NASA returns to the Moon with a robot armada p. 1188

Serotonin reduces the risk for cocaine addiction pp. 1197 & 1252

Strong, transparent, nacre-inspired glass composite p. 1229

Science

$15
10 SEPTEMBER 2021
science.org

AAAS

DOUSING THE STRATOSPHERE

Hydraulic jumps above supercell storms pp. 1194 & 1248

BII | Science Prize for Innovation

Apply for the new BII & Science Prize for Innovation today

Behind every life-changing solution is an entrepreneurial scientist–a creative mind who proved an idea in the lab and dared to carry it out in the world.

To encourage more scientists to translate their research, BioInnovation Institute (BII) & Science present a new annual award.

Our three winners will have their essays published in Science magazine and will be invited into BII's entrepreneurial ecosystem. In addition, the Grand Prize winner will receive a prize of USD 25,000 and each runner-up will receive USD 10,000 at a grand award show celebration in Copenhagen, Denmark.

The call for applications has just opened.
Apply before November 1, 2021.
www.bii.dk/scienceprize

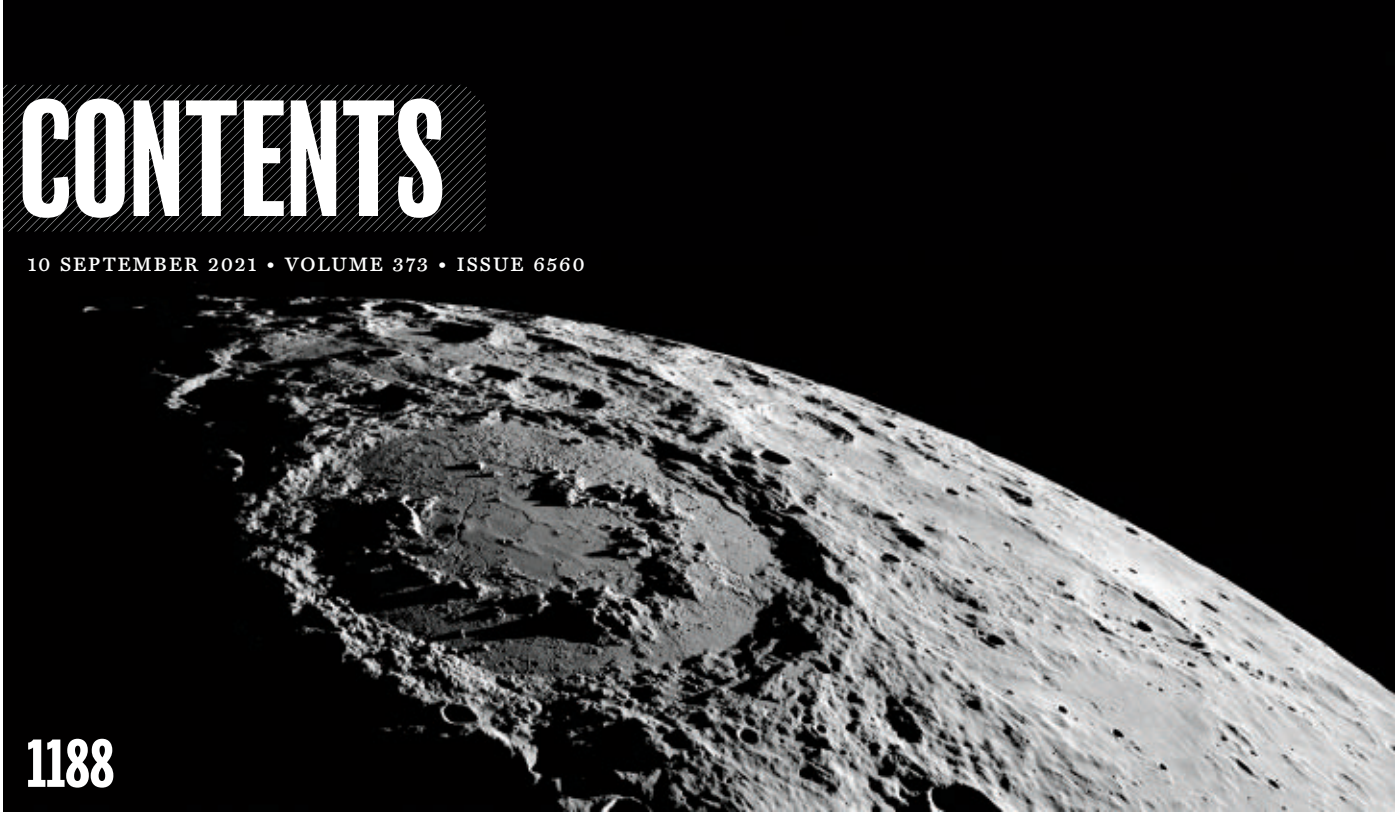
Presented by BII & Science

BII BioInnovation Institute Science

Apply before November 1, 2021
www.bii.dk/scienceprize

CONTENTS

10 SEPTEMBER 2021 • VOLUME 373 • ISSUE 6560



1188

NEWS

IN BRIEF

1180 News at a glance

IN DEPTH

1182 A COVID-19 publishing revolution? Not yet
Preprints, lauded for fast dissemination, remain a small part of pandemic literature
By J. Brainard

1183 A pandemic prevention moonshot
Biden plan would deliver new vaccines in 100 days *By J. Cohen*

1184 U.S. ups coronavirus surveillance
Country has fewer blind spots as variant Mu draws attention, but sequencing challenges remain *By M. Wadman*

1185 A push to see 'under the hood' of 2020 census
Letter asks Census Bureau for "noisy" file to measure the impact of differential privacy
By J. Mervis

1186 How farming shaped Europeans' immunity
Ancient farmers throttled their immune responses to help survive infectious diseases
By A. Gibbons

1187 Vultures face new toxic threat
Concern mounts over veterinary drugs used to treat cattle *By E. Stokstad*

FEATURES

1188 The coming lunar armada
A fleet of NASA-funded startups will soon begin to land on the Moon *By P. Voosen*
PODCAST

INSIGHTS

PERSPECTIVES

1194 Convective hydration of the stratosphere
Cirrus plumes above superstorm anvils are visible manifestations of a hydraulic jump *By J. B. Smith*
REPORT p. 1248

1196 More versatile synthesis of oligonucleotides
A phosphotriester chemistry may enable sustainable oligonucleotide production
By P. Virta
REPORT p. 1265

1197 Increased serotonin prevents compulsion in addiction
Extracellular serotonin decreases cocaine taking in mice
By K. Miyazaki and K. W. Miyazaki
REPORT p. 1252

1198 Getting droplets into shape
Protein clusters at interfaces control sizes and properties of biomolecular condensates *By W. T. Snead and A. S. Gladfelter*
RESEARCH ARTICLE p. 1218

1200 Mitochondrial disease: Replace or edit?
Questions over mitochondrial replacement suggest a role for mitochondrial DNA editing
By E. Y. Adashi et al.

