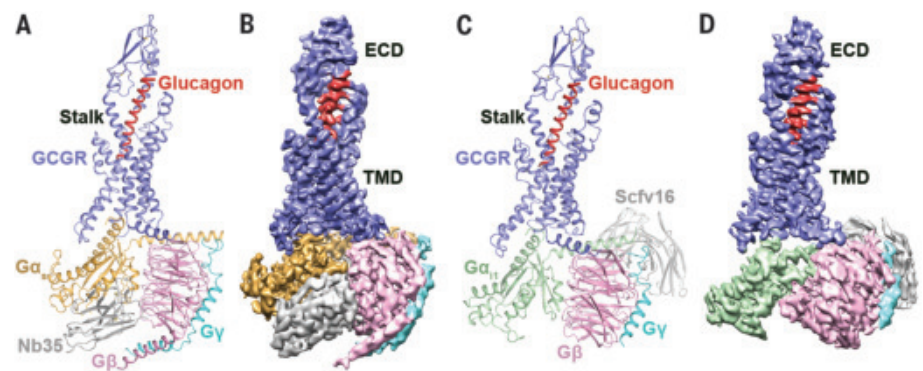


Fig. 1. Overall architectures of glucagon-GCGR-G protein complexes. (A) Cryo-EM structure of the glucagon-GCGR- G_s -Nb35 complex. Nb35 is a nanobody that stabilizes the interface between the $G\alpha_s$ subunit and $G\beta$ subunit. The structure is shown in cartoon representation. GCGR, glucagon, $G\alpha_s$, $G\beta$, $G\gamma$, and Nb35 are colored blue, red, gold, pink, cyan, and gray, respectively. The disulfide bonds are shown as yellow sticks. (B) Cryo-EM map of the glucagon-GCGR- G_s -Nb35 complex, colored according to chains. ECD, extracellular domain; TMD, transmembrane domain. (C) Cryo-EM structure of the glucagon-GCGR- G_{i1} -Scfv16 complex. Scfv16, the single-chain variable fragment of mAb16, stabilizes the GPCR- G_i complex by recognizing an epitope composed of the αN helix of $G\alpha_{i1}$ and the $G\beta$ subunit. $G\alpha_{i1}$ is colored green; Scfv16 is in gray. (D) Cryo-EM map of the glucagon-GCGR- G_{i1} -Scfv16 complex, colored according to chains.

utes 80% of the interaction surface for $G\alpha_i$ (total interface 687 Å², 863 Å² including $G\beta$ interactions). Combined with G protein activation and signaling assays, the G_s - and G_{i1} -bound GCGR structures suggest that the intracellular loops of the receptor play critical roles in G protein engagement and specificity.

Comparison of the two glucagon-GCGR-G protein structures revealed a difference in the position of the $G\alpha$ αN helix relative to the receptor (Fig. 3A). This N-terminal helix shifts toward the receptor in the G_{i1} -bound structure compared to that in the G_s -bound structure. This movement, along with the substitution of A39^{G.hnsl.3} ($G\alpha_s$) with R32^{G.hnsl.3} ($G\alpha_i$) at the interaction interface, is associated with a difference in the conformation of the second intracellular loop (ICL2) that alters the αN -ICL2 interface (Fig. 3A and fig. S7, F and L). In the G_s -bound structure, ICL2 forms extensive interactions with the αN helix, with A256, T257, L258, and E260 forming interactions in the binding groove between the αN helix and the $\beta 1$ strand and $\alpha 5$ helix of $G\alpha_s$ (Fig. 3B). In contrast, when bound to G_{i1} , ICL2 adopts a position farther away from the G protein and makes only limited contact with R32^{G.hnsl.3} in the αN helix of $G\alpha_i$ (Fig. 3C). To investigate the role of ICL2 in activating different G protein subtypes, we assessed glucagon-induced G_s and G_{i1} activation by the wild-type and mutant GCGRs using NanoLuc Binary Technology (NanoBit) (35), which measures the proximal interaction between the α and γ subunits of the G protein. In agreement with the conformational difference of ICL2, mutations L258A and E260A decreased the half-maximal effective concentration (EC₅₀) of glucagon-induced G_s activation by factors of 29 and 16, respectively, whereas they showed a much less pro-

nounced effect on G_i activation (factor of 6 to 8 reduction of EC₅₀) (Fig. 3, F and G; fig. S8, A and D; and table S2). In all previously published GPCR-G protein structures, where the receptors couple to their cognate G protein or a noncognate G protein with comparably high affinity (NTSR1) (36), ICL2 forms extensive interactions with the $G\alpha$ subunit. By contrast, the limited contact between ICL2 and $G\alpha_i$ observed in the G_{i1} -bound GCGR structure most likely contributes to the lower potencies of glucagon in stimulating G_i activation and signaling when compared to G_s . The above data indicate that ICL2 is crucial for the G protein specificity of GCGR.

In contrast to the importance of ICL2 in G_s coupling, other intracellular regions behave as selective determinants for G_i binding. The NanoBit assay showed that the alanine replacement of the residue F263^{4.48b} at the intracellular end of helix IV, which potentially makes contacts with the $G\alpha_i$ αN helix due to the upward shift of this N-terminal helix in G_{i1} relative to that in G_s (Fig. 3, B and C), displayed a notable loss of G_{i1} activation but a wild-type level of G_s activation (Fig. 3, F and G; fig. S8, A and D; and table S2). In association with the movement of the αN helix, the linker region between the $\alpha 4$ helix and $\beta 6$ strand of the $G\alpha_i$ subunit approaches the third intracellular loop (ICL3) of GCGR in the glucagon-GCGR- G_{i1} complex (Fig. 3D). This was reflected by a notable decrease of glucagon potency in G_{i1} activation for the GCGR mutant H339A, which only slightly alters G_s activation (factor of 3 reduction of EC₅₀) (Fig. 3, F and G; fig. S8, A and D; and table S2). Furthermore, accompanying the positional difference of $G\alpha_i$, the $G\beta$ and $G\gamma$ subunits shift closer to the receptor in the glucagon-GCGR- G_{i1} structure relative to the G_s -bound structure,

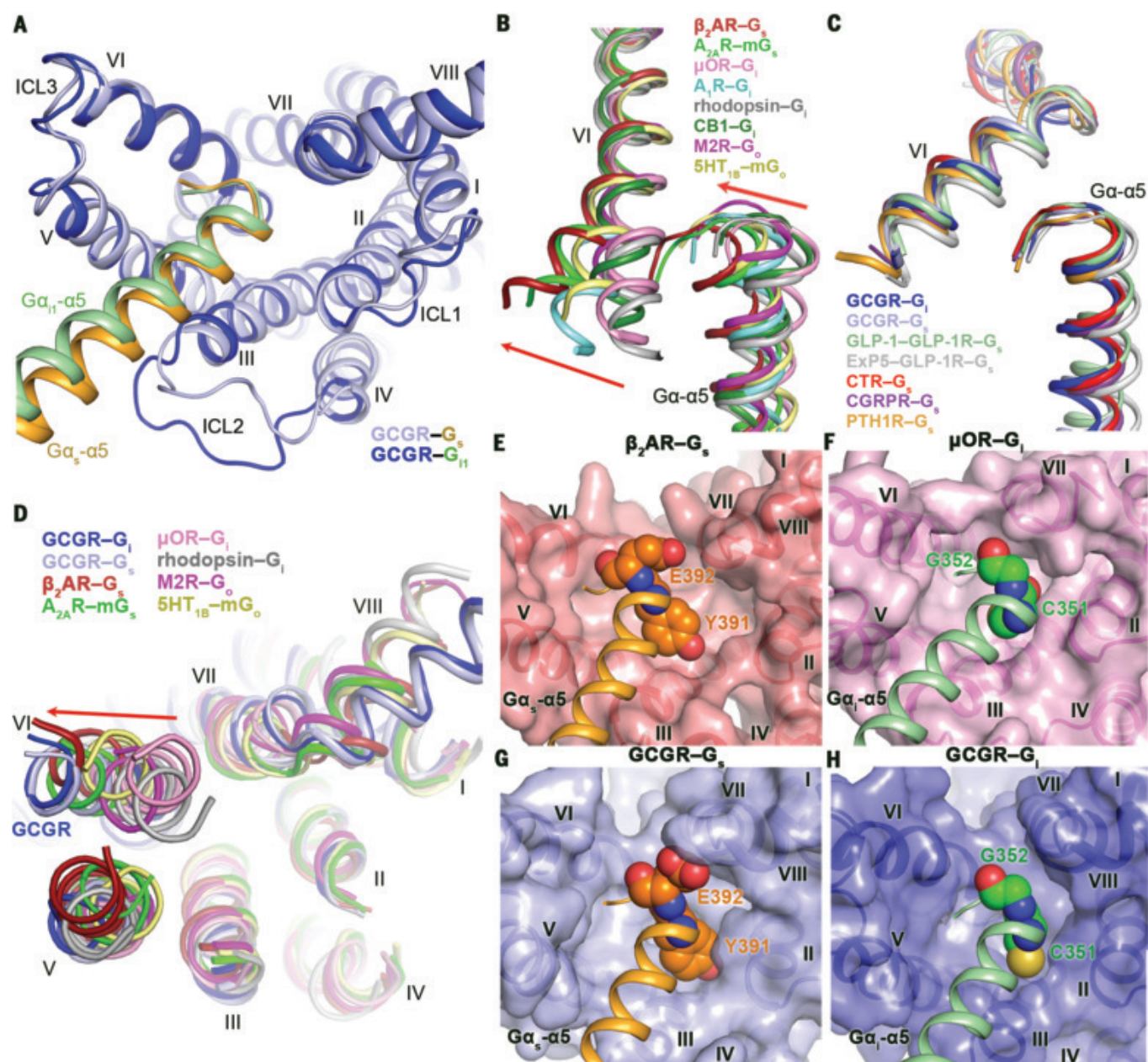


Fig. 2. Structural comparison of G protein-bound GPCR structures.

(A) Comparison of the transmembrane helical bundle conformation and the position of $G\alpha$ $\alpha 5$ helix C terminus in the G protein-bound GCGR structures. The glucagon-GCGR- G_s and glucagon-GCGR- G_{i1} structures are shown in cartoon representation in an intracellular view. The G_s -bound structure is colored light blue (GCGR) and gold ($G\alpha_s$); the G_{i1} -bound structure is colored dark blue (GCGR) and green ($G\alpha_{i1}$). (B) Comparison of the positions of helix VI and the C terminus of $G\alpha$ $\alpha 5$ helix in the G_s - and $G_{i/o}$ -bound class A GPCR structures. The structures of β_2 AR- G_s , A_{2A} R-mini- G_s (mG_s), μ OR- G_i , A_1 R- G_i , rhodopsin- G_i , CB1- G_i , M2R- G_o , and 5HT $_{1B}$ -mini- G_o (mG_o) (PDB IDs: 3SN6, 6GDG, 6DDE, 6D9H, 6CMO, 6N4B, 6OIK, and 6G79) are colored dark red, green, pink, cyan, gray, dark green, magenta, and yellow, respectively. Only helix VI of the receptors and $G\alpha$ $\alpha 5$ helix are shown for clarity. The red arrows indicate the outward tilts of helix VI and the relative shift of the C terminus of $\alpha 5$ helix. (C) Comparison of the positions of helix VI and the C terminus of $G\alpha$ $\alpha 5$ helix in the G protein-bound class B GPCR structures. The structures of glucagon-GCGR- G_{i1} and glucagon-GCGR- G_s and the structures of GLP-1-GLP-1R- G_s , EXP5-GLP-1R- G_s , CTR- G_s , CGRPR- G_s , and

PTH1R- G_s (PDB IDs: 5VAI, 6B3J, 5UZ7, 6E3Y, and 6NBF) are colored dark blue, light blue, light green, light gray, red, purple, and orange, respectively. Only helix VI of the receptors and $G\alpha$ $\alpha 5$ helix are shown for clarity. (D) Comparison of the receptor helical bundles in the G protein-bound GCGR structures and class A GPCR-G protein structures in an intracellular view. The structures of glucagon-GCGR- G_{i1} , glucagon-GCGR- G_s , β_2 AR- G_s , A_{2A} R- mG_s , μ OR- G_i , rhodopsin- G_i , M2R- G_o , and 5HT $_{1B}$ - mG_o are colored dark blue, light blue, dark red, green, pink, gray, magenta, and yellow, respectively. Only the receptors in the structures are shown. The red arrow indicates the larger outward displacement of the intracellular tip of helix VI in GCGR compared to that in the G protein-bound class A GPCR structures. (E to H) Binding pocket for the $G\alpha$ - $\alpha 5$ C terminus. (E) β_2 AR- G_s ; (F) μ OR- G_i ; (G) GCGR- G_s ; (H) GCGR- G_{i1} . The $G\alpha$ residues at positions G.H5.23 and G.H5.24 are shown as spheres. The $\alpha 5$ helices in $G\alpha_s$ and $G\alpha_i$ are colored gold and green, respectively. The receptors are shown in cartoon and surface representations in an intracellular view. Amino acid abbreviations here or elsewhere: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; N, Asn; Q, Gln; R, Arg; T, Thr; V, Val; W, Trp; Y, Tyr.

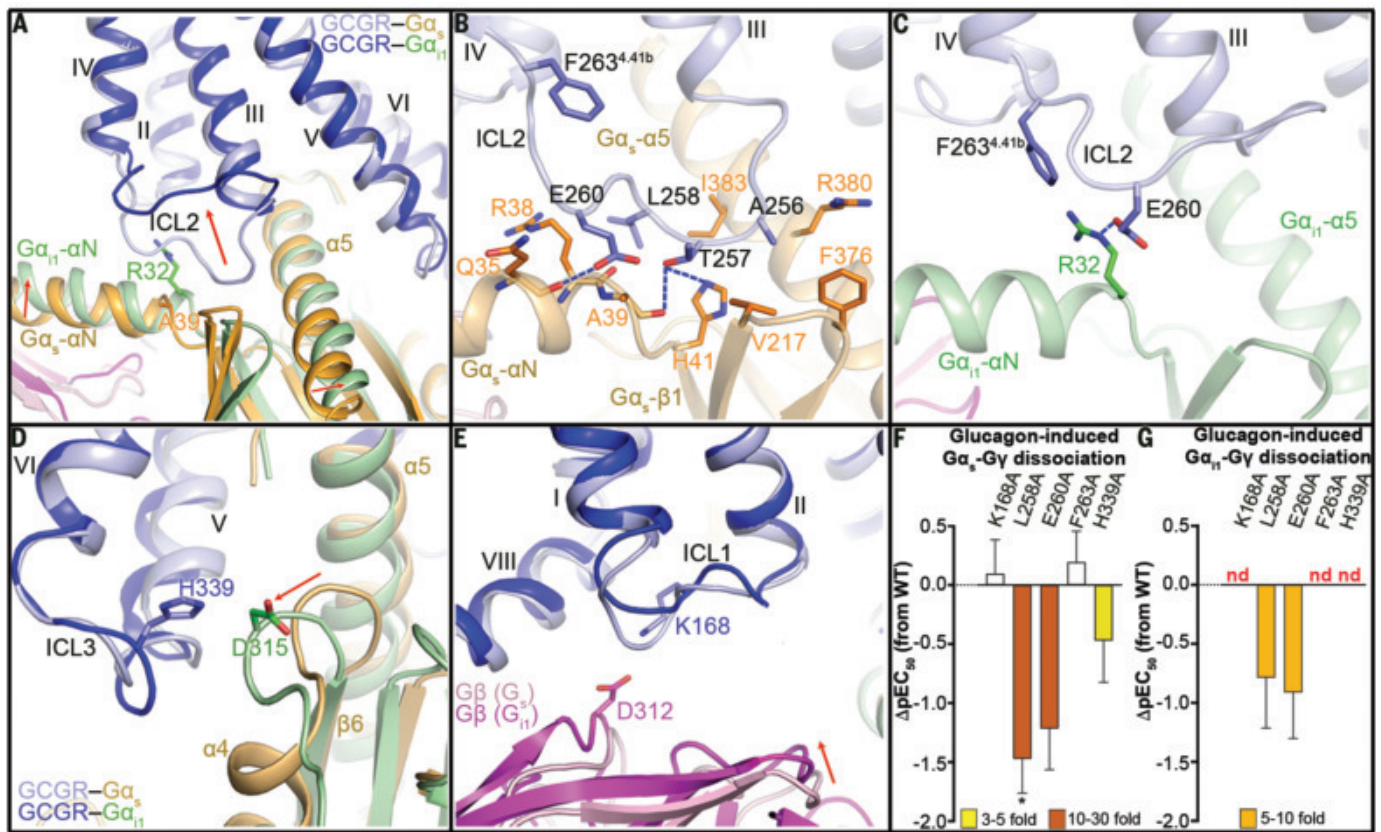


Fig. 3. G protein-binding interface mediated by GCGR intracellular loops.

(A) Comparison of ICL2 conformation in the glucagon-GCGR- G_s and glucagon-GCGR- G_{i1} structures. The glucagon-GCGR- G_s structure is colored light blue (GCGR) and gold ($G\alpha_s$); the glucagon-GCGR- G_{i1} structure is colored dark blue (GCGR) and green ($G\alpha_{i1}$). The $G\alpha_s$ residue A39^{G.hms1.3} and the $G\alpha_i$ residue R32^{G.hms1.3} are shown as sticks. The red arrows indicate the movements of GCGR ICL2, $G\alpha$ $\alpha 5$ helix N terminus, and αN helix in the G_{i1} -bound structure relative to the G_s -bound structure. (B) Interactions between ICL2 and $G\alpha_s$. The residues involved in interactions are shown as sticks and are colored blue (GCGR) and orange ($G\alpha_s$). Polar interactions are shown as blue dashed lines. (C) Interactions between ICL2 and $G\alpha_{i1}$. The residues involved in interactions are shown as sticks and are colored blue (GCGR) and green ($G\alpha_{i1}$). (D) Conformational difference of the linker between the $\alpha 4$ helix and $\beta 6$ strand in $G\alpha$. The GCGR ICL3 residue H339 and the $G\alpha_i$ residue D315 that form a contact in the glucagon-GCGR- G_{i1} structure are shown as sticks. The red arrow indicates the movement

of the $\alpha 4$ - $\beta 6$ linker in the G_{i1} -bound structure relative to the G_s -bound structure. (E) Conformational difference of $G\beta$. The $G\beta$ subunits in the two structures are colored pink (G_s) and magenta (G_{i1}). The GCGR ICL1 residue K168 and the $G\beta$ residue D312 that form a contact in the glucagon-GCGR- G_{i1} structure are shown as sticks. The red arrow indicates the movement of $G\beta$ in the G_{i1} -bound structure relative to the G_s -bound structure. (F and G) Glucagon-induced G_s and G_{i1} activation assays using NanoBiT. (F) G_s activation; (G) G_{i1} activation. Bars represent differences in calculated glucagon potency (pEC_{50}) for each mutant relative to the wild-type receptor (WT). Data are colored according to the extent of effect (yellow, factor of 3 to 5 reduction of EC_{50} ; gold, factor of 5 to 10 reduction of EC_{50} ; orange, factor of 10 to 30 reduction of EC_{50}). Data are means \pm SEM from at least three independent experiments performed in technical triplicate; nd, not determined. * $P < 0.05$ [one-way analysis of variance (ANOVA) followed by Dunnett's posttest, compared with the response of WT]. See table S2 for detailed statistical evaluation and expression level.

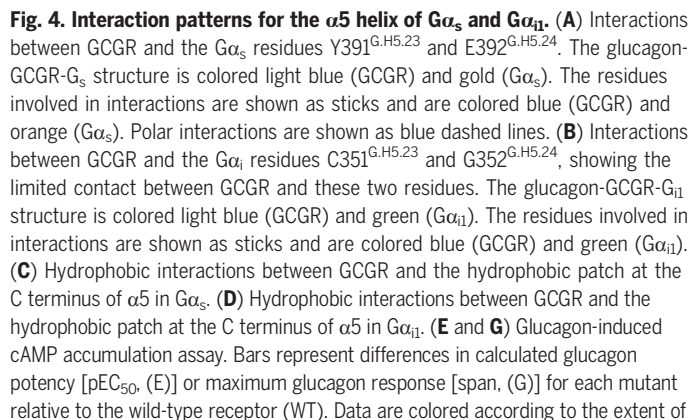
introducing an additional interaction interface between GCGR and G_i mediated by the first intracellular loop (ICL1) of the receptor and $G\beta$ (Fig. 3E and fig. S7, E and K). This binding interface was supported by our mutagenesis studies showing that the K168A mutation abolished G_{i1} activation but had no effect on G_s activation (Fig. 3, F and G; fig. S8, A and D; and table S2). Taken together, the G protein-bound GCGR structures demonstrate that individual intracellular loops play different roles in governing G protein recognition and specificity.

Recognition patterns for the C-terminal $\alpha 5$ helix of $G\alpha_s$ and $G\alpha_{i1}$

Despite the overall similarity in the backbone of the intracellular binding cavity, the

two GCGR complex structures exhibit different molecular details in the recognition patterns for the C termini of $G\alpha_s$ and $G\alpha_{i1}$, whereby $G\alpha_{i1}$ forms more limited interactions that are largely hydrophobic and $G\alpha_s$ forms more extensive interactions, both polar and hydrophobic (Fig. 4, A to D, and fig. S7). Two highly conserved class B GPCR polar networks, the HETX motif (H^{2.50b}, E^{3.50b}, T^{6.42b}, and Y^{7.57b}) and the helix II-VI-VII-VIII network (R^{2.46b}, R/K^{6.37b}, N^{7.61b}, and E^{8.41b}), at the intracellular face of the receptor have been suggested to play critical roles in modulating conformational change upon receptor activation (10, 37). In the G_s -bound GCGR structure, the bulky residue Y391^{G.H5.23} at the $\alpha 5$ helix C terminus binds to a subpocket formed by R173^{2.46b}, H177^{2.50b}, E245^{3.50b}, Y248^{3.53b},

L249^{3.54b}, and Y400^{7.57b} in GCGR (Fig. 4A). In contrast, the residue at position G.H5.23 is a cysteine (C351^{G.H5.23}) in $G\alpha_i$. Without the bulky side chain, this residue only forms weak hydrophobic contacts with R173^{2.46b} and L249^{3.54b} in GCGR (Fig. 4B). Despite the different interaction modes, the mutations R173^{2.46b}A, H177^{2.50b}A, E245^{3.50b}A, and Y400^{7.57b}A all impaired glucagon-induced cAMP production and glucagon-induced inositol phosphate (IP) accumulation using a chimeric $G\alpha$ protein, $G\alpha_{q19}$ (32, 38) (Fig. 4, E to H; fig. S8, G and J; and tables S3 and S4). This latter assay allowed a cAMP-independent interrogation of the effect of mutants on the C-terminal nine amino acids of $G\alpha_i$ that constitute most of the interaction surface in the G_i complex structure (Fig. 4, B



SCIENCE

and D, and fig. S1E). Additionally, the NanoBiT assay revealed that these alanine replacements heavily impaired both G_s and G_{i1} activation (Fig. 4, I to L; fig. S8, B and E; and table S2). The effect of these mutations on G_s and G_i activation could be explained by the disruption of the interaction networks within the receptor and impairment of the global conformational rearrangement that is required for G protein recognition and signaling. Nonetheless, in the NanoBiT G protein activation assay, there was a greater impact of the H177^{2.50b}A mutant on G_s than on G_{i1} (Fig. 4, I to L, and fig. S8, B and E), which may reflect the additional loss of direct interaction that occurs with G_{α_s} Y391^{G.H5.23}.

The interaction patterns of GCGR with G_s and G_i also differ for the residue at position G.H5.24 of the $G\alpha$ $\alpha 5$ helix. The side chain of the G_{α_s} residue E392^{G.H5.24} is within interaction distance of N404^{7.61b} and K405^{7.62b} at the hinge region between helices VII and VIII in the glucagon-GCGR- G_s structure (Fig. 4A and fig. S7B), whereas these interactions are not possible in the G_i complex because of glycine (G352^{G.H5.24}) substitution in G_{α_i} (Fig. 4B and fig. S7H). In the NanoBiT assay, a factor of 31 reduction of the EC₅₀ in glucagon-induced G_s activation occurred for the mutant N404^{7.61b}A, which in contrast had little effect on G_i activation (factor of 4 reduction) (Fig. 4, I and J; fig. S8, B and E; and table S2). However, the mutant K405^{7.62b}A was not different from the wild-type GCGR for either G protein, which suggests that interaction with the side chain of this residue is less critical for engagement with G_s .

In both G_s - and G_{i1} -bound GCGR structures, a G protein-binding cavity formed by a cluster of hydrophobic residues from helices III, V, and VI is observed at the intracellular face of the receptor transmembrane domain. It recognizes different hydrophobic patches at the $\alpha 5$ C terminus in G_{α_s} and G_{α_i} (G_{α_s} : L388^{G.H5.20}, Y391^{G.H5.23}, L393^{G.H5.25}, and L394^{G.H5.26}; G_{α_i} : I344^{G.H5.16}, L348^{G.H5.20}, L353^{G.H5.25}, and F354^{G.H5.26}) (Fig. 4, C and D). The importance of this hydrophobic cavity in receptor signaling was reflected in our mutagenesis studies, where introduction of alanine or tryptophan mutation within the cavity not only decreased glucagon potency on G_s signaling (Fig. 4, E and G; fig. S8, H and I; and table S3) but also reduced G_{q19} -mediated IP production (Fig. 4, F and H; fig. S8, K and L; and table S4). Of note, tryptophan mutations within this pocket were more detrimental for G_s signaling than were alanine mutations (Fig. 4E and fig. S8, H and I); this was not the case for G_i , where some tryptophan mutations such as Y248^{3.53b}W and L328^{5.60b}W were better tolerated than alanine (Fig. 4, F and H, and fig. S8, K and L). The distinct effects of the tryptophan mutants on G_s and G_i activation were also observed in the NanoBiT assay, where L328^{5.60b}W and L329^{5.61b}W exhibited

larger effects on G_s activation [factor of 6 to 8 reduction of potency and 40 to 80% reduction in maximal responses (E_{\max})] than their alanine replacements (factor of 2 to 3 reduction of potency and no reduction in E_{\max}) (Fig. 4, I and K; fig. S8C; and table S2), but again were better tolerated in the assay of G_{i1} activation (Fig. 4, J and L; fig. S8F; and table S2). Consistent with these results, the mutation L354^{6.45b}W substantially reduced both glucagon potency and E_{\max} in G_s activation, whereas it showed much less influence on G_{i1} activation (factor of 10 reduction of potency and no reduction in E_{\max}) (Fig. 4, I to L; fig. S8, C and F; and table S2). Increasing the size of the hydrophobic residues reduces the size of the intracellular pocket, and this is more detrimental to binding of the bulkier and more polar G_s $\alpha 5$ C terminus.

Most of the alanine replacements—Y248^{3.53b}A, L249^{3.54b}A, L253^{3.58b}A, L328^{5.60b}A, L329^{5.61b}A, and L354^{6.45b}A—reduced the maximum level of G_{q19} -mediated IP production by 50 to 90% (Fig. 4H, fig. S8K, and table S4). These alanine mutants retained E_{\max} values in cAMP accumulation that were similar to that of the wild-type receptor, but decreased potency of glucagon was observed (by a factor of 2 to 9) (Fig. 4, E and G; fig. S8H; and table S3). Although some of these alanine mutants had a similar overall effect on G_s and G_i signaling, Y248^{3.53b}A, L249^{3.54b}A, L253^{3.58b}A, and L354^{6.45b}A were more detrimental for G_i signaling (Fig. 4, E to H, and tables S3 and S4). Similarly, in the NanoBiT assay, the mutations L328^{5.60b}A and L329^{5.61b}A reduced glucagon potency by a factor of >30 in G_{i1} activation but decreased the potency of G_s activation by only a factor of 2 to 3 (Fig. 4, I and J; fig. S8, C and F; and table S2). The different behaviors of these mutants indicate that disturbing the hydrophobic contact between the receptor and the $G\alpha$ $\alpha 5$ C terminus has a larger effect on G_i signaling than on G_s signaling. This aligns well with the fact that the interaction interface between GCGR and the $G\alpha$ C terminus (residues G.H5.16 to G.H5.26), which is mainly composed of hydrophobic residues, accounts for 80% of the GCGR- G_{α_i} interface but only about 60% of the total interface between GCGR and G_{α_s} . Thus, the hydrophobic cavity within the receptor intracellular face may play a more critical role for G_i recognition than that of G_s .

Collectively, this work provides a model for the diverse G protein signaling observed with class B GPCRs. The G protein-bound GCGR structures reveal that the less bulky G_i protein is accommodated in the large intracellular cavity but forms less extensive, predominantly hydrophobic, interactions, which account for G protein coupling specificity. Furthermore, there are specific conformational differences in the receptor, which also govern the nature of GCGR-G protein interactions and may mediate

biased agonism, including in the intracellular loops and individual residue side chains. Although there are studies that implicate G_i coupling of GCGR, physiological relevance remains unclear. Nonetheless, our structures of G_s and G_i bound to the same GPCR give an opportunity to study the basis of G protein specificity and offer new insights into the molecular details that govern pleiotropic GPCR-G protein coupling.

REFERENCES AND NOTES

1. A. J. Venkatakrisnan et al., *Nature* **494**, 185–194 (2013).
2. D. Wooten, L. J. Miller, C. Koole, A. Christopoulos, P. M. Sexton, *Chem. Rev.* **117**, 111–138 (2017).
3. J. Van Rampelbergh et al., *Biochim. Biophys. Acta* **1357**, 249–255 (1997).
4. T. Blank et al., *J. Neurosci.* **23**, 700–707 (2003).
5. D. K. Grammatopoulos, H. S. Rande, M. A. Levine, K. A. Kanellopoulou, E. W. Hillhouse, *J. Neurochem.* **76**, 509–519 (2001).
6. K. J. Culhane, Y. Liu, Y. Cai, E. C. Yan, *Front. Pharmacol.* **6**, 264 (2015).
7. K. E. Mayo et al., *Pharmacol. Rev.* **55**, 167–194 (2003).
8. Y. L. Liang et al., *Nature* **546**, 118–123 (2017).
9. Y. Zhang et al., *Nature* **546**, 248–253 (2017).
10. Y. L. Liang et al., *Nature* **555**, 121–125 (2018).
11. Y. L. Liang et al., *Nature* **561**, 492–497 (2018).
12. L. H. Zhao et al., *Science* **364**, 148–153 (2019).
13. K. M. Habegger et al., *Nat. Rev. Endocrinol.* **6**, 689–697 (2010).
14. Y. Xu, X. Xie, *J. Recept. Signal Transduct. Res.* **29**, 318–325 (2009).
15. T. Grady, M. Fickova, H. S. Tager, D. Trivedi, V. J. Hruby, *J. Biol. Chem.* **262**, 15514–15520 (1987).
16. E. C. Aromataris, M. L. Roberts, G. J. Barritt, G. Y. Rychkov, *J. Physiol.* **573**, 611–625 (2006).
17. L. H. Hansen et al., *Am. J. Physiol.* **274**, C1552–C1562 (1998).
18. J. D. Kilts et al., *Circ. Res.* **87**, 705–709 (2000).
19. M. Rossi et al., *J. Clin. Invest.* **128**, 746–759 (2018).
20. J. I. Bagger, F. K. Knop, J. J. Holst, T. Vilsbøll, *Diabetes Obes. Metab.* **13**, 965–971 (2011).
21. D. Wooten, J. Simms, L. J. Miller, A. Christopoulos, P. M. Sexton, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5211–5216 (2013).
22. See supplementary materials.
23. H. Zhang et al., *Nature* **553**, 106–110 (2018).
24. C. J. Draper-Joyce et al., *Nature* **558**, 559–563 (2018).
25. Y. Kang et al., *Nature* **558**, 553–558 (2018).
26. A. Koehl et al., *Nature* **558**, 547–552 (2018).
27. J. Garcia-Nafria, R. Nehmé, P. C. Edwards, C. G. Tate, *Nature* **558**, 620–623 (2018).
28. J. Garcia-Nafria, C. G. Tate, *Mol. Cell. Endocrinol.* **488**, 1–13 (2019).
29. K. Krishna Kumar et al., *Cell* **176**, 448–458.e12 (2019).
30. S. Maeda, Q. Qu, M. J. Robertson, G. Skiniotis, B. K. Kobilka, *Science* **364**, 552–557 (2019).
31. N. Okashah et al., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12054–12059 (2019).
32. B. R. Conklin, Z. Farfel, K. D. Lustig, D. Julius, H. R. Bourne, *Nature* **363**, 274–276 (1993).
33. T. Flock et al., *Nature* **524**, 173–179 (2015).
34. A. S. Rose et al., *J. Am. Chem. Soc.* **136**, 11244–11247 (2014).
35. A. S. Dixon et al., *ACS Chem. Biol.* **11**, 400–408 (2016).
36. H. E. Kato et al., *Nature* **572**, 80–85 (2019).
37. D. Wooten et al., *Biochem. Pharmacol.* **118**, 68–87 (2016).
38. C. Monnier et al., *EMBO J.* **30**, 32–42 (2011).

ACKNOWLEDGMENTS

We thank Y.-L. Liang for expert advice on class B GPCR-G protein complex biochemistry, and T. T. Truong and S. Darabalaie for assistance with the NanoBiT studies. The cryo-EM studies were performed at the Center for Biological Imaging (CBI, <http://cbi.ibp.ac.cn>), Institute of Biophysics, Chinese Academy of Sciences; we thank B.-L. Zhu and X. Huang from CBI for their help on cryo-EM data collection. **Funding:** Supported by the National Key R&D Program of China 2018YFA0507000 (B.W., Q.Z., and M.-W.W.) and 2017YFA0504703 (F.S.); National Science Foundation of China grants 31825010 (B.W.), 81525024 (Q.Z.), 31830020 (F.S.), 81872915 (M.-W.W.), and 81773792 (D.Y.); CAS Strategic Priority Research Program XDB37000000 (B.W.); Shanghai Outstanding Academic Leaders Plan of Shanghai Municipal Science and Technology Committee 18XD1404800 (Q.Z. and S.H.); the National Science & Technology Major Project—Key New Drug Creation and Manufacturing Program, China grant 2018ZX09711002 (L.M., D.Y., and M.-W.W.); the National Mega R&D Program for Drug Discovery

grant 2018ZX09735-001 (M.-W.W.); Shanghai Science and Technology Development Fund 16ZR1407100 (A.D.); and Australian National Health and Medical Research Council (NHMRC) grants 1126857 (D.W.) and 1150083 (P.M.S.). D.W. is an NHMRC Senior Research Fellow and P.M.S. is a Senior Principal Research Fellow. **Author contributions:** A.Q. optimized the constructs, developed the expression and purification procedures, and prepared the protein samples for cryo-EM; S.H. helped with protein sample optimization and performed negative-stain EM data acquisition and analysis, cryo-EM data processing and analysis, model building, and structure refinement; X.L. performed cryo-sample preparation, acquired cryo-EM data, and assisted with data processing and analysis; A.Q. and Z.L. performed signaling assays with assistance from R.C.; P.Z. performed NanoBiT assays; A.D. performed the ligand-binding assay; L.T. performed preliminary cryo-EM screening and assisted with data processing and analysis; Q.T.

helped with negative-stain EM data acquisition/analysis and cryo-EM data collection; X.C. and L.M. expressed the proteins; R.C. and T.S.T. helped with protein purification optimization; D.Y. and M.-W.W. oversaw ligand binding studies; P.M.S. and D.W. oversaw the NanoBiT assay; F.S. oversaw EM data acquisition, analysis, and processing; S.R.-R., D.Y., M.-W.W., P.M.S., D.W., and F.S. helped with data analysis/interpretation and edited the manuscript; and B.W. and Q.Z. initiated the project, planned and analyzed experiments, supervised the research, and wrote the manuscript with input from all co-authors. **Competing interests:** S.R.-R. is an employee and stock owner of Novo Nordisk, a pharmaceutical company focused on treatment of diabetes and obesity. All other authors declare no competing interests. **Data and materials availability:** Atomic coordinates and the cryo-EM density maps for the structures of glucagon-GCGR-G_s and glucagon-GCGR-G₁₁ have been deposited in the RCSB Protein Data Bank (PDB) with

identification codes 6LMK and 6LML, and the Electron Microscopy Data Bank (EMDB) under accession codes EMD-0917 and EMD-0918. All other data are available in the manuscript or the supplementary materials. Reagents are available from the corresponding authors upon reasonable request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1346/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S4
References (39–53)

17 September 2019; accepted 11 February 2020
10.1126/science.aaz5346

SOLAR CELLS

Resolving spatial and energetic distributions of trap states in metal halide perovskite solar cells

Zhenyi Ni^{1*}, Chunxiong Bao^{2*}, Ye Liu^{1,2}, Qi Jiang¹, Wu-Qiang Wu¹, Shangshang Chen¹, Xuezheng Dai¹, Bo Chen¹, Barry Hartweg³, Zhengshan Yu³, Zachary Holman³, Jinsong Huang^{1,2†}

We report the profiling of spatial and energetic distributions of trap states in metal halide perovskite single-crystalline and polycrystalline solar cells. The trap densities in single crystals varied by five orders of magnitude, with a lowest value of 2×10^{11} per cubic centimeter and most of the deep traps located at crystal surfaces. The charge trap densities of all depths of the interfaces of the polycrystalline films were one to two orders of magnitude greater than that of the film interior, and the trap density at the film interior was still two to three orders of magnitude greater than that in high-quality single crystals. Surprisingly, after surface passivation, most deep traps were detected near the interface of perovskites and hole transport layers, where a large density of nanocrystals were embedded, limiting the efficiency of solar cells.

The photovoltaic performance of metal halide perovskites (MHPs) is mainly attributed to their high optical absorption coefficient (1), high carrier mobility (2), long charge-diffusion length (3), and small Urbach energy (4). Defect tolerance in MHPs was initially proposed as one origin for their excellent carrier transport and particular recombination properties, in that most point defects have low formation energy in the bulk of perovskites and do not form deep charge traps (5, 6). Later theoretical studies showed that the structural defects at the material surface and grain boundaries of perovskites can induce deep charge traps, which has guided the development of passivation techniques in perovskite solar cells (7–9), but this was only inferred indirectly. The nonradiative recombination process also leads to the energy loss of the perovskite solar cells, which is closely related to the defect-induced trap states in the perovskites (10, 11). Charge trap states play an important role in

the degradation of perovskite solar cells and other devices (12, 13). Knowledge of the distributions of trap states in space and energy is one of the most fundamental ingredients for understanding the impact of the charge traps on charge transport and recombination in perovskite materials and devices.

Thermal admittance spectroscopy (TAS) and thermally stimulated current methods have been broadly applied to measure the energy-dependent trap density of states (tDOS) in perovskite solar cells (14–16). These methods can generally reach a trap depth of ~ 0.55 eV from the conduction or valence band edge, which is normally deep enough for most low-band gap perovskites that make efficient solar cells. Techniques like surface photovoltage spectroscopy and sub-band gap photocurrent are capable of detecting deeper trap states that exist in wide-band gap perovskites (17–19). Sub-band gap photoluminescence, which was adopted to investigate the properties of luminescent trap states in perovskite (20), and cathodoluminescence were shown to image the nanoscale stoichiometric variations that are related to the traps at the film surface (21). However, these techniques are not readily applied to completed solar cell devices to measure the spatial distribution of trap states. Deep-level defect characterization methods such as deep-level transient

spectroscopy are not readily applicable to perovskite devices, because the long biasing times are affected by ion migration in MHPs. Here, we demonstrate that the drive-level capacitance profiling (DLCP) method, an alternate capacitance-based technique, can provide well-characterized spatial distributions of carrier and trap densities in perovskites. We mapped the spatial and energetic distributions of trap states in perovskite single crystals and polycrystalline thin films. A straightforward comparison of the trap densities and distributions in perovskite single crystals and thin films in typical planar-structured solar cells was then conducted.

Drive-level capacitance profiling of perovskites

The DLCP method was developed to study the spatial distribution of defects in the band gap of amorphous and polycrystalline semiconductors, including amorphous silicon (Si) (22), $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ (23), and $\text{Cu}_2\text{ZnSnSe}_4$ (24). With the junction capacitance measurements, DLCP can directly determine the carrier density that includes both free carrier density and trap density within the band gap of the semiconductors and their distributions in space and energy (Fig. 1A and supplementary materials). The trap density was estimated by subtracting the estimated free carrier density, which was measured at high alternating current (ac) frequencies when the measured carrier densities saturate with the further increase of the ac frequency, from the total carrier density measured at the low ac frequency. This technique allowed us to derive the energetic distribution (E_ω) of trap states by tuning the frequency of the ac bias (δV) or temperature (T) and the position of trap states in real space by changing the direct current (dc) bias that was applied to the depletion region of the junction. As long as the spatial property of the semiconductor did not change dramatically, the differences in the profiling distance closely approximated the actual changes in the position where trap states responded to the capacitance, thus reflecting the change of the trap density in real space. In principle, DLCP can have a high resolution because the depletion edge can be continuously tuned by

¹Department of Applied Physical Sciences, University of North Carolina, Chapel Hill, NC 27599, USA. ²Department of Mechanical and Materials Engineering and Nebraska Center for Materials and Nanoscience, University of Nebraska–Lincoln, Lincoln, NE 68588, USA. ³School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: jhuang@unc.edu

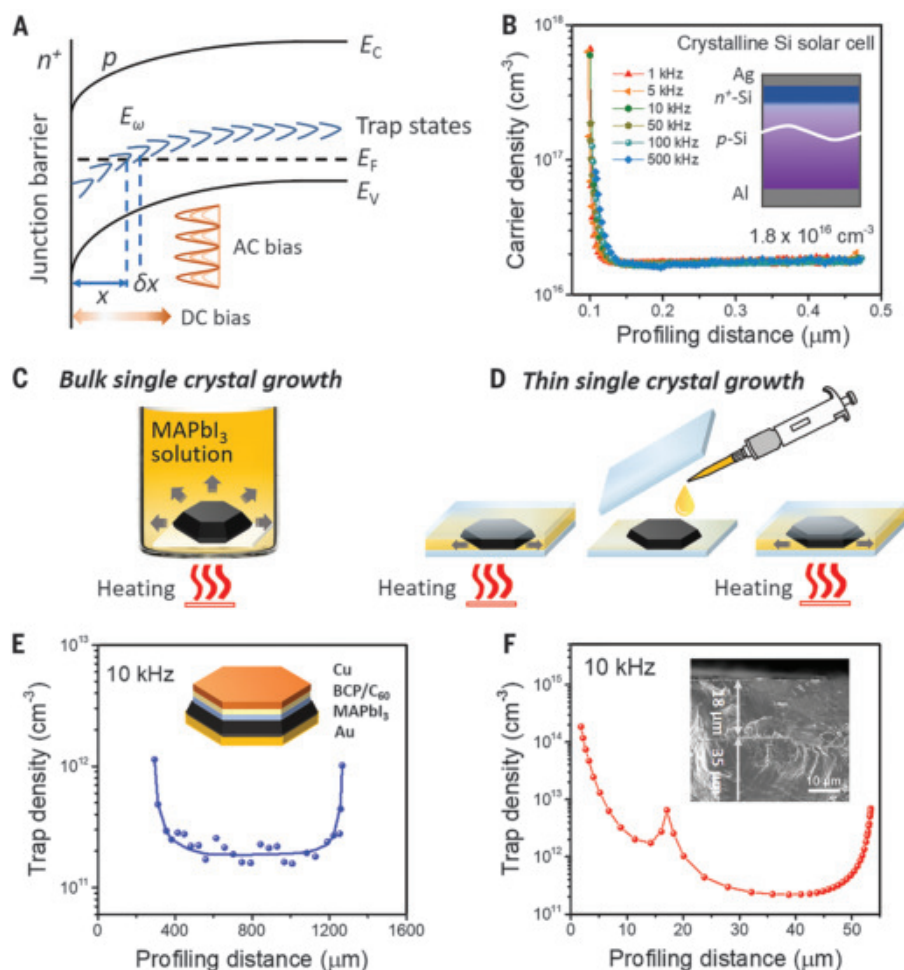


Fig. 1. DLCP technique. (A) Schematic of band bending of a p-type semiconductor with deep trap states in an n^+ -p junction. X denotes the distance from the junction barrier where the traps may be able to dynamically change their charge states with the ac bias δV . δX denotes the differential change of X with respect to δV . E_ω is the demarcation energy determined by $E_\omega = kT \ln(\omega_0/\omega)$ (where k is the Boltzmann's constant). E_C , E_V , and E_F indicate the conduction band edge, valence band edge, and Fermi level, respectively. (B) Dependence of the carrier density on the profiling distance of a Si solar cell at different ac frequencies measured by DLCP. The inset shows the schematic of the device structure. (C) Schematic of the synthesis of a bulk MAPbI₃ single crystal in an open-air solution. (D) Schematic of the synthesis of a double-layer MAPbI₃ thin single crystal using the space-confined growth method. (E) Dependence of the trap density on the profiling distance of a MAPbI₃ single crystal measured by DLCP. The inset shows the device structure. (F) Dependence of the trap density on the profiling distance of a double-layer MAPbI₃ thin single crystal. The inset shows the cross-sectional SEM image of the double-layer MAPbI₃ thin single crystal. The thicknesses of the top and bottom single crystals were 18 and 35 μm , respectively.

the applied dc bias. However, the profiling distance within the real devices was affected by the nonflat depletion interfaces caused by either the roughness or the heterogeneity of the materials, which could compromise the resolution of the profiling distance.