POLICY FORUM

1202 Digital exposure tools: Design for privacy, efficacy, and equity
Apps can cut transmission of SARS-CoV-2—but how do we ensure that they don't exacerbate public health inequities?
By S. Landau

BOOKS *ET AL.*

1205 The keepers of the seas
As oceans warm, an author reveals the centrality of maritime activities in the modern world
By J. W. Smith

1206 Leveraging the collective mind
Shared social identities can better prepare us to confront societal problems
By S. A. Haslam

LETTERS

1208 Extractivist policies hurt Chile's ecosystems
By M. A. Urbina et al.

1209 Preventing urban soil loss in China
By S. Nie et al.

1209 China's carbon emissions in Brazil
By P. M. Fearnside

PHOTO: NASA GSFC SCIENTIFIC VISUALIZATION STUDIO

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Science Advances
AAAS

GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

RESEARCH

IN BRIEF

1212 From *Science* and other journals

REVIEW

1215 Marine conservation
The MPA Guide: A framework to achieve global goals for the ocean
K. Grorud-Colvert et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF0861

RESEARCH ARTICLES

1216 Ribosome
Nucleolar maturation of the human small subunit processome *S. Singh* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ5338

1217 Neuroscience
Complement factor C1q mediates sleep spindle loss and epileptic spikes after mild brain injury
S. S. Holden et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ2685

1218 Molecular biology
Regulation of biomolecular condensates by interfacial protein clusters *A. W. Folkmann* et al.
PERSPECTIVE p. 1198

1225 Respiratory enzymes
The three-spin intermediate at the O–O cleavage and proton-pumping junction in heme–Cu oxidases *A. Jose* et al.

REPORTS

1229 Materials science
Centrifugation and index matching yield a strong and transparent bioinspired nacreous composite *A. Amini* et al.

1235 Superconductivity
Anomalously strong near-neighbor attraction in doped 1D cuprate chains *Z. Chen* et al.

1239 Physics
Pendellösung interferometry probes the neutron charge radius, lattice dynamics, and fifth forces *B. Heacock* et al.

1243 Phase-change memory
Ultralow–switching current density multilevel phase-change memory on a flexible substrate *A. I. Khan* et al.

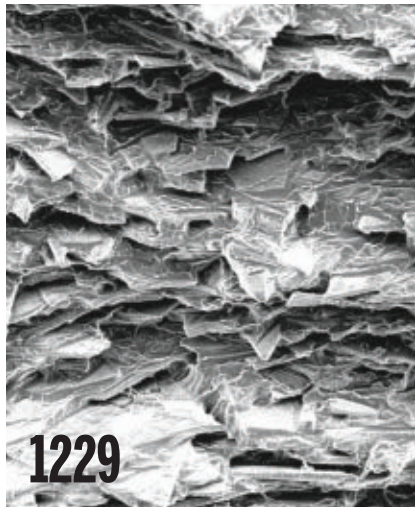
1248 Atmospheric dynamics
Hydraulic jump dynamics above supercell thunderstorms *M. E O'Neill* et al.
PERSPECTIVE p. 1194; VIDEO

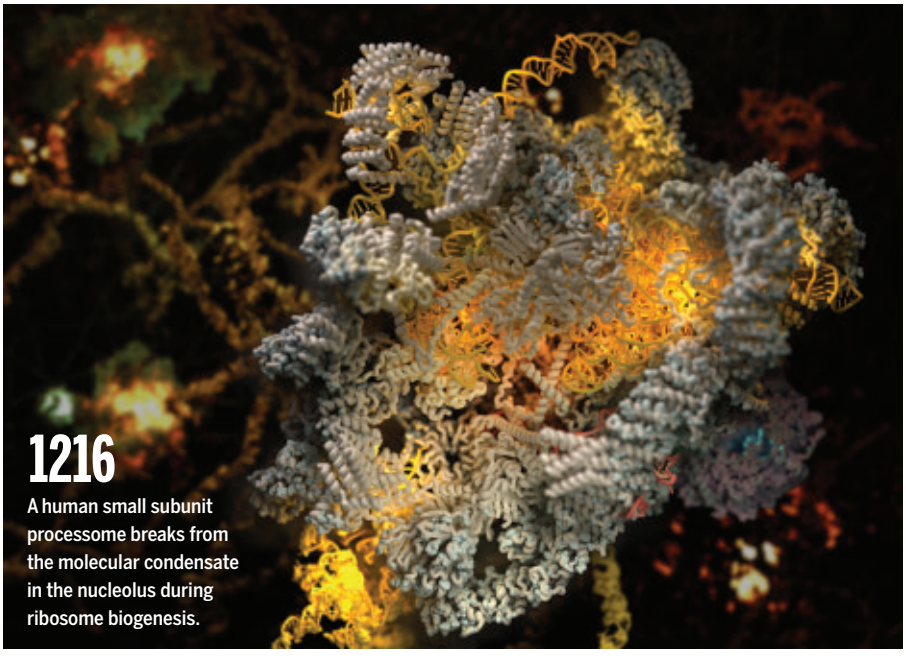
1252 Neuroscience
Synaptic mechanism underlying serotonin modulation of transition to cocaine addiction
Y. Li et al.
PERSPECTIVE p. 1197

1257 2D materials
Correlated electron-hole state in twisted double-bilayer graphene *P. Rickhaus* et al.

1261 Optics
Coherent perfect absorption at an exceptional point *C. Wang* et al.

1265 Oligonucleotides
A P(V) platform for oligonucleotide synthesis *Y. Huang* et al.
PERSPECTIVE p. 1196





CREDITS (LEFT TO RIGHT): ARNAUD VANDEN BROECK; A. AMINI *ET AL.*

DEPARTMENTS

1179 Editorial
Welcome to science.org! *By H. Holden Thorp*

1274 Working Life
A ghost author speaks up *By Karishma Bisht*

ON THE COVER

A powerful tropical thunderstorm exhibits an overshooting top that may inject water vapor into the lower stratosphere. Supercell thunderstorms, in particular, can excite hydraulic jumps at the tropopause, substantially increasing the local hydration of the lower stratosphere. Enhanced understanding of this feature has the potential to improve severe weather forecasts. See pages 1194 and 1248. *Photo: Santiago Borja*

New Products.....1271
Science Careers.....1272

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 © 2021 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.sciencemag.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.

eppendorf

New epT.I.P.S.®
Box 2.0

Clever and Smart

Time for a fresh reboot

Often copied but unrivaled, the epT.I.P.S.® Box has been in laboratories worldwide since 2002. The fresh new epT.I.P.S Box 2.0 design features a bright white tone, clean lines and a redesigned closure button in typical Eppendorf blue, as well as the bluish-tinted transparent lid with indentations on the sides.