To validate the accuracy of the carrier density measured by DLCP, we first performed DLCP measurements on a Si solar cell, which was fabricated based on a p-type ($\sim 0.94 \text{ ohm}\cdot\text{cm}$ with the dopant concentration of $\sim 1.6 \times 10^{16} \text{ cm}^{-3}$) crystalline Si (p-Si) wafer with a heavily n-type diffusion layer Si (n^+) on top (details in the materials and methods section of the supplementary materials). The carrier density was calculated from the derived linear and non-linear capacitive coefficients C_0 and C_1 , respectively, by fitting the C - δV plots at different dc biases and ac frequencies (fig. S1). When the profiling distance is $>0.15 \text{ }\mu\text{m}$, which should reach the interior of the p-Si, the carrier densities measured at different ac frequencies (1 to 500 kHz) were basically the same (Fig. 1B), indicating negligible contributions of the trap states to the junction capacitance. In this case, the measured carrier density should be the free carrier concentration of the p-Si wafer, which

was read to be $\sim 1.8 \times 10^{16} \text{ cm}^{-3}$ from the DLCP measurement. This value was consistent with the dopant concentration of the p-Si wafer derived from the conductivity measurement, validating the accuracy of the carrier density measured by DLCP.

Because the profiling of carrier and trap densities by DLCP relied on the sweeping of a depletion region edge across a device from one electrode to the counter electrode, it was critical to understand the location of the junction(s) in typical planar-structured perovskite solar cells with the device structure of indium tin oxide (ITO)/poly[bis(4-phenyl)(2,4,6-trimethylphenyl)amine] (PTAA) (15 nm)/perovskite/fullerene (C_{60}) (25 nm)/bathocuproine (BCP)/copper (Cu). It was found that these perovskite solar cells essentially had a n^+ -p junction formed between the C_{60} and the perovskites (figs. S2 and S3). Another concern with DLCP measurements of MHPs is the role of ion migration. During the DLCP measurement, a positive dc bias was usually applied onto perovskite devices, which actually partially compensated for the built-in field in the devices. Thus, the field in the device was always less than the built-in field, which should, in

principle, minimize ion migration. In addition, each DLCP scan takes only a few minutes, and we confirmed the negligible influence of the ion migration on the DLCP measurement of these hysteresis-free perovskite solar cells by performing consecutive forward and backward scans of the dc biases (fig. S4).

We synthesized bulk $\text{CH}_3\text{NH}_3\text{PbI}_3$ (MAPbI₃) single crystals using the inverse solubility method (Fig. 1C). The DLCP measurements were performed on a MAPbI₃ single-crystal device with a structure of gold (Au)/MAPbI₃/C₆₀/BCP/Cu, where both sides of the crystal were polished to remove the defective surface layers (fig. S5). A symmetric distribution of the trap density was observed (Fig. 1E), in a good agreement with the structural symmetry of the double-side polished MAPbI₃ single crystal. This result demonstrated the spatial profiling of trap densities in MAPbI₃ single crystals by DLCP. The trap density near the interface region was ~ 10 -fold greater than that inside the MAPbI₃ single crystal. This difference indicated that dangling bonds at the surface of the crystal form charge traps.

To determine whether the profile depth corresponded to the physical material depth, we

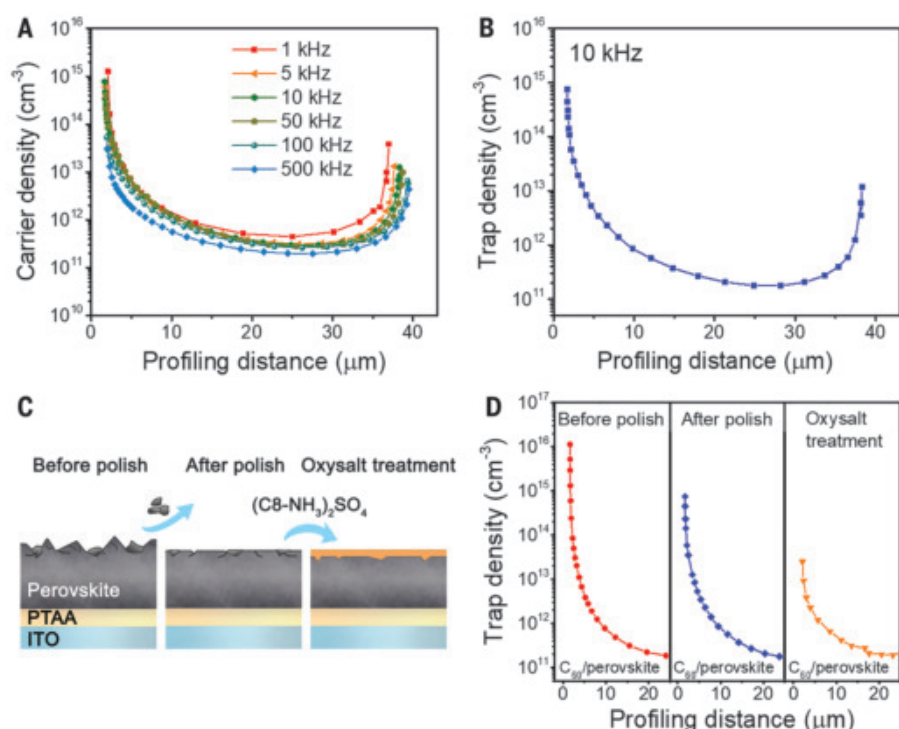


Fig. 2. Spatial distributions of trap states in a MAPbI₃ thin single crystal. (A) Dependence of the carrier density on the profiling distance of a 39-μm-thick MAPbI₃ thin single crystal at different ac frequencies, as measured by DLCP. (B) Dependence of the trap density on the profiling distance of a MAPbI₃ thin single crystal measured at an ac frequency of 10 kHz. The carrier density measured at 500 kHz is regarded as free carriers. (C) Schematics of a MAPbI₃ thin single crystal on a PTAA/ITO substrate before mechanical polish, after mechanical polish, and after oxysalt [(C₈-NH₃)₂SO₄] treatment. (D) Trap density near the junction barrier of a MAPbI₃ thin single crystal before mechanical polish, after mechanical polish, and after oxysalt treatment.

made a device with double-layer MAPbI₃ thin single crystals so that we knew the location of the charge traps (Fig. 1D). We first synthesized a 35-μm-thick MAPbI₃ thin single crystal on a PTAA/ITO substrate using the space-confined growth method (25) and then interrupted the growth by exposing the top surface of the thin single crystal to air for 1 min before continuing the crystal growth. This step created a distinct boundary between the two layers [cross-sectional scanning electron microscope (SEM) image in the inset of Fig. 1F] that should be rich in charge traps. This defective interface was located 18 μm below the surface of the top subcrystal (inset of Fig. 1F). The profiled trap density of this device (Fig. 1F) showed a peak in the trap density at the profiling distance of 18 μm.

Trap distributions in MAPbI₃ thin single crystals

We studied the trap distribution in perovskite single-crystal solar cells. Perovskite solar cells made from a single-crystal perovskite could, in principle, have a power conversion efficiency (PCE) approaching the Shockley-Queisser limit (usually 33.7% for a single junction) because of the extremely low defect density and long carrier diffusion length (3, 26). How-

ever, the highest PCE of the first-reported MAPbI₃ single-crystal solar cell was only 17.9% (25). A more recent study reported 21.1% (27), which is still far lower than that of polycrystalline solar cells. Initial studies indicate that thin crystals formed by the space-confined growth method have a smaller carrier diffusion length (10 to 20 μm) than that of thick bulk crystals (175 μm), which do not suffer from the impact of substrates. (25). However, the underlying mechanism limiting carrier diffusion in thin crystals was not clear.

We conducted DLCP measurements to understand the relationship of trap density and distribution with synthetic-crystal methods. Figure 2A shows the spatial distribution of carrier densities throughout a typical MAPbI₃ thin single crystal that was synthesized by the space-confined growth method in the device with a structure of ITO/PTAA/MAPbI₃ (39 μm)/C₆₀/BCP/Cu at different ac frequencies. The carrier density increased with the decrease in ac frequency, indicating the existence of charge traps in the MAPbI₃ thin single crystal that contributed to the junction capacitance at low ac frequencies (large E_w). Figure 2B shows a representative spatial distribution of trap densities in the MAPbI₃ thin single crystal at the

ac frequency of 10 kHz by subtracting the free carrier density measured at high ac frequencies from the total carrier density measured at the 10-kHz frequency. As the profiling position changed from the interfaces to the interior of the single crystal, the trap density decreased. This result indicated that the majority of the trap states were near the surface of the MAPbI₃ thin single crystal. The free carrier density was also higher near the surface of the MAPbI₃ thin single crystal (Fig. 2A), indicating that both self-doping and trap states are caused by defects, most likely, of different kinds.

To figure out how sensitive DLCP is to the change in the trap density close to the surface of the MAPbI₃ thin single crystals, we varied the trap density at the top surface (C₆₀ side) of the MAPbI₃ thin single crystal by polishing and treating it with (C₈-NH₃)₂SO₄ before performing DLCP measurements (Fig. 2C). Recent work has demonstrated that surface wrapping with oxysalt can effectively passivate the defective surface of perovskites with the wide-band gap oxysalts (28). The trap density was reduced by about one order of magnitude after polishing the top surface of the MAPbI₃ thin single crystal and was further reduced after the surface treatment with (C₈-NH₃)₂SO₄ (Fig. 2D). Because DLCP only measured the carrier density in the junction area, this result also validates the finding that the measured junction is located at the perovskite/C₆₀ interface. Thus, the interface regions of the perovskite/C₆₀ and the perovskite/PTAA were readily distinguished in the spatial distribution profiling of trap states.

The trap density distribution in the MAPbI₃ thin single crystal synthesized by the space-confined method was quite different from that in the bulk crystal. The trap densities varied by up to five orders of magnitude, and the trap density near both surfaces was two to four orders of magnitude higher than that in the bulk crystals. The trap density decreased gradually toward the center of the crystal, and its distribution along the normal direction was not symmetric, despite both surfaces of the thin single crystal contacting PTAA/ITO during the growth process. To understand these differences, we synthesized MAPbI₃ thin single crystals with different thicknesses (10 to 39 μm) and investigated the variation of the trap density distribution with the change in the crystal thicknesses (Fig. 3A). The minimal bulk trap density ($N_{T \min}$) inside the MAPbI₃ thin single crystal, which occurred near the center of the crystal, decreased from $\sim 3.2 \times 10^{12}$ to 1.9×10^{11} cm⁻³ as the thickness of the thin single crystal increased from 10 to 32 μm and began to saturate with increased crystal thickness (Fig. 3B). The saturated $N_{T \min}$ of 1.8×10^{11} cm⁻³ approached that of the bulk MAPbI₃ single crystal synthesized in open-air solution (Fig. 1E). These results indicate that there is a critical crystal thickness for the MAPbI₃ thin single

crystals synthesized by the space-confined method, below which the trap density inside the crystals was substantially higher than that in the bulk crystal. We speculate that the space-confined method may induce defects through the strain imposed by the mismatch of the substrates and the crystals during growth and that the strain inside the crystals may be released with the increase in crystal thickness. Another possible mechanism for defect formation is that the substrates affect the transport of ions for micrometer-scale channels. The microfluid would undergo laminar flow at low flow rates near the substrates (inset of Fig. 3B) (29). For the thinner single crystals, the averaged velocity of the solution flow would be reduced owing to the confinement of the boundary layer by the small space, resulting in not enough ions being delivered to the crystal for growth. Insufficient ion delivery to the crystal surface would create a deficiency for one type of ions, or misfit defects.

Similarly, the trap density near the MAPbI₃/PTAA interface also decreased with the increase in the crystal thickness and was essentially constant with further increases in crystal thickness (Fig. 3A). However, the imposed strains between the two sides of the single crystal and the substrates were not always identical because of the subtle difference in the roughness of the two PTAA/ITO substrates, which led to the different defect density distributions at the two surfaces of the crystals. The side with a higher defect density may have a weaker contact with the PTAA/ITO substrate, making the PTAA/ITO substrate easier to be peeled off from this side. It is this defective side that C₆₀ was deposited on during the device fabrication process that was used. Moreover, the distribution of trap density in MAPbI₃ thin single crystals depended not only on the spacing of the two substrates but also on the substrate upon which the crystals grew. The latter also affects the strain inside the crystals and the microfluid of the precursor solution, as evidenced by the difference in the trap densities measured in the top and bottom subcrystals of the double-layer sample in Fig. 1F. The top subcrystal had a much higher $N_{T\min}$ than the bottom MAPbI₃ thin single crystal grown directly on a PTAA/ITO substrate (Fig. 3A).

We examined further the tDOS in energy space in the MAPbI₃ thin single crystal with a thickness of 39 μm . To verify the effectiveness of the DLCP in determining the tDOS, we derived the tDOS in the MAPbI₃ thin single crystal by both TAS and DLCP methods, because TAS is a well-established method to determine the tDOS in perovskite devices (15). The temperature-dependent differential capacitance spectra ($-f \cdot dC/df$, where f is the frequency of the ac bias) and the Arrhenius plot of the characteristic frequencies with respect to the temperature $[\ln(\omega/T^2) - 1/T]$ of the device

are shown in fig. S6. Figure 3C shows the E_{∞} -dependent N_T for the MAPbI₃ thin single crystal measured by TAS. For the DLCP measurement, the tDOS could be estimated by the derivative of the carrier density with respect to E_{∞} , that is, $N_T(E_{\infty}) = dN/d(E_{\infty})$. As the profiling distance was scanned from one side of the single crystal to the other, the energy distribution and spatial distribution of the trap states in the MAPbI₃ thin single crystal were mapped (Fig. 3D).

The tDOS measured by DLCP exhibited a similar feature to that measured by TAS (Fig. S7). Both tDOS spectra showed three trap bands with E_{∞} values of 0.27 eV (zone I), 0.35 eV (zone II), and greater than 0.40 eV (zone III). Previous studies speculated that the deep trap states were mainly related to the surface defects of the perovskite and that shallower trap states were more likely from inside the perovskite (15). The spatial and energy distributions of the tDOS in perovskites (Fig. 3D) indicated that the deep trap density at the MAPbI₃/PTAA interface was >100-fold higher than inside the

MAPbI₃ thin single crystal, whereas the shallow trap density at the MAPbI₃/PTAA interface was barely higher than those inside the single crystal. Deep traps were mainly located at the surface region of the MAPbI₃ thin single crystals, whereas the shallower traps were prevalent throughout the entire single crystals. This difference indicated their different origins, that is, shallow trap bands I and II may form from point defects in the bulk and deep trap band III originated from the dangling bonds at the material surface. The measured carrier densities near the MAPbI₃/C₆₀ interface at the ac frequencies from 1 to 50 kHz were quite near each other (Fig. 2A), indicating a low deep trap density of states near the MAPbI₃/C₆₀ interface caused by the passivation effect of C₆₀.

Trap distributions in polycrystalline perovskite films

The spatial and energetic distributions of trap states in polycrystalline perovskite thin films are crucial to understanding the performance of those solar cells. We first performed DLCP

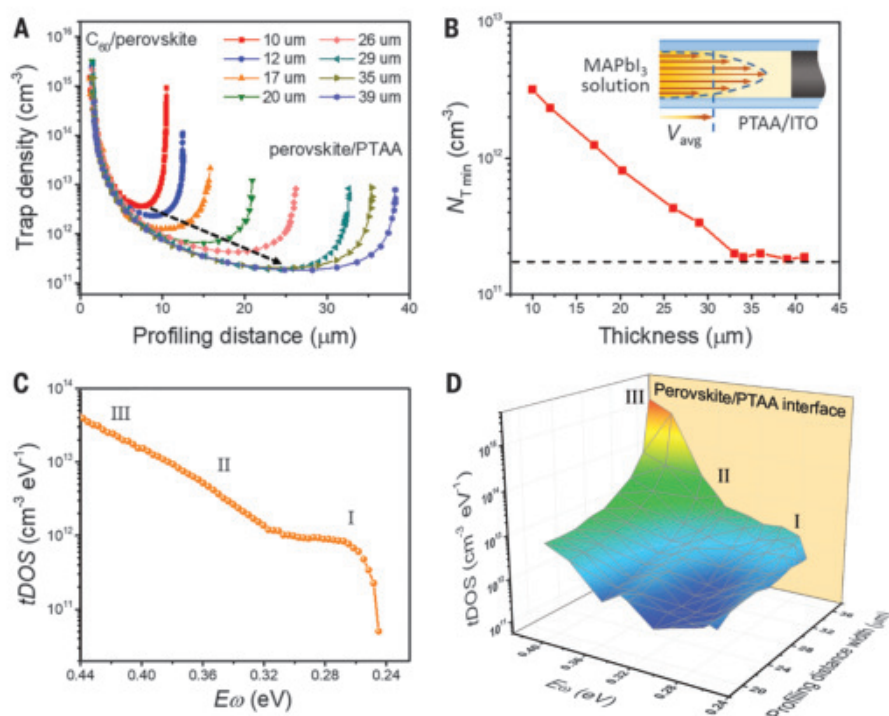


Fig. 3. Thickness-dependent trap density distributions in MAPbI₃ thin single crystals. (A) Dependence of the trap densities on the profiling distances of MAPbI₃ thin single crystals with different crystal thicknesses measured at an ac frequency of 10 kHz. The location of the MAPbI₃/C₆₀ interface for each crystal is aligned for comparison. The black dashed arrow indicates the trend of the change of minimal trap density $N_{T\min}$ in MAPbI₃ single crystals with different thicknesses. (B) Dependence of the $N_{T\min}$ in the MAPbI₃ thin single crystal on the crystal thickness. The horizontal dashed line indicates the $N_{T\min}$ value in a bulk MAPbI₃ single crystal. The inset shows a schematic of the laminar flow of the precursor solution between two PTAA/ITO glasses during the growth of the crystal. The arrows denote the direction of the laminar flow of the precursor solution, and the length of the arrow denotes the laminar flow velocity. (C) tDOS of a MAPbI₃ thin single crystal, as measured by the TAS method. The thickness of the MAPbI₃ thin single crystal was 39 μm . (D) Spatial and energy mapping of the densities of trap states in the MAPbI₃ thin single crystal, as measured by DLCP.

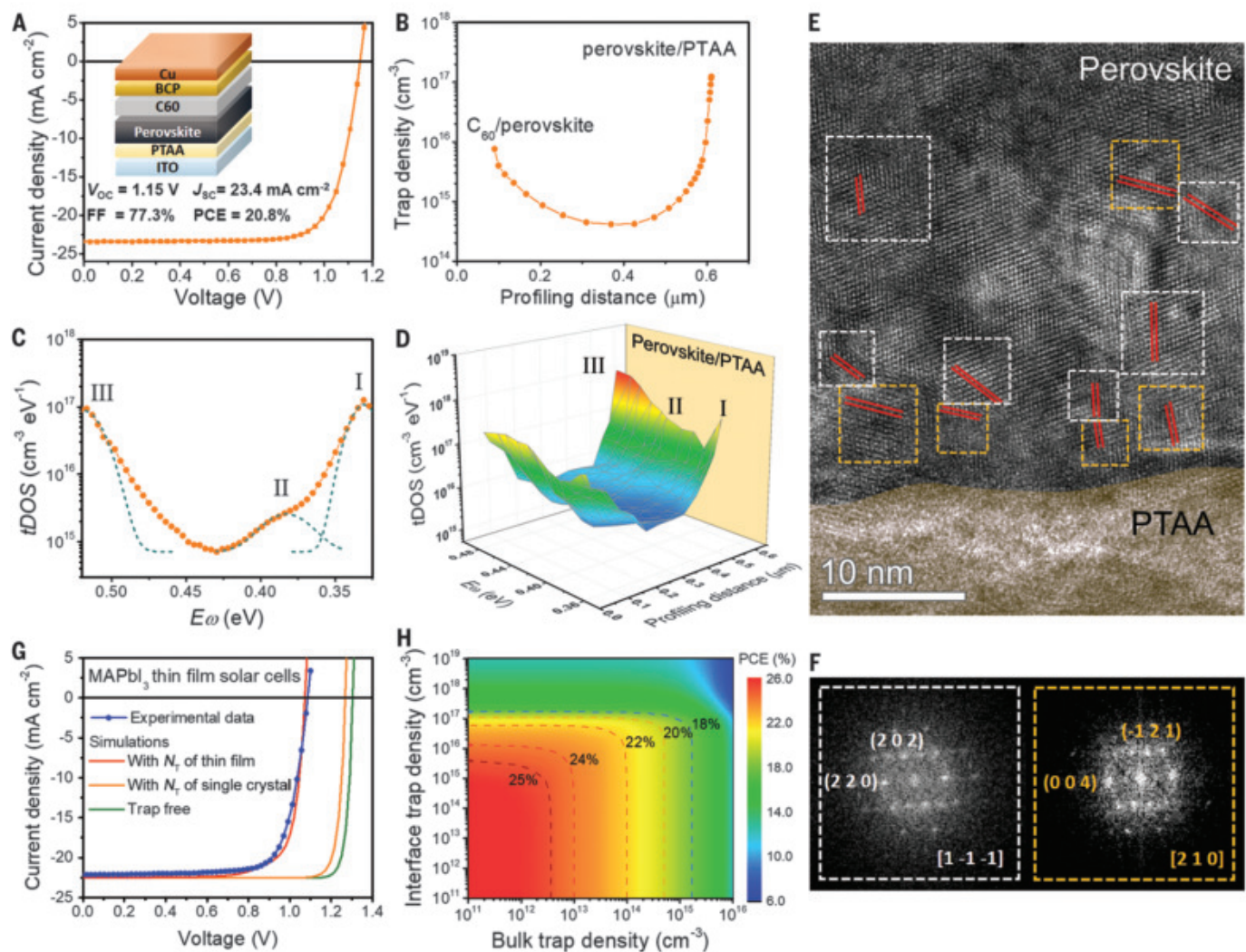


Fig. 4. Spatial and energetic distributions of trap states in perovskite thin films. (A) J-V curve of the Cs_{0.05}FA_{0.70}MA_{0.25}PbI₃ thin-film solar cells. The inset shows the device structure. (B) Dependence of the trap density on the profiling distance for the perovskite thin film in the solar cell measured at an ac frequency of 10 kHz. (C) tDOS of the perovskite thin-film solar cell, as measured by the TAS method. (D) Spatial and energy mapping of the densities of trap states of the perovskite thin film in the solar cell, as measured by DLCP. (E) Cross-sectional HR-TEM image of the stack of perovskite and PTAA. The dashed squares mark the

areas where the fast Fourier transforms of the lattices were performed, with white and yellow indicating zone axes of [1 -1 -1] and [2 1 0], respectively. The red lines denote the orientation of the facets. (F) Fast Fourier transforms of the areas indicated in (E). (G) Measured and simulated J-V curves of planar-structured solar cells based on MAPbI₃ polycrystalline thin films. The thin-film (single crystal) bulk and interface trap densities were adopted for the simulations. (H) Dependence of the PCE of the MAPbI₃ thin-film solar cell on the bulk and interface trap densities. The dashed lines denote the contour lines of certain PCE values, which are noted.

Table 1. Comparison of the minimal bulk trap densities and the interface trap densities between perovskite single crystals and thin films. The interconnected layers are listed in parentheses. The trap densities are calculated at an ac frequency of 10 kHz. The interface trap density values vary depending on different surface and interface conditions.			
Perovskite material	Minimal bulk trap density ($N_{T\min}$) (cm^{-3})	Interface trap density (cm^{-3})	
MAPbBr ₃ single crystal (bulk)	6.5×10^{10}	1.8×10^{12} (C ₆₀)	1.8×10^{12} (Au)
MAPbI ₃ single crystal (bulk)	1.8×10^{11}	1.2×10^{12} (C ₆₀)	1.2×10^{12} (Au)
MAPbI ₃ single crystal (thin)	1.9×10^{11} to 3.2×10^{12}	2.0×10^{13} to 1.1×10^{16} (C ₆₀)	1.5×10^{13} to 1.0×10^{15} (PTAA)
Cs _{0.05} FA _{0.70} MA _{0.25} PbI ₃ film	4.3×10^{14}	8.6×10^{15} (C ₆₀)	1.2×10^{17} (PTAA)
Rb _{0.05} Cs _{0.05} FA _{0.75} MA _{0.15} Pb(I _{0.95} Br _{0.05}) ₃ film	5.7×10^{14}	2.0×10^{16} (C ₆₀)	1.1×10^{17} (PTAA)
FA _{0.92} MA _{0.08} PbI ₃ film	7.9×10^{14}	1.9×10^{16} (C ₆₀)	9.0×10^{16} (PTAA)
MAPbI ₃ film	9.2×10^{14}	2.2×10^{16} (C ₆₀)	1.2×10^{17} (PTAA)
Cs _{0.05} FA _{0.8} MA _{0.15} Pb _{0.5} Sn _{0.5} (I _{0.85} Br _{0.15}) ₃ film	1.2×10^{15}	1.5×10^{16} (C ₆₀)	1.1×10^{17} (PTAA)

measurement on a typical planar-structured perovskite thin-film solar cell with the device structure of ITO/PTAA/MAPbI₃/C₆₀/BCP/Cu, which has a typical PCE of 17.8% (table S1). The measured trap density distribution and tDOS mapping in the MAPbI₃ thin film are shown in fig. S8. The MAPbI₃ polycrystalline thin film shows a similar feature of trap distribution with the thin single crystals in which most of the deep trap states (trap band III) are located close to the MAPbI₃/PTAA interface. Then, we carried out DLCP measurement on a high-performance solar cell with the device structure of ITO/PTAA/Cs_{0.05}[HC(NH₂)₂]_{0.70}(CH₃NH₃)_{0.25}PbI₃(Cs_{0.05}FA_{0.70}MA_{0.25}PbI₃)/C₆₀/BCP/Cu, in which the perovskite thin films were modified with the additive of 1,3-diaminopropane (30). The open circuit voltage (V_{OC}), short-circuit current density (J_{SC}), fill factors (FF), and PCE of the solar cell were 1.15 V, 23.4 mA cm⁻², 77.3%, and 20.8%, respectively (Fig. 4A). The spatial distribution of the carrier densities in the Cs_{0.05}FA_{0.70}MA_{0.25}PbI₃ solar cell is shown in fig. S9. The differences in the carrier densities measured at different ac frequencies revealed the presence of trap states in these perovskite thin films. The trap density at the perovskite/C₆₀ interface was about 10-fold lower than that at the perovskite/PTAA interface (fig. 4B), which might be caused by the passivation of C₆₀ on the surface defects of the perovskite thin film (15, 31). Both interfaces had a higher defect density compared with the interior of the perovskite films.

In the tDOS spectrum of the Cs_{0.05}FA_{0.70}MA_{0.25}PbI₃ solar cell measured by TAS (Fig. 4C), the attempt-to-escape angular frequency ω_0 was derived from the temperature-dependent C - f measurements, as detailed in the materials and methods. The tDOS spectrum contained three different trap centers, dividing the spectrum into three energy bands (marked as I, II, and III). The trap distributions in bands I, II, and III of the perovskite thin film are centered at ~0.33, 0.38, and 0.52 eV, respectively. The spatial mapping of the tDOS in the perovskite thin film (Fig. 4D) revealed that the tDOSs at both interface regions were >100-fold higher than that in the film. In addition, deep traps in band III were more localized at the perovskite/PTAA interface, whereas shallower traps in bands I and II were enriched at both interfaces. This result showed that the perovskite surfaces of polycrystalline films were rather defective (31).

To understand the origin of the high deep trap density at the perovskite/TPAA interface, we examined this region by high-resolution transmission electron microscopy (HR-TEM). As shown in Fig. 4, E and F, the lattice had basically the same orientation with the zone axis of [1 -1 -1] in the grain interior, whereas there were a large number of small crystals with sizes

orientations (zone axis of [2 1 0]) with respect to the grain interior at the region near the PTAA/perovskite interface. At least 10 samples were examined with different perovskite compositions. All of them had very similar morphology, confirming the heterogeneity of the perovskite films in the vertical direction (32), which we believe was caused by the deposition method-related grain-growth behavior that made the interface between the perovskite and the PTAA rather defective and rich in charge trap centers. Given the excellent passivation effect of C₆₀ on the deep trap states at the perovskite surface (33), the remaining deep trap states were mainly located near the perovskite/PTAA interface, which might limit the efficiency of the perovskite thin-film solar cells.

Perovskite single crystals versus polycrystalline films

To find out the differences between the trap density distributions in perovskite single crystals and polycrystalline thin films and in those with different compositions, we measured the $N_{T \text{ min}}$ and the interface trap densities in several perovskite single crystals and polycrystalline thin films with different compositions (Table 1). The bulk MAPbI₃ or MAPbBr₃ single crystal showed a quite low $N_{T \text{ min}}$ ($<2.0 \times 10^{11}$ cm⁻³) which increased to 3.0×10^{12} cm⁻³ in MAPbI₃ thin single crystals, depending on their growth conditions. However, these values are still two to three orders of magnitude lower than that in MAPbI₃ polycrystalline thin films, which are generally formed by a very quick thin-film coating process. In addition, our current results signify the importance of proper surface-modification processes (mechanical polishing and oxysalt treatment) to reduce trap densities in perovskite single crystals. Similar scenarios could be applied to polycrystalline thin films to reduce the interface trap densities.

Table 1 also lists the $N_{T \text{ min}}$ and the interface trap densities of several typical polycrystalline perovskite thin films used in planar-structured solar cells, including Cs_{0.05}FA_{0.70}MA_{0.25}PbI₃, Rb_{0.05}Cs_{0.05}FA_{0.75}MA_{0.15}Pb(I_{0.95}Br_{0.05})₃, FA_{0.92}MA_{0.08}PbI₃, MAPbI₃, and Cs_{0.05}FA_{0.8}MA_{0.15}Pb_{0.5}Sn_{0.5}(I_{0.85}Br_{0.15})₃. The current density-voltage (J - V) curves and trap density distributions of the solar cells fabricated based on these films are shown in fig. S10. The corresponding parameters of device performance are listed in table S1. Among these configurations, the Cs_{0.05}FA_{0.70}MA_{0.25}PbI₃-based solar cell exhibited the highest PCE of 20.8% after optimizing the fabrication processes (30) and had the lowest $N_{T \text{ min}}$ of $\sim 4.0 \times 10^{14}$ cm⁻³, which was still more than two orders of magnitude greater than that in high-quality single crystals. For the perovskite solar cells fabricated based on the other compositions without comprehensive optimizations, the Rb_{0.05}Cs_{0.05}FA_{0.75}MA_{0.15}Pb(I_{0.95}Br_{0.05})₃-based solar cell showed a relatively

high PEC of 19.6%, whereas FA_{0.92}MA_{0.08}PbI₃ and MAPbI₃ showed lower PCEs of ~18.0%. Accordingly, the $N_{T \text{ min}}$ in Rb_{0.05}Cs_{0.05}FA_{0.75}MA_{0.15}Pb(I_{0.95}Br_{0.05})₃ was moderately lower than that in FA_{0.92}MA_{0.08}PbI₃ and MAPbI₃. The tin-incorporated Cs_{0.05}FA_{0.8}MA_{0.15}Pb_{0.5}Sn_{0.5}(I_{0.85}Br_{0.15})₃ film showed the highest $N_{T \text{ min}}$ of $\sim 1.2 \times 10^{15}$ cm⁻³ among all the configurations, reflecting the relative defective nature of tin-containing thin films.

The trend of the variation in the PCE of these solar cells was basically in accordance with the change in the $N_{T \text{ min}}$ in different perovskite thin films. This finding signifies the importance of reducing the trap densities in the perovskite thin films for enhancing the device performances. Our current results demonstrate that intrinsic trap densities in perovskite polycrystalline thin films were closely related to the film compositions as well as the film fabrication process. For all the perovskite thin-film compositions, the interface trap density was in the range of $\sim 9.0 \times 10^{15}$ to 2.0×10^{17} cm⁻³, depending on the type of the charge transport layers. Overall, the trap density at the perovskite/PTAA interface was higher than that at the perovskite/C₆₀ interface because of the formation of large amounts of small crystals near the perovskite/PTAA interface. This morphology points out an important direction to explore for further boosting the performance of perovskite solar cells or other electronic devices by reducing the trap density at the perovskite/PTAA interfaces.

Relationship of trap density and solar cell efficiency

We used the solar cell capacitance simulator to simulate both thin-film and single-crystal perovskite solar cells with varied trap densities. We first used the trap density and distribution measured by DLCP and TAS, which is detailed in fig. S11 and tables S2 and S3, to simulate a MAPbI₃ thin-film solar cell. Here, the capture cross sections of the bulk and interface trap states were determined by a global fitting of the experimental J - V curves (fig. S12). Figure 4G shows the simulated J - V curve of the MAPbI₃ thin-film solar cell with a bulk trap density of 5.0×10^{14} cm⁻³ and interface trap density of 1.0×10^{17} cm⁻³ obtained from the polycrystalline thin films, which was near the measured value. We simulated temperature-dependent J - V curves of the MAPbI₃ thin-film solar cell. As shown in fig. S13, the simulated J - V curves agree well with the measured J - V curves at different temperatures, which indicates that the DLCP measurement range of traps is deep enough to predict the behavior of these solar cells. After reducing only the bulk trap density to 1.0×10^{13} cm⁻³, the value attainable in single-crystalline MAPbI₃, the PCE increased to 20.0% and saturated with any further decrease in the bulk trap density (Fig. 4H and fig.

S14). This saturated PCE was mainly limited by the large interface trap density. If the interface trap density was reduced to that in a MAPbI₃ thin single crystal ($2.0 \times 10^{15} \text{ cm}^{-3}$), the PCE could be further enhanced to 25.4%, which is near the PCE of 26.6% for a trap-free MAPbI₃ thin-film solar cell (Fig. 4G). Simulation of single-crystal solar cells also gave data that were a good match to the experimental data (27), which again showed that the PCE of the MAPbI₃ single-crystal solar cell could be further improved to 26.8% once the interface trap densities are reduced to that of the bulk trap density (fig. S15).

Lower-band gap perovskites are being studied to harvest more sunlight, so we simulated perovskite thin-film solar cells with band gaps of 1.50 and 1.47 eV, which correspond to the compositions of FA_{0.92}MA_{0.08}PbI₃ and FAPbI₃ (if it can be stabilized), respectively (34, 35). Assuming that these materials have the same trap densities (a bulk trap density of $5.0 \times 10^{14} \text{ cm}^{-3}$ and interface trap density of $1.0 \times 10^{17} \text{ cm}^{-3}$) and capture cross sections as regular polycrystalline MAPbI₃ thin films, the devices showed PCEs of 22.5 and 22.8%, respectively (fig. S16). The efficiencies could be further increased to 27.7 and 28.4%, respectively, when the trap densities in the thin film are substantially reduced to be the same as those in single crystals.

REFERENCES AND NOTES

- N.-G. Park, *Mater. Today* **18**, 65–72 (2015).
- L. M. Herz, *ACS Energy Lett.* **2**, 1539–1548 (2017).
- Q. Dong et al., *Science* **347**, 967–970 (2015).
- C. Gehrman, D. A. Egger, *Nat. Commun.* **10**, 3141 (2019).
- S. B. Zhang, S.-H. Wei, A. Zunger, *Phys. Rev. Lett.* **78**, 4059–4062 (1997).
- K. X. Steirer et al., *ACS Energy Lett.* **1**, 360–366 (2016).
- W.-J. Yin, T. Shi, Y. Yan, *Appl. Phys. Lett.* **104**, 063903 (2014).
- C. Eames et al., *Nat. Commun.* **6**, 7497 (2015).
- A. Walsh, D. O. Scanlon, S. Chen, X. G. Gong, S. H. Wei, *Angew. Chem. Int. Ed.* **54**, 1791–1794 (2015).
- G. J. Wetzelaer et al., *Adv. Mater.* **27**, 1837–1841 (2015).
- W. Tress et al., *Adv. Energy Mater.* **5**, 1400812 (2015).
- J. M. Ball, A. Petrozza, *Nat. Energy* **1**, 16149 (2016).
- T. Leijtens et al., *Adv. Energy Mater.* **5**, 1500962 (2015).
- C. Ran, J. Xu, W. Gao, C. Huang, S. Dou, *Chem. Soc. Rev.* **47**, 4581–4610 (2018).
- Y. Shao, Z. Xiao, C. Bi, Y. Yuan, J. Huang, *Nat. Commun.* **5**, 5784 (2014).
- Y. Hu et al., *Adv. Energy Mater.* **8**, 1703057 (2018).
- I. Levine et al., *ACS Energy Lett.* **4**, 1150–1157 (2019).
- C. M. Sutter-Fella et al., *ACS Energy Lett.* **2**, 709–715 (2017).
- A. Musienko et al., *Energy Environ. Sci.* **12**, 1413–1425 (2019).
- E. T. Hoke et al., *Chem. Sci.* **6**, 613–617 (2015).
- O. Hentz, Z. Zhao, S. Gradečak, *Nano Lett.* **16**, 1485–1490 (2016).
- C. E. Michelson, A. V. Gelatos, J. D. Cohen, *Appl. Phys. Lett.* **47**, 412–414 (1985).
- J. T. Heath, J. D. Cohen, W. N. Shafarman, *J. Appl. Phys.* **95**, 1000–1010 (2004).
- H.-S. Duan et al., *Adv. Funct. Mater.* **23**, 1466–1471 (2013).
- Z. Chen et al., *Nat. Commun.* **8**, 1890 (2017).
- D. Shi et al., *Science* **347**, 519–522 (2015).
- Z. Chen et al., *ACS Energy Lett.* **4**, 1258–1259 (2019).
- S. Yang et al., *Science* **365**, 473–478 (2019).
- H.-S. Rao, B.-X. Chen, X.-D. Wang, D.-B. Kuang, C.-Y. Su, *Chem. Commun.* **53**, 5163–5166 (2017).
- W.-Q. Wu et al., *Sci. Adv.* **5**, eaav8925 (2019).
- B. Chen, P. N. Rudd, S. Yang, Y. Yuan, J. Huang, *Chem. Soc. Rev.* **48**, 3842–3867 (2019).
- E. M. Tennyson, T. A. S. Doherty, S. D. Stranks, *Nat. Rev. Mater.* **4**, 573–587 (2019).
- Q. Wang et al., *Energy Environ. Sci.* **7**, 2359–2365 (2014).
- Q. Jiang et al., *Nat. Photonics* **13**, 460–466 (2019).
- T. Niu et al., *Energy Environ. Sci.* **11**, 3358–3366 (2018).

ACKNOWLEDGMENTS

Funding: This work was supported by the Center for Hybrid Organic Inorganic Semiconductors for Energy (CHOISE), an Energy Frontier Research Center funded by the Office of Basic Energy Sciences, Office of Science within the U.S. Department of Energy. The study of the silicon device was supported by the Solar Energy Technologies Office (SETO) within the U.S. Department of Energy under award no. DE-EE0008749. Partial study of the single crystal growth was supported by Defense Threat Reduction Agency under Grant HDTRA1170054. **Author contributions:** J.H., Z.N., and C.B. designed the experiments. Z.N. and Y.L. synthesized the perovskite single crystals. Q.J., W.-Q.W., S.C., and B.C. fabricated the polycrystalline perovskite thin-film solar cells. B.H., Z.Y., and

Z.H. fabricated the silicon solar cells. Z.N. and C.B. carried out the capacitance measurements for the devices. Z.N. conducted the solar cell simulations. S.C. and X.D. carried out electron microscope measurements for the perovskites. J.H. and Z.N. wrote the paper, and all authors reviewed the paper. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1352/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S3
References (36–39)

4 November 2019; accepted 25 February 2020
10.1126/science.aba0893

REPORTS

GEOMORPHOLOGY

Latitudinal effect of vegetation on erosion rates identified along western South America

J. Starke, T. A. Ehlers*, M. Schaller

Vegetation influences erosion by stabilizing hillslopes and accelerating weathering, thereby providing a link between the biosphere and Earth's surface. Previous studies investigating vegetation effects on erosion have proved challenging owing to poorly understood interactions between vegetation and other factors, such as precipitation and surface processes. We address these complexities along 3500 kilometers of the extreme climate and vegetation gradient of the Andean Western Cordillera (6°S to 36°S latitude) using 86 cosmogenic radionuclide-derived, millennial time scale erosion rates and multivariate statistics. We identify a bidirectional response to vegetation's influence on erosion whereby correlations between vegetation cover and erosion range from negative (dry, sparsely vegetated settings) to positive (wetter, more vegetated settings). These observations result from competing interactions between precipitation and vegetation on erosion in each setting.