> No slits on the back of boxes to reduce risk of contamination
> Autoclavable up to 100 x for a long lifetime
> Lid with indentations on sides to improve stackability
> Small silicone feet in the bottom of the box reduce risk of slipping

www.eppendorf.com/epTIPS-News

Eppendorf®, the Eppendorf Brand Design, epT.I.P.S.® and PhysioCare Concept® are registered trademarks of Eppendorf AG, Germany.
All rights reserved, including graphics and images. Copyright © 2021 by Eppendorf AG.

EDITORIAL

Welcome to science.org!

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

Over the past 2 weeks, the *Science* family of journals released its new website. Among the many big changes is its new URL: science.org (sciencemag.org will redirect to the new site). For those of us who remember the launch of sciencemag.org 26 years ago, there is a nostalgic tug in retiring this web address, but science.org represents where we want to go with the journals. It's also a moment to look back at how the American Association for the Advancement of Science (AAAS, publisher of the *Science* journals) has led the way in online scholarly publishing.

Sciencemag.org appeared in 1995 with a minimal menu—a table of contents, abstracts, and a few other items. At the time, there was a lot of squatting on domain names, and science.org was not available. There was strong determination not to choose science.com, an address that would have betrayed *Science*'s nonprofit status. It was also a time when the print version of *Science* was the only publication produced by AAAS, so referring to a digital version as a magazine still made sense. The leaders who deserve credit for how this transition took place are Editor-in-Chief Floyd Bloom, Executive Editor Monica Bradford, Editor Ellis Rubinstein, and Publisher Beth Rosner. They and their colleagues recognized the importance of moving scholarly journals online at a very early stage of publishing in the internet age. When sciencemag.org was launched, the main provider of online scientific publishing was HighWire Press, which we have worked with until 2 weeks ago. The only other major scientific journal that had moved online at HighWire before *Science* was the *Journal of Biological Chemistry*, also a society journal. Most of the commercial scientific journals were still being produced in print form only.

> "...science.org represents where we want to go with the journals."

Bloom clearly saw the power of moving *Science* online. According to Bradford and Rosner, he believed strongly that linking citations in scientific papers to the actual referenced articles would become very important. At the time, an electronic repository of abstracts in the life and biomedical sciences was called MEDLINE; PubMed, which widened access to this archive, evolved in 1996. Providing access to scientific papers through the internet set the stage for posting news and commentary as well. Rubinstein envisioned the release of news daily, as well as the growth of a site that engaged with career seekers and policymakers. These visions collectively led to the concept of sciencemag.org as a living website for a broad scientific community.

These ideals continued to inform how we have designed science.org. An important change for readers is the way the research papers are now displayed. Navigating around the paper, viewing figures, and linking to references and social media are all greatly improved. In addition, the overall environment for news and commentary is more integrated and visually engaging. The expansive volume of news, from daily fast-breaking stories to deep investigations, showcases graphics that are narratives themselves. Videos and podcasts are more discoverable, rounding out the user's experience. The past 18 months of pandemic life have revealed how much appetite there is for a trusted forum for scientific conversation because of the enormous challenges that the scientific community faces in communicating matters of, and related to, science. We plan to make the commentary page a home for all voices of the scientific community and the best place for its members to talk to each other.

When sciencemag.org was launched, there were predictions that didn't come to pass. It took scientists longer than expected to change their behavior and embrace the online world of *Science*. The movement of more journals to online publishing was slower than imagined, and working out the site license concept for university libraries was a long process of trial and error. The move online also triggered predictions that the print magazine would disappear. I am frequently asked when the print magazine will cease to exist, and my answer is that there is no time that I can foresee when that will happen. The print is still highly valued by many readers. Will it vanish one day? Probably—but that's far from our minds right now.

Science.org reflects all that the *Science* family values—outstanding, high-quality information presented with clarity and care. Welcome!

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.abm2245

Hurricane Ida made landfall near Houma, Louisiana, a few kilometers from a marine laboratory, which escaped major damage.

PHOTO: SCOTT OLSON/GETTY IMAGES

CLIMATE

Ida's fury spares Louisiana labs, but staff face challenges

Research laboratories in Louisiana emerged largely unscathed from last week's Hurricane Ida, despite worries that the Category 4 storm would follow in the destructive footsteps of Hurricane Katrina, which 16 years ago caused massive losses of research animals, biospecimens, and equipment (*Science*, 25 November 2005, p. 1267). Ida, which made landfall on 29 August, didn't bring the levee failures of Katrina, but it did trigger flooding that caused scores of deaths in the U.S. South and Northeast. New Orleans lost power, but generators at Tulane University and Louisiana State University (LSU) medical centers kept freezers and incubators running and lab animals cool. By week's end, medical buildings had power again. A coastal science laboratory near Chauvin, operated by the Louisiana Universities Marine Consortium, also avoided major damage. "It's night and day compared to Katrina," says Steve Nelson, dean of the LSU Health New Orleans School of Medicine. Research operations across the region will require time to return to normal, however, because evacuated staff and students need to return. Many are also dealing with damage to their homes. As *Science* went to press, many areas in the state, including parts of greater New Orleans, still lacked power.

Mars rover nabs first rock

PLANETARY SCIENCE | NASA's Perseverance rover drilled its first rock core in Mars's Jezero crater last week, likely capturing a volcanic sample. The bored boulder resembled ancient lava in its appearance, mineral signatures, and grain sizes, NASA scientists say. If the sample can be returned to Earth by a subsequent mission, scientists will date it, helping reveal when water flowed into the crater billions of years ago. During the rover team's first drilling attempt last month, the targeted rock crumbled on contact. This time, the team was more cautious, pausing to make sure the core, not much bigger than a pencil, remained in a tube prior to storing it. The rover will spend the next few weeks confirming the boulder's volcanic origin before moving to its next target.

Research confirms masks work

COVID-19 | The largest randomized controlled study of wearing masks to control COVID-19 has found that a nearly 30% increase in mask wearing reduced symptomatic cases by 9.3% in 600 Bangladeshi villages. The effect was most pronounced in those over age 50, and in villages that used surgical, rather than cloth, masks. Researchers—including hundreds of observers—spent 8 weeks distributing masks and promoting their use to more than 178,000 people in half of the villages; nearly 164,000 people in the other half served as controls. The study, published as a preprint, took place from November

2020 to April, before the Delta variant of SARS-CoV-2 became dominant in many countries.

Studies assess Long Covid toll

COVID-19 | Although the suite of persistent symptoms known as Long Covid remains puzzling, a preprint last week provides clues to how frequently the condition affects children. After surveying more than 50,000 children between the ages of 11 and 17 in the United Kingdom, researchers found about 14% of those previously infected with SARS-CoV-2 had three or more symptoms—such as tiredness, headaches, or shortness of breath—15 weeks after infection. Roughly half of those had five or more symptoms. Meanwhile, a study using self-reported data collected by the United Kingdom's Zoe app found that vaccinated adults who have breakthrough infections have low risk of Long Covid. Only about 5% of those infected after vaccination still had any symptoms at 28 days, compared with 11% of unvaccinated people who were infected, the researchers reported last week in *The Lancet Infectious Diseases*.

Astro journals go open access

PUBLISHING | Stargazers will soon be able to read all papers from the *Astrophysical Journal* and other publications of the American Astronomical Society (AAS) for free on the day of their release—one of the first prominent scientific societies to allow this. AAS announced last week that when all its journals become "open access" in January 2022, the society will end subscriptions; it will replace that revenue with fees paid by authors, based on length, with a maximum charge of $4500. Some research funders have required grantees to publish results in open-access journals and include publication fees in their grants. Many publishers have gone "hybrid," publishing both open and paywalled papers. AAS says it will waive or discount fees for authors unable to pay them.

Brazil museum seeks artifacts

COLLECTIONS | Three years after a devastating fire, the National Museum of Brazil is calling on museums, research institutions, and collectors from all over the world for help in "revitalizing" its collections, 85% of which were destroyed. Paleontologist Alexander Kellner, the museum's director, last week made a public plea for donations of animal and plant samples, fossils, minerals, and ethnographic and archaeological objects collected in Brazil and around the world. The museum promised in a statement to "spare no effort to safeguard the collection we still have and the [objects] we will receive." Germany has already provided €1 million in emergency aid, and 24 German institutions have pledged to help the museum rebuild its collections. The largest of its kind in Latin America and the oldest scientific institution in Brazil, it is scheduled to reopen in 2026.

Big U.S. research boost proposed

FUNDING | The U.S. Department of Energy's (DOE's) national laboratories are big winners in the science section of a $3.5 trillion budget blueprint unveiled last week by Democratic lawmakers in the House of Representatives. Some $13 billion of the additional $15 billion authorized for DOE over the next 5 years would be managed by its Office of Science. The National Science Foundation's budget would grow by $11 billion over 10 years, with $7.6 billion for its research directorates and $3.4 billion for large new facilities. NASA would receive $4 billion more through 2026 to repair and modernize its network of research facilities. And the National Oceanic and Atmospheric Administration would get an additional $4.2 billion over 5 years for programs aimed at understanding and adapting to climate change. This week, the House science panel was expected to approve these components in its $45.5 billion slice of the budget blueprint, a massive bill that would also expand an array of nonscience, social welfare programs; raise revenue; and cut spending in other areas. Appropriations committees would need to approve separate bills to provide the money called for in the plan.

The sharp mandibles of a leafcutter ant can easily slice through human skin.

PHOTO: ALEXANDER WILD

BIODIVERSITY

Zinc-laced mandibles keep ants sharp

Leafcutter ants (*Atta cephalotes*) can bite through plants and invading army ants as if they were butter. Now, scientists know why: Carefully arranged zinc atoms are responsible for their scalpel-sharp teeth. To get a close-up of arthropod "tools" such as ant mandibles, researchers mapped their atomic makeup using a microscopy technique called atom probe tomography. Their images revealed zinc atoms uniformly distributed through the ant teeth, making them sharp enough to cut through food or foe with less than 1% of the force required if they were made of the same material as human teeth, the team writes this month in *Scientific Reports*. The researchers suggest the molecular arrangement of the mandibles could serve as a model for medical cutting devices or other tools that need to retain their sharpness and shape over many uses.

IN DEPTH

THE NEW NORMAL

A COVID-19 publishing revolution? Not yet

Preprints, lauded for fast dissemination, remain a small part of pandemic literature

By **Jeffrey Brainard**

In January 2020, as COVID-19 spread insidiously, research funders and journal publishers recognized their old ways wouldn't do. They needed to hit the gas pedal to meet the desperate need for information that could help slow the disease.

One major funder, the Wellcome Trust, issued a call for changing business as usual. Authors should put up COVID-19 manuscripts as preprints, it urged, because those are publicly posted shortly after they're written, before being peer reviewed. Scientists should share their data widely. And publishers should make journal articles open access, or free to read immediately when published.

Dozens of the world's leading funders, publishers, and scientific societies (including AAAS, publisher of *Science*) signed Wellcome's statement. Critics of the tradition-bound world of scientific publishing saw a rare opportunity to tackle long-standing complaints—for example, that journals place many papers behind paywalls and take months to complete peer review. They hoped the pandemic could help birth a new publishing system.

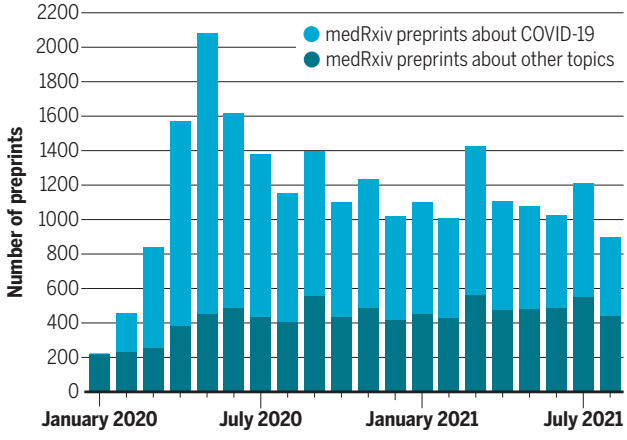
But nearly 2 years later, hopes for a wholesale revolution are fading. Preprints by medical researchers surged, but they remain a small fraction of the literature on COVID-19. Much of that literature is available for free, but access to the underlying data is spotty. COVID-19 journal articles were reviewed faster than previous papers, but not dramatically so, and some ask whether that gain in speed came at the expense of quality. "The overall system demonstrated what could be possible," says Judy Luther, president of Informed Strategies, a publishing consulting firm.

One thing is clear. The pandemic prompted an avalanche of new papers: more than 530,000, released either by journals or as preprints, according to the Dimensions bibliometric database. That fed the largest 1-year increase in all scholarly articles, and the largest annual total ever. That response is "bonkers," says Vincent Larivière of the University of Montreal, who studies scholarly publishing. "Everyone had to have their COVID moment and write something."

Helping drive the explosion was a sudden turn to preprints among medical researchers. Various servers accept papers directly from authors, posting them days later after a few basic checks but no peer review; most later appear in journals. A leading server, medRxiv—itself a newcomer, founded only in 2019—posted about 200 preprints on all topics in January 2020; by May 2020, the monthly tally had swelled to more than 2000, about three-quarters of them about the pandemic coronavirus. Traffic to medRxiv skyrocketed in tandem, with abstracts viewed a dizzying 6.1 million times in July, up from 30,000 in August 2019. Some preprints delivered findings that physicians immediately put to use at patients' bedsides. News that the corticosteroid dexamethasone reduced deaths in severe illness was published as a preprint in June 2020, but not in a peer-reviewed journal until February. By then, doctors had been giving the steroid for months.

Storm surge

Medical research preprints about COVID-19 grew especially fast in the pandemic's early months.