The impact of vegetation on the shape and evolution of Earth's surface ranges (for example) from the microscopic scale of Mycorrhiza weathering for plant nutrition to macroscopic scales where plants retard hillslope erosion, stabilize environments for sediment deposition, and affect precipitation through evapotranspiration, interception, and leaf phenology (1–6). However, defining the influence of vegetation on catchment-averaged erosion rates has proven difficult because of, among other things, nonlinear interactions between vegetation type and cover with precipitation, temperature, and solar radiation (7–10). One approach for disentangling the effects of vegetation and climate on landscape evolution requires quantifying catchment erosion rates over a large range of climate and biogeographic conditions. The production of cosmogenic radionuclides in the upper ~2 m of Earth's surface

provides one means for quantifying millennial time scale catchment-averaged erosion rates (11, 12). We investigate the relationships between catchment-averaged erosion rates with vegetation cover, climate, and topographic slope along the climate and ecological gradient of the Andean Western Cordillera, South America.

The Andean Western Cordillera between 6°S and 36°S latitude extends 3500 km (Fig. 1A) and crosses six climate zones, from hyperarid to temperate (13), and four distinct biogeographic regions (Fig. 1C and 2A) (14). Cosmogenic radionuclide-derived erosion rates and their controlling factors have been investigated along the Andean Western Cordillera, often with conflicting results [e.g., (15, 16)]. The emphasis of previous studies ranged from quantification of erosion rates in the vegetation-limited Atacama Desert (16, 17) and sediment storage in hyperarid environments (18) to the rates of canyon incision and hillslope erosion in the Andean Western Cordillera (19). Few studies (15, 20) have previously looked at systematic

Department of Geosciences, University of Tuebingen, 72074, Germany.

*Corresponding author. Email: todd.ehlers@uni-tuebingen.de

latitudinal variations in erosion rates along this gradient.

We measured cosmogenic radionuclide concentrations of ^{10}Be (22) from 12 samples and combined this data with 74 published samples from Peru and Chile (Fig. 1, A and B, and tables S1 and S2). The 86 catchments are adjacent to a similar tectonic plate boundary with 13 (broadly defined) catchment lithologies, including Oligo-Miocene, Plio-Pleistocene volcanoclastic deposits and ignimbrites, Jurassic and Cretaceous sedimentary rocks, Paleozoic and Cretaceous granodiorites, and Precambrian gneiss (21). The total lithological-weighted quartz content for each catchment varies between 15 to 49% (fig. S2). We used ^{10}Be concentrations to recalculate erosion rates using the same sea level high-latitude production rates and production-rate scaling (22) (fig. S1A). We determined from MODIS, TRMM, CHELSA, WorldClim, SRTM, and GLiM datasets (Figs. 1C and 2 and fig. S3) the 2σ range in vegetation cover, mean annual precipitation (MAP) and temperature (MAT), mean solar radiation (MSR), catchment-averaged slope, local relief, and lithologic quartz content for each catchment. We found no large differences between the TRMM, CHELSA, and WorldClim (MAT and MAP values) datasets (23) (tables S1 to S6).

The catchment-averaged erosion rates varied between 1.4 and 150 m per million years (m/Myr) (solid line, Fig. 1B). Starting in the north (6°S to 12°S), erosion rates displayed increasing values between 0 and 150 m/Myr. From 12°S to 20°S , the erosion rates decreased (150 to 0 m/Myr). The lowest erosion rates were located between 20°S to 30°S (0 to 50 m/Myr). In the south (30°S to 36°S), erosion rates ranged between 0 to 140 m/Myr and showed increasing values from 30°S to 33.5°S and decreasing values from 33.5°S to 36°S . In general, the quartz content for each catchment shows no latitudinally dependent variation (fig. S2C). Vegetation cover and MAP were the highest in the north from 6°S to 10°S (50 to 85% and 200 to 700 mm/year, respectively; solid lines in Fig. 2, A and B). Vegetation cover and MAP then decreased to a minimum (5% and <50 mm/year) at the latitudes of the Atacama Desert (20°S to 30°S). Further south (30°S to 36°S), the vegetation cover and MAP increased to a southern maximum (40% and ~700 mm/year). Catchment-averaged slopes had increasing values up to 30° from 6°S to 12°S . Slopes gradually decreased toward the south (12°S to 20°S) and varied between 25° to 10° . The lowest slopes (5° to 10°) were situated between 20°S to 30°S and increased to the south (30°S to 36°S) up to 28° (solid line, Fig. 2C).

Results from a multivariate factor analysis indicated that catchment erosion rate, vegetation cover, slope, MAT, MAP, MSR, local relief, quartz content, and lithology result in four factors that can explain 62% of the variance

(tables S4 and S5). Factor 1 explains 23% of the variance and contains MAP, vegetation cover, erosion rate (factor loadings of 0.9, 0.6, and 0.5, respectively). We interpreted factor 1 as representing the combined interactions of vegetation, precipitation, and erosion. Factor 2 (20% of the variance) contains vegetation cover, MAT (factor loading >0.6), and acid volcanic rocks ("va" in table S4, factor loading 0.7). We interpreted factor 2 as representing interactions between vegetation, temperature, and substrate composition. Both the factor and Pearson statistical analyses showed no correlation or covariance of erosion rates with quartz content or rock type (tables S4 to S6 and figs. S1 and S6).

We report our results for the Pearson correlation coefficients (R) averaged over n catchments within 2° latitudinal increments (Fig. 2, D and E). The Pearson R values are a metric for the degree of linear dependence between individual parameters and the erosion rate (e.g., fig. S9). We chose the 2° increments as the minimum spatial scale over which a sufficient number of catchments ($n > 5$) are available for analysis (e.g., fig. S9 and table S5). The cor-

relation coefficients are reported for a range of values in each bin using a Monte Carlo analysis of the 2σ range of uncertainties in a catchment (Figs. 1 and 2). We classified the absolute values of the correlation coefficients into regions with very weak (0.00 to 0.19), weak (0.20 to 0.39), moderate (0.40 to 0.59), strong (0.60 to 0.79), and very strong (0.80 to 1.0) R values (23).

We found that precipitation had a clear latitudinal gradient, but the correlation coefficients between precipitation and erosion rate do not follow this gradient. Instead, they oscillate between very weak to moderate (Fig. 2D). Similarly, no clear systematic latitudinal variation in the correlation coefficients existed between slope and erosion rates (except between 6°S and 12°S). By contrast, vegetation-erosion correlations show a pattern of latitudinal variations to the north and south of the arid region (region A, Fig. 2E, 18°S to 32°S). More specifically, in the arid (<100 mm/year) and sparsely vegetated region A (vegetation cover <20%), the vegetation-erosion correlation indicated a very weak to moderate negative relationship. To the north and south of region A, the correlation between

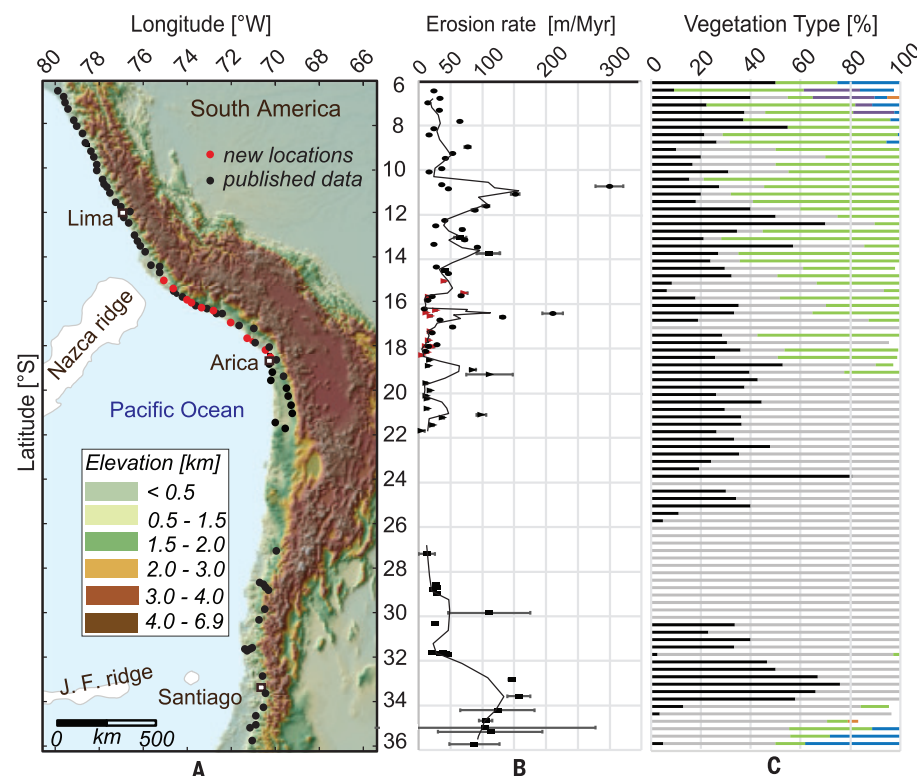


Fig. 1. Topography, erosion rates, and vegetation types along the western Andean margin. (A) Topographic map showing the catchment-averaged erosion rate sample locations of river sediments from the Andean Western Cordillera. Black dots indicate published ^{10}Be concentrations. Red dots are new data presented in this study (tables S1 and S2). (B) Calculated catchment-averaged erosion rates (m/Myr) with 1σ uncertainty versus latitude ($^\circ\text{S}$). The black line represents the three-point moving average. All catchment-averaged erosion rates were calculated with the same procedure (materials and methods). (C) Percent vegetation type versus latitude across a 100-km-wide latitudinal profile in the Andean Western Cordillera. Gray lines represent barren or sparsely vegetated areas. Black lines denote shrublands. Green lines represent grasslands, and purple lines indicate woody savannas. Blue lines represent mixed forests and orange lines, evergreen forest. Values are derived from MODIS 2012 vegetation continuous field data.

Fig. 2. Latitudinal variations in vegetation cover, precipitation, slope, and their correlation with erosion rates from the catchments shown in Fig. 1A. (A) Vegetation cover and type plotted versus latitude. Vegetation types shown in colored zones are from MODIS 2012 vegetation continuous field data (see also Fig. 1C). The vegetation types represent mixed forest (I), grassland (II), open shrubland (III), and barren or sparsely vegetated areas (IV). (B) TRMM2b precipitation versus latitude. (C) Mean catchment slope (90 m window) versus latitude. All dashed lines in (A) to (C) represent the three-point moving averages of the 2σ variation from the mean of each value. Solid lines in (A) to (C) represent the three-point moving averages of values. (D) Correlation coefficients (R) versus latitude for erosion rate and slope (red) and erosion rate and precipitation (blue). Dots and color-shaded regions show the mean and 1σ uncertainty within each bin based on a Monte Carlo analysis of the variability in erosion rates and precipitation or slope within each 2° bin (see tables S3 to S5 for all values plotted). (E) Correlation coefficient versus latitude for erosion rates and vegetation cover. Dots and color-shaded regions show mean and 1σ uncertainty, as in (D.). The dashed green line represents trends in the correlation coefficient with latitude.

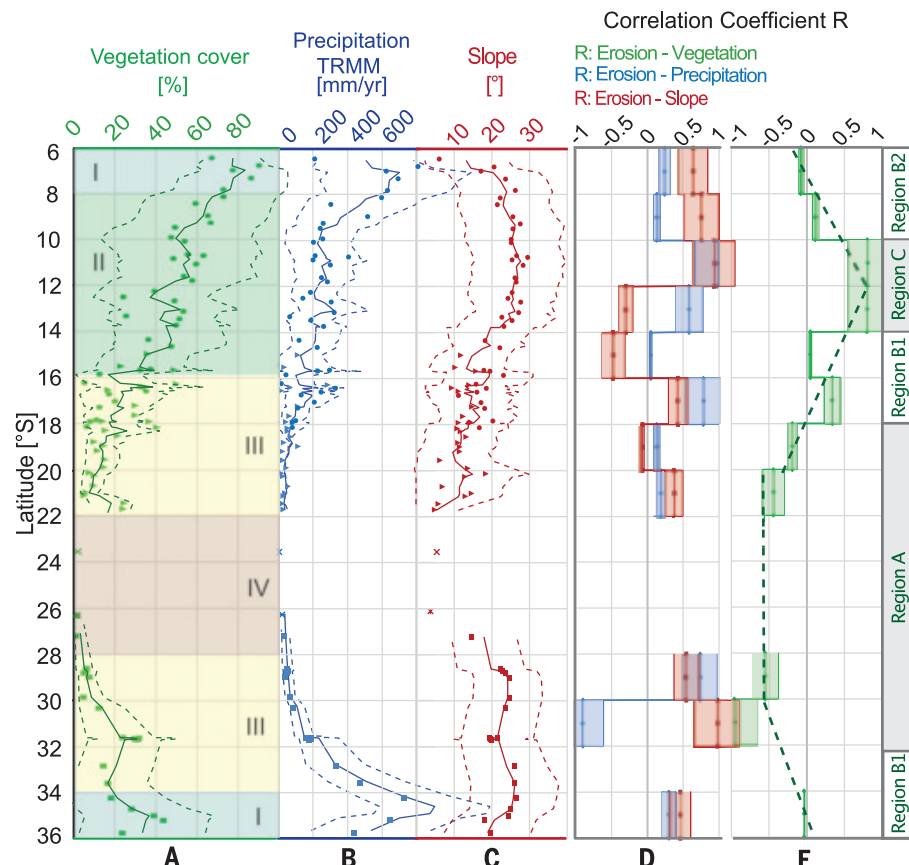
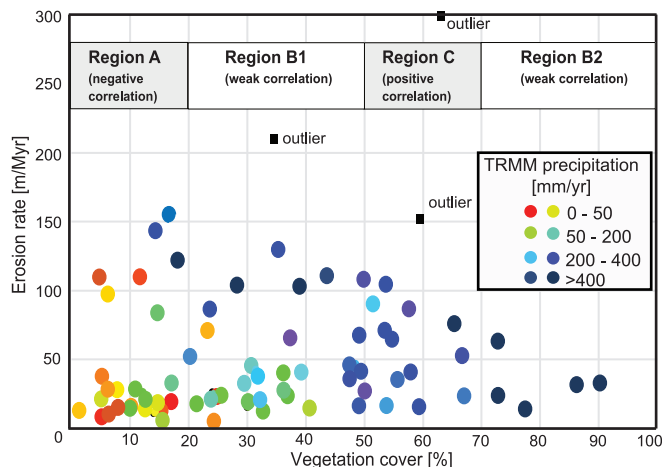


Fig. 3. Observed relationship between erosion rate, vegetation cover, and mean annual precipitation rate (colors) for each catchment in Figs. 1 and 2.

Black squares represent outliers that are possibly biased by glaciation in the catchment. For comparison, fig. S6A is identical but color-coded by slope. Regions labeled correspond to those identified in Fig. 2E.



vegetation and erosion rate was near 0 and very weak to moderate (region B1, Fig. 2E). Further north (region C, 14°S to 10°S), in the more vegetated (50 to 60% vegetation cover) and wetter (200 to 300 mm/year precipitation) latitudes, a positive strong to very strong correlation was found. We could not document a similar maximum in the vegetation-erosion correlation (or region C) in the south from 36°S to 40°S because we lacked samples from the region. In the most heavily vegetated (50 to 80% cover) and wettest (300 to 700 mm/year precipitation) area north of 10°S (region B2),

the correlation strength is very weak and similar to that in region B1 (Fig. 2E).

We interpreted latitudinal changes in the vegetation-erosion correlation strength (regions A to C, Fig. 2E) qualitatively on the basis of abiotic and biotic factors that influence catchment erosion. Region A (18°S to 32°S) has sparse vegetation cover ($<20\%$) and arid conditions. Stochastic variations in precipitation occur in this region over decadal time scales such that we observed, with one exception, a positive correlation with erosion rates and precipitation (Fig. 2D). A moderate negative vegetation-

erosion correlation occurred in region A, which indicated that the sparse vegetation present was sufficient to influence erosion rates by increasing surface roughness for overland flow (Fig. 2E). Region B1 (14°S to 18°S and 32°S to 36°S) has a vegetation cover of 20 to $\sim 50\%$ and represents a transition zone where a very weak to weak erosion-vegetation correlation occurs. This weak vegetation-erosion correlation could be due to competing processes such that increased biotic regolith production associated with higher vegetation cover is offset by the ability of vegetation cover to retard the physical transport of sediment. Alternatively, the weak correlation between erosion rates and vegetation in region B1 (and also B2) could reflect areas where landscapes were closer to steady state with the rock uplift rate. In steady-state landscapes, correlations between erosion rate and vegetation, or precipitation, are expected to be weak. In this case, erosion rates should be positively correlated with slope. This interpretation is partially supported in regions B1 and B2. In region C (14°S to 10°S , Fig. 2E), the observed maximum in the erosion rate and vegetation correlation indicated that vegetation contributed to enhanced weathering and erosion but was ineffective at stabilizing hillslopes from erosion under the higher (~ 300 mm/year) precipitation rates. Finally, the decrease in the correlation to near zero in the northernmost

area (region B2) indicated that dense vegetation cover of 50 to 85% hindered erosion such that the correlation coefficient between vegetation and erosion decreased (24). We also found that vegetation cover greater than 50% leads to a maintenance of steep mean slopes (25° to 30°; fig. S6). By contrast, regions A and B1 had lower mean slope angles between 5° and 25°.

Our observations are consistent with coupled vegetation-landscape evolution modeling work that investigated the effects of varying precipitation and vegetation cover on catchment erosion in Chile (25). More specifically, Schmid *et al.* [see figure 17 in (25)] applied a nonlinear parameterization for vegetation cover-dependent fluvial erodibility and hillslope diffusivity. They found an inverse relationship (negative correlation) between catchment erosion and vegetation cover for sparsely vegetated areas (~10% vegetation cover) due to small increases in precipitation in an arid setting resulting in an increase in vegetation cover that reduced erosion. The negative vegetation-erosion correlation and positive precipitation-erosion correlation that we observed in region A (Fig. 2, D and E) are comparable to this modeling result (25). By contrast, in more vegetated regions (~70% vegetation cover), the model results suggest that precipitation and vegetation cover are positively correlated with catchment erosion. This trend is caused by the high precipitation rates required to sustain dense vegetation cover having a stronger impact on runoff-related erosion and biotic weathering than the stabilizing effects of vegetation cover on erosion. This prediction is consistent with the positive vegetation-erosion and precipitation-erosion correlations that we observed in regions C (~10°S to 14°S, 50 to 70% vegetation cover, Fig. 2E). The northward decrease in the vegetation-erosion correlation (region B2) that we observed reflects that above ~50 to 70% vegetation cover, vegetation effects on inhibiting erosion outpace increases in the precipitation rates that promote erosion. Thus, both the observations presented here and landscape evolution model results (25) confirm a bidirectional response of vegetation cover and precipitation effects on catchment erosion. Similar findings have been reported for smaller geographic areas in both East Africa (26) and the Himalaya (27), as well as globally for differing amounts of tree cover (20).

Complications associated with our interpretations could result from latitudinal variations in tectonic activity, paleoclimate, and paleovegetation cover. However, the regions of vegetation-erosion interactions that we identified (Fig. 2) do not correspond to known patterns of upper plate seismicity or subducting oceanic ridges (28) (fig. S7). The main phases of mountain building in the Andean Western Cordillera were from 20 to 10 million years ago (Ma) and terminated around 9 Ma (29, 30), near the time when river knick-points initiated and migrated

upstream (31) (fig. S8). Another potential caveat is that paleoprecipitation rates and paleovegetation cover could differ in magnitude from the modern values. However, paleo-precipitation gradients (from Pliocene to modern time) in the region are similar to modern precipitation gradients along the Andean Western Cordillera (32) (fig. S5), and latitudinal variations in paleovegetation cover in north-central Chile are estimated to be within 5 to 15% of the present-day values (33). Finally, this study focuses on regional-scale variations in vegetation cover and catchment erosion. This simplified approach highlights the need to evaluate how individual plant functional types (e.g., trees, shrubs, grasses), and not just total vegetation cover, could affect biotic weathering and erosional processes [e.g., (2, 20, 33)].

Our observations have broader implications for the vegetation cover effects on erosion. The latitudinal variations that we identified in the correlation strength between vegetation, precipitation, and erosion imply that studies conducted on a smaller spatial scale could poorly resolve vegetation-erosion interactions. For example, Fig. 3 shows that over the entire study area, the magnitude, and variance, of erosion rates observed decrease with increased vegetation cover. If we take a subset of samples from the catchments shown in Fig. 3, then recovering correlations between vegetation and erosion would be difficult, particularly in regions with low (<60%) vegetation cover, where the variance in erosion rates is high. This conclusion helps explain the diversity of vegetation-erosion relationships synthesized in previous work (34, 35). Results from previous studies (15, 26, 36) have shown both positive and negative correlations between vegetation cover, precipitation, and erosion rates. These seemingly conflicting results may have occurred in areas that are located at, or straddle, the diverse range of climate and vegetation cover regions identified here (Figs. 2E and 3).

In conclusion, we found that correlations between catchment erosion and vegetation cover spanning 30° latitude varied in their direction (positive or negative) and strength (very weak to very strong). These observations, taken together with previous modeling results, indicated a bidirectional relationship between vegetation cover and catchment erosion. The source of this nonlinearity is due to competing, and latitudinally varying, interactions between precipitation and vegetation on erosion. Results presented here provide a regional context for future (smaller scale) studies investigating similar interactions over a narrower range of vegetation cover and precipitation.

REFERENCES AND NOTES

1. R. A. Berner, *Science* **276**, 544–546 (1997).
2. D. W. Schwartzman, T. Volk, *Nature* **340**, 457–460 (1989).
3. G. J. Retallack, *Science* **276**, 583–585 (1997).
4. L. A. Derry, *Science* **311**, 1386–1387 (2006).
5. W. E. Dietrich, J. T. Perron, *Nature* **439**, 411–418 (2006).
6. R. Buitenerwerf, L. Rose, S. I. Higgins, *Nat. Clim. Chang.* **5**, 364–368 (2015).
7. A. Ganopolski, K. Kubatzki, M. Claussen, V. Brovkin, V. Petoukhov, *Science* **280**, 1916–1919 (1998).
8. R. D. Alward, J. K. Detling, D. G. Milchunas, *Science* **283**, 229–231 (1999).
9. N. Zeng, K. Hales, J. D. Neelin, *J. Clim.* **15**, 3474–3487 (2002).
10. W. B. Langbein, S. A. Schumm, *Trans. Am. Geophys. Union* **39**, 1076–1084 (1958).
11. D. E. Granger, M. Schaller, *Elements* **10**, 369–373 (2014).
12. F. von Blanckenburg, J. K. Willenbring, *Elements* **10**, 341–346 (2014).
13. M. C. Peel, B. L. Finlayson, T. A. McMahon, *Hydrol. Earth Syst. Sci. Discuss.* **4**, 439–473 (2007).
14. S. Rivas Martínez, G. Navarro Sánchez, Á. Penas Merino, M. Costa Talens, *Int. J. Geobot. Res.* **1**, 21–40 (2011).
15. S. Carretier *et al.*, *Geology* **41**, 195–198 (2013).
16. J. Starke, T. A. Ehlers, M. Schaller, *J. Geophys. Res. Earth Surf.* **122**, 1949–1971 (2017).
17. F. Kober *et al.*, *Earth Surf. Process. Landf.* **34**, 398–412 (2009).
18. C. J. Placzek, A. Matmon, D. E. Granger, J. Quade, S. Niedermann, *Earth Planet. Sci. Lett.* **295**, 12–20 (2010).
19. D. McPhillips, P. R. Bierman, T. Crocker, D. H. Rood, *J. Geophys. Res. Earth Surf.* **118**, 2488–2499 (2013).
20. A. K. Mishra, C. Placzek, R. Jones, *PLOS ONE* **14**, e0211325 (2019).
21. J. Hartmann, N. Moosdorf, *Geochim. Geophys. Geosyst.* **13**, Q12004 (2012).
22. T. J. Dunai, *Earth Planet. Sci. Lett.* **176**, 157–169 (2000).
23. K. G. Renard, J. R. Freimund, *J. Hydrol. (Amst.)* **157**, 287–306 (1994).
24. M. Schaller, T. A. Ehlers, K. A. H. Lang, M. Schmid, J. P. Fuentes-Espoz, *Earth Planet. Sci. Lett.* **489**, 111–122 (2018).
25. M. Schmid, T. A. Ehlers, C. Werner, T. Hickler, J.-P. Fuentes-Espoz, *Earth Surf. Dynam.* **6**, 859–881 (2018).
26. V. T. Acosta *et al.*, *Lithosphere* **7**, 408–420 (2015).
27. S. M. Olen, B. Bookhagen, M. R. Strecker, *Earth Planet. Sci. Lett.* **445**, 57–67 (2016).
28. L. Audin *et al.*, *J. Quat. Sci.* **18**, 681–694 (2003).
29. G. Wörner, K. Hammerschmidt, F. Henjes-Kunst, J. Lezaun, H. Wilke, *Rev. Geol. Chile* **27**, 205–240 (2000).
30. T. E. Jordan *et al.*, *Tectonics* **29**, TC5007 (2010).
31. L. M. Abbuhl *et al.*, *Earth Surf. Process. Landf.* **36**, 1464–1473 (2011).
32. S. G. Mutz *et al.*, *Earth Surf. Dynam.* **6**, 271–301 (2018).
33. C. Werner *et al.*, *Earth Surf. Dyn.* **6**, 859–881 (2018).
34. C. S. Riebe, J. W. Kirchner, D. E. Granger, R. C. Finkel, *Geology* **29**, 447–450 (2001).
35. J. T. Perron, *Annu. Rev. Earth Planet. Sci.* **45**, 561–591 (2017).
36. Y. Garcin *et al.*, *Earth Planet. Sci. Lett.* **459**, 58–69 (2017).

ACKNOWLEDGMENTS

We thank D. Kost and L. Michel for laboratory and field assistance, respectively. We also thank three anonymous reviewers and R. Drews for their thoughtful comments. **Funding:** This study was funded by a European Research Council (ERC) consolidator grant (CoG 615703) and the German Science Foundation (DFG) priority research program *EarthShape* (EH329/17-2) to T.A.E. **Author contributions:** T.A.E. and J.S. planned the study. J.S. was responsible for all sample collection and calculations. J.S. and M.S. performed the laboratory analysis. All authors contributed to manuscript and figure preparation. **Competing interests:** The authors have no competing financial conflicts of interest with this study. **Data and materials availability:** Data reported in the paper are presented in the supplementary materials.

SUPPLEMENTARY MATERIAL

science.sciencemag.org/content/367/6484/1358/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S6
Google Earth kmz file of catchment properties
References (37–68)

10 August 2019; accepted 24 February 2020
10.1126/science.aaz0840

NEUROSCIENCE

Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work

A. Westbrook^{1,2,3*}, R. van den Bosch^{2,3}, J. I. Määttä^{2,3}, L. Hofmans^{2,3}, D. Papadopetaki^{2,3}, R. Cools^{2,3,†}, M. J. Frank^{1,4,†}

Stimulants such as methylphenidate are increasingly used for cognitive enhancement but precise mechanisms are unknown. We found that methylphenidate boosts willingness to expend cognitive effort by altering the benefit-to-cost ratio of cognitive work. Willingness to expend effort was greater for participants with higher striatal dopamine synthesis capacity, whereas methylphenidate and sulpiride, a selective D2 receptor antagonist, increased cognitive motivation more for participants with lower synthesis capacity. A sequential sampling model informed by momentary gaze revealed that decisions to expend effort are related to amplification of benefit-versus-cost information attended early in the decision process, whereas the effect of benefits is strengthened with higher synthesis capacity and by methylphenidate. These findings demonstrate that methylphenidate boosts the perceived benefits versus costs of cognitive effort by modulating striatal dopamine signaling.

Cognitive control is effortful, causing people to avoid demanding tasks (1) and to discount goals (2, 3). Striatal dopamine invigorates physical action by mediating cost-benefit tradeoffs (4). In corticostriatal loops, dopamine has opponent effects on D1- and D2-expressing medium spiny neurons, which modulate sensitivity to the benefits versus the costs of actions (5). Given that similar mechanisms may govern cognitive action selection (6–8), we hypothesized that striatal dopamine could promote willingness to exert cognitive effort, enhancing attention, planning, and decision-making (8–11).

Converging evidence on cognitive motivation in Parkinson's disease (12–15) provides

an initial basis for this conjecture. Moreover, catecholamine-enhancing psychostimulants alter cognitive effort in rodents (10) and humans (16). This raises the question of whether so-called “smart drugs” act by enhancing the willingness rather than the ability to exert cognitive control. Indeed, the dominant interpretation is that stimulants improve cognitive processing by direct cortical effects, noradrenaline transmission (17, 18), and/or concomitant working memory improvements (19). We instead hypothesized that methylphenidate (a dopamine and noradrenaline reuptake blocker) boosts cognitive control by increasing striatal dopamine and, accordingly, sensitivity to the benefits versus costs of cognitive effort.

Fifty healthy, young adults (ages 18 to 43, 25 men) completed a cognitive effort-discounting paradigm (2) quantifying subjective effort costs as the amount of money required to make participants equally willing to perform a hard (N = 2, 3, 4) versus an easier (N = 1, 2) level of the N-back working memory task. We defined the subjective value of an offer to complete a harder task (N = 2 to 4) as the amount offered for the task minus subjective costs.

Subjective values decreased with N-back load, indicating rising subjective costs (Fig. 1A). Critically, greater willingness to expend cognitive effort correlated with higher dopamine synthesis capacity (measured using [¹⁸F]DOPA positron emission tomography) in the caudate nucleus [independently defined (20); Fig. 1, A to C, and fig. S1]. A mixed-effects model confirmed that on placebo, subjective values increased with larger offer amounts (€4 versus €2 offers; $\beta = 0.022$, $P = 0.011$), smaller relative load ($\beta = -0.15$, $P = 8.9 \times 10^{-15}$), and higher dopamine synthesis capacity ($\beta = 0.064$; $P = 0.022$). These individual difference effects were selective to the caudate nucleus (figs. S1 and S2), consistent with human imaging studies on cognitive motivation (7, 21, 22). Although N-back performance decreased with load,

¹Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA. ²Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Nijmegen, Netherlands. ³Radboud University Medical Centre, Department of Psychiatry, Nijmegen, Netherlands. ⁴Carney Institute for Brain Science, Brown University, Providence, RI, USA.

*Corresponding author. Email: andrew.westbrook@brown.edu

†Senior authors contributed equally to this work and are listed in alphabetical order.

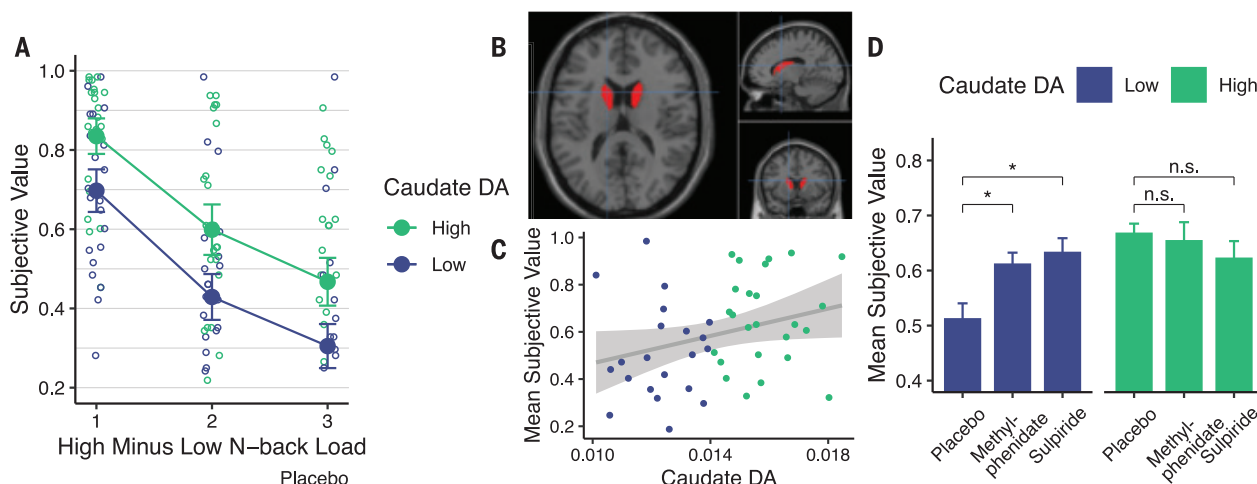


Fig. 1. Participants discounted offers as a function of cognitive load, dopamine synthesis capacity (DA), and drug. (A) Offers were discounted more for high- versus low-load levels and more by participants with below- versus above-median dopamine synthesis capacity. Circles show individual's indifference points. Filled circles show group mean ± SEM. **(B)** Caudate nucleus mask. Crosshairs indicate Montreal Neurological

Institute (MNI) coordinates of [−14, 10, 16]. **(C)** Participant-averaged subjective values correlated with synthesis capacity on placebo (Spearman's $r = 0.32$, $P = 0.029$). **(D)** Methylphenidate [$t_{\text{paired}(22)} = 2.29$; $P = 0.032$] and sulpiride [$t_{\text{paired}(22)} = 2.36$; $P = 0.028$] increased subjective values for participants with low but not high synthesis capacity ($P \geq 0.021$ for both). Error bars indicate within-subject SEM.

dopamine effects on discounting could not be attributed to performance changes (see the supplementary results). Moreover, there were no drug effects on performance because drugs were administered after the N-back task.

If dopamine mediates cognitive effort, then it should be possible to increase motivation pharmacologically. Indeed, both methylphenidate and sulpiride increased subjective values for participants with low, but not high, dopamine synthesis capacity (Fig. 1D and fig. S2, B and C). A mixed-effects model revealed that both methylphenidate ($\beta = -0.069$, $P = 0.0042$) and sulpiride ($\beta = -0.10$, $P = 8.3 \times 10^{-4}$) interacted with dopamine synthesis capacity to increase subjective values. Neither drug showed main effects (both $P \geq 0.37$).

The converging effects of synthesis capacity and two separate drugs strongly implicate striatal dopamine. Methylphenidate blocks re-

uptake, increasing extracellular striatal dopamine tone (23), and can amplify transient dopamine signals (24). Sulpiride is a D2 receptor antagonist that at low doses can increase striatal dopamine release by binding to presynaptic autoreceptors, enhancing striatal reward signals and learning (6, 25). Although sulpiride can block postsynaptic D2 receptors at higher doses (26), both drugs increase behavioral vigor [reaction times and saccade velocities; compare (6, 26)], especially in participants with low dopamine synthesis capacity, corroborating that both drugs increase dopamine release (see the supplementary results).

To assess whether dopamine amplifies subjective benefits versus costs, we made a series of offers, in a second decision task, centered around participants' indifference points (Fig. 2A). To generate specific predictions, we simulated psychometric choice functions with a computational model of striatal dopamine effects

on decision-making (5). With higher dopamine, the model predicts enhanced sensitivity to benefits and reduced sensitivity to costs. This manifests as a steeper choice function to the right of indifference, where the ratio of benefits to costs (of the high- versus low-effort option) is larger, but a shallower choice function to the left, where the benefits-to-costs ratio is smaller (Fig. 2B).

Choice behavior supported model predictions. Simulated effects were mirrored by effects of dopamine synthesis capacity (Fig. 2C) and of methylphenidate and sulpiride versus placebo (Fig. 2D). Formally, high effort selection was sensitive to both benefits (offer amount differences; $\beta = 2.30$, $P = 1.2 \times 10^{-9}$) and costs (load differences; $\beta = -1.07$, $P = 2.2 \times 10^{-16}$). The effect of benefits increased with synthesis capacity ($\beta = 0.65$, $P = 0.0024$) and on methylphenidate ($\beta = 1.34$, $P = 0.0048$), whereas the effect of costs was attenuated on

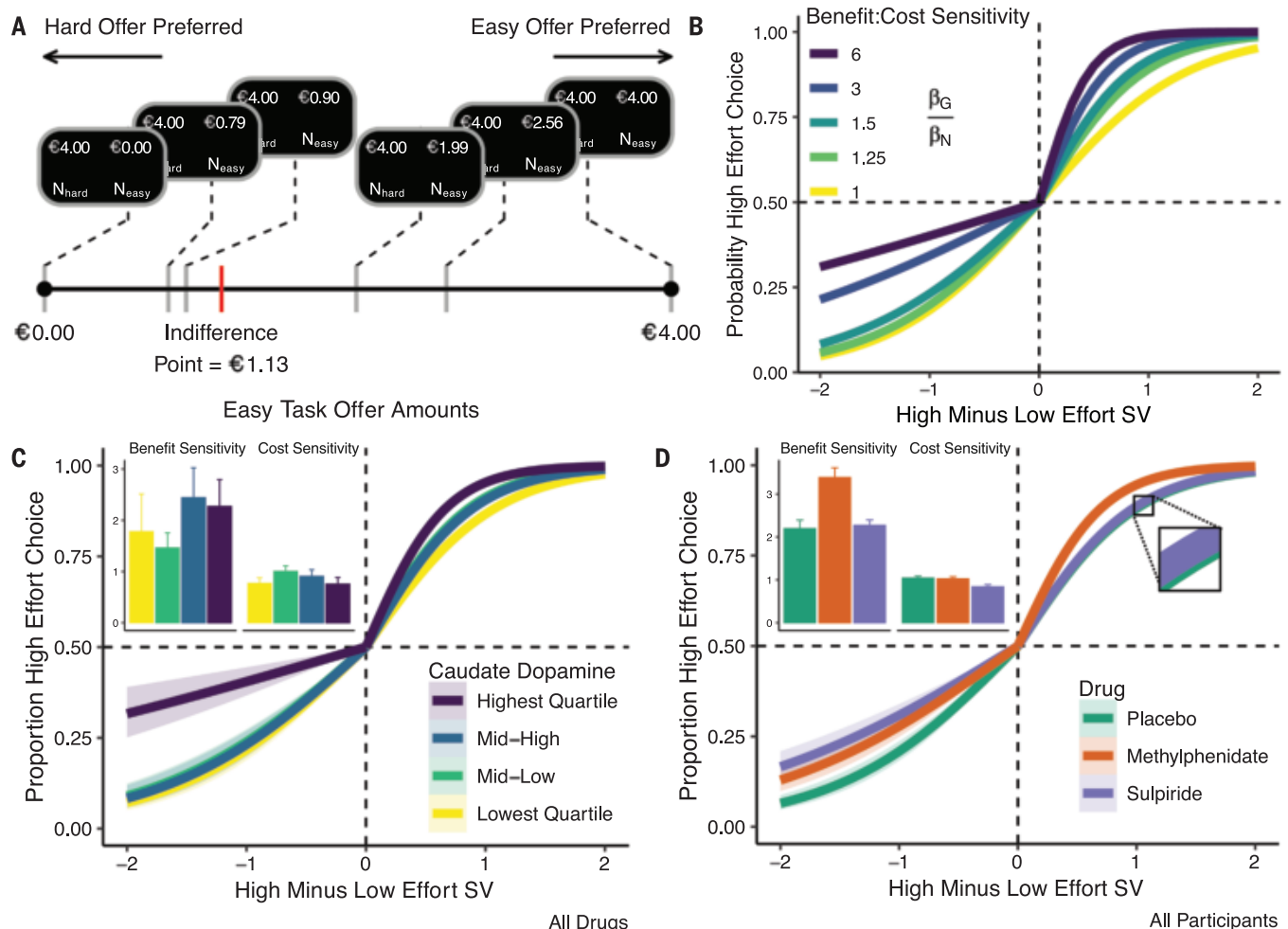


Fig. 2. Dopamine alters valuation by reweighting the benefits versus costs of cognitive work. (A) Low-effort (N_{easy}) offers were paired with a €4 offer for a high-effort N-back task (N_{hard}). (B to D) Simulated effects of dopamine on benefit-versus-cost sensitivity (B) are mirrored by empirical effects of dopamine synthesis capacity (C) and pharmacological agents

(D). Mixed-effects logistic regression curves and 95% confidence intervals (CI) fit across all drugs for each synthesis capacity quartile (C) or all participants for each drug (D). SV, subjective value. Insets show the estimated effect of benefits and costs on choice across participants in each quartile \pm SEM (C) and on each drug \pm within-subject SEM (D).

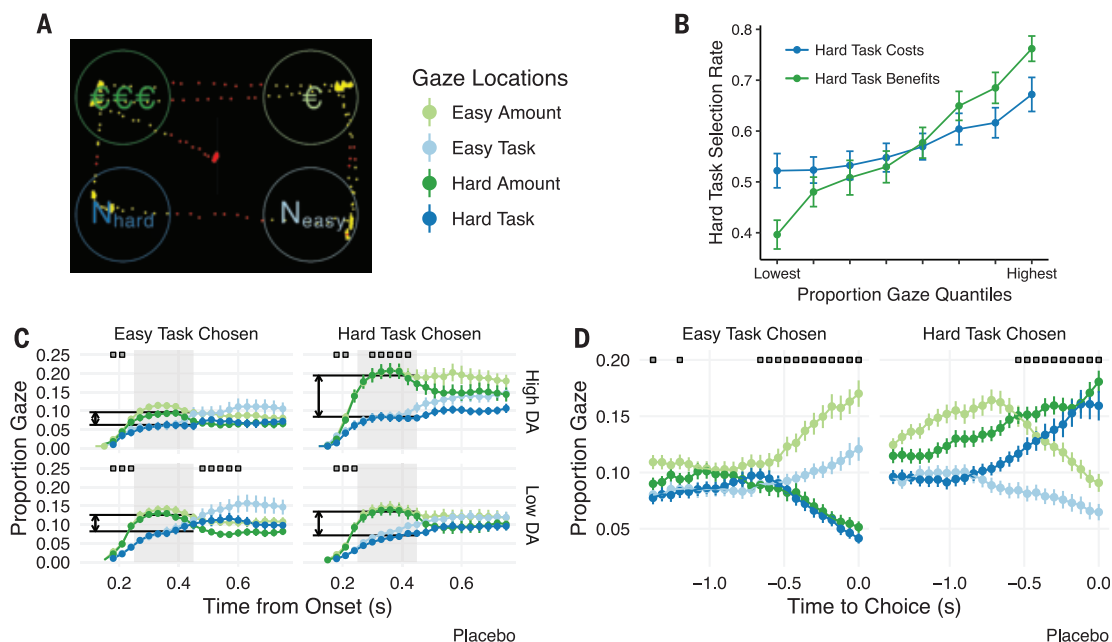


Fig. 3. Effect of gaze, value, and dopamine synthesis capacity on effort selection. (A) Participants decided between offers with costs (N-back load) and benefits (Euros) separated in space. Dots indicate gaze at (yellow) and away from (red) offers. (B) Proportion gaze at the high-effort offer predicted high-effort selection, and more so with gaze at benefits versus costs. (C and D) Proportional (cross-trial) gaze at the four information quadrants

after offer onset and leading up to response. In (C), early gaze (250 to 450 ms after offer onset) is indicated by gray shading, and boxes indicate time points at which participants gazed reliably more at either benefits or cost information (paired *t* tests, $P < 0.05$). In (D), boxes indicate time points at which participants gazed reliably more at the selected offer (one-tailed paired *t* tests, $P < 0.05$). Error bars indicate \pm SEM.

sulpiride ($\beta = 0.24$, $P = 0.036$). Participants also selected high-effort choices more often with higher dopamine synthesis capacity ($\beta = 1.02$, $P = 3.1 \times 10^{-4}$) and on methylphenidate ($\beta = 1.75$, $P = 0.0016$) versus placebo, but not reliably so for sulpiride ($\beta = 0.46$, $P = 0.12$). No other interactions or main effects were significant (all $P \geq 0.47$).