CREDITS: (ILLUSTRATION) KATTY HUERTAS; (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) MEDRXIV

By embracing preprints, medical researchers played fast catch-up to physicists, who have relied on them for decades—the original preprint server, arXiv, began to post physics papers in 1991.

A running concern about preprints in medical science had been their lack of vetting—but the pandemic made clear that peer review was not infallible either. High-profile, peer-reviewed journal articles about COVID-19 were retracted; suspicions arose whether data were fabricated for one that asserted hydroxychloroquine, an antimalarial drug promoted by former President Donald Trump, was not only ineffective, as many studies have shown, but could harm COVID-19 patients. "The border between validated scientific knowledge"—the turf traditionally claimed by peer-reviewed journals—"and not-yet-validated scientific knowledge [in the form of preprints] has become more blurred" during the pandemic, Larivière says.

COVID-19 also accelerated fledgling efforts to create a middle ground for preprints by putting some through a kind of rapid peer review. But those efforts remain small in scale. *Rapid Reviews: COVID-19*, a journal launched by MIT Press in August 2020, has quickly published online reviews of about 150 COVID-19 preprints.

But early predictions that preprints would dominate the COVID-19 literature—and reshape all scientific publishing after the pandemic—so far appear incorrect. Through May, only about 5% of all peer-reviewed journal articles about COVID-19 papers started out as a preprint, according to an estimate by Ludo Waltman and Stephen Pinfield of the Research on Research Institute, based on the Dimensions and Unpaywall bibliometric databases. (Most COVID-19 preprints have not appeared in peer-reviewed journals, at least so far.) And last year's preprint surge is subsiding, with medRxiv's monthly total now at about 600 papers, roughly half about COVID-19 (see graphic, p. 1182).

"Many researchers are not used to preprints and don't get rewarded for them," Waltman says. "Perhaps funders could have really encouraged or even mandated preprinting. If we really want [more preprints], the different stakeholders need to work together. … Otherwise, it will remain a small-scale phenomenon."

Complementing the spurt of preprints, editors of traditional journals promised to speed up reviews of COVID-19 manuscripts. They had some success: In January 2020, manuscripts on COVID-19 and related topics were accepted about 130 days after submission, on average. In July 2020, the lag had dropped to about 90 days, according to an analysis by Daniel Hook and colleagues at Digital Science, based on its Dimensions database. (The study examined the 20% of all papers in which these dates were provided.)

Did that speed up blunt quality? Although comprehensive evidence is lacking, only about 0.03% of journal articles on COVID-19 have been retracted—slightly less than in all disciplines before the pandemic.

Serge Horbach of Aarhus University in Denmark dove deeper, analyzing the rigor of reviewers' comments from two journals, *The BMJ* and *eLife*, which routinely post the comments with published articles. He examined 30 papers that appeared through late April 2020. Ten were about COVID-19; the rest covered other topics.

He found that reviewers of the COVID-19 papers frequently did not comment on whether the data or experiments were sufficient to support the conclusions, as they did for the other papers. When the reviewers did suggest improvements to COVID-19 articles, they tended to do so gently, inviting the authors to add data or experiments only if it could be done "easily" or "quickly."

It's "laudable," Horbach says, "that peer review, which tends to have a high level of inertia, was able to adapt so quickly in these crisis conditions. But indeed, it does raise questions about the quality of the material" published.

As for Wellcome's call to make COVID-19 articles and data free for all, the results are mixed. In early 2020, many publishers pledged to exempt COVID-19 papers from their paywalls, and as of last week, 77% of all such articles were free to read, according to the Dimensions database. That's a decline from a high of 85% in May 2020, but still well above the 50% for newly published scientific literature before the pandemic.

Early analyses suggest less progress toward providing "open data," the sharing of data sets with other scientists to speed the reuse of data and increase transparency of research. Among more than 7000 medRxiv preprints about COVID-19 published during 2020, authors of only about half offered to readily share the underlying data, according to an analysis by Larivière and colleagues.

Even though the changes were incremental, observers say COVID-19 did show that in a crisis, scholarly communications can be sped up and opened up. Larivière hopes the progress made for COVID-19 publishing can serve as a model, and be replicated for other global crises. One possibility: climate science, where only about 50% of new articles are now open access.

"Many people consider climate change a much more important threat than COVID, over the long term," Larivière says. "We need to treat it in the same manner." ■

U.S. HEALTH POLICY

A pandemic prevention moonshot

Biden plan would deliver new vaccines in 100 days

By **Jon Cohen**

An ambitious new plan from the White House that has a $65.3 billion price tag and an Apollo program design would transform the way the United States responds to pandemics, in part by vastly accelerating vaccine development, testing, and production.

U.S. President Joe Biden's administration hopes to launch the effort, announced last week, with $15 billion set aside for public health in a budget bill now before Congress. It calls for creating an Apollo-like "mission control" center to coordinate the many branches of government already involved in pandemic preparedness. Nearly 40% of the money would go toward vaccines, followed by just under 20% for treatments. The rest would support new diagnostics, early warning systems, improved public health and biosafety measures, and global health efforts.

The current U.S. response to the COVID-19 pandemic rests on scientific advances made over the previous 5 years, notes Eric Lander, director of the White House Office of Science and Technology Policy, which helped write the plan. "Five years from now, we ought to be able to do dramatically better," he says. The plan would lay the groundwork to develop, test, and approve vaccines against new pandemic pathogens in just 100 days, three times faster than for COVID-19.

The plan "is music to my ears," says Nahid Bhadelia of the Center for Emerging Infectious Diseases Policy and Research at Boston University, though she would like more specifics. Epidemiologist Michael Osterholm of the University of Minnesota, Twin Cities, feels the blueprint is not ambitious enough. "It's a good down payment, but it hardly will provide enough resources for a real plan."

The White House notes that the COVID-19 pandemic has so far cost the United States alone at least $16 trillion, making $65.3 billion for preparedness seem paltry. "It's hard to imagine," the plan notes, "a higher economic—or human—return on national investment." ■

A girl was tested for COVID-19 last week in North Miami, Florida, near where the coronavirus variant Mu was spotted in July.

PHOTO: AP PHOTO/MARTA LAVANDIER

COVID-19

U.S. ups coronavirus surveillance

Country has fewer blind spots as variant Mu draws attention, but sequencing challenges remain

By **Meredith Wadman**

Seven months ago, White House COVID-19 response coordinator Jeffrey Zients declared the scale of the U.S. effort to sequence the pandemic coronavirus "totally unacceptable." At that point, the country had managed to read the genetic blueprints of just 0.36% of all the coronavirus samples collected from U.S. noses and throats—far too few to detect rare, potentially dangerous mutations early and prepare to fight them.

But today, even as new U.S. coronavirus cases exceed 160,000 a day—a number not seen since late January—public health experts have a much better sense of their foe's changing genetics. By July, aided by $1.7 billion in funding from the American Rescue Plan, expanded sequencing operations were reading out 8.6% of new coronavirus cases.

Weekly submissions to the leading international sequence repository, GISAID, "right now are well north of 40,000," says Duncan MacCannell, who heads the Centers for Disease Control and Prevention (CDC) program that coordinates U.S. genomic surveillance. By comparison, it took 11 months—until November 2020—for the country's first 40,000 sequences to be deposited. "We are doing amazingly" in numbers of samples sequenced, says Trevor Bedford, a computational biologist at the Fred Hutchinson Cancer Research Center. "I have no concerns about a new variant slipping by unnoticed."

At the moment, sequencers around the country are seeing a deluge of Delta. The highly contagious variant accounted for 99.1% of cases in the United States by late August, up from 7.5% in late May, according to CDC. "We saw Delta sweep in in the blink of an eye," says integrative biologist Lauren Ancel Meyers of the University of Texas, Austin. "But that can change."

For example, coronavirus watchers noted in July that another variant, Mu, accounted for 9% of 101 sequenced cases in Florida's Jackson Memorial Health System and the University of Miami's UHealth Tower. Mu was first spotted in Colombia, where it made up 69% of cases in early July; last week the World Health Organization (WHO) declared it a variant of interest (VOI). Mu carries a mutation, E484K, that likely helps SARS-CoV-2 blunt vaccine and infection-induced immunity.

But Delta seems to readily outcompete Mu. In the Miami hospitals "by early August we were at approximately 97% Delta" and remain there, says pathologist David Andrews of the University of Miami. In Colombia, as of late August, Mu was down to 63% whereas Delta had surged from 6% to 19%. In Andrews's samples, Delta "has displaced virtually all other lineages," he says.

"The virus could surprise us but I would bet that Delta is going to fully sweep and [become] the common ancestor of all viruses in the next few months," Bedford says. That's likely "where the next variant will emerge from."

Some states might still be slow to spot a new variant. Nineteen, all but one in the Midwest, South, and Southeast, have sequenced fewer than 2% of their cumulative cases. Oklahoma, where hospitals are overwhelmed with COVID-19 cases, is last among them: It has uploaded 1617 sequences, or 0.3% of cases, to public repositories.

Still, in most states, the percentage of samples sequenced is orders of magnitude higher than it was in February. Wyoming, the leader, has sequenced 20% of its cases. "Our lab is essentially sequencing every positive sample they receive," says Alexia Harrist, the state epidemiologist. Maine, where cumulative sequencing is at 7.34%, sequenced 21% of new positive cases in the 3 weeks ending 14 August.

Those sequences "are a true random assortment," from around the state, says Nick Matluk of the Maine State Health and Environmental Testing Laboratory. The lab's staff has grown from 15 to 50 during the pandemic, including National Guard members deployed to do data entry and dispose of biowaste, and its facilities include a contracted 18-wheeler containing an enhanced biosafety level 2 lab.

In Mississippi, surveillance now reaches 75% of counties, and more than 99% of the samples are Delta. But, "That 1% is interesting to keep an eye on because that might be what seeds the next wave," says Ashley Robinson, a microbiologist at the University of Mississippi Medical Center genomics core facility. For example, in late August a WHO-designated VOI named Lambda popped up in a single sample from a small town in southwestern Mississippi. Enabled by CDC funding, "I threw in a couple of extra samples from that little town" for last week's analysis, Robinson says.

Yet problems remain in the U.S. scale-up. They include the fragmented health care system, in which states have different rules about data sharing. For example, the public CDC dashboard shows data for California only as a whole, not county by county, restricting data analysis by outside scientists.

Bedford says connecting the country's plethora of sequences to patient data remains a key problem, one that countries like the United Kingdom and Israel, with interconnected patient data, do not face. "What proportion of breakthrough cases are Delta versus Alpha? What proportion of hospitalized cases are different variants?" he asks. Answering such questions in the United States "is much harder to do." ■

STATISTICAL SCIENCE

A push to see 'under the hood' of 2020 census

Letter asks Census Bureau for "noisy" file to measure the impact of differential privacy

By **Jeffrey Mervis**

The 2020 U.S. census numbers released last month offer a snapshot of the country's 331 million residents. The Census Bureau says they also protect the identity of every person whose information is embedded within the billions of pieces of demographic data.

To meet that second legal obligation, the agency applied a new mathematical tool called differential privacy that subtly distorts all the data it collected on age, race/ethnicity, gender, and household status. Census officials counseled the public to view the census results as they would an Impressionist painting: blurry up close, but clear once you take a few steps back.

But that suggestion isn't very helpful for researchers who sweat the statistical details to understand the country's ever-changing demographics. So more than 50 of them have asked the Census Bureau to release an interim file that would help researchers analyze exactly how differential privacy affected data used not just for research, but also to distribute more than $1 trillion a year in federal funding and to help states redraw voting districts, an exercise now underway.

"It is a way to let researchers look under the hood," says Joe Salvo, former chief demographer for New York City and one of the signers of a 12 August letter to acting Census Director Ron Jarmin. (An earlier version of the letter, drafted by Cynthia Dwork, Gary King, and Ruth Greenwood of Harvard University, appeared in *The Boston Globe* on 26 July.) Census officials declined requests to discuss the issue, but a spokesperson said the agency "is considering" the request.

Differential privacy essentially allows people to see patterns among groups within a data set while holding back information about individuals. The agency can decide how much "noise" to add to any data set—the more noise, the greater the degree of privacy. But more privacy comes at the cost of accuracy, and perfect privacy would block the release of any information. Census officials believe they have struck the proper balance; they have also disclosed how much noise they have injected.

The privacy-protected noisy file, however, also contains negative values, fractions, and other statistical oddities, such as a household headed by a 2-year-old. Such oddities don't bother researchers, says Dwork, a theoretical computer scientist who co-authored a 2006 paper that describes the theory of differential privacy, but they would confound most public officials or citizens trying to understand their community's demographics.

So before issuing the final set of numbers on the 2020 census, officials created a cleaner version of the privacy-altered data, called the postprocessed file. Measuring and correcting for that second round of changes requires "considerable statistical effort and expertise," the letter explains, and the Census Bureau can lighten the load by releasing the initial "noisy" file. "Making available the noisy measurements file is … the key to ensuring that analysts can easily use the data," they write.

Posters displayed last month in the Mississippi Capitol show how population shifts found in the 2020 census will affect redistricting maps.

PHOTO: AP PHOTO/ROGELIO V. SOLIS

Getting access to the noisy file would remove another impediment for researchers like Michael McDonald, a political scientist at the University of Florida, who want to understand how the application of differential privacy to the 2020 census might have affected the demographics of voting districts. To avoid tampering with the state population totals, which are used to apportion the 435 seats in the U.S. House of Representatives, while still protecting privacy, census officials had to alter the demographic profile of each state. In practice, they subtracted from the largest demographic subgroups: Someone from a metropolitan area might be added to a rural area, for example, or a white resident assigned to a majority minority neighborhood. (Those adjustments led to media reports of "ghost" residents inflating the tally of a hamlet or an underpopulated urban neighborhood.)