These results clearly implicate dopamine in choice, but they do not uncover how decision-making is altered. Dopamine could increase attention to benefits versus costs. Alternatively, it could alter the impact of these attributes on choice without affecting attention itself. We thus tracked eye gaze to quantify attention to attributes and how it interacted with dopamine. Proportion gaze at an offer (either costs or benefits) strongly predicted offer selection [Fig. 3B; $\beta = 0.30$, $P = 7.6 \times 10^{-6}$; cf. (27, 28)]. However, gaze at benefits predicted steeper increases in hard task selection than gaze at costs (gaze by dimension interaction: $\beta = 0.41$, $P = 1.1 \times 10^{-5}$).

Gaze patterns implicated dopamine in enhancing the impact of attention to benefits versus costs on the decision to engage in cognitive effort. Early in a trial, participants fixated on benefits (of either offer) more than on costs, and this asymmetry was larger in trials in which they chose the high-effort option (choice effect: $\beta = 0.41$, $P = 0.0017$; Fig. 3C). Moreover, this

effect was stronger in participants with higher dopamine synthesis capacity (choice by synthesis capacity interaction: $\beta = 0.37$, $P = 0.0045$; top versus bottom row, Fig. 3C). For those with lower synthesis capacity, methylphenidate strengthened this relationship (interaction between drug, synthesis capacity, and choice: $\beta = -0.36$, $P = 0.012$), although sulpiride did not ($\beta = -0.041$, $P = 0.78$). Drugs and synthesis capacity did not affect gaze patterns themselves ($P \geq 0.10$ for main effects), indicating that dopamine did not alter attention to benefits but rather strengthened the impact of attention to benefits versus costs on choice.

Gaze may correlate with choice because attention amplifies the perceived value of attended offers, causally biasing choice (27). Alternatively, reversing this causality, participants may simply look more at offers that they have already implicitly chosen (28). We found evidence for both: Early in a trial, attention influenced choice, whereas later, choice influenced attention. To address this, we fit drift diffusion models (29) in which cost and benefit information accumulate in a decision variable rising to a threshold. This variable is the instantaneous difference in the perceived value of the high- versus the low-valued offer. We considered “attention-biasing choice” models with multiplicative effects (i.e., gaze multiplies the effects of value information) and “choice-

biasing attention” models in which gaze has a simple, additive effect (i.e., gaze correlates with choice but does not amplify value) (28). The best-fitting model [Fig. 4, A to C, and Eq. 1 (30)] included both additive and multiplicative effects (see the supplementary results).

We next considered the possibility that the gaze-value interactions changed dynamically across the trial. Indeed, ~ 775 ms before responding, participants began committing their gaze toward the to-be-chosen offer (Fig. 3D). Thus, whereas early gaze appears to influence choice formation (Fig. 3C), later gaze appears to reflect latent choices once formed. On this basis, we investigated whether early attention causally amplifies attended attributes (a multiplicative combination) whereas late gaze simply correlates with choice (additive; Fig. 4A). To test our hypothesis, we split trials according to when participants began committing their gaze to the to-be-chosen offer (the “bifurcation”) for each participant and session and refit our model to gaze data from before or after this time point. The result supported our hypothesis. Multiplicative terms were reliably positive before bifurcation but near zero after bifurcation, with the opposite pattern for additive terms (Fig. 4, F and G). These results support the idea that although early attention appeared to amplify the effect of benefits

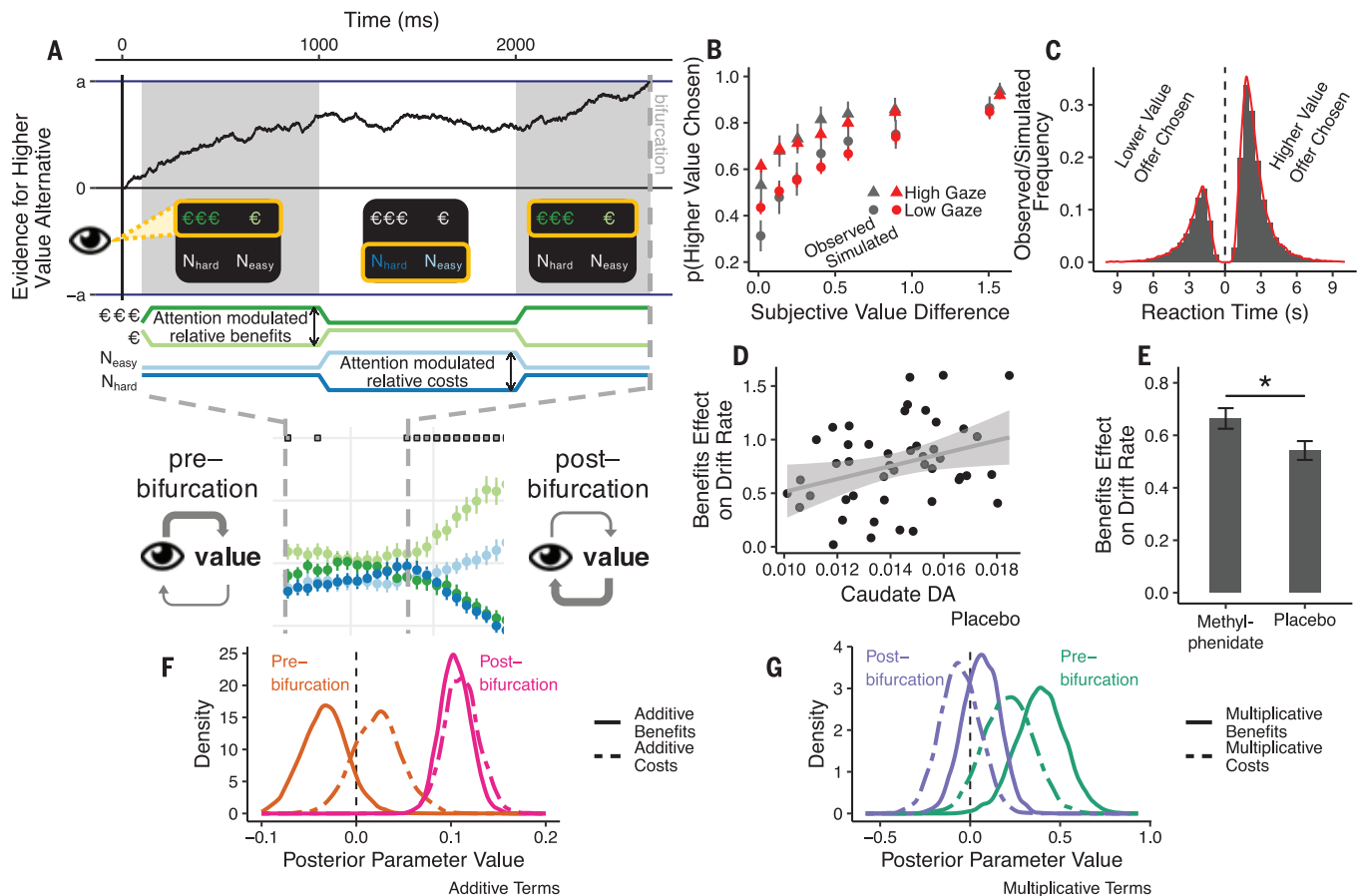


Fig. 4. Gaze dynamically biases and then reflects implicit choice. (A) Gaze attribute model. Early gaze amplified the effect of attended versus unattended attributes on choice during evidence accumulation to a decision threshold (a). Late gaze reflected the to-be-selected response. (B and C) Model simulations (red) predicted choice (gray) (B, split by median proportion gaze at the higher value offer) and reaction time (C) distributions. (D) Benefits effect on drift rate correlate with dopamine synthesis capacity (95% CI shown). (E) Methylphenidate enhances the benefit effect. (F and G) Posterior parameter densities from models fit alternately

with pre- or postbifurcation gaze on placebo. (F) Additive benefit ($\beta_1 = -0.030$; $P = 0.076$) and cost ($\beta_2 = 0.020$; $P = 0.81$) gaze terms were approximately zero before bifurcation and reliably positive after bifurcation ($\beta_1 = 0.10$; $P < 2.2 \times 10^{-16}$ and $\beta_2 = 0.11$; $P = 0.0031$). (G) Multiplicative interaction terms reveal that the effects of benefits ($\beta_3 - \beta_5 = 0.40$; $P = 0.0024$) and costs (at trend level; $\beta_4 - \beta_6 = 0.12$; $P = 0.060$) were larger when fixating the respective attribute before bifurcation, whereas neither term was different from zero after bifurcation ($\beta_3 - \beta_5 = 0.07$; $P = 0.27$ and $\beta_4 - \beta_6 = -0.060$; $P = 0.70$). Error bars indicate \pm SEM.

versus costs, later gaze simply reflected a latent choice.

Finally, we tested whether the effects of dopamine on choice could be attributed to these dynamic decision processes. Indeed, both higher dopamine synthesis capacity [on placebo; Eq. 1 (30): ($\beta_3 + \beta_5$)/2; Pearson $r = 0.30$, $P = 0.039$; Fig. 4D] and methylphenidate [$t_{\text{paired},45} = 2.54$, $P = 0.015$; Fig. 4E] increased the effect of benefits on evidence accumulation. The corresponding effect of sulpiride on cost was not significant [Eq. 1 (30): ($\beta_4 + \beta_6$)/2; $t_{\text{paired},45} = -1.41$; $P = 0.17$; see the supplementary discussion]. We further found that methylphenidate amplified the effects of benefits on drift rate even when only modeling pre-bifurcation gaze [$t_{\text{paired},45} = 2.44$; $P = 0.019$] before the latent choice. Collectively, our results support that striatal dopamine enhances motivation for cognitive effort by amplifying the

effects of benefits versus costs attended early in a decision.

REFERENCES AND NOTES

- W. Kool, J. T. McGuire, Z. B. Rosen, M. M. Botvinick, *J. Exp. Psychol. Gen.* **139**, 665–682 (2010).
- A. Westbrook, D. Kester, T. S. Braver, *PLOS ONE* **8**, e68210 (2013).
- M. A. Apps, L. L. Grima, S. Manohar, M. Husain, *Sci. Rep.* **5**, 16880 (2015).
- J. D. Salamone et al., *Behav. Processes* **127**, 3–17 (2016).
- A. G. E. Collins, M. J. Frank, *Psychol. Rev.* **121**, 337–366 (2014).
- M. J. Frank, R. C. O'Reilly, *Behav. Neurosci.* **120**, 497–517 (2006).
- E. Aarts, M. van Holstein, R. Cools, *Front. Psychol.* **2**, 163 (2011).
- A. Westbrook, T. S. Braver, *Neuron* **89**, 695–710 (2016).
- N. D. Volkow et al., *Mol. Psychiatry* **16**, 1147–1154 (2011).
- P. J. Cocker, J. G. Hosking, J. Benoit, C. A. Winstanley, *Neuropsychopharmacology* **37**, 1825–1837 (2012).
- R. Cools, *Curr. Opin. Behav. Sci.* **4**, 152–159 (2015).
- E. Aarts et al., *Neuropsychologia* **62**, 390–397 (2014).
- S. G. Manohar et al., *Curr. Biol.* **25**, 1707–1716 (2015).
- M. H. M. Timmer, E. Aarts, R. A. J. Esselink, R. Cools, *Eur. J. Neurosci.* **48**, 2374–2384 (2018).
- S. McGuigan et al., *Brain* **142**, 719–732 (2019).
- M. I. Froböse et al., *J. Exp. Psychol. Gen.* **147**, 1763–1781 (2018).
- J. G. Hosking, S. B. Floresco, C. A. Winstanley, *Neuropsychopharmacology* **40**, 1005–1015 (2015).
- R. C. Spencer, D. M. Devilbiss, C. W. Berridge, *Biol. Psychiatry* **77**, 940–950 (2015).
- R. Cools, M. D'Esposito, *Biol. Psychiatry* **69**, e113–e125 (2011).
- P. Piray, H. E. M. den Ouden, M. E. van der Schaaf, I. Toni, R. Cools, *Cereb. Cortex* **27**, 485–495 (2017).
- L. Schmidt, M. Lebreton, M.-L. Cléry-Melin, J. Daunizeau, M. Pessiglione, *PLOS Biol.* **10**, e1001266 (2012).
- W. M. Pauli, R. C. O'Reilly, T. Yarkoni, T. D. Wager, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1907–1912 (2016).
- N. D. Volkow et al., *Am. J. Psychiatry* **155**, 1325–1331 (1998).
- N. D. Volkow et al., *J. Neurosci.* **21**, RC121 (2001).
- G. Jocham, T. A. Klein, M. Ullsperger, *J. Neurosci.* **31**, 1606–1613 (2011).
- C. Eisenegger et al., *Neuropsychopharmacology* **39**, 2366–2375 (2014).
- I. Krajbich, C. Armel, A. Rangel, *Nat. Neurosci.* **13**, 1292–1298 (2010).
- J. F. Cavanagh, T. V. Wiecki, A. Kocher, M. J. Frank, *J. Exp. Psychol. Gen.* **143**, 1476–1488 (2014).
- T. V. Wiecki, I. Sofer, M. J. Frank, *Front. Neuroinform.* **7**, 14 (2013).
- $v_i = \beta_0 + \beta_1(g_{\text{BenA}} - g_{\text{BenB}}) + \beta_2(g_{\text{CostA}} - g_{\text{CostB}}) + \beta_3 g_{\text{BenAVBen}} + \beta_4 g_{\text{CostAVCost}} + \beta_5 g_{\text{CostAVBen}} + \beta_6 g_{\text{BenAVCost}}$ (Eq. 1), where

the rate at which participants accumulate evidence in favor of offer A versus B (v) on trial (i) is given by proportion gaze at benefits (g_{Ben}) and its interaction with the benefits of A versus B (Δv_{Ben}), proportion gaze at costs (g_{Cost}) and its interaction with costs (Δv_{Cost}), as well as additive contributions of gaze at offer A for both benefits ($g_{\text{BenA}} - g_{\text{BenB}}$) and costs ($g_{\text{CostA}} - g_{\text{CostB}}$).

ACKNOWLEDGMENTS

We thank the individuals who participated in this study and J. Wilmott for eye-tracking code and consultation. **Funding:** This work was supported by NWO VICI grant 453-14-005

(2015/01379/VI) to R.C., NIH grant F32MH115600-01A1 to A.W., and NIH grant R01MH080066 to M.J.F. **Author contributions:** Conceptualization: R.C., A.W., M.J.F., Data curation: J.I.M., R.v.d.B., L.H., D.P., Formal analysis: A.W., R.v.d.B., R.C., M.J.F., Funding acquisition: R.C., A.W., Investigation: J.I.M., R.v.d.B., L.H., D.P., Project administration: J.I.M. and R.C., Software: A.W., Writing: A.W., M.J.F., R.C., Supervision: R.C. and M.J.F. **Competing interests:** The authors declare no competing interests. **Data and materials**

availability: Data and analysis scripts will be made publicly available at the conclusion of the parent study at <http://hdl.handle.net/11633/aac2qvfx>.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1362/suppl/DC1
Materials and Methods
Supplementary Results
Figs. S1 to S5
Tables S1 to S7
References (31–37)

21 September 2019; accepted 20 January 2020
10.1126/science.aaz5891

CELL BIOLOGY

Golgi-derived PI(4)P-containing vesicles drive late steps of mitochondrial division

Shun Nagashima¹, Luis-Carlos Tábara^{1*}, Lisa Tilokani^{1*}, Vincent Paupe¹, Hanish Anand¹, Joe H. Pogson², Rodolfo Zunino², Heidi M. McBride^{2†}, Julien Prudent^{1†}

Mitochondrial plasticity is a key regulator of cell fate decisions. Mitochondrial division involves Dynamin-related protein-1 (Drp1) oligomerization, which constricts membranes at endoplasmic reticulum (ER) contact sites. The mechanisms driving the final steps of mitochondrial division are still unclear. Here, we found that microdomains of phosphatidylinositol 4-phosphate [PI(4)P] on trans-Golgi network (TGN) vesicles were recruited to mitochondria–ER contact sites and could drive mitochondrial division downstream of Drp1. The loss of the small guanosine triphosphatase ADP-ribosylation factor 1 (Arf1) or its effector, phosphatidylinositol 4-kinase IIIβ [PI(4)KIIIβ], in different mammalian cell lines prevented PI(4)P generation and led to a hyperfused and branched mitochondrial network marked with extended mitochondrial constriction sites. Thus, recruitment of TGN-PI(4)P-containing vesicles at mitochondria–ER contact sites may trigger final events leading to mitochondrial scission.

Mitochondrial division is initiated at sites where the endoplasmic reticulum (ER) contacts mitochondria, which marks the site of constriction and subsequent recruitment of the large guanosine triphosphatase (GTPase) Dynamin-related protein-1 (Drp1) (1). At these sites, Drp1 oligomerization further enhances mitochondrial constriction driven by GTP hydrolysis (2). It has been suggested that the GTPase Dynamin 2 (Dnm2) is required downstream of Drp1-mediated constriction to terminate membrane scission (3); however, its precise contribution and the molecular details of late events are currently unclear (4, 5). A growing body of evidence supports the role of other factors regulating mitochondrial division, including phospholipids, calcium, and lysosomes (6). Furthermore, a recent study revealed that loss of the small GTPase ADP-ribosylation factor 1 (Arf1) led to alterations in mitochondrial morphology with hyperfusion in *Caenorhabditis*

elegans (7). GTP-bound Arf1 is recruited primarily to the Golgi apparatus, where it is canonically known for its role in the generation of COPI-coated vesicles. GTP-specific effector proteins of Arf1 include phosphatidylinositol 4-kinase IIIβ [PI(4)KIIIβ], which mediates the phosphorylation of phosphatidylinositol to generate phosphatidylinositol 4-phosphate [PI(4)P] (8). This generates lipid microdomains enriched for PI(4)P that are required for membrane-remodeling events (9–12). Given the primary role for these enzymes in membrane dynamics (7, 13), we investigated the mechanisms that underlie the contribution of PI(4)P pools in the regulation of mitochondrial morphology.

Silencing of both Arf1 and PI(4)KIIIβ led to mitochondrial hyperfusion in HeLa cells (Fig. 1, A to D). In contrast to Drp1-silenced cells, loss of PI(4)KIIIβ and Arf1 induced mitochondrial elongation and mitochondrial branching, leading to a highly interconnected network and an increase of mitochondrial intersections called junctions (Fig. 1E). These results were confirmed in two other mammalian cell lines, Cos-7 and U2OS (fig. S1, A to K). We further quantified mitochondrial interconnectivity using a photo-activatable GFP probe targeted to the mitochondrial matrix (OCT-PAGFP) (14) (Fig. 1, F and G). Mitochondrial hyperfusion induced by

PI(4)KIIIβ silencing was rescued upon reexpression of the bovine wild-type (WT) PI(4)KIIIβ (PI4K-HA), but not with the kinase-dead mutant (PI4K-KD-HA) (15) (Fig. 1, H and I, and fig. S1, L and M). Treatment of HeLa or Cos-7 cells with the selective PI(4)KIIIβ inhibitor PIK93 also resulted in mitochondrial hyperfusion and branching (fig. S2). In addition, among the PI(4)K family, only PI(4)KIIIβ silencing induced mild mitochondrial hyperfusion in HeLa cells (fig. S3A–G), but not in Cos-7 cells (fig. S3H–N), which may be coincident with a cell-type-specific decrease of Drp1 and PI(4)KIIIβ protein levels upon silencing (fig. S3F). Finally, cells silenced for ceramide transfer protein (CERT), another Arf1 effector, did not lead to mitochondrial hyperfusion (fig. S4). Thus, both the kinase activity of the specific effector PI(4)KIIIβ and the GTPase Arf1 are required to modulate mitochondrial dynamics.

In transmission electron microscopy (TEM), PI(4)KIIIβ- and Arf1-silenced cells displayed exaggerated mitochondrial hyperfusion and branching (Fig. 2, A to D, and fig. S5A). We observed an accumulation of unusual mitochondrial-hyperconstricted sites in cells lacking either Arf1 or PI(4)KIIIβ (Fig. 2, E to G, and fig. S5, B and C). These sites were characterized by a long and narrow neck, where the inner membrane was observed running parallel with the constricted outer membrane (Fig. 2, A, E, and F, and fig. S5B). In addition, the ER was in close apposition along these constricted sites (Fig. 2G), suggesting that mitochondria–ER contacts (MERCs) were maintained. A similar level of mitochondrial hyperconstriction has been reported in cells silenced with Dnm2 (3), where it has been suggested that this dynamin may act downstream of Drp1 to drive fission. However, silencing Dnm2 in U2OS and HeLa cells failed to recapitulate the mitochondrial hyperfusion and branching phenotype induced by the loss of PI(4)P pools (fig. S6), as recently reported (4, 5), suggesting that Arf1 and PI(4)KIIIβ may not be required for Dnm2 recruitment.

Loss of Arf1 in yeast results in an accumulation of the fusion GTPase Fzo1 at mitochondria and an alteration in mitochondrial morphology (7). However, we found no major changes in the levels of the main pro-fission and pro-fusion regulators after 48 or 72 hours of silencing for

¹Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0XY, UK. ²Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec H3B 2B4, Canada.

*These authors contributed equally to this work.

†Corresponding author. Email: heidi.mcbride@mcgill.ca (H.M.M.); julien.prudent@mrc-mbu.cam.ac.uk (J.P.)

Fig. 1. Arf1 and PI(4)KIII β silencing lead to mitochondrial hyperfusion and branching.

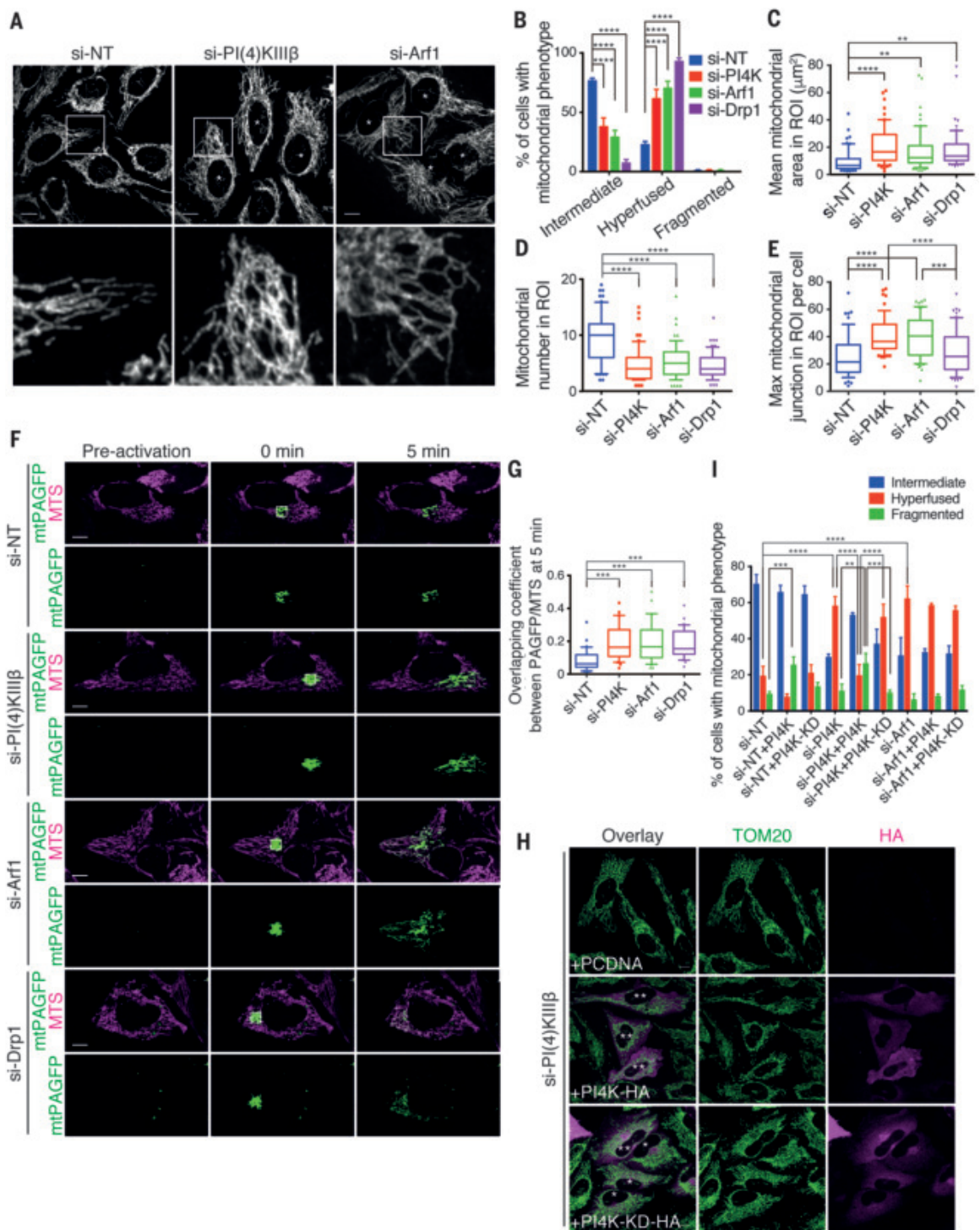
(A) Representative confocal images of mitochondrial morphology in HeLa cells treated with the indicated small interfering RNAs (siRNAs). Mitochondria were labeled using an anti-TOM20 antibody. Asterisks indicate cells with elongated and/or branched mitochondria.

(B) Quantification of mitochondrial morphology from (A). (C to E) Mitochondrial morphology quantified for (C) mean mitochondrial area per mitochondrion, (D) mitochondrial number, and (E) mitochondrial branching measured by maximum mitochondrial junction number for each region of interest (ROI).

(F) Live-cell imaging of HeLa cells treated with indicated siRNAs overexpressing the ornithine carbamoyl-transferase (OCT)-photoactivatable GFP (mt-PAGFP) probe and the mitochondrial marker MTS-Scarlet. (G) Quantification of the OCT-PAGFP probe diffusion over a 5-min period from (F) using the overlapping Mander's coefficient.

(H) Representative confocal images of mitochondrial morphology in HeLa cells silenced with PI(4)KIII β siRNA and transiently overexpressing empty vector (vehicle), WT-PI(4)KIII β -HA (PI4K-HA), and kinase-dead mutant PI(4)KIII β -HA (PI4K-KD-HA). Shown are HA-positive transfected cells with elongated and/or branched mitochondria (*) and intermediate mitochondria (**).

(I) Quantification of mitochondrial morphology from (H) and fig. S1L. All scale bars, 10 μ m. All data are shown as mean \pm SD of at least three independent experiments. For (B) and (I), two-way ANOVA and Tukey's multiple-comparisons test were used; for (C) to (E) and (G), ordinary one-way ANOVA and Tukey's multiple-comparisons test were used. See also table S1.



Arf1 or PI(4)KIII β , respectively, in HeLa (Fig. 2H), Cos-7 (fig. S5D), and U2OS (fig. S5E) cells. Although Arf1 silencing for 3 days led to ER morphology aberrations and increased levels of cell death (fig. S7), we still did not observe

changes in the main fission and fusion regulator levels (fig. S7D). Immunofluorescence analysis of endogenous Mfn1 and Mfn2 in PI(4)KIII β - and Arf1-silenced HeLa cells also did not reveal any aggregation or mislocaliza-

tion (fig. S5, F and G). In addition, subcellular distribution analyses of Drp1 revealed no mitochondrial recruitment defects (Fig. 2, I to K) and the presence of Drp1 foci specifically at mitochondrial superconstrictions induced by

loss of PI(4)KIIIβ in the human fibroblast line MCH64 (Fig. 2K). Furthermore, stimulated Drp1-dependent mitochondrial fission induced by mitochondrial-anchored protein ligase (MAPL) (16) overexpression (fig. S8, A and B) or by carbonyl cyanide chlorophenylhydrazone (CCCP) treatment (fig. S8, C and D) was significantly reduced in PI(4)

KIIIβ- and Arf1-silenced cells. Silencing of the key component involved in stress-induced mitochondrial hyperfusion, SLP-2 (17), as well as the pro-fusion factors Mfn1 and Mfn2, also failed to reverse mitochondrial hyperfusion in PI(4)KIIIβ-silenced cells (fig. S9). Finally, compared with Drp1 silencing, which leads to drastic peroxisomal elongation (18, 19),

loss of Arf1 and PI(4)KIIIβ only induced a subtle peroxisomal elongation in HeLa cells, not in Cos-7 cells (fig. S10). Thus, these data potentially support a specific role for these enzymes in the regulation of mitochondrial fission downstream of Drp1 recruitment.

PI(4)KIIIβ mainly localized to the Golgi apparatus (fig. S11A) but PI(4)KIIIβ foci were also

Fig. 2. Loss of Arf1 and PI(4)KIIIβ induces mitochondrial superconstriction sites and does not alter fusion and/or fission machinery.

(A) Representative TEM images of HeLa cells treated with indicated siRNAs, showing (i) hyperfused mitochondria, (ii) branched mitochondria, and (iii) mitochondrial superconstriction sites with ER contacts. Scale bars, 500 nm. (B to G) Quantification of TEM images from (A) showing (B) mitochondrial area, (C) distribution of mitochondrial length, (D) percentage of branched mitochondria with indicated branch count, (E) percentage of mitochondria harboring mitochondrial superconstrictions, (F) distribution of mitochondrial superconstriction length (width <100 nm), and (G) percentage of mitochondrial superconstriction with ER contacts. (H) Levels of proteins relevant to mitochondrial fission (left panel) and fusion (right panel) from HeLa cells treated with the indicated siRNAs. (I) Representative confocal images of mitochondrial morphology and Drp1 localization in HeLa cells treated with the indicated siRNAs. Scale bars, 10 μm. (J) Subcellular fractionation analysis of Drp1 distribution in HeLa cells treated with the indicated siRNAs. Total cell lysates (whole cell) were fractionated into crude mitochondrial (heavy membrane) and cytosolic (cytosol) fractions. (K) Representative confocal images of Drp1 accumulating at mitochondrial superconstriction sites (white arrows) in human fibroblasts silenced for PI(4)KIIIβ. Scale bar, 10 μm. For (B), ordinary one-way ANOVA and Tukey's multiple-comparisons test were used in two independent experiments.

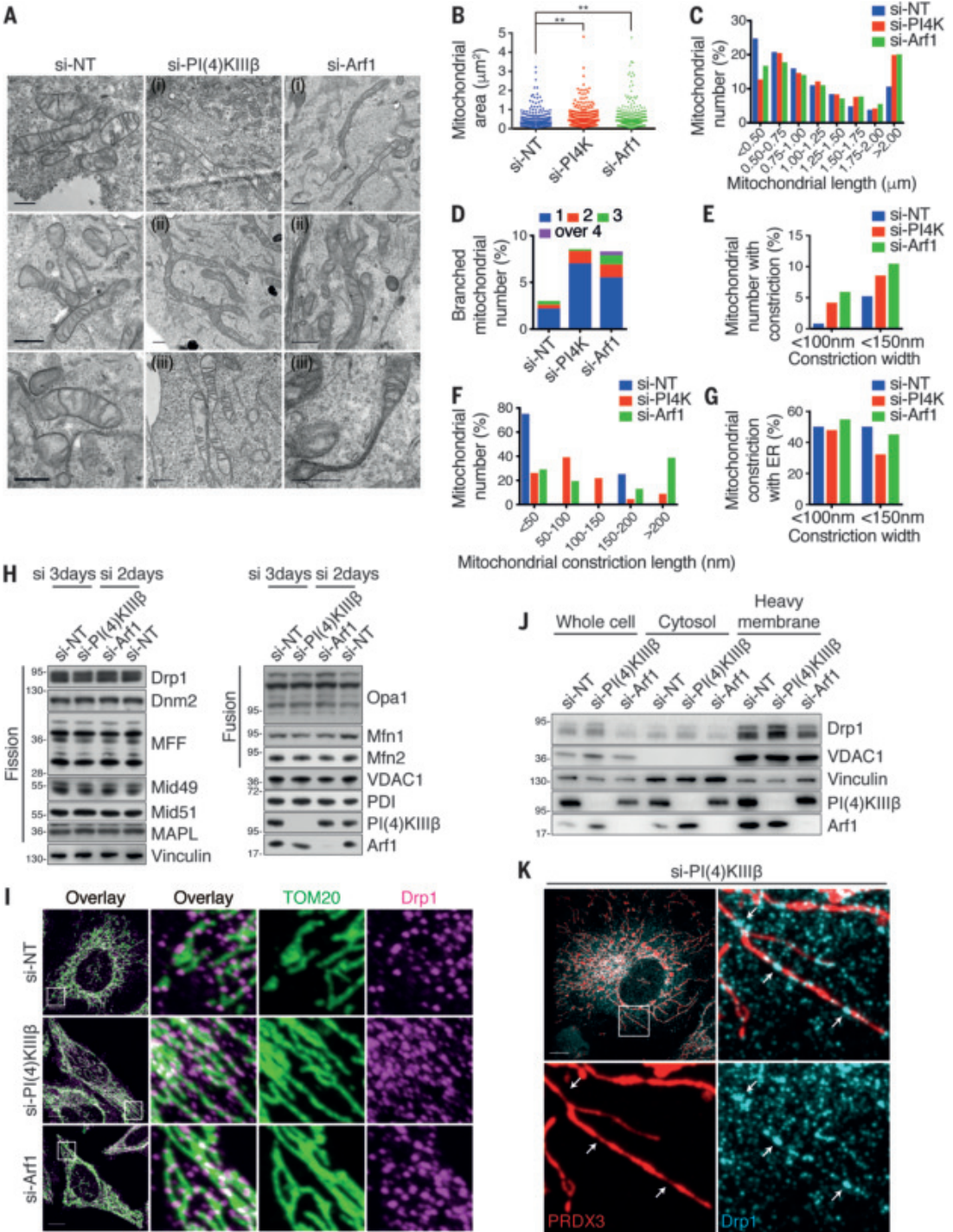


Fig. 3. PI(4)KIII β and Arf1 localized on TGN vesicles are recruited to mitochondrial constrictions and ER contacts before division. (A) Representative images of PI(4)KIII β punctae localization at mitochondrial constriction sites and ER contacts in HeLa cells (white arrows). Line-scan analysis of relative fluorescence intensity from the dashed line are shown. **(B)** PI(4)KIII β and Arf1 localization analysis by subcellular fractionation from HeLa cells. Total cell lysates (whole cell) were fractionated into cytosolic, heavy membrane (crude mito), purified mitochondrial (pure mito), mitochondria-associated membranes (MAM), and microsomal (microsomes) fractions. **(C)** Live-cell imaging of HeLa cells transiently expressing Arf1-GFP and TOM20-mCherry. **(D)** Quantification of mitochondrial fission events marked by Arf1-GFP before division (left panel) and Arf1-GFP dynamics on mitochondria after division (right panel). **(E)** Live-cell imaging of HeLa cells transiently expressing Arf1-GFP and Drp1-Scarlet with mitochondria labeled using MitoTracker deep red. **(F)** Quantification of mitochondrial fission events marked by Arf1-GFP downstream of Drp1-Scarlet recruitment. **(G)** Live-cell imaging of HeLa cells transiently expressing Arf1-GFP and LAMP1-mCherry with mitochondria labeled using MitoTracker deep red. **(H)** Quantification of mitochondrial fission events marked by Arf1-GFP, LAMP1-mCherry, or double Arf1-GFP/LAMP1-mCherry before division (left panel) and Arf1-GFP/LAMP1-mCherry dynamics before recruitment to division sites (right panel). **(I)** Live-cell imaging of HeLa cells transiently expressing Arf1-GFP and TGN46-mCherry with mitochondria labeled using MitoTracker deep red. **(J)** Quantification of mitochondrial fission events marked by Arf1-GFP, TGN46-mCherry, or double Arf1-GFP/TGN46-mCherry before division (left panel) and Arf1-GFP/TGN46-mCherry dynamics before recruitment to division sites (right panel). In (C), (E), (G), and (I), white and yellow arrows indicate Arf1-GFP puncta before and after a fission event, respectively. All scale bars, 10 μ m. All data are shown as mean \pm SEM of at least three independent experiments.

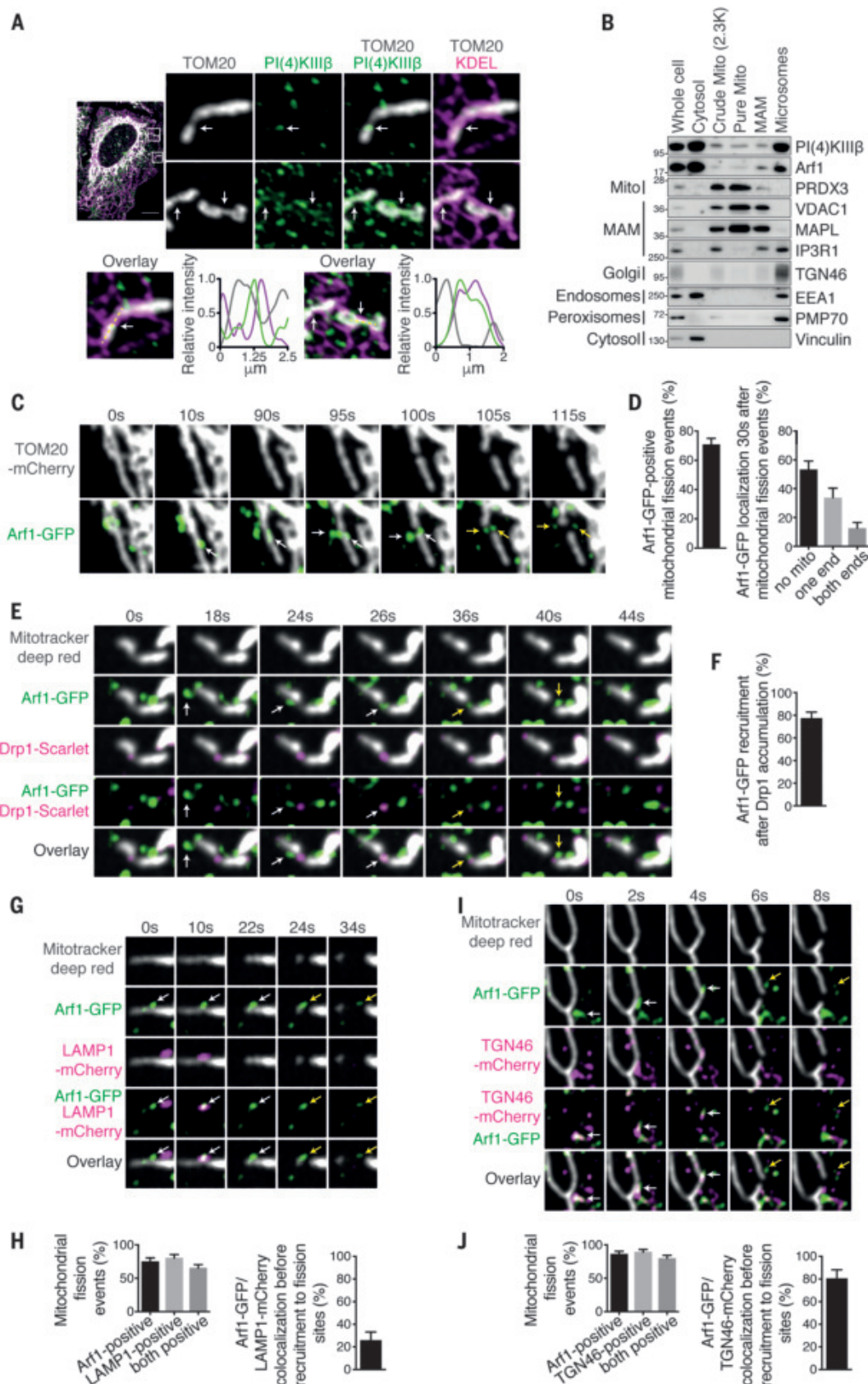


Fig. 4. Arf1- and PI(4)KII β -dependent PI(4)P formation on TGN vesicles at mitochondrial constrictions and ER contacts drive mitochondrial fission.

(A) Representative confocal images of HeLa cells transfected with GFP-PH^{FAPP1} showing GFP-PH^{FAPP1} at mitochondrial constriction sites and ER contact localization (white arrows). Line-scan analysis of relative fluorescence intensity from the dashed line is shown.

(B) Live-cell imaging of HeLa cells transiently expressing GFP-PH^{FAPP1} and TOM20-mCherry.

(C) Quantification of mitochondrial fission events marked by GFP-PH^{FAPP1} before division (left panel) and 30 s after division (right panel).

(D) Live-cell imaging of HeLa cells transiently expressing GFP-PH^{FAPP1} and Drp1-Scarlet with mitochondria labeled using MitoTracker deep red.

(E) Quantification of mitochondrial fission events marked by GFP-PH^{FAPP1} downstream of mitochondrial Drp1-Scarlet recruitment.

(F) Live-cell imaging of HeLa cells transiently expressing GFP-PH^{FAPP1} and TGN46-mCherry with mitochondria labeled using MitoTracker deep red.

(G) Quantification of mitochondrial fission events marked by GFP-PH^{FAPP1}, TGN46-mCherry, or double GFP-PH^{FAPP1}/TGN46-mCherry before division (left panel) and GFP-PH^{FAPP1}/TGN46-mCherry dynamics before recruitment to mitochondrial division sites (right panel).

(H) Live-cell imaging of HeLa cells transiently expressing GFP-PH^{FAPP1} and LAMP1-mCherry with mitochondria labeled using MitoTracker deep red.

(I) Quantification of mitochondrial fission events marked by GFP-PH^{FAPP1}, LAMP1-mCherry, or double GFP-PH^{FAPP1}/LAMP1-mCherry before division (left panel) and GFP-PH^{FAPP1}/LAMP1-mCherry dynamics before recruitment to mitochondrial division sites (right panel). For (B), (D), (F), and (H), white and yellow arrows indicate GFP-PH^{FAPP1} puncta before and after a fission event, respectively. All scale bars, 10 μ m. All data are shown as mean \pm SEM of at least three independent experiments.

detected at ER-induced mitochondrial constriction sites (Fig. 3A and fig. S11B). Similar results were obtained for subcellular localization analysis of Arf1-GFP (fig. S11C). The presence of PI(4)KIII β and Arf1 at the MERC compartment was confirmed by subcellular fractionation experiment (Fig. 3B). We then performed live-cell imaging to determine whether Arf1-GFP was recruited to mitochondrial constriction sites before division (Fig. 3C and movies S1 and S2). About 71% of mitochondrial division events analyzed were marked with Arf1-GFP punctae before fission (Fig. 3D) and 77% of division events showed the recruitment of these punctae at constriction sites after Drp1 recruitment (Fig. 3, E and F, and movie S3). Although Arf1-GFP foci were preferentially found on ER-induced mitochondrial constriction sites (fig. S11D and movies S4 and S5), Arf1-GFP foci were not localized at mitochondrial tip ends after division (Fig. 3, C and D, and movies S1 and S2) and they did not localize perfectly with ER markers (fig. S11D). This suggested that Arf1 was primarily recruited to the MERC during division from a ternary compartment. Previous work uncovered a role for lysosomes at sites of mitochondrial division (20), so we first explored whether Arf1-GFP foci may reflect these structures. However, whereas Arf1-GFP foci converged with lysosomes at fission sites, their recruitment was distinct from lysosomes (Fig. 3, G and H, and movie S6). Instead, we found that Arf1-GFP foci were recruited to constriction sites upon trans-Golgi network (TGN) vesicles (Fig. 3, I and J, and movie S7). Indeed, TGN46-mCherry vesicles were recruited to mitochondrial constriction just before division in 85% of fission events analyzed, which was correlated with a colocalization with Arf1-GFP punctae before and during this process in 80% of division events (Fig. 3, I and J, and movie S7).