McDonald, who signed the letter, says knowing how many white residents might have been added to a congressional district that is predominantly Black could determine whether a newly drawn district violates federal laws designed to assure equal access to voting and protect minority representation.

The need to understand how differential privacy could affect redistricting highlights a larger problem with the census data, says Erica Groshen, an economist at Cornell University and a former head of the Bureau of Labor Statistics. "Because we live in a society where data, lawsuits, and special interest groups have become so pervasive, the 'accuracy' of the Decennial Census is now much harder to define and much easier to challenge," she writes in an upcoming paper with Daniel Goroff of the Sloan Foundation about how researchers can make good use of privacy-protected data.

That new reality is why former Census Director John Thompson chose not to sign the letter despite being a staunch supporter of the agency's use of differential privacy. "It's a good idea, scientifically," Thompson says. "But I'm worried that [releasing a second set of numbers] could lead to lawsuits over redistricting. So I'm against it for political reasons."

Groshen signed onto the letter despite concerns that it would be "a heavy lift" for the Census Bureau to prepare the noisy file for release. "I'm very reluctant to make their lives more difficult," she says. "But it would send a message that providing researchers with the information they need is a high priority." ■

HUMAN EVOLUTION

How farming shaped Europeans' immunity

Ancient farmers throttled their immune responses to help survive infectious diseases

By **Ann Gibbons**

When early farmers of the Vinca culture first sowed barley and wheat 7700 years ago in the rich soil of the Danube River and its tributaries, they changed more than their diet: They introduced a new way of life to the region. Crowding together in mud huts, they lived cheek by rump with aurochs, cows, pigs, and goats—and their poop—in settlements that eventually swelled to thousands of people. Togetherness brought a surge in diseases as new strains of influenza, tuberculosis, and other maladies spread from animals to people and through early farming communities.

Now, a new study of ancient DNA shows how the immune systems of those early farmers responded to this new, pathogen-ridden environment. The Neolithic Revolution was a "turning point" in the evolution of immune responses to infectious disease, according to a paper published in *eLife* this week. The study suggests that in Europeans, evolution favored genes that throttled back the immune response to pathogens like influenza, restraining a hyperalert inflammatory reaction that can be deadlier than the pathogen itself.

"This study does a great job of showing that our immune system has continued to evolve in response to pathogen pressure," says population geneticist Joseph Lachance of the Georgia Institute of Technology. But he notes that the paper relies on an unproven method of predicting ancient immune responses. "I buy it, but it needs to be studied [more] when we have more ancient DNA."

Researchers have long suspected that early farmers got sick more often than nomadic hunter-gatherers. Studies suggest farmers in large Neolithic sites such as Çatalhöyük in Turkey faced a flurry of new zoonotic diseases such as influenza and salmonella, as well as new strains of diseases like malaria and tuberculosis. "If farmers got sick more, how did their immune systems change?" asked infectious disease specialist Mihai Netea of Radboud University Nijmegen Medical Centre, who led the study.

To approach that question, his team first studied genetically based variation in the immune responses of living people. They took blood samples from more than 500 people in the Human Functional Genomics Project, a biobank based in Nijmegen, Netherlands, and challenged the samples with various pathogens. Then they measured levels of specific cytokines—immunoregulatory proteins such as interleukin and interferon that are secreted by immune cells—and looked for correlations between those levels and a suite of immune gene variants.

In the new study, the team used those results to come up with what's called a polygenic risk score that predicts the strength of the inflammatory response in the face of specific diseases, based on an individual's combination of immune gene variants. The researchers then applied their technique to the past: From existing databases they downloaded ancient DNA sequences from 827 remains found across Europe, including Vinca farmers from today's Romania. They calculated the cytokine levels ancient people would likely have produced and their polygenic risk scores for inflammation.

Early farmers lived in crowded settlements where disease could spread rapidly, as seen in this artist's conception of 9000-year-old Çatalhöyük in Turkey.

The remains dated from between 45,000 and 2000 years ago, enabling the team to look for changes over time. They found that when faced with infections, Europeans who lived after agriculture likely produced dramatically lower levels of systemic cytokines than earlier hunter-gatherers. Those lower levels were likely adaptive, Netea says. "When people first encountered new pathogens, some overreacted and died, like we see with COVID today," he says. "The children of the people who survived didn't produce as many cytokines, so the whole population becomes more resistant."

The study also revealed a flip side: When infected with the fungus *Candida* and *Staphylococcus* bacteria—pathogens that tend to start as localized infections—early farmers likely mounted more robust inflammatory responses than the hunter-gatherers before them. A strong inflammatory response can quell a localized infection before it spreads, but a robust systemic response, as sparked by the flu or malaria, can spiral out of control.

The study is exciting because it shows how "the change in the way of life in the Neolithic exerted a considerable selective pressure on humans," says biological anthropologist Conchi de la Rúa of the University of the Basque Country, Bilbao.

But Lachance questions whether polygenic risk scores developed for modern people can predict inflammation for people in other places and times. Pathogens have evolved over time, he notes, and modern risk prediction might not apply to ancient disease strains. Population geneticist Luis Barreiro of the University of Chicago agrees, saying the authors "don't formally demonstrate the predictive value of these polygenic risk scores."

More samples of ancient DNA from people and pathogens, especially on other continents, is needed to test whether evolution scaled back the production of inflammatory cytokines in farmers everywhere. But the study clearly demonstrates that somehow or other, European's inflammatory responses to pathogens did change dramatically during the Neolithic, Lachance says.

To Netea, the findings suggest the ancient burst of evolution may have an impact even today: If a coronavirus like SARS-CoV-2 had swept through Europe before agriculture, he says, "more people would have died than today because they produced more proinflammatory cytokines." ■

ILLUSTRATION: DEAGOSTINI/GETTY IMAGES

Populations of white-rumped vultures plummeted about 20 years ago in Asia and haven't recovered.

PHOTO: MIKE LANE/MINDEN PICTURES

CONSERVATION

Vultures face new toxic threat

Concern mounts over veterinary drugs used to treat cattle

By **Erik Stokstad**

Two decades after a now-banned drug nearly wiped out vultures in several Asian countries, conservationists have identified another pharmaceutical threat to these ecologically vital scavengers. Two recent studies indicate vultures can die from eating the carcasses of cattle that were given nimesulide, an anti-inflammatory drug that is increasingly popular in Asia as a veterinary treatment. "It's quite worrying news," says Vibhu Prakash, deputy director of the Bombay Natural History Society (BNHS).

Environmental groups hope the new evidence will help persuade governments to ban the use of nimesulide in cattle. They are embracing a report last month that an anti-inflammatory drug already used in cattle, tolfenamic acid, is safe for vultures, only the second such alternative identified.