We confirmed the predominant Golgi localization for PI(4)P (fig. S12A) using the established probe GFP-PH^{FAPP1} (27), but we also observed PI(4)P enriched foci crossing ER-induced mitochondrial constriction sites (Fig. 4A) in an Arf1- and PI(4)KIII β -dependent manner (fig. S12, B and C). Loss of Drp1 also significantly decreased mitochondrial GFP-PH^{FAPP1} punctae, suggesting that Drp1 activity was required for the recruitment of TGN-derived PI(4)P vesicles (fig. S12, B and C). Video analysis revealed that pools of PI(4)P accumulated and extended toward mitochondria at sites of constriction (Fig. 4B and movies S8 and S9) in ~73% of division events analyzed (Fig. 4, B and C). Similar to Arf1-GFP, PI(4)P foci were recruited to MERCs downstream of Drp1 (Fig. 4, D and E; fig. S13; and movies S10 and S11) and remained on TGN46 vesicles throughout the fission event (Fig. 4, F and G, and movie S12). Moreover, we observed no colocalization with lysosomes that converged at the site of division (Fig. 4,

H and I, and movie S13). These results were confirmed using an additional PI(4)P probe, mCherry-P4M, which recognizes PI(4)P pools in multiple endomembranes (22) (fig. S14 and movies S14 to S16). Finally, consistent with the assembly of the mitochondrial fission machinery and the coordination of PI(4)P accumulation, endogenous TGN46, PI(4)KIII β , and GFP-PH^{FAPP1} foci colocalized with endogenous Drp1 at ER-induced mitochondrial constrictions (fig. S15). Thus, Arf1 and PI(4)KIII β enable the accumulation of PI(4)P punctae on TGN vesicles, driving late steps of mitochondrial division.

Mitochondrial fission is a complex process that requires many factors, including the ER, which is involved in the early steps of organelle constriction (1, 23), and the lysosomes, which were recently identified at division sites (20). However, the functional contribution of these organelles to the process of membrane fission remains unclear. We now add a further layer of complexity by identifying a key role for Arf1 and PI(4)KIII β on Golgi vesicles in driving late steps of mitochondrial division. These data reveal a four-way contact among mitochondria, ER, TGN, and lysosomal vesicles occurring at >80% of fission sites. It is unclear why so many organelles are required to drive mitochondrial division. In considering the contribution of PI(4)P-enriched vesicles (24), we envision a potential role in the recruitment of adaptors that drive Arp2/3-dependent actin polymerization at transient and localized microdomains that could allow the dynamic assembly of force-generating machineries essential for the final steps of mitochondrial division (25–27). We now suggest that the intimate contacts between mitochondria and Golgi-derived PI(4)P-containing vesicles are key modulators of mitochondrial dynamics.

REFERENCES AND NOTES

- J. R. Friedman *et al.*, *Science* **334**, 358–362 (2011).
- E. Smirnova, L. Griparic, D. L. Shurland, A. M. van der Bliek, *Mol. Biol. Cell* **12**, 2245–2256 (2001).
- J. E. Lee, L. M. Westrate, H. Wu, C. Page, G. K. Voeltz, *Nature* **540**, 139–143 (2016).
- S. C. Kamerkar, F. Kraus, A. J. Sharpe, T. J. Pucadyil, M. T. Ryan, *Nat. Commun.* **9**, 5239 (2018).
- T. B. Fonseca, Á. Sánchez-Guerrero, I. Milosevic, N. Raimundo, *Nature* **570**, E34–E42 (2019).
- L. Tilokani, S. Nagashima, V. Paupe, J. Prudent, *Essays Biochem.* **62**, 341–360 (2018).
- K. B. Ackema *et al.*, *EMBO J.* **33**, 2659–2675 (2014).
- A. Godi *et al.*, *Nat. Cell Biol.* **1**, 280–287 (1999).
- B. Mesmin *et al.*, *Cell* **155**, 830–843 (2013).
- J. Moser von Filseck *et al.*, *Science* **349**, 432–436 (2015).
- J. Moser von Filseck, S. Vanni, B. Mesmin, B. Antonny, G. Drin, *Nat. Commun.* **6**, 6671 (2015).
- J. Chung *et al.*, *Science* **349**, 428–432 (2015).
- J. H. Pogson *et al.*, *PLOS Genet.* **10**, e1004815 (2014).
- G. H. Patterson, J. Lippincott-Schwartz, *Science* **297**, 1873–1877 (2002).
- X. H. Zhao, T. Bondeva, T. Balla, *J. Biol. Chem.* **275**, 14642–14648 (2000).
- E. Braschi, R. Zunino, H. M. McBride, *EMBO Rep.* **10**, 748–754 (2009).

- D. Tondera *et al.*, *EMBO J.* **28**, 1589–1600 (2009).
- X. Li, S. J. Gould, *J. Biol. Chem.* **278**, 17012–17020 (2003).
- A. Koch *et al.*, *J. Biol. Chem.* **278**, 8597–8605 (2003).
- Y. C. Wong, D. Ysselstein, D. Krainc, *Nature* **554**, 382–386 (2018).
- S. Dowler *et al.*, *Biochem. J.* **351**, 19–31 (2000).
- G. R. Hammond, M. P. Machner, T. Balla, *J. Cell Biol.* **205**, 113–126 (2014).
- F. Korobova, V. Ramabhadran, H. N. Higgs, *Science* **339**, 464–467 (2013).
- R. Dong *et al.*, *Cell* **166**, 408–423 (2016).
- E. Derivery *et al.*, *Dev. Cell* **17**, 712–723 (2009).
- S. Li *et al.*, *J. Cell Biol.* **208**, 109–123 (2015).
- N. H. Hong, A. Qi, A. M. Weaver, *J. Cell Biol.* **210**, 753–769 (2015).
- H. Lochmüller, T. Johns, E. A. Shoubridge, *Exp. Cell Res.* **248**, 186–193 (1999).
- R. Zunino, A. Schauss, P. Rippstein, M. Andrade-Navarro, H. M. McBride, *J. Cell Sci.* **120**, 1178–1188 (2007).
- E. Braschi *et al.*, *Curr. Biol.* **20**, 1310–1315 (2010).
- R. N. Day, M. W. Davidson, *Chem. Soc. Rev.* **38**, 2887–2921 (2009).
- D. S. Bindels *et al.*, *Nat. Methods* **14**, 53–56 (2017).
- J. Chun, Z. Shapovalova, S. Y. Deigaard, J. F. Presley, P. Melançon, *Mol. Biol. Cell* **19**, 3488–3500 (2008).
- A. Sugiura, S. Mattie, J. Prudent, H. M. McBride, *Nature* **542**, 251–254 (2017).
- J. Prudent *et al.*, *Mol. Cell* **59**, 941–955 (2015).
- J. Schindelin *et al.*, *Nat. Methods* **9**, 676–682 (2012).
- A. S. Moore, Y. C. Wong, C. L. Simpson, E. L. Holzbaur, G. K. Voeltz, *Nat. Commun.* **7**, 12886 (2016).
- I. Arganda-Carreras, R. Fernández-González, A. Muñoz-Barrutia, C. Ortiz-De-Solorzano, *Microsc. Res. Tech.* **73**, 1019–1029 (2010).
- C. D. Williamson, D. S. Wong, P. Bozidis, A. Zhang, A. M. Colberg-Poley, *Curr. Protoc. Cell Biol.* **68**, 3.27.1–3.27.33 (2015).

ACKNOWLEDGMENTS

We thank S. Mattie for contributions to TEM sample preparation and image acquisition. **Funding:** This work was supported by the Canadian Institutes of Health Research Operating Grants Program (CIHR grant 68833 to H.M.M.), the Medical Research Council (MRC grants MC_UU_00015/7 and RG89175 to J.P.), the Isaac Newton Trust (grant RG89529 to J.P.), and the Wellcome Trust Institutional Strategic Support Fund (grant RG89305 to J.P.). H.M.M. is a Canada Research Chair. S.N. and L.-C.T. are recipients of Daiichi Sankyo Foundation of Life Science and Ramon Areces postdoctoral fellowships, respectively. L.T. was supported by an MRC-funded graduate student fellowship. V.P. was supported by the European Union's Horizon 2020 research and innovation program (MITODYN-749926). **Authors contributions:** S.N. performed the experiments; L.-C.T. and L.T. contributed to quantitative confocal imaging and immunoblots analysis; V.P. performed organelle fractionation; H.A. provided technical assistance; J.H.P. contributed intellectually to the initial development of the project; R.Z. contributed to biochemical analysis; H.M.M. and J.P. conceived the study, designed the experiments, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1366/suppl/DC1
Materials and Methods
Table S1
Figs. S1 to S15
Movies S1 to S16
References (28–39)

9 April 2019; resubmitted 9 December 2019
Accepted 27 February 2020
10.1126/science.aax6089

BIOTECHNOLOGY

Genetically targeted chemical assembly of functional materials in living cells, tissues, and animals

Jia Liu^{1*}, Yoon Seok Kim^{2*}, Claire E. Richardson^{3*}, Ariane Tom^{2*}, Charu Ramakrishnan², Fikri Birey⁴, Toru Katsumata¹, Shucheng Chen¹, Cheng Wang⁵, Xiao Wang², Lydia-Marie Joubert⁶, Yuenwen Jiang¹, Huiliang Wang², Lief E. Fenno^{2,4}, Jeffrey B.-H. Tok¹, Sergiu P. Pasca⁴, Kang Shen^{3,7}, Zhenan Bao^{1,†}, Karl Deisseroth^{2,4,7,†}

The structural and functional complexity of multicellular biological systems, such as the brain, are beyond the reach of human design or assembly capabilities. Cells in living organisms may be recruited to construct synthetic materials or structures if treated as anatomically defined compartments for specific chemistry, harnessing biology for the assembly of complex functional structures. By integrating engineered-enzyme targeting and polymer chemistry, we genetically instructed specific living neurons to guide chemical synthesis of electrically functional (conductive or insulating) polymers at the plasma membrane. Electrophysiological and behavioral analyses confirmed that rationally designed, genetically targeted assembly of functional polymers not only preserved neuronal viability but also achieved remodeling of membrane properties and modulated cell type-specific behaviors in freely moving animals. This approach may enable the creation of diverse, complex, and functional structures and materials within living systems.

The complex properties of living systems arise from the structure and function of constituent cells, exemplified by the roles of neurons (1) within nervous systems (2, 3).

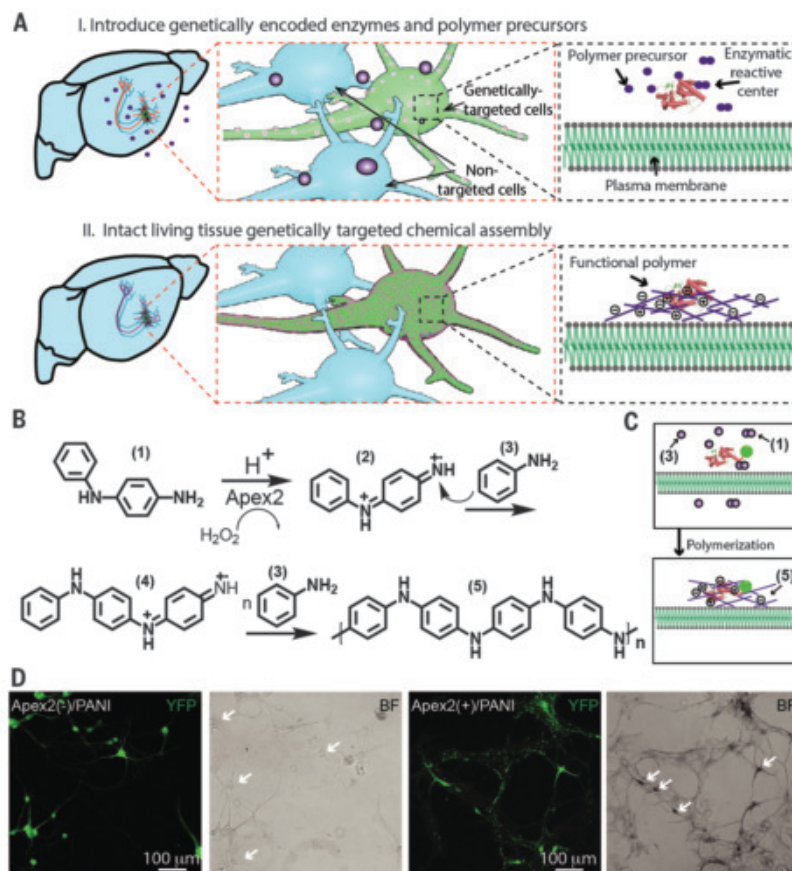
We considered whether specific cells within intact biological systems may be genetically co-opted to build new physical structures with desired form and function. Incorporation of min-

iatized electrical components onto membranes can change cellular activity (4–6), although without the capability for genetic targeting of cells. In another approach, electroactive (such as conductive) polymers have been directly synthesized through electrochemical polymerization in living tissue to reduce impedance (7), although without the genetic targeting of cells or cell types.

To achieve biocompatible in vivo synthesis of electroactive polymers within genetically specified cells of living animals, we began with conductive polymers from polyaniline (PANI) and poly(3,4-ethylenedioxythiophene) (PEDOT). These polymers were chosen for aqueous synthesis (which is important for biological system compatibility) and for dual conduction of electrons and ions, which reduces local electrochemical impedance (δ) when interfacing electronics with living cells. We designed a single-enzyme-facilitated polymerization using chemically modified monomers (Fig. 1A) for which polymerization is triggered by an enzyme that can be expressed in specific cells. Perfusion of small-molecule conductive-polymer precursors capable of diffusion through intact tissue (step I) was followed by oxidative radical cation polymerization steps at the genetically targeted enzyme's reactive center. Because of the short

Fig. 1. Genetically targeted chemical assembly of functional materials in cells. (A) Specific

instantiation shown is enzyme/ H_2O_2 -catalyzed functional polymerization in brain. Blue indicates non-enzyme-targeted cells. (B) Reaction of Apex2-mediated polymerization from precursor reagents containing aniline monomer-dimer mixture. Labels 1 to 5 show chemical structures of N-phenylenediamine (aniline dimer, 1), aniline dimer radical cations (2), aniline monomer (3), aniline trimer radical cations (4), and polyaniline (PANI, 5), respectively. (C) Schematic of Apex2-mediated polymerization and deposition of PANI on targeted cells. (D) In situ genetically targeted synthesis and incorporation of conductive polymer. Shown are epifluorescence (YFP) and BF images of fixed rat hippocampal neurons. Arrows indicate individual neurons.



¹Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. ²Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. ³Department of Biology, Stanford University, Stanford, CA 94305, USA. ⁴Department of Psychiatry, Stanford University, Stanford, CA 94305, USA. ⁵Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley CA 94720, USA. ⁶Cell Sciences Imaging Facility, Stanford University, Stanford, CA 94305, USA. ⁷Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: deisseroth@stanford.edu (K.D.); zbao@stanford.edu (Z.B.)

mean diffusion length of radical cations in aqueous solution and low solubility of the resulting polymers, the synthesized conductive polymers were expected to be deposited onto targeted cells at juxtamembranous locations (a design feature for limiting adverse effects on native intracellular chemistry) (step II).

Peroxidases can catalyze synthesis of conductive polymers in the presence of hydrogen peroxide (H_2O_2) in vitro under harsh conditions: high concentrations of hydrogen peroxide (>1 mM), low pH (pH = 1 to 5), and high monomer concentrations (>10 mM) (9). Therefore, we sought a biocompatible synthesis by enabling polymerization in pH-neutral and biocompatible conditions. We first expressed a humanized version of ascorbate peroxidase Apex2 (10); cultured rat hippocampal neurons were transduced with adeno-associated virus (AAV) vectors containing Apex2 and in some cases fused with a 13-amino acid peptide (selected

in a screen for expression-enhancing tags) (fig. S1) (11) and/or enhanced yellow fluorescent protein (YFP) (for tracking localization) (Fig. 1A). We first selected aniline as the monomer for its relatively low oxidation potential (12), but Apex2 was unable to polymerize aniline monomers in phosphate-buffered saline (fig. S2A). Because *N*-phenylenediamine (an aniline dimer) would further reduce oxidation potential (13), an aniline monomer-dimer mixture (0.5 mM, 1:1 molar ratio) (Fig. 1B) was added to an aqueous solution of 0.1 mM H_2O_2 and applied to fixed cultured neurons (Fig. 1C). Epifluorescence and bright-field (BF) phase images confirmed that Apex2(+) but not Apex2(-) neurons exhibited a dark-colored reaction product (Fig. 1D and fig. S2, B and C). Confocal images revealed a formation of deposited aggregates (fig. S3).

To test whether these deposits consisted of PANI, we used ultraviolet-visible-near infra-

red (UV-vis-NIR) absorption spectroscopy to compare with spectra previously reported for PANIs (Fig. 2A). The shorter absorption peak wavelength of Apex2(+)/PANI (~574 nm versus ~620 nm for commercial 50 kDa PANI) indicated that the synthesized polymer was of lower molecular weight (Fig. 2B). We then treated the PANI-fixed neurons with 100 mM *p*-toluenesulfonic acid (termed Apex2(+)/dPANI), which resulted in increased conductivity and red-shift in the UV-vis spectrum (Fig. 2A), as expected for doped PANI (14). For Apex2(-)/PANI, Apex2(+)/PANI, and Apex2(+)/dPANI neurons, we observed expected color changes (fig. S4A). The UV-vis-NIR spectrum showed a red-shifted peak at ~615 nm for doped PANI (Fig. 2C), indicating transition to the emeraldine salt form (14). Peak absorption-wavelength was maintained across different reaction times, suggesting increased PANI deposition over time (fig. S4, B and C). X-ray

Fig. 2. Chemical and electrical characterization of synthesized conductive polymer.

(A) Structures shown are PANI (red) conversion to doped PANI (dPANI, green), by means of acid (HX) treatment. (B) Normalized UV-vis-NIR spectra. Pure PANI, purple; Apex2(-) neurons black; Apex2(-)/PANI neurons, blue; and Apex2(+)/PANI neurons, red. Arrows indicate absorption peak. (C) UV-vis-NIR spectra of Apex2(-)/PANI and Apex2(+)/PANI before and after *p*-toluenesulfonic acid treatment. Dashed arrows indicate red-shift of absorption peak in UV-vis from ~574 to ~615 nm. (D to I) Variable-pressure SEM images of (D) nonreacted wild-type, (E) Apex2(-)/PANI, (F) Apex2(+)/PANI, and (G) Apex2(+)/dPANI neurons. Zoomed-in images of (H) blue- and (I) red-boxed regions from (G) show polymer deposition. (J) Schematic of electrical interface to fixed neurons (blue) with PANI coating, for conductivity measurements. Acid doping (green) was used to test presence of deposited conductive polymer. (K and L) BF image of postreacted neurons on the glass substrate with gold electrodes for current-voltage (*I*-*V*) measurement. (M and N) Representative *I*-*V* curves (M) and summary of resistance changes (N) (log-scale violin plots of resistance, $n = 20$ electrode-pairs per category, $***P < 0.001$, $****P < 0.0001$, unpaired two-tailed *t* test) between electrodes for cultured Apex2(-) and Apex2(+) neurons on slides before and after acidic vapor treatment (HCl). Reduction in Apex2(-)/dPANI sample likely because of ionic conductivity from evaporated HCl solution.

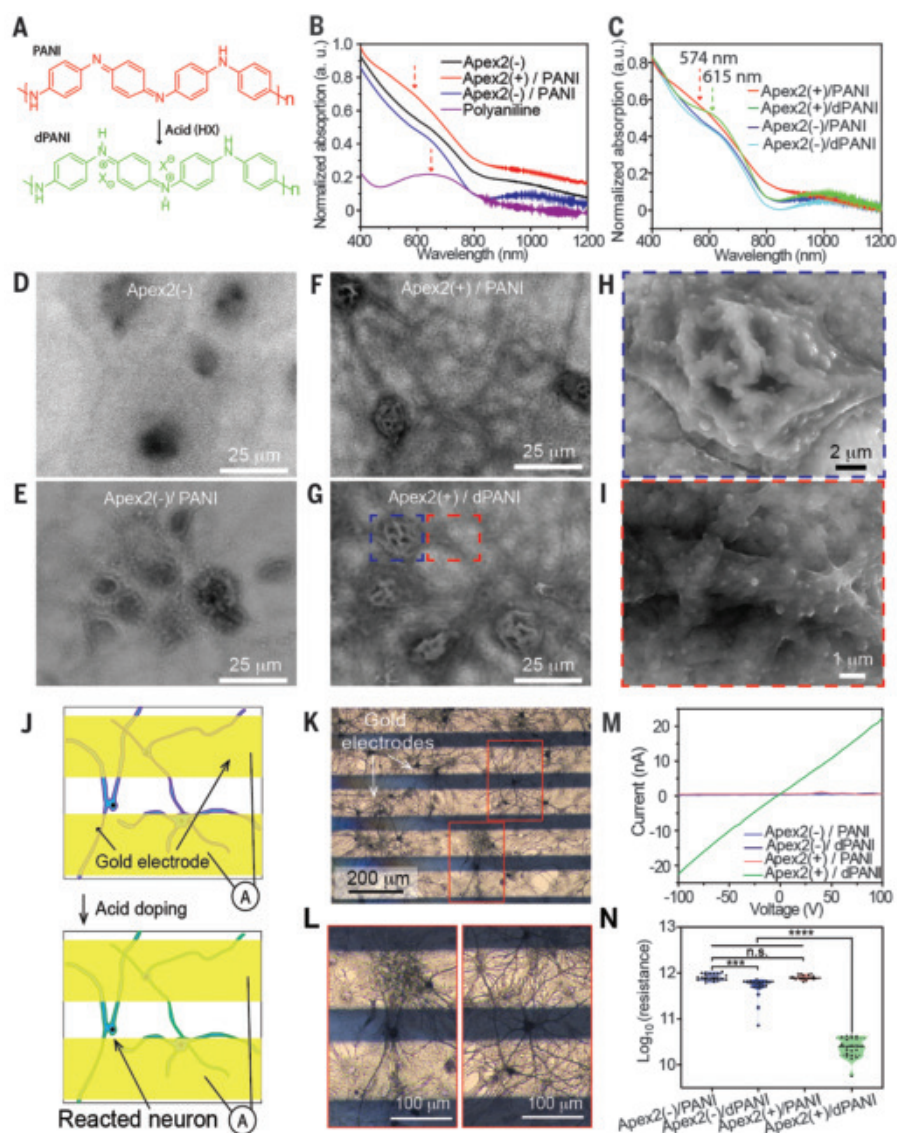
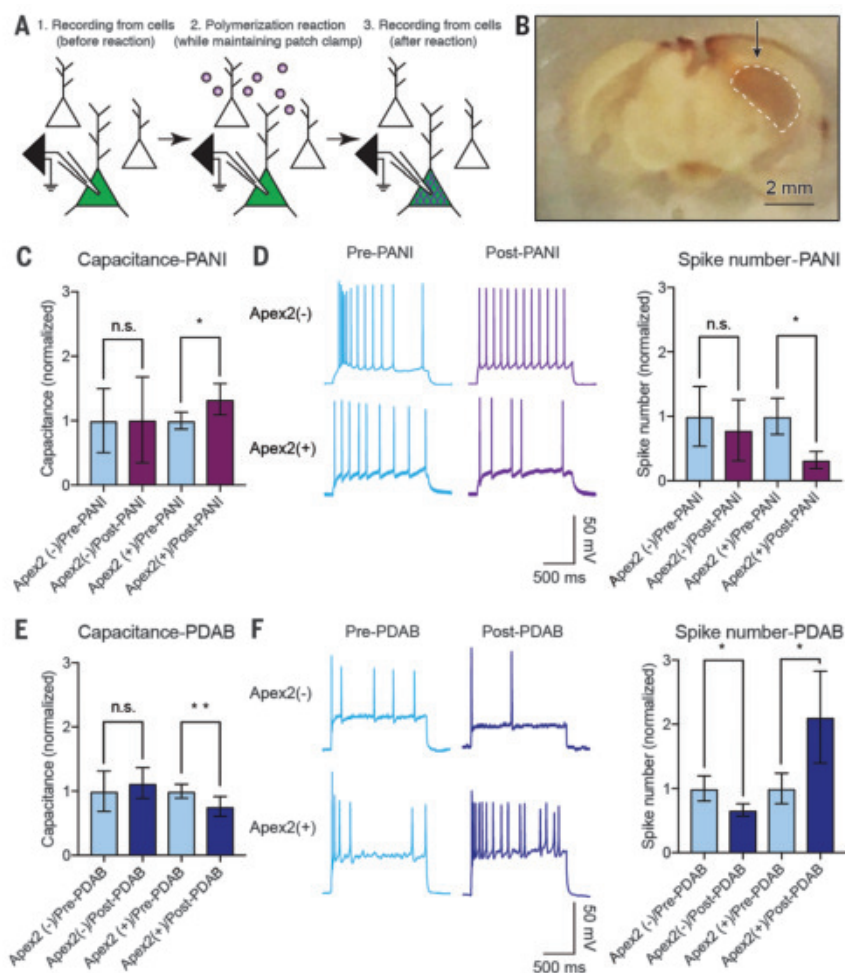


Fig. 3. Electrophysiological characterization: conductive polymers in living brain slices.

Light blue, before reaction; purple, after PANI; dark blue, after PDAB. (A) Slice physiology workflow. (B) Photomicrograph of brain slice after polymerization reaction. Arrow indicates injection site of Apex2 virus; dashed line indicates hippocampus. (C) Membrane capacitance and (D) current-injection-evoked spikes before and after PANI polymerization [mean \pm SEM, $n = 7$ Apex2(+), $n = 4$ Apex2(-), (D) $n = 3$ Apex2(-); all individual cells were maintained in the whole-cell patch clamp configuration across pre-reaction and post-reaction time points for direct comparison; ratio-paired t tests: $*P < 0.05$]. All postconditions here and in (E) and (F) were normalized to corresponding preconditions for comparison; mean capacitance values were 20 to 45 pF. (E) Membrane capacitance and (F) current-injection-evoked spikes before and after PDAB polymerization. Increased spiking can be seen with Apex2(+)/PDAB despite mild rundown from PDAB-only Apex2(-) reaction conditions [mean \pm SEM, (E) $n = 10$ Apex2(+), (F) $n = 8$ Apex2(+), (E) $n = 5$ Apex2(-), (F) $n = 4$ Apex2(-). Ratio-paired t tests: $*P < 0.05$, $**P < 0.01$].



photoelectron spectroscopy (XPS) showed enhanced S-element signal only in Apex2(+)/dPANI neurons, confirming incorporation of *p*-toluenesulfonic acid (fig. S5A). Near-edge x-ray absorption fine structure (NEXAFS), for identifying different types of C–N or C=N features from amines and imines (15), confirmed the chemical composition of the deposited material (fig. S5, C to G).

Variable-pressure scanning electron microscopy (VP-SEM) imaging of neurons provided initial qualitative comparison of conductivity before and after reaction (Fig. 2, D to G, and fig. S6). Apex2(+)/PANI neurons in liquid showed higher contrast, which is consistent with a more conductive outer layer (Fig. 2F), with contrast further enhanced through acidic doping (Fig. 2G). Both soma and neurites could be directly observed, suggesting substantial surface-conductivity enhancement (16) from doped PANI (Fig. 2, H and I). In addition, transmission electron microscopy (TEM) confirmed deposition of polymers on neuronal membranes (fig. S7).

We further investigated the conductive nature of PANI-fixed neurons by depositing gold electrodes onto air-dried, fixed neurons (Fig. 2, J to L); electrical conduction between electrodes

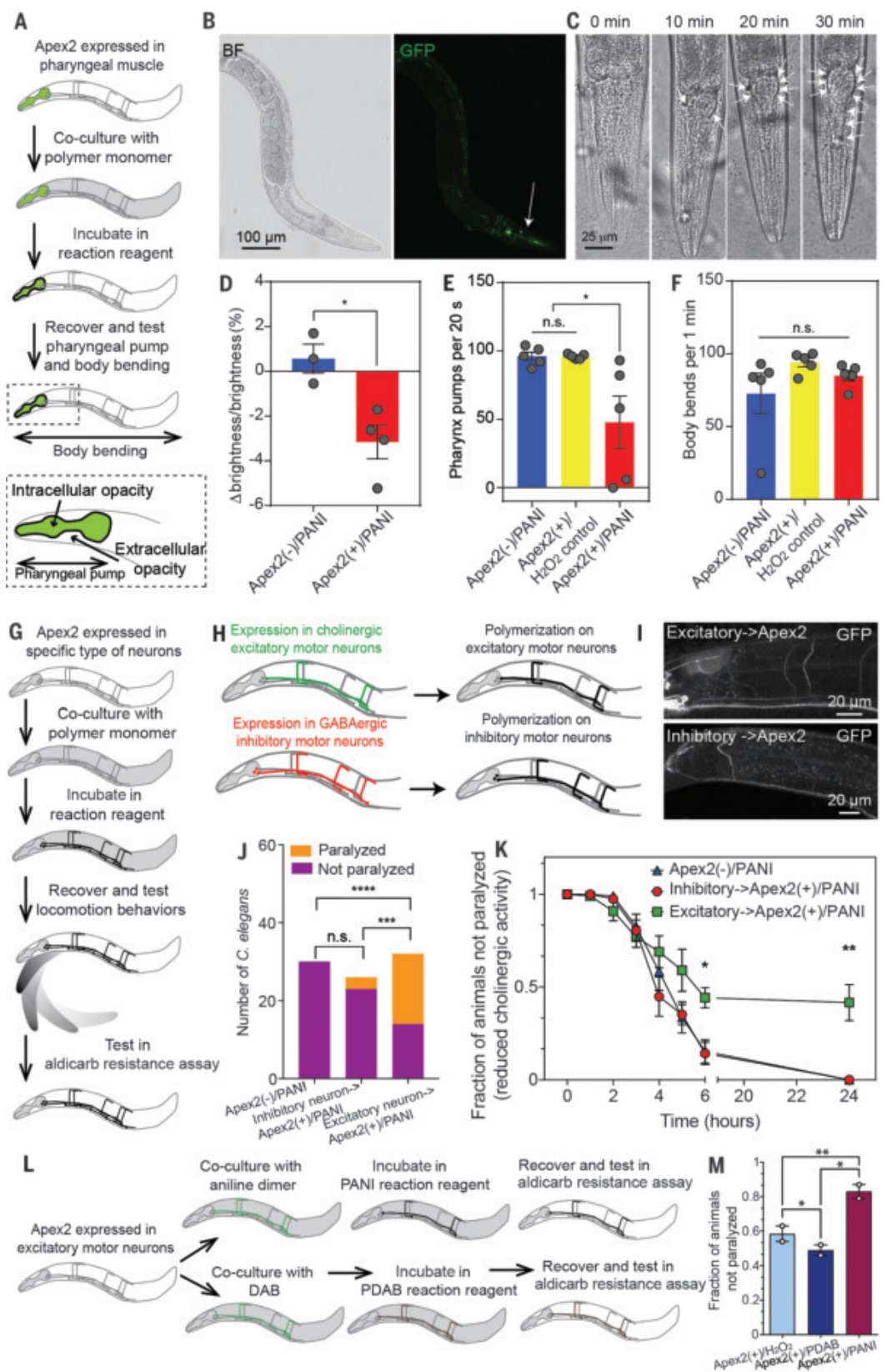
was expected to only arise from conductive polymer on the neurons. To prevent delamination between gold electrodes and polymers during solution-doping, HCl vapor was used to dope the polymer. Apex2(+)/dPANI showed the lowest resistance, as expected (Fig. 2, M and N). We tested other polymers, including a poly(3,4-ethylenedioxythiophene/PEDOT) derivative, sodium 4-((5,7-di(thiophen-2-yl)-2,3-dihydrothieno[3,4-b][1,4]dioxin-2-yl)methoxy)butane-1-sulfonate (termed TETs) (17), and a nonconductive polymer, poly(3,3'-diaminobenzidine) (PDAB) (18). Apex2(+)/PANI-PTETs neurons showed higher conductivity than that of Apex2(-)/PANI-PTETs neurons (fig. S8), whereas Apex2(+)/PDAB showed no conductivity change compared with that of Apex2(-)/PDAB (fig. S9). To further verify conductivity, we cultured human embryonic kidney 293T cells in a confluent sheet suited for conductivity measurements (fig. S10). Apex2(+)/PANI without acidic doping exhibited an approximately two orders of magnitude reduction in resistance versus that of Apex2(-)/PANI control. We next investigated Apex2-catalyzed polymerization in human cortical spheroids (hCS), a human stem cell-derived three-dimensional (3D) or-

ganoid (19, 20). We observed coloration (fig. S11A) and particle deposits within 30 min of reaction-treated Apex2(+)/PANI hCS at locations corresponding to YFP signal (fig. S11, B to E). In another 3D preparation (brain slices), the dark-colored reaction product could be visualized ~60 μ m and 110 μ m deep after 30- and 60-min reactions, respectively (fig. S12).

We further explored application of this method to living systems. Neurons remained viable after exposure to the aniline and its dimer in 0.05 mM H₂O₂ (fig. S15A)—a reaction condition sufficient for polymer deposition (fig. S13, A and B) [verified by means of UV-vis-NIR absorption (fig. S13, C to E)]. The same reaction condition in living mice elicited no reactive gliosis over weeks (fig. S14). We also performed whole-cell patch clamp in Apex2(+)- and Apex2(-)-cultured rat hippocampal neurons before and after PANI or PDAB polymerization (fig. S15). Current injection in Apex2(+)/PDAB neurons elicited robust action potentials both before and after polymerization, decreased capacitance consistent with the expected juxtamembranous localization of this insulating polymer, and increased charge-separation distance across the membrane; by contrast, Apex2(+)/PANI neurons showed

Fig. 4. Cell type-specific polymerization in *C. elegans*.

(A) Schematic of targeting polymerization to pharyngeal muscle. (B) BF (left) and fluorescence images (right) of *C. elegans* expressing *Pmyo-2::Apex2::mcd8::gfp* labeled Apex2(+) versus wild-type controls labeled Apex2(-). Arrow indicates GFP-labeled pharyngeal muscle and Apex2 expression. (C) BF time-course images of pharyngeal muscle of Apex2(+) worms in 30-min PANI reaction. Arrows indicate increased black reaction-product in Apex2(+) between pharyngeal muscle and epidermis. (D) Brightness change. Shown is Apex2(-) versus Apex2(+) after reaction (mean \pm SEM, $n = 3$ to 4 animals, $*P < 0.05$, two-tailed, unpaired *t* test). (E) Body-bending rate and (F) pharyngeal-pumping rate for Apex2(-)/PANI, Apex2(+)/H₂O₂ control, and Apex2(+)/polymerization (mean \pm SEM, $*P < 0.05$, $n = 5$ animals each condition, two-tailed unpaired *t* test). (G) Schematic of motor-neuron testing. (H) Cell type-specific polymerization of GABAergic (inhibitory) neurons [Inhibitory→Apex2(+)] or cholinergic (excitatory) neurons [Excitatory→Apex2(+)]. Black lines indicate cell type-specific polymer. (I) Inhibitory→Apex2(+) (top) or Excitatory→Apex2(+) (bottom) motor neurons expressing *Apex2::mcd8::gfp* under *Punc-47* or *Punc-17* promoters, respectively. (J) Excitatory→Apex2(+) worms show reduced locomotion after polymerization, whereas Inhibitory→Apex2(+) worms show negligible paralysis. ($n = 30$ animals for Apex2(-), $n = 26$ animals for Inhibitory→Apex2(+), and $n = 32$ animals for Excitatory→Apex2(+); $***P < 0.001$, $****P < 0.0001$, one-sided Fisher's exact test). (K) Aldicarb resistance assay after polymerization [50 animals per strain, two replicates, one-way analysis of variance (ANOVA)/Tukey correction, $*P < 0.05$, $**P < 0.01$]. (L) Schematic of polymerization of conductive (PANI) and insulating (PDAB) polymers in worm cholinergic motor neurons. (M) Summary of fraction-resistant *C. elegans* after 5 hours in aldicarb resistance assay (mean \pm SEM, $*P < 0.05$, $**P < 0.01$, one-way ANOVA test, Tukey correction).



decreased action-potential firing with increased capacitance (fig. S15, B to I), which is consistent with reported capacitance effects resulting from conducting-polymer introduction (21, 22).

To allow rigorous testing of the same cells before and after polymerization, we also conducted recordings in acute brain slices (Fig. 3A), which allowed holding the same cells in whole-cell patch clamp throughout the polymerization reaction. Four weeks after Apex2 virus injection, we observed robust Apex2-driven polymerization (Fig. 3B and fig. S12), with increased capacitance after PANI reaction and decreased capacitance after PDAB reaction (Fig. 3, C and E). Little effect was observed on other passive membrane properties (fig. S16), and patched cells were healthy in terms of input resistance and resting potential under all conditions. We next studied action potentials (Fig. 3, D and F); whereas Apex2(−) neuron firing rates were unchanged after treatment, Apex2(+)/PANI neurons exhibited decreased current-injection-evoked firing, and Apex2(+)/PDAB neurons showed increased firing (Fig. 3, D and F). The stability of resting potential and input resistance coupled with the bidirectionality of this effect would not have been expected from non-specific cell-health mechanisms for altered firing. By contrast, experimental and theoretical studies have demonstrated an inverse correlation between spike firing and capacitance (supplementary materials) (23–25), which is consistent with our slice physiology that shows increased capacitance after conductive-polymer deposition on the dielectric lipid bilayers of living neurons and decreased capacitance after insulating-polymer deposition (Fig. 3, C and E).

Last, we tested behavior in freely moving animals upon assembling genetically targeted electroactive polymers in vivo. We expressed Apex2–green fluorescent protein (GFP) on the membrane of worm (*Caenorhabditis elegans*) pharyngeal muscle cells (Fig. 4, A and B) and observed robust localized polymerization (Fig. 4, C and D, and fig. S18, A and B). Apex2(+)/PANI worms exhibited reduced pumping frequency of pharyngeal muscle (Fig. 4E) consistent with the inhibition of targeted cells observed in cultured neuron and brain slice electrophysiology, but no quantitative alteration in other body movements, such as bending (Fig. 4F). Because liquid-state atomic force microscopy showed no clear changes in Young's modulus of cellular membranes after polymerization (fig. S17), altered pharyngeal pumping was unlikely owing to changed elasticity of muscle membranes, and viability assays confirmed long-term biocompatibility of PANI in worms (fig. S18, A and C).

We next expressed Apex2-GFP in γ -aminobutyric acid (GABA)–ergic (inhibitory) or cholinergic (excitatory) motor neurons (Fig. 4, G to I, Inhibitory→Apex2(+) and Excitatory→Apex2(+), respectively). After polymerization (Fig. 4J), Excitatory→Apex2(+)/PANI worms displayed

impaired sinusoidal forward locomotion (both spontaneous and aversive-stimulus-evoked), which is concordant with prior observations from optogenetic inhibition of worm excitatory neurons (26). Sinusoidal forward locomotion in Apex2(−)/PANI and Inhibitory→Apex2(+)/PANI was unaffected. On the other hand, Inhibitory→Apex2(+)/PANI worms exhibited increased reversal frequency (fig. S18, D to G) and increased sharp ($<90^\circ$) turns versus Apex(−)/PANI worms (fig. S18H), which is consistent with prior results from optogenetic manipulation of inhibitory neurons that also induced sharper turns (27). Inhibitory→Apex2(+)/PANI worms maintained the capability to move forward in sinusoidal waves of unchanged amplitude (fig. S18I) and minimally reduced wavelength (fig. S18J), but when inhibitory neurons were ablated (*unc25-null*), sinusoidal wave amplitude was greatly reduced (fig. S18, K and L) (28).

Consistent with this pattern, Excitatory→Apex2(+)/PANI worms became resistant to the acetylcholinesterase-inhibitor aldicarb, suggesting that this treatment causes reduced acetylcholine release, but Inhibitory→Apex2(+)/PANI and Apex2(−)/PANI worms did not (Fig. 4K and fig. S19, A and B). Moreover, Excitatory→Apex2(+)/PDAB showed reduced resistance to aldicarb, compared with Excitatory→Apex2(+)/PANI (Fig. 4, L and M, and fig. S19C), pointing to enhanced cholinergic activity with insulating-polymer assembly—a specific gain of function in living animals and an opposite-direction effect compared with conducting-polymer assembly, both of which are concordant with the electrophysiology.

We have achieved chemical assembly of electroactive polymers on genetically specified cellular elements within living cells, tissues, and animals. Future work may address potential limitations and opportunities; for example, reaction products could over time occupy substantial space in and near targeted cells, which may be useful in some contexts but also could result in cytotoxicity. Distinct strategies for the targeting and triggering of chemical synthesis could extend beyond the oxidative radical initiation shown here while building on the core principle of assembling within cells (as reaction compartments) genetically and anatomically targeted reactants (such as monomers), catalysts (such as enzymes or surfaces), or reaction conditions (through modulators of pH, light, heat, redox potential, electrochemical potential, and other chemical or energetic signals). Diverse cell-specific chemical syntheses may thus be explored and developed for a broad array of functional characteristics in assembled structures.

REFERENCES AND NOTES

1. A. L. Hodgkin, A. F. Huxley, *J. Physiol.* **117**, 500–544 (1952).
2. K. Deisseroth, *Nat. Neurosci.* **18**, 1213–1225 (2015).

3. V. Gradinaru, J. Treweek, K. Overton, K. Deisseroth, *Annu. Rev. Biophys.* **47**, 355–376 (2018).
4. T. Dvir et al., *Nat. Nanotechnol.* **6**, 720–725 (2011).
5. G. Cellot et al., *Nat. Nanotechnol.* **4**, 126–133 (2009).
6. J. Niu et al., *Nat. Chem.* **9**, 537–545 (2017).
7. L. Ouyang, C. L. Shaw, C. C. Kuo, A. L. Griffin, D. C. Martin, *J. Neural Eng.* **11**, 026005 (2014).
8. L. Pan et al., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9287–9292 (2012).
9. R. Cruz-Silva et al., *Eur. Polym. J.* **41**, 1129–1135 (2005).
10. S. S. Lam et al., *Nat. Methods* **12**, 51–54 (2015).
11. T. Scherf et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6629–6634 (2001).
12. A. S. Pavitt, E. J. Bylaska, P. G. Tratnyek, *Environ. Sci. Process. Impacts* **19**, 339–349 (2017).
13. Y. Wei et al., *J. Polym. Sci. C* **28**, 81–87 (1990).
14. S. E. Moulton, P. C. Innis, L. A. P. Kane-Maguire, O. Ngamna, G. G. Wallace, *Curr. Appl. Phys.* **4**, 402–406 (2004).
15. C. Henning, K. H. Hallmeier, R. Szargan, *Synth. Met.* **92**, 161–166 (1998).
16. D. J. Stokes, *Principles and Practice of Variable Pressure/Environmental Scanning Electron Microscopy* (John Wiley, 2008).
17. E. Stavrinidou et al., *Sci. Adv.* **1**, e1501136 (2015).
18. I. E. Mulazimoglu, *Asian J. Chem.* **22**, 8203–8208 (2010).
19. A. M. Paşca et al., *Nat. Methods* **12**, 671–678 (2015).
20. S. A. Sloan, J. Andersen, A. M. Paşca, F. Birey, S. P. Paşca, *Nat. Protoc.* **13**, 2062–2085 (2018).
21. R. C. Van Lehn et al., *Nano Lett.* **13**, 4060–4067 (2013).
22. R. P. Carney et al., *ACS Nano* **7**, 932–942 (2013).
23. M. J. Gillespie, R. B. Stein, *Brain Res.* **259**, 41–56 (1983).
24. D. K. Hartline, D. R. Colman, *Curr. Biol.* **17**, R29–R35 (2007).
25. B. Howell, L. E. Medina, W. M. Grill, *J. Neural Eng.* **12**, 056015–056015 (2015).
26. H. E. Kato et al., *Nature* **561**, 349–354 (2018).
27. J. L. Donnelly et al., *PLOS Biol.* **11**, e1001529 (2013).
28. S. L. McIntire, E. Jorgensen, J. Kaplan, H. R. Horvitz, *Nature* **364**, 337–341 (1993).

ACKNOWLEDGMENTS

Funding: K.D. was supported by NIH and NSF. J.L. was supported by Stanford Bio-X, K.D., Z.B., and S.P.P. were supported by the Wu-Tsai Neuroscience Institute. Y.S.K. was supported by the Kwanjeong International Fellowship and Stanford Bio-X. X.W. was supported by the Life Sciences Research Foundation and the Gordon and Betty Moore Foundation. We acknowledge resources of the Advanced Light Source U.S. Department of Energy (DOE) Facility (DE-AC02-05CH11231). Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), supported by NSF (ECCS-1542152). **Author contributions:** Z.B. and K.D. conceived and initiated the project with implementation by the experimental team of J.L., Y.S.K., A.T., and C.R.; K.D. and C.R. designed the Apex2 molecular strategy. J.L., C.R., A.T., and Z.B. developed the polymerization reactions. J.L. and A.T. performed UV-vis characterizations. J.L. and Y.J. performed conductivity measurements. Y.S.K., L.E.F., and J.L. performed electrophysiology. C.E.R., Y.S.K., and J.L. conducted *C. elegans* work, guided by K.S.; F.B. and S.P.P. developed hCS. T.K. synthesized the TETs monomer. S.C. and J.L. conducted XPS. C.W. and J.L. conducted NEXAFS. L.-M.J., H.W., and J.L. conducted EM imaging. X.W. optimized tissue imaging. J.L., Y.S.K., A.T., Z.B., and K.D. prepared figures and wrote the manuscript with edits from all authors. Z.B. and K.D. supervised all aspects of the work. **Competing interests:** A patent application has been filed by Stanford related to this work; all methods and protocols are freely available. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1372/suppl/DC1
Materials and Methods
Figs. S1 to S19
References (29–33)

22 June 2019; accepted 21 January 2020
10.1126/science.aay4866

HYBRIDIZATION

Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates

Catherine Chen* and Karin S. Pfennig*

Hybridization—interbreeding between species—is generally thought to occur randomly between members of two species. Contrary to expectation, female plains spadefoot toads (*Spea bombifrons*) can increase their evolutionary fitness by preferentially mating with high-quality males of another species, the Mexican spadefoot toad (*Spea multiplicata*). Aspects of Mexican spadefoot males' mating calls predict their hybrid offspring's fitness, and plains spadefoot females prefer Mexican spadefoot males on the basis of these attributes, but only in populations and ecological conditions where hybridization is adaptive. By selecting fitness-enhancing mates of another species, females increase hybridization's benefits and exert sexual selection across species. Nonrandom mating between species can thereby increase the potential for adaptive gene flow between species so that adaptive introgression is not simply happenstance.

Mating between species (hybridization) is widespread and has important evolutionary and ecological consequences (1–3). Although hybridization is often considered deleterious, it is increasingly recognized as potentially fitness enhancing if mates of one's own species (i.e., conspecifics) are limited (4) or if hybrids are better adapted to their environment than pure-species types (5, 6).

When hybridization is adaptive, selection can favor the evolution of traits that foster hybridization. Such traits can include mating preferences for members of a different species (i.e., heterospecifics) (4, 7, 8). However, the quality of heterospecific mates might vary such that mating with some heterospecifics versus others might result in higher fitness. Thus, mate choice among heterospecifics could allow individuals to enhance their fitness in the same way as mate choice among conspecifics (9). In such cases, sexual selection could operate between species—a possibility that has not been previously considered.

We tested the hypothesis that females exercise adaptive mate choice among members of another species by using plains spadefoot toads, *Spea bombifrons*. In the southwestern United States, this species hybridizes with Mexican spadefoot toads, *Spea multiplicata* (10). Although the resulting F₁ hybrid males are sterile and F₁ hybrid females have reduced fecundity, F₁ hybrid females can breed back to either parent species (10).

Hybridization is maladaptive for *S. multiplicata* females, but it is sometimes adaptive for *S. bombifrons* females. Spadefoot tadpoles develop in ephemeral desert ponds that often dry before the tadpoles metamorphose, resulting

in their deaths (8). Because hybrid tadpoles develop faster than pure *S. bombifrons* tadpoles, hybrids are more likely to escape drying ponds (8), thereby potentially passing on alleles from their parents to future generations. Consequently, female *S. bombifrons* have evolved mate preferences for *S. multiplicata* males but only in shallow, ephemeral ponds (8, 10); when

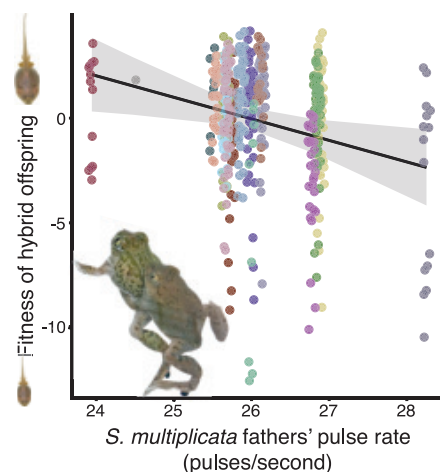


Fig. 1. Hybrid-tadpole fitness as a function of father's call. Tadpoles resulting from *S. bombifrons* females crossed with *S. multiplicata* males (inset) had the highest fitness when their fathers had calls with slow pulse rates. The fitness of hybrid offspring is the first principal component from a principal components analysis of body size, body mass, and developmental stage; all variables loaded positively on the principal component, so higher values correspond to increased fitness (table S1). The line is the predicted fitness at given pulse rates using the averaged model (Table 1) while holding all other variables at mean values; intervals are 95% confidence intervals. Different colors denote different families, and points are jittered horizontally. Tadpole photos are strictly visual representations of the variation in tadpole fitness.

a pond is deep and likely to last long enough for *S. bombifrons* tadpoles to successfully metamorphose, *S. bombifrons* females prefer males of their own species (8).

To determine if *S. bombifrons* females could exercise adaptive mate choice among *S. multiplicata* males, we first determined if the fitness consequences vary for *S. bombifrons* females mated with different *S. multiplicata* males. We then evaluated whether any aspects of *S. multiplicata* male calls (the sexual signals that females evaluate) predicted hybrid-offspring fitness. We reasoned that this may occur because attributes of *S. multiplicata* calls predict offspring fitness in conspecific matings (10).

We bred 20 *S. bombifrons* females with 20 *S. multiplicata* males that differed in mating call characteristics and reared their offspring (10). We measured tadpole body size [snout-to-vent length (SVL)], body mass, and Gosner stage (a generalized system describing anuran development) as fitness components (10). We then combined these measurements into a single principal component that explained 91.6% of the variation in our data (table S1).

This combined measure of fitness varied among families of hybrid offspring (Fig. 1 and table S2). Body size of the mothers positively predicted this fitness measure in hybrid offspring, but the pulse rate of the fathers' calls was an even better predictor of our fitness measure in the hybrid offspring (Table 1). Specifically, males with slower pulse rates sired the largest, heaviest, and most rapidly developing hybrid offspring (Table 1 and Fig. 1).

Because slower pulse rates of *S. multiplicata* calls predicted higher hybrid-offspring fitness, we next investigated whether *S. bombifrons* females preferred *S. multiplicata* male calls with slower versus faster pulse rates. We further predicted that if any such preferences are plastic [sensu (8)], they would be expressed only in shallow water, where hybridization is adaptive. Using previous methods (8, 10), we presented *S. bombifrons* females with a choice of *S. multiplicata* calls with either slow or fast pulse rates. The same females were given this choice under two conditions: shallow water (simulating ephemeral ponds where hybridization is advantageous) and deep water (simulating longer-lasting ponds where hybridization is disadvantageous).

As predicted, in shallow water, *S. bombifrons* females preferred *S. multiplicata* calls with slower pulse rates (exact binomial test, $P = 0.0026$; Fig. 2A). By contrast (and also as predicted), *S. bombifrons* females showed no such preference in deep water (exact binomial test, $P = 0.15$; Fig. 2A). The pattern of preference in deep water differed significantly from that in shallow water (McNemar binomial exact test, $P = 0.0081$; Fig. 2A). Thus, *S. bombifrons*

Department of Biology, Campus Box 3280, Coker Hall, University of North Carolina, Chapel Hill, NC 27599, USA.
*Corresponding author. Email: catherine.chen@unc.edu (C.C.); kpfennig@unc.edu (K.S.P.)

females distinguish between different *S. multiplicata* male calls and prefer slower pulse rate calls that are associated with greater fitness in hybrid offspring (Fig. 1), but they exhibit this preference only when ponds are shallow and it is adaptive to hybridize.

Sympatric *S. bombifrons* females may prefer *S. bombifrons* males with slow pulse rates, and such preferences among conspecifics may be simply expressed when choosing among *S. multiplicata* males. However, we found that *S. bombifrons* females do not discriminate between conspecific calls varying in pulse rate in either shallow water (exact binomial test, $P = 1$) or deep water (exact binomial test, $P = 0.14$; Fig. 2A). Moreover, their behavior was unchanged between the two water conditions (McNemar binomial exact test, $P = 0.30$; Fig. 2A). Thus, sympatric *S. bombifrons* female preferences for males of their own species are not simply generalized to *S. multiplicata* males.

S. multiplicata females may prefer *S. multiplicata* males with slow pulse rates, and this preference might have been transferred from *S. multiplicata* into *S. bombifrons* via gene flow between the species (i.e., introgression). However, when we tested *S. multiplicata* females for their preferences for *S. multiplicata* calls that differed in pulse rate, we found no preferences in either shallow water (exact binomial test, $P = 0.49$; table S6) or deep water (exact binomial test, $P = 0.21$; table S6). Water level also did not alter female preferences (McNemar binomial exact test, $P = 0.31$). Thus, our results cannot be explained by introgression of mate preferences from *S. multiplicata* into *S. bombifrons*.

Our finding that *S. bombifrons* females prefer fitness-enhancing males of a different species suggests that these preferences evolved in sympatry (where hybridization occurs) via selection acting to optimize the fitness of their hybrid offspring. To assess whether this behavior did evolve in sympatry, we determined whether the same preferences are present in *S. bombifrons* females from allopatric populations (i.e., where they occur in the absence of *S. multiplicata*). Our rationale for doing so was that allopatric females would possess ancestral preferences that predate contact with *S. multiplicata* in the southwestern United States. The center of origin for *S. bombifrons* is the grasslands of the Great Plains, and they have apparently expanded their range into the desert Southwest, possibly because of adaptive hybridization with desert-adapted *S. multiplicata* (11, 12) (Fig. 2). Male *S. bombifrons* from the two regions produce two different call types: Males from the Great Plains produce a slow call that is more similar to *S. multiplicata* calls than to the fast call that is produced by males in the desert Southwest (Fig. 2, fig. S1, and audio S2 and S3). We presented allopatric

S. bombifrons females with conspecific calls of their own slow-call type that differed in pulse rate and found that they preferred slow pulse rate calls in shallow water (exact binomial test, $P = 0.029$) but not in deep water (exact binomial test, $P = 0.18$), but patterns of preference in the two environments were similar (McNemar binomial exact test, $P = 0.80$; Fig. 2B). In contrast with sympatric *S. bombifrons* females, allopatric *S. bombifrons* females did not differentiate between *S. multiplicata* calls

that differed in pulse rate (exact binomial tests, shallow water: $P = 1$, deep water: $P = 0.28$; McNemar binomial exact test, $P = 0.42$; Fig. 2B). Thus, allopatric *S. bombifrons* females potentially use pulse rate to discriminate among conspecifics, but this preference does not account for *S. bombifrons* preferences among *S. multiplicata* in sympatry. Instead, sympatric *S. bombifrons* female preferences for *S. multiplicata* calls with slow pulse rates have apparently evolved in sympatry after

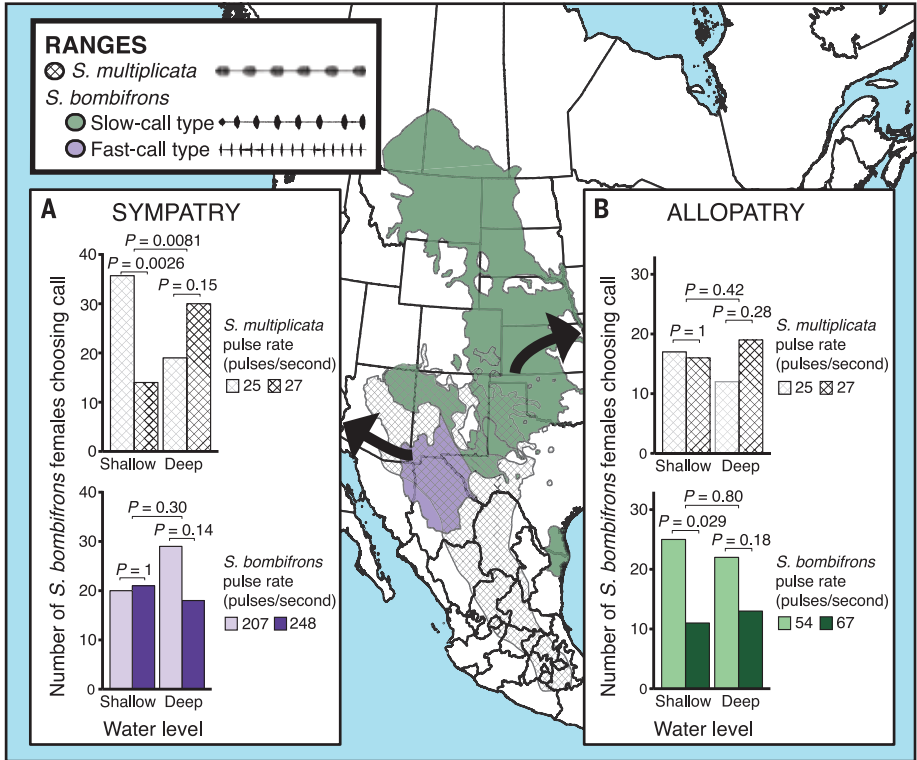


Fig. 2. Spea ranges, calls, and female preferences. *S. bombifrons* originate from the Great Plains of the United States. They have undergone contact and hybridize with *S. multiplicata* in the southwestern United States. Male *S. bombifrons* calls differ between the two regions as shown. (A) Sympatric *S. bombifrons* females prefer *S. multiplicata* male calls with slower pulse rates in shallow water (where hybridization is adaptive) but not in deep water (where hybridization is not adaptive). Sympatric *S. bombifrons* females do not distinguish between fast-call type *S. bombifrons* male calls that differ in pulse rate. (B) Allopatric *S. bombifrons* females do not distinguish between *S. multiplicata* calls that differ in pulse rate, but they do prefer slower pulse rates in the slow-call type *S. bombifrons* calls. See also table S6.

Table 1. Standardized coefficients of model predictors for offspring fitness, after model averaging of the top nine candidate models. See table S3 for further details. SE, standard error.					
Parameter	Relative importance	Estimate	SE	z value	P value
Male pulse rate	0.95	-1.44	0.563	2.56	0.0106
Female SVL	0.90	-1.34	0.704	1.91	0.0562
Male condition	0.69	0.814	0.679	1.20	0.231
Male call duration	0.16	0.176	0.480	0.367	0.714
Male genotype	0.01	-0.00917	0.110	0.0830	0.934

contact with *S. multiplicata*, possibly as a modification of their preexisting allopatric preference (9, 13).

Spadefoots are unlikely to be the only group in which females can optimize the fitness consequences of hybridization by preferring heterospecific males that sire hybrid offspring with relatively higher fitness. Across taxa, individuals vary in their propensity to hybridize (14) and even prefer certain heterospecific males over others (15). Behavior to optimize hybridization may be especially likely in recently diverged groups with similar mating behaviors and for which hybridization could confer fitness benefits (4–6). However, more work is needed to determine how common this phenomenon is.

Our findings have two general implications. First, they suggest that members of one species might be able to exert sexual selection on another species. Such a pattern could affect the evolution and distribution of sexual signals, local mate competition, and even the extent to which species do or do not diverge where they co-occur (9, 16). Second, nonrandom hybridization can bias gene flow between species. To date, adaptive introgression has been considered a happenstance occurrence in which random, or possibly deleterious, hybridization is followed by the retention of adaptive alleles (17, 18). If, however, mate preferences result in nonrandom production of fitter hybrid genotypes, then nonrandom mating can enhance both the chances of adaptive introgression and the transfer of alleles that are particularly well suited to a given habitat. In a rapidly changing world where hybridization could become increasingly common, understanding when and how adaptive introgression occurs could be key to population rescue, adaptation, or the replacement of one species by another (6, 19–21). Our results indicate that sexual selection and mate choice should be considered as integral components of these processes.

REFERENCES AND NOTES

- R. Abbott et al., *J. Evol. Biol.* **26**, 229–246 (2013).
- K. S. Pfennig, A. L. Kelly, A. A. Pierce, *Proc. R. Soc. B* **283**, 20161329 (2016).
- S. Lamichaney et al., *Science* **359**, 224–228 (2018).
- T. Veen et al., *Nature* **411**, 45–50 (2001).
- M. R. Jones et al., *Science* **360**, 1355–1358 (2018).
- E. M. Ozolator et al., *Science* **364**, 455–457 (2019).
- I. Lesna, M. W. Sabelis, *Nature* **401**, 581–584 (1999).
- K. S. Pfennig, *Science* **318**, 965–967 (2007).
- G. G. Rosenthal, *Mate Choice: The Evolution of Sexual Decision Making from Microbes to Humans* (Princeton Univ. Press, 2017).
- Materials and methods are available as supplementary materials.
- A. M. Rice, D. W. Pfennig, *J. Evol. Biol.* **21**, 696–704 (2008).
- A. A. Pierce, R. Gutierrez, A. M. Rice, K. S. Pfennig, *Proc. R. Soc. B* **284**, 20170007 (2017).
- M. J. Ryan, *Science* **281**, 1999–2003 (1998).
- G. G. Rosenthal, *J. Evol. Biol.* **26**, 252–255 (2013).
- M. J. Ryan, W. E. Wagner Jr., *Science* **236**, 595–597 (1987).
- T. M. Panhuis, R. Butlin, M. Zuk, T. Tregenza, *Trends Ecol. Evol.* **16**, 364–371 (2001).
- M. L. Arnold, K. Kunte, *Trends Ecol. Evol.* **32**, 601–611 (2017).
- S. A. Taylor, E. L. Larson, *Nat. Ecol. Evol.* **3**, 170–177 (2019).
- J. A. Hamilton, J. M. Miller, *Conserv. Biol.* **30**, 33–41 (2016).
- R. P. Kovach, G. Luikart, W. H. Lowe, M. C. Boyer, C. C. Muhlfeld, *Conserv. Biol.* **30**, 428–430 (2016).
- K. S. Pfennig, *Science* **364**, 433–434 (2019).
- We are grateful to D. Pfennig, M. Noor, S. Alberts, M. Servedio, C. Willett, P. Kelly, G. Calabrese, N. Levis, E. Harmon, E. Laub, I. Hamid, T. Maier, S. Marion, F. Xu, C. Shoben, and four anonymous reviewers for discussion and comments. We thank P. Kelly and G. Calabrese for help recording males in the field, G. Calabrese for analyzing male calls, J. Morton for animal collection, A. Kelly and A. McNear for genotyping, and J. Urbanhowar for assistance with analyses. **Funding:** A grant from the U.S. NSF (IOS-1555520) supported this work; C.C. was additionally supported by the NSF GRFP (DGE-1650116) and grants from the Society for the Study of Evolution, Sigma Xi, and the Southwestern Association of Naturalists. **Author contributions:** C.C. and K.S.P. conceived of the project and its design. C.C. reared hybrid tadpoles, conducted behavioral trials, and analyzed data. C.C. and K.S.P. jointly wrote the paper, and each approved and edited the final version. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1377/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Tables S1 to S8
References (22–41)
Audio S1 to S3
Data S1 and S2

20 September 2019; accepted 28 January 2020
10.1126/science.aaz5109

PYROPTOSIS

cFLIP_L protects macrophages from LPS-induced pyroptosis via inhibition of complex II formation

Hayley I. Muendlein¹, David Jetton², Wilson M. Connolly³, Keith P. Eidell², Zoie Magri², Irina Smirnova³, Alexander Poltorak^{3,4*}

Cell death and inflammation are interdependent host responses to infection. During pyroptotic cell death, interleukin-1 β (IL-1 β) release occurs through caspase-1 and caspase-11-mediated gasdermin D pore formation. In vivo, responses to lipopolysaccharide (LPS) result in IL-1 β secretion. In vitro, however, murine macrophages require a second “danger signal” for the inflammasome-driven maturation of IL-1 β . Recent reports have shown caspase-8-mediated pyroptosis in LPS-activated macrophages but have provided conflicting evidence regarding the release of IL-1 β under these conditions. Here, to further characterize the mechanism of LPS-induced secretion in vitro, we reveal an important role for cellular FLICE-like inhibitory protein (cFLIP) in the regulation of the inflammatory response. Specifically, we show that deficiency of the long isoform cFLIP_L promotes complex II formation, driving pyroptosis, and the secretion of IL-1 β in response to LPS alone.

Inflammatory responses to infection are mediated via nuclear factor κ B (NF- κ B) and mitogen-activated protein kinase (MAPK) signaling cascades downstream of Toll-like receptors (TLRs) and are crucial for host survival. These responses up-regulate various effectors, including cytokines, chemokines, and prosurvival factors (1). Many pathogens have evolved to block host signaling cascades, promoting pathogen survival (2). The *Yersinia* species bacteria rely on the effector protein YopJ to block activation of the level 3 MAPK TAK1 (transforming growth factor β -activated

kinase) (3, 4). In response, host cells limit infection via initiation of cell death pathways (5), such as pyroptosis, which is accompanied by interleukin-1 β (IL-1 β) release.

Pyroptosis is mediated via the inflammasome-driven activation of caspase-1 (CASP1) and caspase-11 (CASP11), resulting in cleavage of the pore-forming protein gasdermin D (GSDMD) (6–8). In macrophages, IL-1 β maturation requires two signals to up-regulate pro-IL-1 β and to induce NOD, LRR and pyrin domain-containing protein 3 (NLRP3)-mediated maturation of IL-1 β (9). Interaction of NLRP3 with apoptosis-associated speck-like protein (ASC) recruits pro-CASP1, which cleaves and releases mature IL-1 β via pores formed by GSDMD (10).

Recently, we and others reported caspase-8 (CASP8)-mediated pyroptosis in response to *Yersinia* infection. Pyroptosis was dependent on YopJ and could be mimicked with lipopolysaccharide (LPS) and the small-molecule

¹Graduate Program in Genetics, Tufts Graduate School of Biomedical Sciences, Boston, MA 02111, USA. ²Graduate Program in Immunology, Tufts Graduate School of Biomedical Sciences, Boston, MA 02111, USA.

³Department of Immunology, Tufts University School of Medicine, Boston, MA 02111, USA. ⁴Laboratory of Genetics of Innate Immunity, Petrozavodsk State University, Petrozavodsk, Republic of Karelia 185910, Russia.

*Corresponding author. Email: alexander.poltorak@tufts.edu

inhibitor of TAK1, 5Z-7-oxozeaenol (5z7) (11, 12). Cell death required the kinase activity of receptor-interacting serine/threonine-protein kinase 1 (RIP1) and GSDMD (Fig. 1A) but was accompanied by low IL-1 β levels (fig. S1A),

likely owing to the inhibition of MAPK- and NF- κ B-mediated pro-IL-1 β production. Bone marrow-derived macrophages (BMDMs) pre-primed with LPS before 5z7 treatment produced significantly higher levels of IL-

1 β compared with cells concurrently treated with LPS and 5z7 (Fig. 1B). IL-1 β release in LPS-pre-primed BMDMs was dependent on TIR-domain-containing adapter-inducing interferon- β (TRIF), RIP1 kinase activity, and

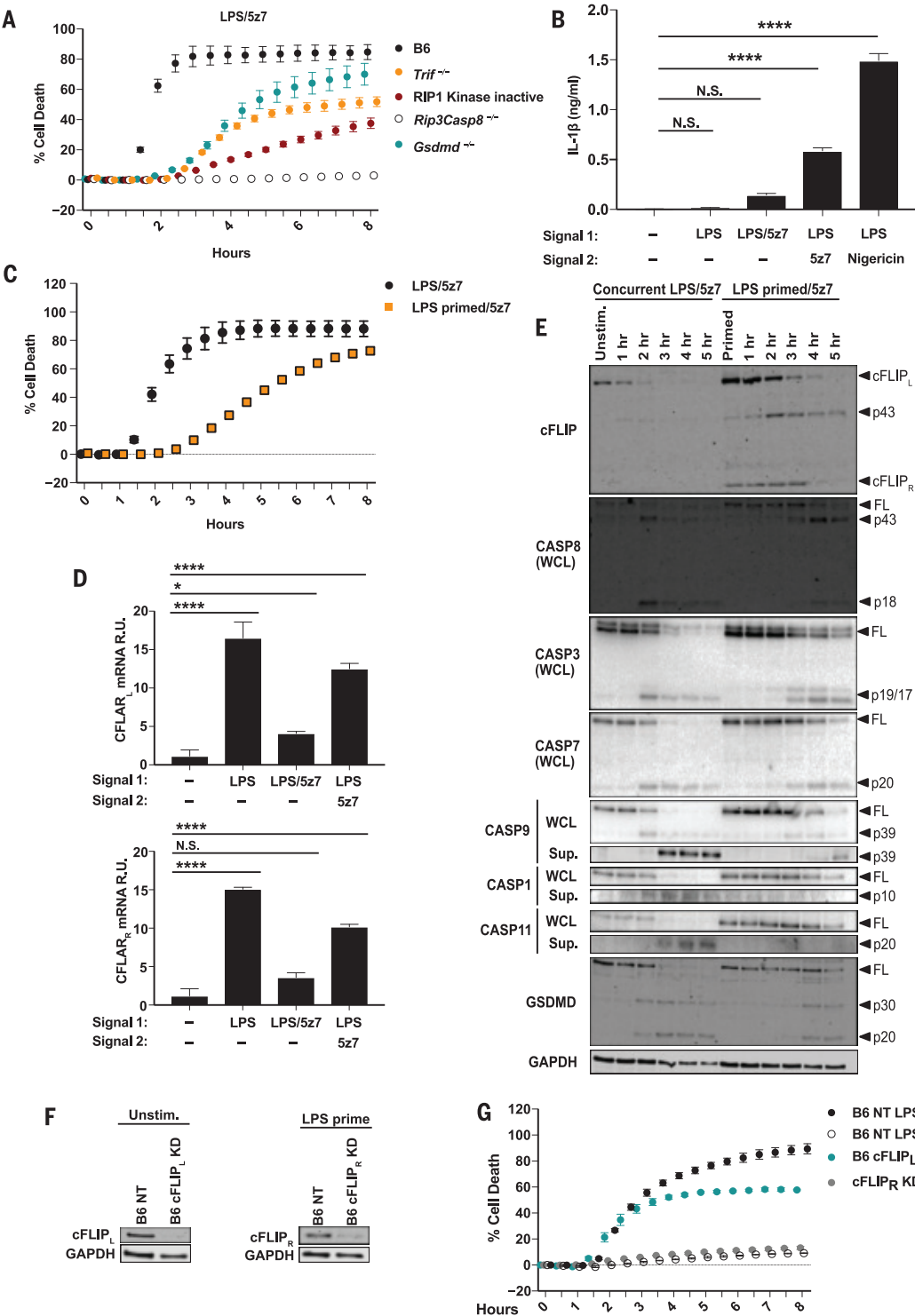


Fig. 1. LPS priming before TAK1 inhibition drives IL-1 β release, inhibits cell death, and regulates cFLIP levels. (A) Cell death in BMDMs from B6, RIP1 kinase-inactive (RIP1Ki), *Trif*^{-/-}, *Rip3*^{-/-}*Casp8*^{-/-}, and *Gsdmd*^{-/-} mice treated concurrently with LPS/5z7. (B) IL-1 β release from B6 BMDMs 6 hours after indicated treatments. LPS pre-priming (10 or 100 ng/ml) occurred 4 hours before the addition of 5z7 or nigericin, respectively. (C) Cell death in B6 BMDMs stimulated concurrently with LPS/5z7 or LPS-pre-primed/5z7. (D) Relative *Cflar* mRNA levels, normalized to *Gapdh* (glyceraldehyde phosphate dehydrogenase) after 1 hour of treatment with LPS, LPS/5z7, or LPS-pre-primed/5z7 treatment in B6 BMDMs. R.U., relative units. (E) Full-length (FL) and cleaved products of cFLIP, GSDMD, and indicated caspases from whole-cell lysates (WCL) or precipitated from the supernatant (Sup.) of B6 BMDMs treated concurrently with LPS/5z7 or LPS-pre-primed/5z7 for indicated times. Unstim., unstimulated; hr, hours. (F) cFLIP protein levels in B6 BMDMs knocked down for cFLIP_L or cFLIP_R or transduced with a nontargeting (NT) control. (G) Cell death in B6 BMDMs knocked down for cFLIP_L or cFLIP_R and stimulated as indicated [extent of KD shown in (F)]. Data from cell death assays and immunoblots are representative of three or more independent experiments, and cell death data are presented as the mean \pm SD of triplicate wells. IL-1 β release data are presented as the mean \pm SD for triplicate wells from three or more independent experiments. Analysis of variance (ANOVA) was used for comparison between groups: N.S., nonsignificant ($P > 0.05$); * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

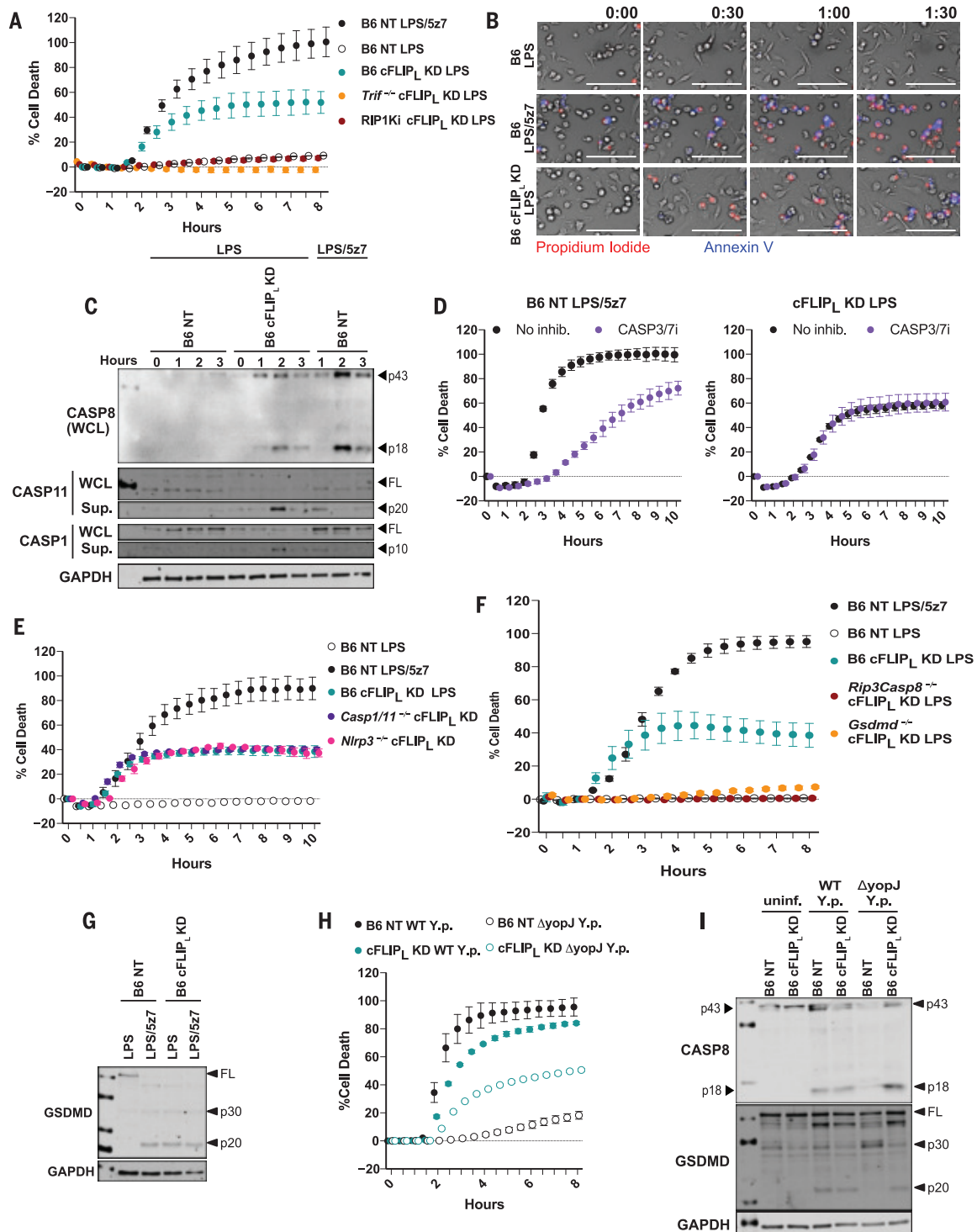


Fig. 2. LPS induces caspase-8-mediated pyroptosis in the absence of cFLIP_L.

(A) Cell death in B6, *Trif*^{-/-}, and RIP1Ki BMDMs knocked down for cFLIP_L and stimulated as indicated (extent of KD shown in fig. S3A). (B) Kinetic 20× magnification imaging of PI incorporation and annexin V staining in B6 and B6 cFLIP_L-KD BMDMs stimulated with LPS or LPS/5z7 for up to 1 hour and 30 min (extent of KD shown in fig. S3B). Scale bars: 100 μm. (C) Full-length and cleaved products of indicated caspases from whole-cell lysates (WCL) or precipitated from the supernatant of B6 NT control or B6 cFLIP_L-KD BMDMs (extent of KD shown in fig. S3C). (D) Inhibition of cell death with CASP3/7 inhibitor in LPS/5z7-stimulated B6 macrophages or LPS-treated cFLIP_L-KD BMDMs (extent of KD shown in fig. S3G).

(E) Cell death in B6, *Nlrp3*^{-/-}, and *Casp1*^{-/-} *Casp11*^{-/-} macrophages stimulated with LPS or LPS/5z7 (extent of KD shown in fig. S3H). (F) Cell death in B6, *Rip3*^{-/-} *Casp8*^{-/-}, and *Gsdmd*^{-/-} BMDMs knocked down for cFLIP_L and stimulated as indicated (extent of KD shown in fig. S3I). (G) Full-length and cleaved GSDMD from B6 NT control or B6 cFLIP_L-KD BMDMs stimulated as indicated (extent of KD shown in fig. S3J). (H) Cell death and (I) cleavage of CASP8 and GSDMD in B6 NT and B6 cFLIP_L-KD BMDMs stimulated with WT or YopJ deficient (ΔyopJ) *Y. pseudotuberculosis* (extent of KD shown in fig. S3L). uninf., uninfected. All immunoblots and cell death data are representative of three or more independent experiments. Cell death data are presented as the mean ± SD of triplicate wells.

CASP8 (fig. S1B). Notably, LPS-induced pro-IL-1 β production was entirely dependent on TRIF. Additionally, at later time points, pro-IL-1 β levels were dependent on CASP8 and RIP1 kinase activity (fig. S1C). IL-1 β production was independent of GSDMD, suggesting an alternate mechanism for IL-1 β release (fig. S1B). Finally, IL-1 β secretion in LPS-pre-primed

BMDMs was mediated by NLRP3 inflammatory components (fig. S1D).

LPS pre-priming delayed cell death (Fig. 1C), which hinted at a mechanism in which LPS stimulation drives the expression of pro-survival factors that are inhibited by 5z7. We identified these factors by comparing genome-wide mRNA levels in BMDMs treated

with LPS or both LPS and 5z7 (fig. S2A). Among many differentially expressed transcriptional modulators, we identified *Cflar*, a gene encoding the enzymatically inactive homolog of CASP8, cFLIP (fig. S2A). The 25-kDa short isoform cFLIP_R blocks CASP8 activation entirely, and the 55-kDa long isoform cFLIP_L blocks it partially (13, 14). LPS-pre-primed

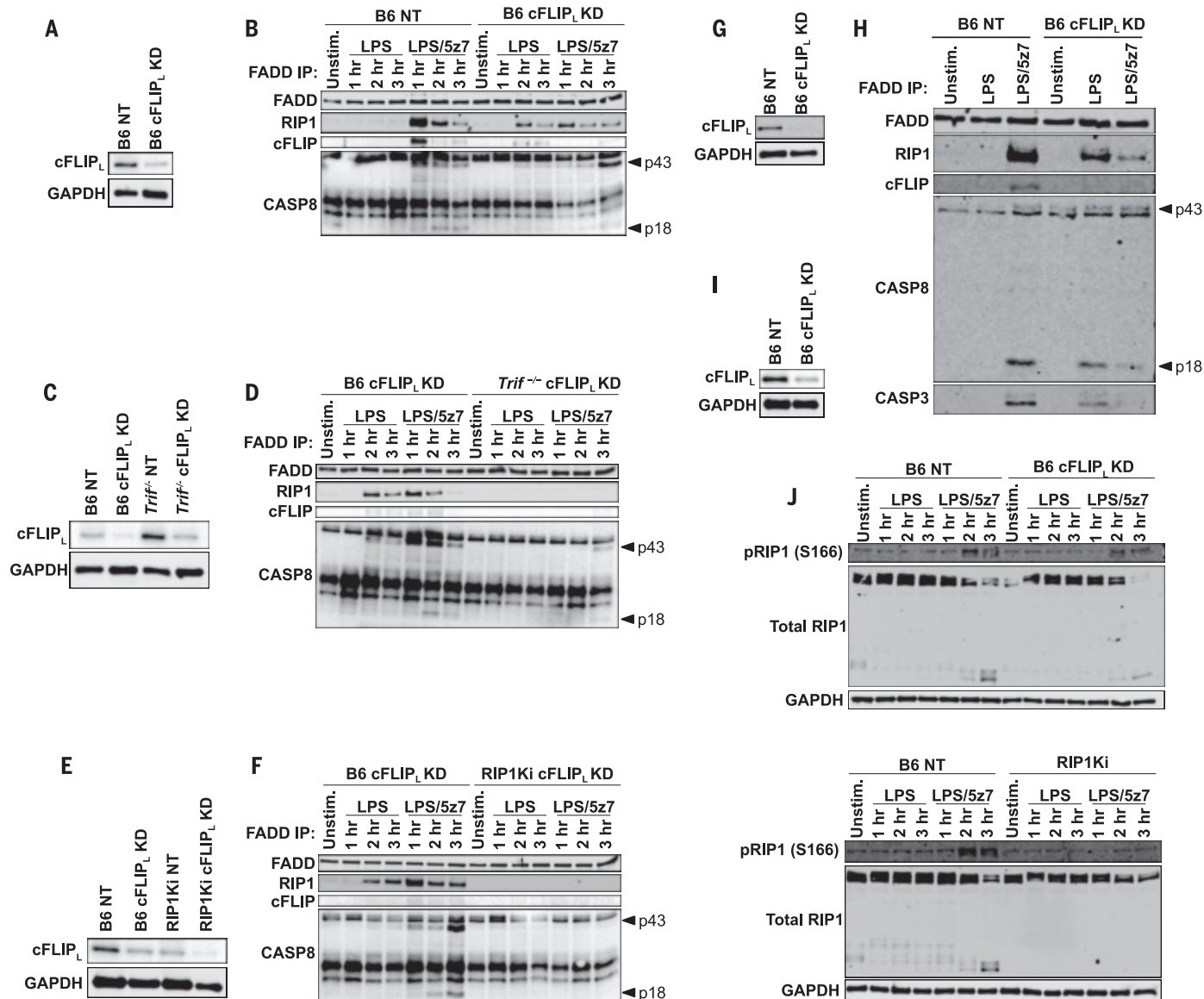


Fig. 3. cFLIP_L deficiency promotes complex II formation downstream of TRIF in response to LPS. (A) cFLIP_L protein levels in B6 BMDMs knocked down for cFLIP_L or transduced with a NT control. (B) FADD immunoprecipitation (IP) in B6 NT and B6 cFLIP_L-KD BMDMs stimulated as indicated and probed for complex II components [extent of KD shown in (A)]. (C) cFLIP_L protein levels in B6 and *Trif*^{-/-} BMDMs knocked down for cFLIP_L or transduced with a NT control. (D) FADD IP in B6 and *Trif*^{-/-} BMDMs knocked down for cFLIP_L, stimulated as indicated, and probed for complex II components [extent of KD shown in (C)]. (E) cFLIP_L protein levels in B6 and RIP1Ki BMDMs knocked down for cFLIP_L or transduced with a NT control.

(F) FADD IP in B6 and *Trif*^{-/-} BMDMs knocked down for cFLIP_L, stimulated as indicated, and probed for complex II components [extent of KD shown in (E)]. (G) cFLIP_L protein levels in B6 BMDMs knocked down for cFLIP_L or transduced with a NT control. (H) FADD IP in B6 NT and B6 cFLIP_L-KD stimulated for 2 hours as indicated and probed for complex II components [extent of KD shown in (G)]. (I) cFLIP_L protein levels in B6 BMDMs knocked down for cFLIP_L or transduced with a NT control. (J) Total and pRIP1 (S166) levels in LPS- and LPS/5z7-stimulated cFLIP_L silenced or NT control B6 and RIP1Ki BMDMs [extent of KD shown in (I)]. All immunoblots are representative of three or more independent experiments.

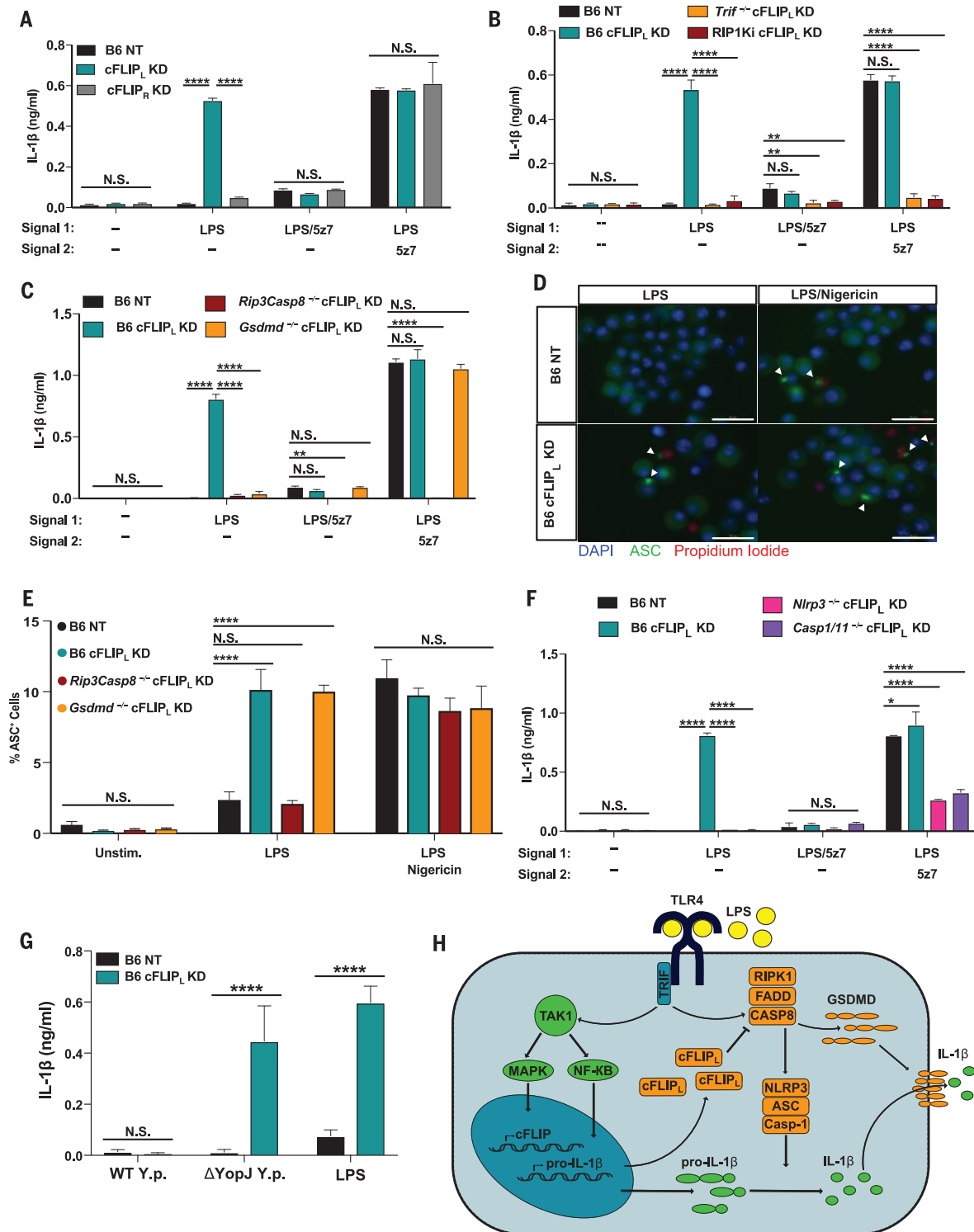


Fig. 4. cFLIP_L deficiency drives IL-1 β maturation and release in response to a single signal from LPS. (A to C) IL-1 β release from (A) B6 NT control, B6 cFLIP_L-KD, and B6 cFLIP_R-KD BMDMs; (B) RIP1Ki and *Trif*^{-/-} BMDMs knocked down for cFLIP_L; and (C) RIP3/CASP8 and GSDMD-deficient BMDMs knocked down for cFLIP_L after 6 hours of indicated treatments (extent of KD shown in fig. S4, A to C). (D) Representative 60 \times images of ASC specks in B6 NT control and B6 cFLIP_L-KD BMDMs stimulated for 4 hours with LPS, or 2 hours with LPS/nigericin (extent of KD shown in fig. S4D). Scale bar: 30 μ m. (E) Percentage of ASC⁺ cells in B6, *Rip3*^{-/-}*Casp8*^{-/-}, and *Gsdmd*^{-/-} BMDMs knocked down for cFLIP_L (extent of KD

shown in fig. S4F). (F) IL-1 β release in B6, *Nlrp3*^{-/-}, and *Casp1*^{-/-}*Casp11*^{-/-} macrophages silenced for cFLIP_L and stimulated as indicated (extent of KD shown in fig. S4G). (G) B6 BMDMs silenced for cFLIP_L and stimulated with WT or YopJ-deficient (Δ yopJ) *Y. pseudotuberculosis* for 6 hours (extent of KD shown in fig. S4L). (H) Model of LPS-driven pyroptosis and IL-1 β production, as regulated by cFLIP_L. IL-1 β release and ASC percentage data are presented as the mean \pm SD for triplicate wells from three or more independent experiments. Analysis of variance (ANOVA) was used for comparison between groups: N.S., nonsignificant ($P > 0.05$); * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

BMDMs maintained *Cflar* mRNA levels after the addition of 5z7 (Fig. 1D). Both cFLIP_L and cFLIP_R protein levels were elevated in LPS-pre-primed BMDMs (Fig. 1E). In LPS/5z7-activated BMDMs, cFLIP_L was cleaved within 2 hours, whereas in LPS-pre-primed BMDMs this loss was delayed. cFLIP_L cleavage coincided with the kinetics of CASP8 cleavage and the onset of cell death (Fig. 1, C and E). In agreement with the protective role of the partially active CASP8-cFLIP_L heterodimer, we observed enhanced CASP8 cleavage to the locally active p43 fragment in LPS-pre-primed BMDMs, whereas cleavage to p18 was nearly abolished (Fig. 1E). LPS pre-priming delayed the activation of CASP1, CASP3, CASP7, CASP9, and CASP11, and the cleavage of GSDMD, suggesting that cellular stores of cFLIP, regulated by TAK1, determine the rate and extent of LPS/5z7-induced death (Fig. 1E).

Notably, a knockdown (KD) of cFLIP_L but not cFLIP_R sensitized cells to death after stimulation with LPS alone (Fig. 1, F and G), with comparable kinetics of LPS/5z7-induced cytotoxicity (Fig. 1G). The silencing effect was gene-specific and dependent on TRIF but not on myeloid differentiation primary response protein MyD88, reinforcing the role of cFLIP_L as a regulator of cell death downstream of TAK1 inhibition (fig. S2, B to F).

Ablation of TRIF and RIP1 kinase activity reversed the effect of cFLIP_L KD (Fig. 2A and fig. S3A). Annexin V and propidium iodide (PI) co-staining revealed the early appearance of exclusively PI⁺ cells in LPS-activated cFLIP_L-deficient BMDMs (Fig. 2B and fig. S3B), which suggested that the mechanism of cell death in LPS-activated cFLIP_L-deficient BMDMs differed from LPS/5z7-driven death. Unlike LPS/5z7-induced cell death, cFLIP_L deficiency-mediated cell death lacked CASP3, CASP7, or CASP9 activation (fig. S3, C and D). Instead, CASP1 and CASP11 were fully activated in LPS-stimulated cFLIP_L-deficient BMDMs (Fig. 2C and fig. S3C). cFLIP_L-deficient BMDMs stimulated with LPS exhibited robust CASP8 activation, and CASP1 and CASP11 cleavage was completely dependent on CASP8 (Fig. 2C and fig. S3, C, E, and F). Inhibition of CASP3 and CASP7 delayed death after LPS/5z7 treatment, but not in LPS-activated cFLIP_L-KD BMDMs (Fig. 2E and fig. S3G). Furthermore, LPS-driven death required CASP8 and GSDMD but not NLRP3, CASP1, or CASP11 (Fig. 2, E to G, and fig. S3, H to J). This suggests that cFLIP_L deficiency strictly promotes pyroptosis upon LPS activation. By contrast, both apoptosis and pyroptosis are activated in the context of LPS/5z7 (II, 12).

The cFLIP_L KD removed the requirement for TAK1 inhibition for the induction of pyroptosis. Similarly, YopJ was not required for *Yersinia*-induced death. YopJ-deficient *Yersinia pseudotuberculosis* (ΔyopJ) induced cell death and the cleavage of CASP8 and GSDMD in

cFLIP_L-KD BMDMs alone (Fig. 2, H and I, and fig. S3L). Thus, silencing of cFLIP_L recapitulates the effects of TAK1 inhibition (either by YopJ or 5z7 treatment), implicating cFLIP_L as one of the main regulators of pyroptosis in response to LPS.

cFLIP_L deficiency was sufficient to drive complex II formation in response to LPS. RIP1 and CASP8 recruitment to the FAS-associated death domain (FADD) occurred as early as 2 hours after LPS addition (Fig. 3, A and B). Notably, LPS/5z7 also drove complex II formation, and cFLIP_L was detected in the complex at early time points. LPS-induced complex II formation in cFLIP_L-KD cells was dependent on TRIF and RIP1 kinase activity (Fig. 3, C to F). LPS/5z7 induced phosphorylation of RIP1 at Ser¹⁶⁶ (S166) and promoted CASP3 binding in complex II. However, these modifications were not required for complex II formation and LPS-induced cell death in cFLIP_L-KD cells (Fig. 3, G to J). Thus, cFLIP_L and the kinase activity of RIP1 regulate complex II formation downstream of TRIF signaling and determine the extent and mode of cell death.

A single signal from LPS elicited robust IL-1β secretion in cFLIP_L-KD BMDMs (Fig. 4A and fig. S4A). Similar to death, the effect of cFLIP_L-silencing on IL-1β production was dependent on TRIF, CASP8, GSDMD, and the kinase activity of RIP1 (Fig. 4, B and C, and fig. S4, B and C). This implicates GSDMD as the sole effector of pyroptosis and IL-1β release in cFLIP_L-deficient BMDMs. To confirm inflammasome activation in response to LPS, we observed a substantially higher percentage of ASC speck positive cells in cFLIP_L-deficient BMDMs compared with wild-type (WT) (Fig. 4D and fig. S4, D and E). Furthermore, cFLIP_L-deficiency-driven ASC speck formation was CASP8-dependent but GSDMD-independent (Fig. 4E and fig. S4F), and NLRP3 and CASP1/11 deficiency abrogated IL-1β release in cFLIP_L-deficient BMDMs (Fig. 4F and fig. S4G). Thus, CASP8 plays a critical role in inflammasome activation and IL-1β maturation, whereas GSDMD is required for the release of IL-1β. Finally, LPS-induced IL-1β production in cFLIP_L-deficient BMDMs required potassium efflux, as excess extracellular potassium inhibited ASC speck formation and IL-1β release (fig. S4, H to J). Confirming the specificity of the short hairpin RNA-based approach, small interfering RNA-induced silencing of cFLIP_L resulted in death (fig. S3K) and IL-1β production (fig. S4K) in response to LPS alone, both of which were dependent on TRIF, CASP8, GSDMD, and the kinase activity of RIP1.

BMDMs deficient in cFLIP_L released IL-1β in response to ΔyopJ but not WT *Yersinia* (Fig. 4G and fig. S4L), further supporting LPS-induced IL-1β production (Fig. 2H). Notably, the same titer of ΔyopJ *Yersinia* elicited both IL-1β secretion (Fig. 4G) and cell death (Fig. 2H). These

data show a crucial role for cFLIP_L in regulating CASP8 activation and complex II formation, protecting macrophages against LPS-induced pyroptosis. Indeed, skewing toward the increased production of cFLIP_L confers resistance to LPS cytotoxicity in vivo (15). This underscores the importance of cFLIP_L as a key regulator of cell death and inflammation.

In macrophages, and perhaps in other cells, if levels of cFLIP_L are sufficiently high, CASP8 activation and pyroptosis are inhibited (Fig. 4H). When cFLIP_L levels are low, CASP8 homodimers form readily. Fully active CASP8 cleaves and activates distant targets, and LPS-activated macrophages rapidly undergo pyroptosis and secrete IL-1β. CASP3, CASP7, and CASP9 are dispensable for CASP8-driven pyroptosis in the absence of cFLIP_L. Instead, CASP8 likely directly activates GSDMD to drive pyroptosis and the NLRP3 inflammasome to drive IL-1β maturation and release.

REFERENCES AND NOTES

1. L. A. O'Neill, D. Golenbock, A. G. Bowie, *Nat. Rev. Immunol.* **13**, 453–460 (2013).
2. L. E. Reddick, N. M. Alto, *Mol. Cell* **54**, 321–328 (2014).
3. S. Mukherjee et al., *Science* **312**, 1211–1214 (2006).
4. N. Paquette et al., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12710–12715 (2012).
5. J. M. Park, F. R. Greten, Z. W. Li, M. Karin, *Science* **297**, 2048–2051 (2002).
6. T. Bergsbaken, S. L. Fink, B. T. Cookson, *Nat. Rev. Microbiol.* **7**, 99–109 (2009).
7. R. A. Aglietti et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7858–7863 (2016).
8. W. T. He et al., *Cell Res.* **25**, 1285–1298 (2015).
9. J. E. Vince, J. Silke, *Cell. Mol. Life Sci.* **73**, 2349–2367 (2016).
10. X. Liu et al., *Nature* **535**, 153–158 (2016).
11. J. Sarhan et al., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10888–E10897 (2018).
12. P. Orning et al., *Science* **362**, 1064–1069 (2018).
13. Y. Tsuchiya, O. Nakabayashi, H. Nakano, *Int. J. Mol. Sci.* **16**, 30321–30341 (2015).
14. T. M. Fu et al., *Mol. Cell* **64**, 236–250 (2016).
15. D. R. Ram et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1606–1611 (2016).

ACKNOWLEDGMENTS

We thank K. Fitzgerald and A. Degterev for sharing various mouse strains for this study. We thank A. Tai and the Tufts University Genomics Core for help with RNA sequencing and data analysis. We thank S. Bunnell for access to and advice for confocal imaging and K. Munger for access to the Amara Nucleofector System. **Funding:** This work was supported by NIH grants AI135369 and AI056234 to A.P. **Author contributions:** Conceptualization: A.P., H.I.M.; Validation: H.I.M.; Formal analysis: H.I.M.; Investigation: H.I.M., D.J., W.M.C., K.P.E., Z.M., I.S.; Writing: A.P., H.I.M.; Visualization: H.I.M.; Supervision: A.P.; Project administration: A.P.; Funding acquisition: A.P. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1379/suppl/DC1
Materials and Methods
Figs. S1 to S4
MDAR Reproducibility Checklist

18 June 2019; resubmitted 16 October 2019
Accepted 20 February 2020
10.1126/science.aay3878

CHEMICAL ENGINEERING

A sustainable wood biorefinery for low-carbon footprint chemicals production

Yuhe Liao^{1*}, Steven-Friso Koelewijn¹, Gil Van den Bossche¹, Joost Van Aelst¹, Sander Van den Bosch¹, Tom Renders¹, Kranti Navare², Thomas Nicolai³, Korneel Van Aelst¹, Maarten Maesen⁴, Hironori Matsushima⁴, Johan M. Thevelein³, Karel Van Acker^{2,5}, Bert Lagrain¹, Danny Verboeckend^{1†}, Bert F. Sels^{1,*}

The profitability and sustainability of future biorefineries are dependent on efficient feedstock use. Therefore, it is essential to valorize lignin when using wood. We have developed an integrated biorefinery that converts 78 weight % (wt %) of birch into xylochemicals. Reductive catalytic fractionation of the wood produces a carbohydrate pulp amenable to bioethanol production and a lignin oil. After extraction of the lignin oil, the crude, unseparated mixture of phenolic monomers is catalytically funneled into 20 wt % of phenol and 9 wt % of propylene (on the basis of lignin weight) by gas-phase hydroprocessing and dealkylation; the residual phenolic oligomers (30 wt %) are used in printing ink as replacements for controversial *para*-nonylphenol. A techno-economic analysis predicts an economically competitive production process, and a life-cycle assessment estimates a lower carbon dioxide footprint relative to that of fossil-based production.

Photosynthetic carbon capture by plant biomass, as evidenced by the global tree cover potential of 4.4 billion hectares of canopy, is likely to be among the most effective strategies for climate change mitigation (1). With an average annual production of ~10 metric tons of dry biomass per hectare (2), such nonedible biomass represents an abundant feedstock of renewable carbon worldwide and is a prime candidate to sustainably produce fuels, chemicals, and materials (3, 4). Climate change mitigation through global forest restoration has the potential to capture more than 200 billion tons of additional carbon at maturity, thereby reducing atmospheric carbon by about 25% (1). Together with the exploitation of underused biomass, reforestation will increase future lignocellulose availability and offers great potential for an abundant and inexpensive supply of renewable carbon, provided that production and processing are sustainable.

Petrochemicals are set to become the largest driver of global oil consumption in the future (5, 6). A shift from fossil to renewable carbon resources can decouple chemical production from fossil resources and the resulting CO₂ emissions. However, to be cost and environmentally competitive with fossil-based processes, it

is imperative to maximize feedstock use (7). Thus, there is a need for holistic biorefinery concepts that offer biomass valorization with low energy requirements and high carbon (and mass) efficiency, thereby providing existing and new markets with multiple products. The heterogeneous composition of lignocellulose, comprising entangled carbohydrate and lignin biopolymers, complicates its refining into value-added products. Strategies that extract high-value platform chemicals from lignin—a methoxylated phenylpropanoid biopolymer—are particularly challenging because of lignin's inherent recalcitrance and heterogeneity (8–11). Functionalized aromatics such as phenol, rather than hydrocarbons, are among the most suggested products of lignin conversion, but product yields on lignin weight basis are currently low (supplementary text ST1 and figs. S1 and S2).

To address this need, we propose an integrated biorefinery that simultaneously produces phenol, propylene, and useful phenolic oligomers from in planta wood lignin as well as a carbohydrate pulp that is amenable to bioethanol production (Fig. 1), thereby achieving a high carbon (and mass) efficiency. This work discloses the feedstock, process, and catalysis requirements (and challenges) and validates the techno-economic feasibility of producing (drop-in) chemicals (e.g., phenol and propylene) from lignin. We also demonstrate the application and value proposition of the phenolic oligomers.

The first step of our approach rests on a specific type of lignin-first biorefining, termed reductive catalytic fractionation (RCF) (12–16). RCF of lignocellulose yields a solid carbohydrate pulp and a lignin oil by the cleavage of ester and ether bonds as a result of tandem high-temperature solvolysis, hydrogenation,

and hydrogenolysis either in batch or in (semi-)continuous mode over a metal catalyst in the presence of a reducing agent, such as hydrogen. The general consensus is that stabilization of the reactive intermediates formed by depolymerization of in planta lignin prevents formation of unreactive condensed lignin derivatives (14). Near-complete delignification of hardwoods, such as birch and poplar, can be achieved without notable carbohydrate degradation (16). Besides low-molecular weight oligomers, the lignin oil contains few phenolic monomers in close-to-theoretical yields, namely 50 wt % for hardwoods (16). However, maximal valorization of lignin oil into high-value products, such as phenol, by technology that is profitable—but more importantly, sustainable—is key in demonstrating the potential of wood biorefineries.

The high yield of structurally similar phenolic monomers from the conversion of wood lignin prompted us to design a process for their transformation toward phenol and propylene by catalytic funneling (fig. S4 and supplementary text ST2). A typical composition of phenolic monomers (50.5 wt % on lignin weight basis; Fig. 2A and details in table S1) after RCF of birch wood in methanol over commercial Ru/C includes 4-*n*-propylguaiaicol (PG; 19 wt %) and 4-*n*-propylsyringol (PS; 67 wt %) as major components, and others including 4-ethylguaiaicol (EG) and 4-ethylsyringol. Pine wood gives a yield of 14.1 wt % of monomers because of a lower delignification and depolymerization efficiency. Although *para*-alkyl substituents are dominant in the monomers, considerably more polar groups, such as primary alcohols, remain in the oligomers (figs. S5 and S6, tables S2 and S3, and supplementary text ST3). This polarity difference facilitates the isolation of distinct monomers through a simple extraction in *n*-hexane under reflux (supplementary text ST4). This work demonstrates that a less-than-sixfold mass of *n*-hexane to lignin oil allows the cost-efficient extraction of more than 90 wt % of the lignin monomers (fig. S7). This procedure provides an optimum trade-off between extraction efficiency, solvent usage, and oligomer coextraction. Additional separations of individual phenolic monomers are not necessary because the crude monomeric extract can be catalytically funneled to the two products of interest, phenol and propylene.

To do so, the crude mixture of monomers was first chemo-catalytically hydroprocessed into *n*-propylphenols (PPs) and ethylphenols (EPs). In contrast to previously reported approaches using (batch) liquid-phase and/or sulfided catalysts on pure compounds (17–20), we pursued a solvent- and sulfur-free, continuous catalytic gas-phase hydroprocessing step. This procedure avoids product contamination as well as additional costs related to solvent loss and recovery. To establish the catalytic requirements

¹Center for Sustainable Catalysis and Engineering, KU Leuven, Celestijnenlaan 200F, 3001 Heverlee, Belgium.

²Department of Materials Engineering, KU Leuven, Kasteelpark Arenberg 44, 3001 Leuven, Belgium. ³Laboratory of Molecular Cell Biology, KU Leuven, and Center for Microbiology, VIB, Kasteelpark Arenberg 31, 3001 Heverlee, Belgium. ⁴Lawter bvba, Ketenislaan 1C, Haven 1520, 9130 Kallo, Belgium. ⁵Center for Economics and Corporate Sustainability, KU Leuven, Warmoesberg 26, 1000 Brussels, Belgium.

*Corresponding author. Email: bert.sels@kuleuven.be (B.F.S.); yuhe.liao@kuleuven.be or yuhe.liao20@gmail.com (Y.L.)

†Present address: Zeopore Technologies NV, Lelielaan 4, 3061 Bertem, Belgium.

for this selective hydroprocessing step, we initially studied commercially available PG, a representative monomer of the RCF lignin oil. The catalytic study explored several commercial metal catalysts (fig. S8). Non-noble metal Ni catalysts showed the highest PPs selectivity against other metals (figs. S8 and S9 and supplementary text ST5). Given the absence of a selectivity loss upon increased Ni content (Fig. 2, fig. S8B, and supplementary text ST5), highly loaded, well-dispersed Ni catalysts are preferred because of their high catalytic activity (Fig. 2B). Acidic supports (e.g., silica-alumina) led to more undesirable (propyl)cresols (*vide infra*), whereas redox-active supports (e.g., anatase TiO₂) favored fully deoxygenated products such as *n*-propylbenzene and *n*-propylcyclohexane (fig. S8). Therefore, Ni is preferably supported on inert materials such as silica (fig. S8 and supplementary text ST5). After optimization, a 64 wt % of Ni on silica (64 wt % Ni/SiO₂) catalyst reached 84% yield for PPs and EPs at a productivity of 4.5 kg kg⁻¹ hour⁻¹ (figs. S8 and S10 to S12 and supplementary text ST6). Side products included mainly *n*-propylbenzene and propylcresols and minor amounts of other compounds, such as cresols and *n*-propylanisole (fig. S9). Ni/SiO₂ (64 wt %) showed slight deactivation but without loss of selectivity after 72 hours at 285°C (Fig. 2D). The catalytic performance can be restored by a reduction treatment (fig. S13).

Next, we investigated the hydroprocessing of analytically pure representatives of lignin oil monomers other than PG, such as EG, isoeugenol, and PS (the most abundant monomer). For each compound, we observed high selectivity (75 to 85%) toward PPs and EPs at (near) complete conversion (Fig. 2E and fig. S14). Removal of both methoxy moieties in PS demanded a longer contact time at a higher temperature, achieving a selectivity for PPs and EPs of 77% at full conversion. This notable ver-

satility in substrates is pivotal to the concept of funneling and, hence, to the proposed lignin-to-phenol strategy, that is, maximal conversion of different methoxylated alkylphenols to phenol (and propylene or ethylene). Kinetic studies showed that PG and 3-methoxy-5-*n*-propylphenol were the key intermediates of PS hydroprocessing (Fig. 2E and fig. S14D). Furthermore, a detailed study on the dominant reaction pathways revealed the involvement of both demethoxylation and tandem demethylation-dehydroxylation pathways (figs. S15 to S18 and supplementary text ST7).

We ultimately moved to the hydroprocessing of a crude, unseparated mixture of monomers derived from RCF of pine and birch wood (Fig. 2E). At close-to-full conversion (>90%), the selectivity to PPs and EPs was similarly high for both crude monomer mixtures, yielding a quasi-identical products distribution compared with the reactions on pure compounds under the same conditions (Fig. 2E, fig. S19, and table S4). Thus, 64 wt % Ni/SiO₂ is robust to impurities (e.g., 4-methylsyringol) related to biomass feedstock. Gas chromatographic analysis showed that methoxy cleavage formed methane/H₂O and no CO/CO₂ (fig. S20). Analysis of the liquid condensate, obtained after condensing the gas-phase hydroprocessing products, confirmed the presence of mainly PPs and EPs with minor side products, such as (propyl)cresols and *n*-propylbenzene in addition to water (table S4 and fig. S21). This crude liquid condensate is used directly in the next dealkylation step without intermediate separation or purification.

We previously reported stable continuous gas-phase dealkylation of analytically pure alkylphenols (i.e., 4-*n*-propylphenol and 4-ethylphenol) to phenol and olefins over a commercial microporous ZSM-5 zeolite (21). Cofeeding of water was crucial to maintain robust catalytic activity (22), and hence the presence of water in

the liquid alkylphenol condensate—formed during hydroprocessing—is beneficial. Given the higher complexity of the crude alkylphenol stream (e.g., impurities and bulkier molecules; table S4), we anticipated that an identical commercial ZSM-5 would be inadequate because of site-access restriction and coke formation (fig. S22 and supplementary text ST8.1). To overcome these concerns, we developed a tailor-made hierarchical ZSM-5 (Z140-H) catalyst with a balanced network of micro- and mesopores (figs. S22 and S24 and table S5). With this catalyst, we observed near-quantitative and selective dealkylation of the crude alkylphenol condensates, giving a combined yield for phenol and olefins of 82% at high temperatures (Fig. 2, F and G, and figs. S25 and S26). We assessed the stability of Z140-H (deliberately at incomplete conversion) for biomass-derived crude alkylphenol streams (Fig. 2F and fig. S27). Side products were cresols, benzene, and trace amount of a few others (figs. S28 and S29 and Fig. 2G). Cresols after separation can be selectively converted to phenol over USY (rather than ZSM-5) through bimolecular reactions (fig. S30 and supplementary text ST8.2). Investigation of the product formation routes revealed the involvement of carbenium chemistry, including isomerization, disproportionation, transalkylation, and C-C cracking (fig. S31). Detailed kinetic studies (on 4-*i*MPMP, PPs, EPs, and *n*-propylbenzene) demonstrated that zeolite hierarchization is key for the activity and/or stability (figs. S32 to S35 and supplementary text ST8). Zeolites with large micropores, such as USY, although capable of converting sterically demanding alkylphenols, lack the (transition-state) pore confinement for shape-selective conversion. Confinement of the micropores, such as in Z140-H, is thus essential to achieve high selectivity.

This gas-phase technology enables the catalytic funneling of crude (unseparated) mixture

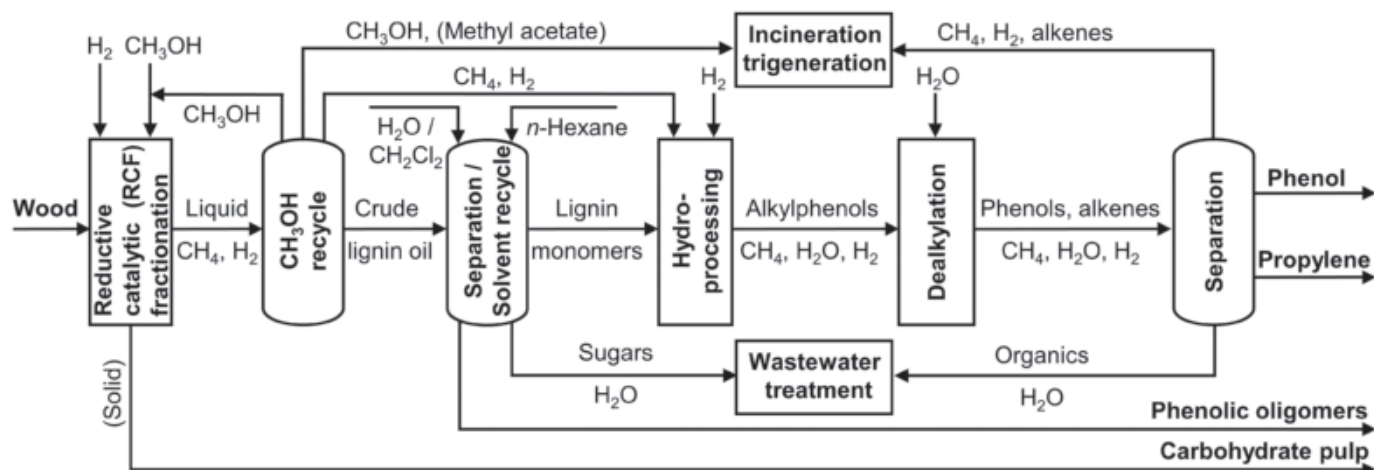


Fig. 1. Proposed integrated biorefinery process for xylochemicals production from wood. Flow diagram of the chemical process to produce carbohydrate pulp, phenol, propylene, and phenolic oligomers from wood.

of lignin monomers (extracted from the RCF lignin oil) into bio-phenol and bio-propylene with 20 and 9 wt % yields, respectively, on birch wood lignin weight basis (fig. S37). Using lignin monomers of pine wood yields lower amounts of phenol and propylene—6 and 3 wt %, respectively—because of a lower delignification and depolymerization efficiency of RCF with softwoods. Therefore, hardwoods, such as birch, are the preferred feedstock for producing phenol and propylene. The markets for these two (drop-in) xylochemicals are well established (fig. S1), so they can be directly supplied with renewable substitutes. Currently, the largest share of phenol flows into bisphenol A (BPA) production. Nonetheless, anticipating a future, post-BPA era, bio-phenol may be better employed for producing bio-aniline (via ammonolysis) and bio-caprolactam in exist-

ing facilities. Given the current uncertainty regarding final product purity, bio-propylene may be better suited to produce chemicals, such as isopropanol (fig. S1).

As mentioned above, RCF of birch also produces a carbohydrate pulp (65 wt % on wood weight basis, composed of <10 wt % of lignin, 60 wt % of cellulose, and 19 wt % of hemicellulose) and phenolic oligomers (30 wt % on lignin weight basis). To ferment both glucose and xylose, the carbohydrate pulp was subjected to a semisimultaneous saccharification-fermentation process, reaching a 40.2 g liter⁻¹ ethanol titer using an enzyme mixture for saccharification and an engineered yeast strain (MDS130) under nonoptimized conditions (Fig. 3A and supplementary text ST9). Note that the presence of Ru/C, originating from the RCF unit, was tolerated during this biological conversion.

Although we chose conversion of pulp into bioethanol for demonstration, other applications such as (news)paper, cardboard (23), insulation materials (24), and other chemicals (e.g., isosorbide, 2,5-furandicarboxylic acid, and 1-butanol) are possible as well.

A market for RCF phenolic oligomers (as obtained as a residue after extraction) is currently nonexistent. Still, these oligomers have a high functionality content (3.46 mmol phenolic OH g⁻¹ and 2.48 mmol aliphatic OH g⁻¹) (fig. S6 and table S3), and they lack the original phenolic interunit ether linkages (fig. S5 and table S2). To improve the overall profitability and sustainability of our proposed biorefinery, we investigated the potential of these oligomers to substitute for fossil-based *para*-nonylphenol [a debated endocrine disruptor (25)] in lithographic printing ink. Ink production typically

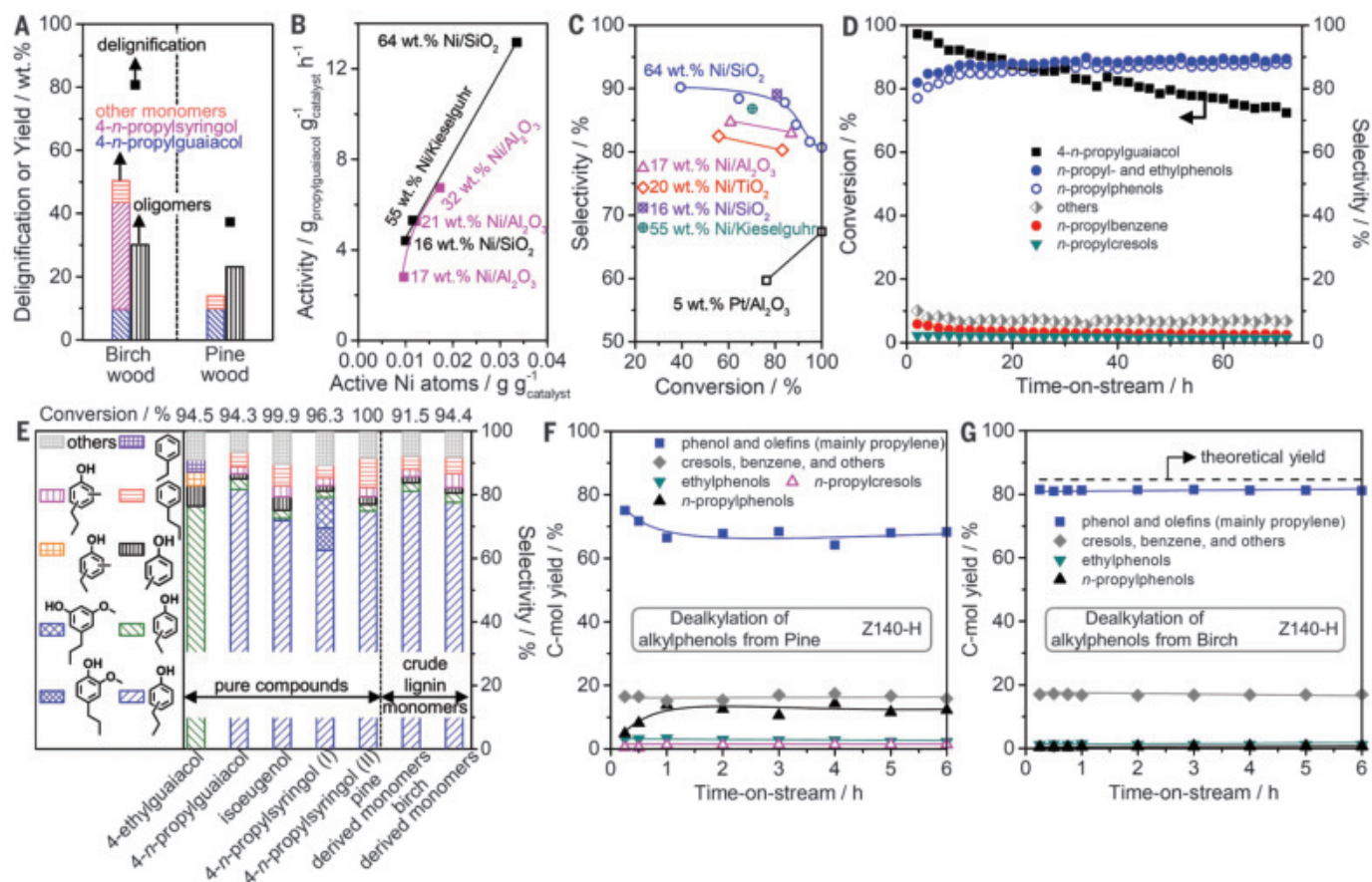


Fig. 2. RCF of wood and catalytic funneling of lignin monomers to phenol and propylene. (A) RCF of birch and pine wood to lignin monomer, oligomers, and carbohydrate pulp (details in supplementary materials). (B) Activity of selected supported Ni catalysts for hydroprocessing of PG (285°C with low conversion <20%, the data are taken at time-on-stream of 3 hours). (C) Selectivity to PPs versus PG conversion [285°C at different weight hourly space velocities (WHSVs)]. (D) Evolution of conversion and products selectivity with time-on-stream over 64 wt % Ni/SiO₂ for hydroprocessing of PG (285°C and 6.0 hour⁻¹ WHSV). (E) Hydroprocessing of different lignin-derived phenolics (over 64 wt % Ni/SiO₂: EG, PG, isoeugenol, and pine-derived monomers at 285°C and 8.2, 6.0, 4.4, and

6.0 hour⁻¹ WHSV, respectively; PS(I), PS(II), and birch-derived monomers at 305°C and 7.1, 5.3, and 5.3 hour⁻¹ WHSV, respectively). The data in (C) and (E) are taken at time-on-stream of 5 hours. Hydroprocessing constant reaction conditions: 1 bar of total pressure (0.4 bar of H₂ partial pressure). Dealkylation of the hydroprocessing products from extracted (unseparated) mixture of monomers of (F) pine and (G) birch wood lignin oils at 410°C over Z140-H with time-on-stream at WHSV of 3.7 hour⁻¹ and 2.8 hour⁻¹, respectively. C-mol yield in (F) and (G) represents the carbon molar yield in the product stream. The theoretical yield (84.7%) in (G) is the maximum combined yield of phenol and olefins on the basis of the substrate composition (table S4).

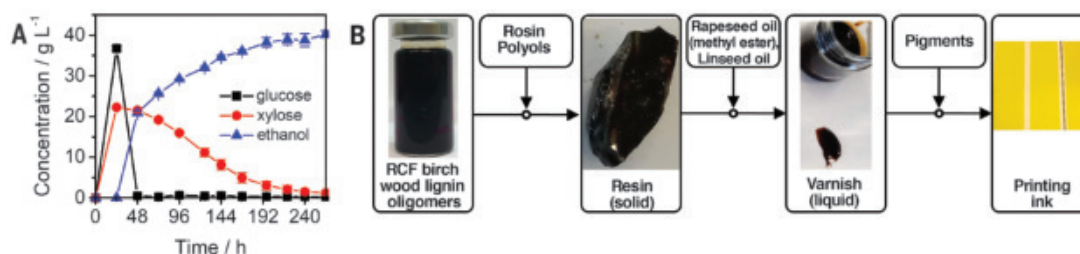


Fig. 3. Valorization of RCF birch wood carbohydrate pulp and phenolic oligomers. (A) Semisimultaneous saccharification-fermentation of carbohydrate pulp (containing Ru/C catalyst) obtained after RCF of birch wood. (B) Stepwise synthesis of ink from RCF birch wood lignin oligomers (details in supplementary materials).

involves a three-step procedure: (i) resin formation from rosin, polyols, and (nonyl)phenols; (ii) varnish production by adding rape- and linseed oil; and (iii) coloration by admixing pigments (Fig. 3B and figs. S38 and S39). The intermediate resin made from RCF birch wood lignin oligomers did meet industrial specifications, such as vacuum time and residue on filter (table S6, figs. S40 to S42, and supplementary text ST10). Next, the oligomer-based varnish formed a stable emulsion and showed similar water balance compared to *para*-nonylphenol-based as well as commercial resin-based ink varnish (table S7, figs. S43 and S44, and supplementary text ST10). Finally, yellow-colored lithographic printing ink was made by admixing the renewable RCF oligomer-based varnish with pigments (Fig. 3B). RCF oligomers outperformed other lignin derivatives, such as methanol-soluble birch wood lignin and commercial acetosolv spruce wood lignin. Substitution of nonylphenol with acetosolv spruce wood lignin failed because of phase incompatibility and the formation of observable black aggregates at the resin stage (supplementary text ST10). This case study underlines the unexplored market potential of RCF phenolic oligomers in high-quality printing ink, in which they could serve as a renewable substitute for fossil-based nonylphenol.

On the basis of the experimental data, we designed a process model to perform a techno-economic analysis (TEA) (Fig. 1 and fig. S45). The process model integrates the three catalytic steps: (i) RCF of wood, (ii) hydroprocessing of crude monomers extract, and (iii) dealkylation of the crude alkylphenol product stream. In the first catalytic step, RCF of birch wood produces a carbohydrate pulp and a lignin oil, the latter of which is obtained by liquid-solid separation and solvent recuperation. From the lignin oil, monomers are readily isolated in a liquid *n*-hexane extraction unit, followed by flash distillation to remove *n*-hexane. The crude monomers extract and the RCF off-gas—containing methane, which originates from the limited MeOH conversion in RCF, and H₂—are fed to the second catalytic step. This gas-phase fixed-bed reactor contains the hydroprocessing catalyst, Ni/SiO₂, to yield alkylphenols. In the third catalytic step, this crude alkylphenol mix-

ture, containing water, hydrogen, and methane impurities, is fed without intermediate purification to the second fixed-bed reactor. This setup contains the dealkylation catalyst (Z140-H) to yield phenol and olefins. The presence of the remaining hydrogen had no effect on the olefin formation (fig. S36). Next, product separation in a gas-liquid separator produces a liquid phenol stream and a gaseous mixture of water, olefins, H₂, and CH₄. Finally, to obtain high purity phenol and propylene, impurities such as cresols and benzene (in the phenol fraction) and H₂/CH₄ (in the olefin fraction) can be removed by distillation. In this model, side streams related to sugar solubilization (during RCF) and benzene and cresols formation end up in a wastewater stream. Methyl acetate, formed by methanolysis of the acetyl groups in (birch wood) hemicellulose, is largely separated in the methanol recovery distillation. Together with the excess H₂, CH₄, C₂H₄, and small amounts of methanol (also from distillation), methyl acetate is incinerated to provide heating, cooling, and electricity through a trigeneration system. The addition of external energy is not required to operate the integrated biorefinery. Overall, this process model design converts 1000 kg of birch wood into 653 kg of raw carbohydrate pulp (for bioethanol), 64 kg of lignin oligomers (for printing ink), 42 kg of phenol, and 20 kg of propylene (>99%), which corresponds to a conversion of 78 wt % of the initial biomass into targeted products (figs. S46 and S47 and table S8). Possible solvent losses were studied, indicating a maximum loss of 1.4% of methanol due to (i) distillation, (ii) hydrogenolysis during RCF, and (iii) incorporation into products (supplementary text ST11).

The TEA of our proposed biorefinery was calculated for an annual production of 100 kilotons of bio-phenol (i.e., the average scale for fossil-based phenol production). Among the different process units, RCF and incineration-trigeneration are the highest contributors toward capital expenditures because of the high cost of pressure reactors and energy integration, respectively (fig. S48 and supplementary text ST12). Investing in an incineration-trigeneration unit is justified, however, by its positive effect on the manufacture cost because of markedly

reduced energy costs. The highest contribution to the manufacturing cost is the cost of feedstock (birch wood, 158 euros per ton; tables S9 and S10). Given the current pricing (table S9) of phenol (1300 euros per ton), propylene (830 euros per ton), and crude pulp (400 euros per ton), and using an estimate for the oligomers (1750 euros per ton, approaching that of nonylphenol), this results in an internal rate of return of 23% and a payout time of ~4 years for a plant with a lifetime of 20 years (table S11). A sensitivity study indicates that feedstock and product pricing have the largest economic effect (fig. S49 and supplementary ST12), whereas the influence of catalyst cost is negligible as long as the catalyst is sufficiently recyclable or reusable. In terms of RCF process parameters, shorter contact times and higher biomass concentrations are crucial factors to improve the profitability of this biorefinery, which implies the need to design a dedicated reactor.

The production of chemicals from biomass makes sense only if a lower CO₂ footprint is achieved. Thus, in addition to a TEA, we performed a life-cycle assessment (LCA). Our proposed integrated birch wood biorefinery showed reduced global warming potentials (GWPs) for phenol (0.736 kg of CO₂-equivalent per kilogram of phenol) and propylene (0.469 kg of CO₂-equivalent per kilogram of propylene) compared with their fossil-based counterparts (1.73 and 1.47 kg of CO₂-equivalent per kilogram of phenol and propylene, respectively; open and red symbols in Fig. 4, A and B, supplementary text ST13, and tables S12 to S14). Moreover, the GWP of the oligomers (proposed as a substitute for *para*-nonylphenol with a GWP of >1.58 kg CO₂-equivalent per kilogram of nonylphenol) and the carbohydrate pulp were calculated to be -0.949 and -0.217 kg of CO₂-equivalent per kilogram of oligomers and carbohydrate pulp, respectively (open symbols in Fig. 4, A and B). These negative values indicate a net consumption of CO₂, that is, a net carbon-capturing effect for their production. Finally, to indicate opportunities for sustainability improvement, additional scenarios were analyzed, such as, (i) the substitution of nonrenewable H₂, which has a high CO₂ contribution, by renewable H₂ and (ii) the inclusion of more sustainable forest

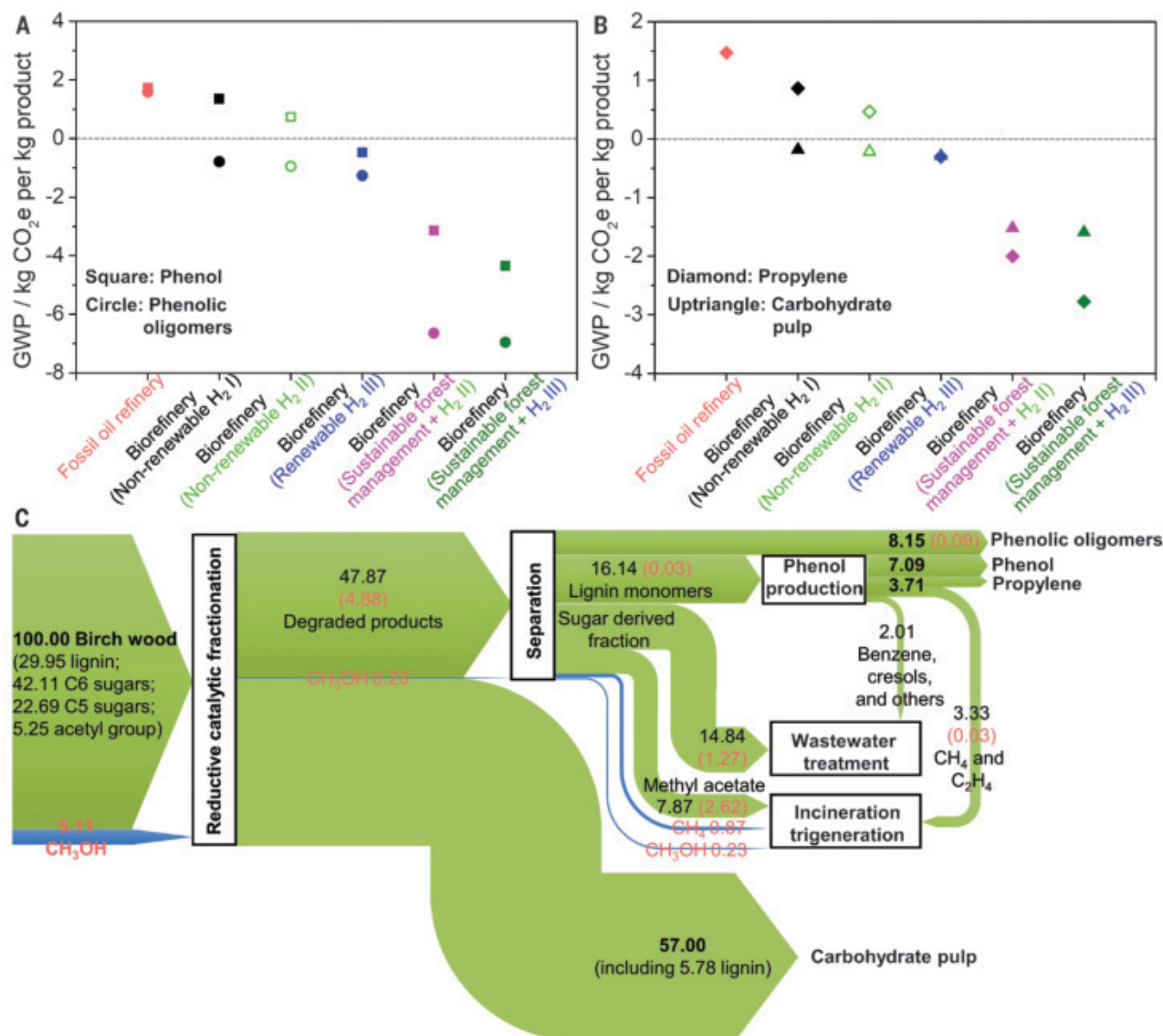


Fig. 4. LCA and carbon flow for the proposed integrated biorefinery based on birch wood. (A and B) GWPs [in kilograms of CO₂-equivalent per kilogram of product (kg CO₂e per kg product)] of phenol, propylene, phenolic oligomers, and carbohydrate pulp in this birch wood biorefinery with different scenarios (i.e. several hydrogen sources and/or forest management strategies). The GWPs

are 11.89, 8.20, and 0.97 kg of CO₂-equivalent per kilogram of H₂ for nonrenewable H₂ I, nonrenewable H₂ II, and renewable H₂ III, respectively (see supplementary materials). The GWP of phenolic oligomers from oil refining is the GWP of fossil-based nonylphenol (>1.58 kg of CO₂-equivalent per kilogram of nonylphenol). (C) Carbon flow of this birch wood biorefinery.

management (Fig. 4 and fig. S52). Such integration scenarios unveil the possibility for CO₂-neutral wood biorefining with a total net consumption of CO₂ (i.e., negative GWP values) for each targeted product.

On the basis of the proposed integrated biorefinery, 78 and 76%, respectively, of the initial mass and carbon content of birch wood can be economically and sustainably valorized into four high-value end-products, namely phenol, propylene, oligomers, and pulp (Fig. 4C and fig. S47). In our opinion, this process will constitute a clear incentive to make profitable, renewable, low-carbon footprint chemicals via the holistic biorefining of sustainable wood.

REFERENCES AND NOTES

1. J.-F. Bastin *et al.*, *Science* **365**, 76–79 (2019).
2. K. Van Meerbeek, B. Muys, M. Hermy, *Renew. Sustain. Energy Rev.* **102**, 139–149 (2019).
3. T. Werpy, G. Petersen, "Top value added chemicals from biomass: Volume I—Results of screening for potential candidates from sugars and synthesis gas" (National Renewable Energy Laboratory, 2004); www.osti.gov/biblio/15008859.
4. J. E. Holladay, J. F. White, J. J. Bozell, D. Johnson, "Top value-added chemicals from biomass: Volume II—Results of screening for potential candidates from biorefinery lignin" (Pacific Northwest National Laboratory, 2007); www.osti.gov/biblio/921839.
5. A. Kätelhön, R. Meys, S. Deutz, S. Suh, A. Bardow, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11187–11194 (2019).
6. International Energy Agency, "The future of petrochemicals: Towards more sustainable plastics and fertilisers" (IEA Publications, 2018); www.connaissancesenergies.org/sites/default/files/pdf-actualites/the_future_of_petrochemicals.pdf.
7. D. M. Alonso *et al.*, *Sci. Adv.* **3**, e1603301 (2017).
8. A. J. Ragauskas *et al.*, *Science* **344**, 1246843 (2014).
9. C. Li, X. Zhao, A. Wang, G. W. Huber, T. Zhang, *Chem. Rev.* **115**, 11559–11624 (2015).
10. Z. Sun, B. Fridrich, A. de Santi, S. Elangovan, K. Barta, *Chem. Rev.* **118**, 614–678 (2018).
11. J. G. Linger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12013–12018 (2014).
12. I. Kumaniaev *et al.*, *Green Chem.* **19**, 5767–5771 (2017).
13. E. M. Anderson *et al.*, *Joule* **1**, 613–622 (2017).
14. W. Schutyser *et al.*, *Chem. Soc. Rev.* **47**, 852–908 (2018).
15. Q. Song *et al.*, *Energy Environ. Sci.* **6**, 994–1007 (2013).
16. S. Van den Bosch *et al.*, *Energy Environ. Sci.* **8**, 1748–1763 (2015).
17. A. Vuori, J. B. Bredenberg, *Holzforschung* **38**, 253–262 (1984).
18. N. Joshi, A. Lawal, *Ind. Eng. Chem. Res.* **52**, 4049–4058 (2013).
19. H. L. Chum, D. K. Johnson, S. Black, M. Ratcliff, D. W. Goheen, in *Advances in Solar Energy: An Annual Review of Research and Development*, K. W. Böer, Ed. (Springer, 1988), pp. 91–200.

20. X. Liu *et al.*, *ChemSusChem* **12**, 4791–4798 (2019).
21. D. Verboekend, Y. Liao, W. Schutyser, B. F. Sels, *Green Chem.* **18**, 297–306 (2016).
22. Y. Liao *et al.*, *ACS Catal.* **8**, 7861–7878 (2018).
23. H. Sixta, *Handbook of Pulp* (Wiley, 2006).
24. S. Schiavoni, F. D'Alessandro, F. Bianchi, F. Asdrubali, *Renew. Sustain. Energy Rev.* **62**, 988–1011 (2016).
25. A. Soares, B. Guieysse, B. Jefferson, E. Cartmell, J. N. Lester, *Environ. Int.* **34**, 1033–1049 (2008).

ACKNOWLEDGMENTS

We thank W. Vermandel for his assistance in producing lignin oil and R. Ooms and J. Maes for technical assistance during the catalytic testing. **Funding:** Y.L., G.V.d.B., D.V., J.V.A., and B.L. acknowledge funding from the China Scholarship Council (201404910467), Catalisti-ICON project MAIA (Flemish government), Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) (postdoc), Flanders Innovation & Entrepreneurship (innovation mandate - postdoc), and the Industrial Research Fund KU Leuven (IOF fellow), respectively. S.-F.K. acknowledges Catalisti cSBO ARBOREF and BIO-HArT (Interreg VA Vlaanderen-Nederland) funding from the Flemish government. K.V.Ae. acknowledges funding from FWO-SBO project Biowood (S003518N). T.R.

acknowledges KU Leuven internal research funds for a postdoctoral mandate (PDM). S.V.d.B. acknowledges funding from KU Leuven internal research funds for a PDM, the Flemish government for the FWO-SBO project Biowood, and Flanders Innovation & Entrepreneurship (innovation mandate - postdoc). T.N. and J.M.T. acknowledge ARBOREF. Funding by BIOFACT (FNRS – FWO EOS project G0H0918N), supporting lignin conversion, is also acknowledged. **Author contributions:** Y.L. and B.F.S. conceived the idea and designed the experiments. Y.L. carried out the experimental work of the main catalysis research, and catalyst characterization and interpretation was assisted by D.V. S.-F.K. performed the liquid-liquid extraction separation work. G.V.d.B. and T.R. performed and interpreted the wood composition analysis and the processing experiments, according to the lignin-first concept, and the preparation and analysis of a large batch of isolated lignin-first phenolic monomers. J.V.A., S.V.d.B., and B.L. composed and focused on the TEA with the kind assistance of Exyte. K.N. and K.V.Ac. performed the LCA. T.N. and J.M.T. performed the fermentation of the lignin-first carbohydrate pulp to bioethanol. K.V.Ae. conducted the characterization and analysis of phenolic oligomers. M.M. and H.M. performed the production and characterization of the renewable resin and ink varnish. The text was initially composed by B.F.S., Y.L., S.-F.K., and

G.V.d.B., and all authors further contributed to the discussion of the experimental work and the final version of the manuscript.

Competing interests: Y.L., B.F.S., J.V.A., and S.V.d.B. are inventors on a patent application [attorney docket number 292-P15233US (ZL919134)], held and submitted by KU Leuven, that covers lignocellulose refinery. **Data and materials availability:** All data needed to support the conclusions of this manuscript are included in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1385/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S53
Tables S1 to S14
References (26–82)

21 May 2018; resubmitted 17 October 2019
Accepted 4 February 2020
Published online 13 February 2020
10.1126/science.aau1567



UCSD Biophysicist Elizabeth Villa uses a cryo-EM technique, cryo-ET, to visualize proteins in situ.

Democratizing cryo-EM: Broadening access to an expanding field

Cryo-electron microscopy (cryo-EM) yields atomic-level structures of megacomplexes and tiny compounds. How can your lab get access to this versatile method? By Chris Tachibana

For decades, says Bridget Carragher, cryo-EM was a “niche, hole-in-the-wall” field. But in 2017, cryo-EM passed nuclear magnetic resonance (NMR) spectroscopy for number of annual entries in the Protein Data Bank, the world’s sole repository for 3D structural data on proteins, nucleic acids, and large biological molecules. And now it’s gaining on the granddaddy of structural methods, X-ray crystallography.

Carragher leads a cryo-EM facility at the **New York Structural Biology Center**, which is supported by the U.S. National Institutes of Health (NIH) and the Simons Foundation. Two other NIH-funded centers are at **Stanford University** and **Oregon Health & Science University (OHSU)**. “The trend everywhere is for national cryo-EM facilities,” says Poul Nissen, structural biologist at **Aarhus University**, which is the Danish national facility, together with the University of Copenhagen.

National centers serve a cryo-EM community that is rapidly expanding as software and hardware breakthroughs, especially in electron detectors, demonstrate how cryo-EM can advance basic research, drug development, and even solar-cell technology (1,2).

Crystallographic resolution—without crystals, but at a cost

Unlike X-ray crystallography, cryo-EM does not require crystallized samples. This eliminates a time-consuming step and allows atomic-level reconstructions of lumpy complexes and integral membrane proteins that have resisted crystallization. It can show conformational changes, such as ribosomes flexing their structure as they go through protein synthesis (3).

Cryo-EM works with unstained, aqueous samples. For single-particle analysis (SPA), its most common application, researchers drop samples onto a grid that is flashcooled by being plunged into liquid ethane. This freezing—or rather

vitrifying—is so rapid that sample molecules are immobilized with their structure preserved and without ice crystals that interfere with transmission electron microscopy (TEM). Researchers then take thousands of TEM images by beaming electrons through the sample. Molecules caught in random orientations scatter the electrons, creating patterns used to generate 3D models.

Craig Yoshioka, codirector of the NIH cryo-EM center at OHSU, points out a promising development: Crystallographers who had truncated or mutated proteins to coax them into crystals can now study full-length wildtype proteins using cryo-EM. “This should better represent targets in their native states,” he says, “including with posttranslational modifications like glycosylation.”

Currently, SPA works best with large samples around 200 kDa, so researchers with smaller proteins might turn to microcrystal electron diffraction (microED), a cryo-EM method with a larger size range. Another issue with SPA is that it uses cell extracts; but inside cells, says **University of California, San Diego** biophysicist **Elizabeth Villa**, “proteins aren’t floating in water. They’re packed with other components, interacting with them, or forming networks that break up during extraction.” Villa uses cryo-electron tomography (cryo-ET), which images sections of cells or even tissues, to visualize components in situ.

And cryo-EM has an overarching drawback: cost. Top-of-the-line, 300-kiloelectron volt (keV) cryo-EM machines are around USD 5–7 million, with added costs for space, service contracts, and experienced staff. Pharmaceutical companies may have in-house facilities or use a company like **Nanolmaging Services**. Most cryo-EM clients are from pharma or biotech, says Carragher, a cofounder. Example projects include analyzing vaccines, antibodies, and drug targets. The company is rare among cryo-EM contractors in owning its own equipment, with others often using instruments at partner institutes.

Major research institutions also invest in cryo-EM facilities, but smaller universities can’t afford them. However, scientists including Gabriel Lander’s group at Scripps Research have revealed single-angstrom (Å) structures of proteins using less-powerful 100-keV or 200-keV microscopes **cont. >**

Upcoming features

Genetics—October 9 ■ Immuno-Oncology—November 20

(4) that cost USD 1–2 million. These results encourage scientists who call for democratizing cryo-EM with more affordable, workhorse instruments (5).

Community service

Cryo-EM access should increase thanks to the new NIH centers, which have cutting-edge equipment and a focus on service and training—center personnel are not allowed to be coauthors on users' publications. At full capacity, the OHSU site will be collecting data 24/7 on 200–300 active projects at a time and training 50-plus visiting scientists a year, Yoshioka estimates. He expects up to hundreds of reconstructions per year per center.

And services are free. "You write a proposal," Carragher says, "and if it's accepted based on criteria, such as scientific merit, feasibility, and need, you get cryo-EM time." This model is similar to national synchrotron facilities, and many, such as the United Kingdom's Electron Bio-Imaging Centre at the Diamond Light Source (Oxfordshire) and the **Brazilian Nanotechnology National Laboratory** (LNNano) of the Brazilian Center for Research in Energy and Materials (Campinas), are located close to synchrotrons.

LNNano is the only cryo-EM facility in Latin America, and is supported by government and State of São Paulo funding. Industrial clients are charged for services, but service and training are free for academic researchers after project evaluation, says LNNano researcher Rodrigo Portugal.

LNNano Senior Scientist Marin van Heel says cryo-EM is a powerful tool for structure-based drug and vaccine design, so it is essential in the region because of "big needs, like in neglected diseases such as Zika." SPA research is underway at LNNano with collaborators in Brazil, Peru, Uruguay, and Argentina.

Besides cost, the major burden at LNNano facilities is brain drain. Despite holding multiple workshops and the annual Brazil School for Single Particle Cryo-EM, "people get headhunted away to a center or pharma company in another country," van Heel says.

"It's THE issue"

Cryo-EM software and hardware have "advanced amazingly," Yoshioka says, "but it can still be difficult to reliably take any protein from a gene to a structure." Cryo-EM doesn't require large crystals, but sample purity, heterogeneity, and concentration are still important.

"Sample prep isn't an issue," says Carragher, "it's THE issue." During vitrification, "particles glue themselves together, stick to the air–water interface, adopt sulky conformations, or fall apart." Commercial automated systems make sample preparation more reliable. However, a downstream challenge is caused by terabytes of data that require dedicated workstations.

Digital developments

Cryo-EM users uniformly praise software groups for advancing data analysis and structure resolution. Open-source software, such as RELION from Sjors Scheres at the UK's Medical Research Council Laboratory of Molecular Biology, and work by others including Niko Grigorieff at Janelia Research Campus and the University of Massachusetts, have been instrumental to the field, says Yoshioka. An up-and-coming computational advance, he notes, is real-time processing and reconstruction as data are collected.

That's what cryoSPARC Live does. Currently in beta testing, the software comes from University of Toronto spinout **Structura Biotechnology**, run by brother-and-sister team Ali Punjani and Saara Virani. CryoSPARC Live adds to the cryoSPARC package of SPA tools, including 2D image curation and 3D reconstruction without prior structural knowledge.

CryoSPARC Live, Virani says, shows initial images after a few minutes, 6-Å to 8-Å 3D structures in about an hour, and refined high-resolution structures a few hours later. Researchers can make real-time adjustments, such as moving the sample to focus on the best areas and deciding how much data to collect, saving time and money, she says. With demand for cryo-EM growing rapidly, the field is wrestling with commercialization issues. Punjani explains that cryoSPARC is free for academic users, while commercial clients such as pharma companies must buy a license.

A computational angle on cryo-EM democratization, Punjani says, is to modify algorithms to get better images from lower-end microscopes. Also, cloud-hosted computation would let labs rent processing time as needed instead of investing in dedicated hardware.

Full tilt on innovations

"Single-particle will be the bread-and-butter method for high-resolution cryo-EM for a while," Yoshioka says. Advances in other areas extend the size range for resolving structures and allow views of the cell interior.

Getting high-resolution images of proteins smaller than 100 kDa pushes the limit of current SPA. MicroED, developed by **Tamir Gonen's** group, achieves atomic resolution for size ranges of complexes larger than 200 kDa to organic molecules under 10 carbon atoms. MicroED uses crystals one-billionth the size needed for X-ray crystallography, explains Gonen, now at the **University of California, Los Angeles** (UCLA). In microED, vitrification protects samples so that diffraction patterns are generated by rotating a single microcrystal through an electron beam, capturing all angles for 3D reconstruction of its molecules.

Gonen used microED to visualize structural changes in a channel as a sodium ion passed through. "Because we used crystals containing only about 1,000 units," he says, "we could tease out smaller differences and capture a transition state" (6).

Medicinal chemists, forensic scientists, and drug developers are excited about the "powder-to-structure" application of microED. Gonen's group and others published methods for 30-min identification of small molecules such as ibuprofen or biotin by structure, including in mixtures (7).

Gonen has worked with **Thermo Fisher Scientific** to develop relatively easy-to-use microED hardware and software. "You don't need to know much now to get a sample into a 'scope and collect data. It could make microED more available to the community," he says. Steve Reyntjens, Thermo Fisher's director of product marketing, says the microED package is easy to add as an optional item on new microscopes or as a retrofit to existing instruments.

The David Geffen School of Medicine at UCLA has a microED center that works with academic and industry scientists and offers microED training, including at an annual summit coming up in October 2020 (8).

Cryo-ET reveals cellular contents not as they appear in textbooks or videos "with empty space, a particle, then empty space," says Villa. It shows cells jam-packed with molecules. **cont. >**

Cryo-EM microscopes help scientists visualize biological molecules at an atomic scale, such as this LRRK2 protein in Parkinson's disease.

Featured participants

Aarhus University
international.au.dk

Brazilian Nanotechnology National Laboratory
lnnano.cnpem.br

Gonen Lab, University of California, Los Angeles
cryoem.ucla.edu

Nanolmaging Services
www.nanoimagingsservices.com

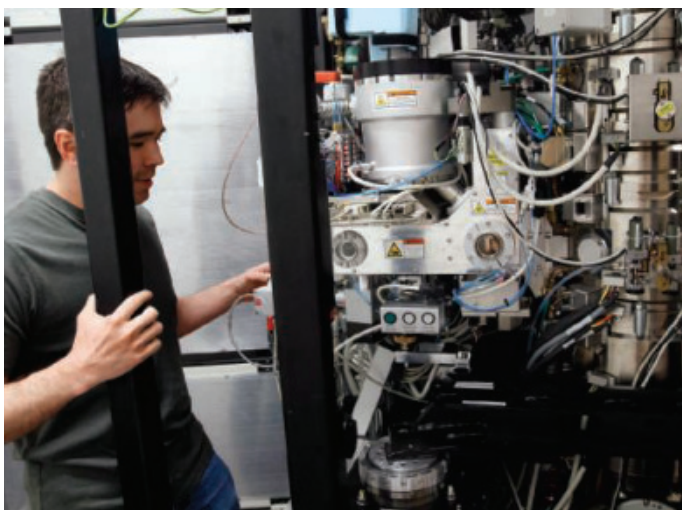
New York Structural Biology Center
nysbc.org

Oregon Health & Science University
www.ohsu.edu

Structura Biotechnology
structura.bio

Thermo Fisher Scientific
www.ThermoFisher.com

Villa Lab, University of California, San Diego
villalab.ucsd.edu



Craig Yoshioka, codirector of the NIH cryo-EM center at Oregon Health & Science University, stands in front of one of his facility's cryo-EM instruments.

Instead of analyzing images of extracted proteins or complexes as SPA does, cryo-ET collects data on one cell, getting multiple views by tilting it on a grid. Averaging the images in a process similar to SPA results in 3D reconstructions.

TEM requires samples thinner than most cells, so researchers like Villa pair cryo-ET with focused ion-beam (FIB) micromachining. Villa credits Mike Marko at the New York State Department of Health for showing that this typical materials science method is feasible for vitrified samples. As a postdoc in Wolfgang Baumeister's lab at the Max Planck Institute of Biochemistry in Munich, Germany, Villa helped advance the technique.

Cryo-FIB uses a special cryo-EM instrument that aims a beam of large ions, such as gallium, across a vitrified sample to plane it down. "You blast it on the top and bottom until you have a window sustained by the sides of the cell that is thin enough for cryo-ET," Villa says. If a cell's tomography shows multiple copies of a protein or complex, subtomogram averaging of individual images of that molecule can reconstruct its 3D structure.

Villa's lab added correlative light and electron microscopy (CLEM) to cryo-ET to determine the 14-Å structure of LRRK2. The researchers tagged the protein, which is involved in Parkinson's disease, to locate it in cells using CLEM. Then they used FIB to mill down the cells for cryo-ET to obtain in situ 3D structures of LRRK2, including its close association with microtubules (9). LRRK2 can't be crystallized, "but with cryo-ET and subtomogram averaging, we solved the structure while it was still in the cell," Villa says.

According to Villa, improvements ahead for cryo-ET include developing specialized TEM grids for growing cells before inducing a change, such as stimulating neurons or exposing human cells to medicines. Researchers using

light microscopy could choose when to vitrify the sample for cryo-ET, Villa says, "to see at high resolution what happened at the point you did something to the cell."

Sample preparation for cryo-ET is low throughput, but a variation of CLEM with multiplexed fluorescent markers, developed by John Briggs' group at the European Molecular Biology Laboratory in Germany, could more quickly identify cells for cryo-ET. Cryo-ET will allow observation of molecules in their native environment and in whole tissues, such as molecular views of "the connectome" of neuron-to-neuron interactions, Nissen says. The Thermo Fisher second-generation FIB instrument, notes Reyntjens, has cryo-liftout capability for manipulating miniscule samples cut from vitrified tissue thinned to 100 nm–150 nm and transferred to a cryo-EM instrument for tomography.

NIH will soon launch national cryo-ET centers. The current cryo-EM centers, Carragher explains, collect tomograms if users have samples ready for cryo-ET. The national centers will provide access to equipment, plus assistance with tricky cryo-ET specimen preparation.

In addition to developments like CLEM, structural analysis that combines cryo-EM with data from multiple sources is on the rise. "Increasingly, people use crystallography, NMR, CLEM, mass spec—everything out there—to get an answer," Carragher says. "But if we want these tools in everyone's toolkit, they need to be more accessible."

New ambitions

Along with solving the access problem, Nissen observes, the field should shift its perspective from focusing only on structures to "asking what the structure is doing in the cell in its native state. Getting label-free, time-resolved structures in natural contexts is the ultimate goal and also a new level of ambition to instill in students and postdocs."

Nissen and others predict increasing industry use of cryo-EM for developing antibody therapeutics, small molecule drugs, and diagnostics. "We should also work with the medical community on unmet diagnostic needs," he says, "where histology doesn't show good differences between disease and healthy tissue. We might find molecular differences in tissues [by also] using cryo-ET."

van Heel, who helped develop cryo-EM and has watched its use grow, says about working in the field, "It's challenging at the moment, but it's a great time to be alive. There's no time for vacation."

References

1. Y. Cheng, *Science* **361**, 876–880 (2018), doi: 10.1126/science.aat4346.
2. A. A. Herzog, D. M. DeLongchamp, *Matter* **1**, 1106–1107 (2019), <https://doi.org/10.1016/j.matt.2019.10.015>.
3. J. Frank, *Curr. Opin. Struct. Biol.* **22**, 778–785 (2012), <https://doi.org/10.1016/j.sbi.2012.08.001>.
4. M. A. Herzik, M. Wu, G. C. Lander, *Nat. Comm.* **10**, 1032 (2019), <https://doi.org/10.1038/s41467-019-08991-8>.
5. E. Hand, *Science* **367**, 354–358 (2020), <https://doi.org/10.1126/science.367.6476.354>.
6. S. Liu, T. Gonen, *Commun. Biol.* **1**, 38 (2018), <https://doi.org/10.1038/s42003-018-0040-8>.
7. C. G. Jones *et al.*, *ACS Cent. Sci.* **4**, 1587–1592 (2018), <https://doi.org/10.1021/acscentsci.8b00760>.
8. UCLA First Annual MicroED Summit, <https://cryoem.ucla.edu/summit>.
9. R. Watanabe *et al.*, *bioRxiv* 1–22 (2019), <https://doi.org/10.1101/837203>.

Chris Tachibana is a freelance writer who specializes in life sciences.

new products: microscopy

**Cryo-EM Pucks**

The Cryo-EM Puck basic lab starter kit is ideal for those just starting to use cryo-electron microscopy (cryo-EM) and for smaller projects or labs. The kit helps with organizing, tracking, and shipping your samples on grids for single-particle analysis and other cryo-EM applications. It includes 10 MiTeGen 2nd Generation Cryo-EM Pucks; one cane for storing your pucks; one cane for shipping them; one grasping tong for

handling them safely and securely; and Cryo-EM Puck barcoding and serialization for organizing and tracking your samples. Our kits can be customized as needed.

MiTeGen

For info: 877-648-3436
www.mitegen.com

Imaging Filter

The BioContinuum imaging filter combines high-performance post-column energy filtration with direct detection technology. Because the filter is ideally suited for low-dose imaging applications, researchers can maximize their cryo-EM and cryo-electron tomography (cryo-ET) capabilities to gain further insight into system function and disease progression at the molecular level. The filter offers a 24-megapixel (5,760 × 4,092) field of view and works at 1,500 full frames per second. The BioContinuum comes with an optional, inline, GPU-based motion correction that saves researchers the need to save terabytes of raw frames.

Gatan

For info: 925-463-0200
www.gatan.com

Cryo-Electron Microscope

JEOL announces the CRYO ARM 300 Field Emission Cryo-Electron Microscope for automatic, unattended acquisition of image data for single-particle analysis. The system is composed of an autoloading specimen stage (of up to four grids) cooled to liquid nitrogen temperatures and a cryo-storage device for long-term storage of up to 12 frozen-hydrated specimens. A cold field-emission gun produces a high-brightness electron beam with a very small energy spread, thus offering high temporal coherency. Equipped with an improved in-column energy filter (Omega filter), the CRYO ARM 300 acquires energy-filtered images and energy-loss spectra. The CRYO ARM 300 incorporates JEOL Automated Data Acquisition System (JADAS) automated software for single-particle analysis workflows. The hole-free phase plate consists of a thin, continuous carbon film placed in the back focal plane of the objective lens, providing

substantial improvement in image contrast of frozen-hydrated specimens.

JEOL

For info: 978-535-5900
www.jeolusa.com

Cryo-EM Kit

Our "Introduction to Cryo-EM" Consumables Kit provides the basic requirements for getting started with grid preparation, and is perfect for both new and existing users looking to try different techniques, test suitability, and explore other product ranges. It includes a pack of R2/2 Quantifoils on 400-mesh copper grids. A variety of cryo grid boxes are supplied for grid storage; the round boxes are compatible with most plunge freezers, and the type of grid box chosen is usually based on personal preference. The AGG3729 squared box has also been included, which is compatible with the Gatan CT3500. The kit also includes three types of tweezers, electrostatic discharge (ESD) soft grip with extra fine tips, ESD epoxy-coated stainless steel with antimagnetic properties, and the component handling tweezers, which are useful for handling grid boxes with the screw type lid.

Agar Scientific

For info: +44-(0)-1279-215-506
www.agarscientific.com/intro-to-cryo-consumable-kit

Cryo Grid Box Tweezers

Rave Scientific offers a variety of high-quality, Swiss-made, precision cryo grid box-handling tweezers, specially designed for handling the small cryo grid boxes used for storing transmission electron microscopy (TEM) grids under liquid nitrogen. These tweezers can be used with the EM-Tec GB-4 cryo grid box and other cryo grid boxes of similar design with a rotating lid and a central screw to hold the lid. They are designed for .05-mm screw heads but can also be used for screw-head sizes .03 mm–.06 mm. The length of these tweezers is 145 mm.

Rave Scientific

For info: 732-898-3828
ravescientific.com/rave-shop/category/cryo-grid-box-tweezers

Cryo-EM Grid Box

The SWISSCI Cryo-EM Grid Box is made from a special polymer formulation that ensures no static charging occurs at cryogenic temperatures. It has four diamond-shaped slot positions, each having its own number indicated on the side. No special tools are required, since the lid is rotated with standard laboratory tweezers. Individually numbered Pressure Adhesive Cryo Labels are supplied with the boxes, giving the consumer a convenient alternative to marking up each box themselves. The Cryo-EM Grid Box fits for all common cryo-EM sample mounting and storage devices. Its low cost enables single use, giving each stored sample a unique identification number and preventing sample mix-ups.

SWISSCI

For info: +41-(0)-41756-0037
www.swissci.com/cryo-em-grid-box

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.

HIGH-LEVEL GLOBAL TALENTS RECRUITMENT

Welcome back to hometown.

Thousands of academic job vacancies are in fast-developing China.

2020 Global Online Job Fair

March 25, 2020 Northeast China Doctoral Talents Recruitment

April 09, 2020 Southwest China Doctoral Talents Recruitment

April 17, 2020 Southeast China Doctoral Talents Recruitment

April 24, 2020 Specialty Session (Engineering)

May 08, 2020 North and Northeast China Doctoral Talents Recruitment

May 09, 2020 Hong Kong, Macao and Taiwan Doctoral Talents Recruitment

May 15, 2020 High-level Global Talents Recruitment

Qualification for Applicants

Global scholars, Doctor and Post-doctor

Key Disciplines

Life Sciences, Medicine, Material Sciences, Physical Sciences

Participating Approach

Please send your CV to consultant@acabridge.edu.cn for

2020 Global Online Job Fair



Scan the QR code to apply for
2020 Global Online Job Fair

Job Vacancies in China's Universities and Institutes

Please visit <https://www.acabridge.edu.cn/>

Contact consultant@acabridge.edu.cn

SPECIAL JOB FOCUS:

Cancer Research

Issue date: April 10

Book ad by March 26

Ads accepted until April 3 if space allows

Deliver your message to a global audience of targeted, qualified scientists.

129,566

subscribers in print every week

26,776

yearly unique active job seekers searching for cancer research jobs

25,690

yearly applications submitted for cancer research positions

To book your ad, contact:
advertise@sciencecareers.org

The Americas
202 326 6577

Europe
+44 (0) 1223 326527

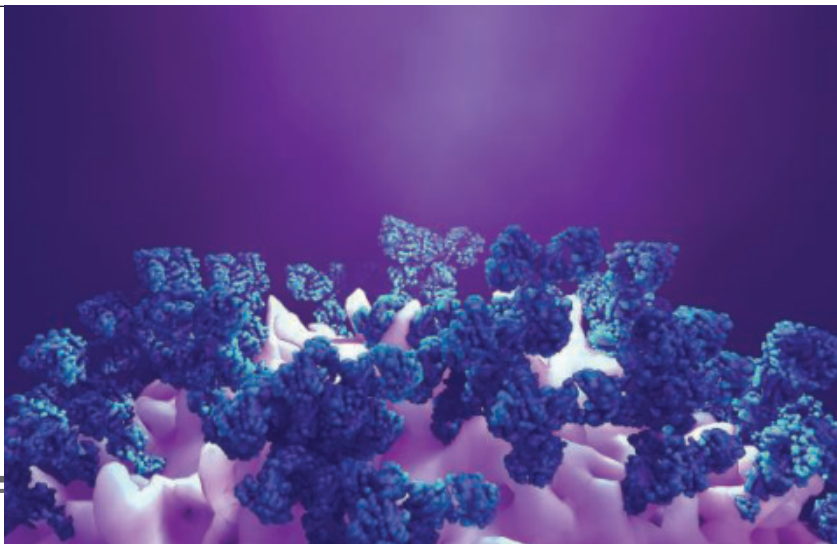
Japan
+81 3 6459 4174

**China/Korea/Singapore/
Taiwan**
+86 131 4114 0012

**Science
Careers**

AAAS

SCIENCECAREERS.ORG



What makes *Science* the best choice for recruiting?

- Read and respected by 400,000 readers around the globe
- Your ad dollars support AAAS and its programs, which strengthens the global scientific community.

Why choose this job focus for your advertisement?

- Relevant ads lead off the career section with a special cancer research banner
- Bonus distributions:
American Association for Cancer Research,
April 24-29, San Diego, CA
AACR Career Fair, April 25, San Diego, CA.

Expand your exposure by posting your print ad online:

- Additional marketing driving relevant job seekers to the job board.

Produced by the *Science*/AAAS Custom Publishing Office.



FOR RECRUITMENT IN SCIENCE, THERE'S ONLY ONE SCIENCE.

Where Science Gets Social.

AAAS.ORG/COMMUNITY



Member
COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE



The USDA, Agricultural Research Service, Grand Forks Human Nutrition Research Center in Grand Forks, ND, is seeking a full time **POSTDOCTORAL RESEARCH ASSOCIATE**, (Research Chemist) for a **TWO-YEAR APPOINTMENT**. Ph.D. is required. Salary is commensurate with experience (\$62,236- \$80,912per annum) plus benefits. Citizenship restrictions apply.

The incumbent will develop and implement mass spectrometry-based analytical methods for lipidomic characterization of dairy products. Knowledge of the principles, theories and practices of chemistry, analytical chemistry and a working knowledge of analytical instrumentation including mass spectrometers are desirable. Send application materials including curriculum vitae, transcripts, and references to:

Dr. Matthew Picklo,
Grand Forks Human Nutrition
Research Center
2420 2nd Ave N
Grand Forks, ND 58203
or
e-mail matthew.picklo@usda.gov

*USDA/ARS is an Equal Opportunity
Provider and Employer.*

**myIDP: A career plan customized
for you, by you.**



For your career in science, there's only one

Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.



Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers | AAAS In partnership with:



By Kai Liu

How I faced my coronavirus anxiety

In early February, I was working from home when I received a message informing me—and all the other professors at my university in China—that courses would be taught online because of the novel coronavirus. I was already feeling anxious about the mounting epidemic, and my university had locked its doors a few days earlier. Then, when I realized I'd have to teach students online, my anxiety level grew. I didn't have any experience with online teaching platforms. I was also skeptical about how effective they'd be. "How will I gauge the students' reactions to my lectures through a computer screen?" I wondered. "Will they learn anything?"

I live in Xuzhou, China—roughly 500 kilometers from Wuhan, the epicenter of the COVID-19 pandemic. Unlike Wuhan, my city isn't on lockdown, but residents have been discouraged from going outside and many businesses and institutions are closed. I've spent most of the past 2 months at home, along with my wife and daughter, fearful of the future and wondering when life will get back to normal.

Thankfully, none of my family members, friends, or colleagues have tested positive for the novel coronavirus. Working from home is also possible for me because my research doesn't involve lab work. But the spread of the virus and the rapidly rising death toll have weighed heavily on my mind. I've found it difficult to sleep. I've also had trouble focusing on work. One day early in the outbreak, I sat down at my computer intending to write a grant proposal. But all I could do was stare at the screen.

Years ago, I'd heard that Taoism philosophies were helpful for finding internal peace. So, I decided to listen to a few recordings. One instructed listeners to "govern [yourself] by doing nothing that goes against nature." That resonated with me because I realized that I'd been trying to push my anxieties aside and force myself to concentrate on work—an approach that wasn't working because it didn't feel natural. From then on, I told myself that it was OK to feel anxious, even if it impeded my work. That helped to lessen my internal struggles.

Over the past 2 months, I've also learned how to teach courses online, and I have found unexpected joy in that process—even though I struggled at first. There were multiple online teaching platforms to choose from, and I didn't know which one was best or how to use it. I opted for a platform that had a large server, thinking that it would



"I've spent most of the past 2 months at home ... wondering when life will get back to normal."

cope better with heavy usage. My university provided some helpful guidance, and I also learned through trial and error.

My first lecture was especially difficult because I couldn't see the students' faces. I was accustomed to lecturing in front of an audience. Online, I felt like I was speaking at my students but not getting anything in return. I communicated with a few of them afterward to get their feedback and they agreed with me, saying that I needed to find a way to make my lectures more interactive. So, I started to encourage my students to leave questions for me in the platform's comment section during my lectures.

Almost immediately, my students started peppering me with questions. I was surprised by the level of engagement. In a normal classroom setting, they are afraid to raise their hands; most wait until after the lecture is over to approach me and ask a question. But online, students were more comfortable sharing their questions in front of the entire class. That was a great outcome because if one student has a question, it's likely that another student has the same question and would benefit from hearing the answer. I've also been pleased to see from the homework assignments that they are following my teaching well.

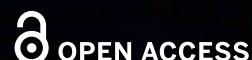
China was the first country to close its universities, but over the past month, universities in Italy, the United States, and elsewhere have made similar moves. I hope that my story can provide inspiration for academics who are fearful of what's to come. It's OK to feel anxious. But I'd also recommend staying open to change. You never know what you'll learn. ■

Kai Liu is an associate professor at Jiangsu Normal University in Xuzhou, China. Send your career story to SciCareerEditor@aaas.org.

CALL FOR PAPERS



Advanced Devices & Instrumentation



Advanced Devices & Instrumentation is an online-only open access journal published in affiliation with **Beijing Institute of Aerospace Control Devices (BIACD)** and distributed by the **American Association for the Advancement of Science (AAAS)**. *Advanced Devices & Instrumentation* aims to publish the latest investigations on novel ideas, methods, and techniques for the development and manufacture of advanced devices and instrumentation, as well as novel and practical solutions for existing applications.

Submit your research to *Advanced Devices & Instrumentation* today!

Learn more at spj.sciencemag.org/adi

The Science Partner Journals (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the non-profit publisher of the *Science* family of journals. The SPJ program features high quality, online-only, editorially independent open-access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS expands its efforts to communicate science broadly and for the benefit of all people by providing a top-tier international research organization with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest. Visit us at spj.sciencemag.org



@SPJournals



@SPJournals

ARTICLE PROCESSING CHARGES WAIVED UNTIL 2023

REDESIGNED WITH ALL OF US IN MIND

Introducing the Stericup® E and Steritop® E sterile filtration devices—**evolved with an eco-conscience.**

This progressive rethinking of filter design reduces your lab's environmental impact by eliminating the need for a receiver funnel, significantly decreasing packaging and biohazardous waste.

Expect the same faultless filtration you trust from Stericup® devices—and leave a smaller footprint.



Up to
48%
Reduced
Plastics

Up to
69%
Reduced
Packaging



SigmaAldrich.com/Stericup-E

*Up to 48% plastic reduction and 69% packaging reduction (depending on receiver volume), derived from comparison to traditional Stericup® sterile filters.

© 2019 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved. MilliporeSigma, the vibrant M, Millipore and Stericup are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources.

2019 - 24573 08/2019



The life science business
of Merck KGaA, Darmstadt,
Germany operates as
MilliporeSigma in the
U.S. and Canada.

Millipore
SIGMA



Stericup® E family of sterile
filters thread directly onto
virtually any media bottle

Millipore®

Preparation, Separation,
Filtration & Monitoring Products