In the early 1990s, tens of millions of vultures lived in India alone. The large scavengers quickly stripped the carcasses of the country's many cows and other animals that died in the open, helping prevent the spread of disease. Then the population of vultures crashed, especially numbers of the common white-rumped vulture (*Gyps bengalensis*). They were suffering kidney failure after ingesting diclofenac, a drug commonly given to cattle to treat pain, fever, and inflammation.

India banned the veterinary use of diclofenac in 2006, and neighboring nations followed. Vulture numbers stabilized, but the damage was done: Populations had dropped by up to 99%. Cattle carcasses accumulated and sustained a booming population of feral dogs, increasing the need to vaccinate people against rabies.

The bans did not eliminate the threat of anti-inflammatory drugs, and in India the vulture population has not rebounded. It is still legal for pharmacies to sell diclofenac for human use, and covert studies of retail pharmacies have revealed the drug is being diverted to treat cattle. One reason is the speed with which it works. Diclofenac is called a "cup of tea" drug because sick cows improve by the time the veterinarian finishes a drink with the owner.

Other drugs still legally sold cause problems, too. Aceclofenac, for example, breaks down in cattle into diclofenac. Ketoprofen is also toxic to vultures, causing Bangladesh in February to completely ban its use in order to protect "vulture safe zones." Widely available and fast-acting nimesulide, meanwhile, has been growing in popularity. Researchers first raised concerns about the drug in 2016, after they discovered it in the tissues of several dead vultures.

The case against nimesulide is now strengthening. In June, researchers led by ecotoxicologist Subramanian Muralidharan of the Sálim Ali Centre for Ornithology and Natural History reported in *Environmental Science and Pollution Research* that four dead white-rumped vultures found in India's Gujarat state had nimesulide throughout their tissues, but not other drugs or pesticides. The birds also had renal failure and gout. "That's a red flag," says Julia Ponder, a veterinarian at the University of Minnesota, Twin Cities.

Toby Galligan, formerly with the Royal Society for the Protection of Birds (RSPB), and colleagues clinched the case with an experiment. After injecting the drug into cattle in South Africa, they slaughtered the animals and calculated a realistic exposure for a vulture feeding on a carcass. Then they gave a solution of nimesulide to two Cape vultures (*G. coprotheres*). Both died within 30 hours, and necropsies revealed gout and signs of renal failure, the team reports in a preprint posted last week on bioRxiv. Two vultures given water as a control remained healthy. (All the birds came from a wildlife sanctuary and were unfit to return to the wild.) The work is an "important, well-conducted study," says ecotoxicologist Barnett Rattner of the U.S. Geological Survey.

To co-author John Mallord of RSPB, the evidence is compelling that nimesulide should now be banned for use in cattle. Other anti-inflammatory and pain relief medications are available, they note. Meloxicam, for example, is safe for vultures. So is tolfenamic acid, the RSPB researchers found in a study done with scientists at the Indian Veterinary Research Institute and BNHS and reported in a bioRxiv preprint posted last month. "Having more than one drug is a real breakthrough," says Jemima Parry-Jones, director of the International Centre for Birds of Prey. "It means that vets have more choice to offer their clients."

So far, the pace of regulatory reform has been frustratingly slow, says Rhys Green, a conservation biologist at the University of Cambridge who heads Saving Asia's Vultures from Extinction (SAVE), an alliance of conservation groups. "The weight of evidence required by governments to achieve these bans is immense," notes biologist Campbell Murn of the Hawk Conservancy Trust.

Yet Green has hope for more immediate efforts to encourage sales of drugs that don't harm the birds. Two companies are marketing meloxicam as "vulture safe" in Bangladesh, he says. And SAVE's members in India are reaching out to cattle sanctuaries and state agencies that supply villages with drugs for their cattle to encourage them to buy safer drugs, which cost about the same. Countries in South Asia and elsewhere have also agreed to create recovery plans for vulture species as part of the global Convention on the Conservation of Migratory Species of Wild Animals. Despite the drug threats to vultures, says Roger Safford of BirdLife International. "There's a lot to be optimistic about." ■

THE COMING LUNAR ARMADA

A fleet of NASA-funded startups will soon begin to land on the Moon

By **Paul Voosen**

Next year, NASA's first mission to the lunar surface in 50 years will be run from an unlikely place: a low-slung building wedged between fast-food joints just off the Ohio River.

This unassuming former data center in Pittsburgh is the new home of Astrobotic, one of a few startups that NASA has selected to ferry scientific instruments to the lunar surface as part of the agency's $2.6 billion Commercial Lunar Payload Services (CLPS) program. Starting next year, CLPS landers will reach the Moon's surface at least twice a year, the agency hopes. It's an astonishing pace after a decadeslong drought for U.S. science on the lunar surface, says Brett Denevi, a planetary scientist at Johns Hopkins University's Applied Physics Laboratory. "We've been talking about this so long. It's almost shocking to see it happen."

Even without a single launch, CLPS has already radically changed the face of lunar science, says Stuart Bale, a space scientist at the University of California, Berkeley. "It feels like the Wild West. It's lean, mean, and cheap." Instruments that are less proven can be developed, flown, and tried out in a matter of a few years. They promise to bring a bounty of new science: maps of the subsurface water recently detected from lunar orbit, the first probes of the lunar interior since the Apollo landings, visits to mysterious magnetic anomalies, and unprecedented views of Earth's magnetic field and deep space.

One day in June, Astrobotic was buzzing during a visit by its ultimate VIP: Thomas Zurbuchen, head of NASA science. Armed with the Swiss accent of his birth and an alarming directness, he had arrived at Astrobotic's headquarters only to find much of it a construction site—work had been delayed by the pandemic. But a mission control room and a high bay, to assemble landers, were complete. The company's youthful CEO, John Thornton, assured Zurbuchen that everything was on track for the launch of the company's Peregrine spacecraft in 2022, 1 year later than originally planned.

CLPS, modeled after the commercial cargo and crew programs that pay SpaceX and others to fly to the International Space Station, is Zurbuchen's brainchild. It pays firms to carry low-cost scientific instruments at fixed prices on their lunar landers, while keeping NASA oversight at a minimum. He proposed the program in 2017, in a canny bid to add to NASA's science budget while accommodating the incoming Trump administration, which was eager to return astronauts to the Moon.

But there are no guarantees this new model will work. Some scientists complain that the process for choosing instruments and landing sites is opaque, without a traditional peer review from outside scientists. "We want more information on how these sites are selected," says Amy Fagan, a planetary scientist at Western Carolina University, Cullowhee, and chair of the Lunar Exploration Analysis Group, which advises NASA. She and others also worry the program lacks a long-term plan for its science campaign. "What is the big picture goal?" she asks.

The technical challenges are daunting as well. Engineers and space scientists alike are haunted by 2019 missions to the lunar surface by an Israeli company and the Indian government that ended in dramatic failure, underscoring how difficult it can be to land on the Moon. "We all got scared by the SpaceIL and Indian failures," says Robert Grimm, a planetary scientist at the Southwest Research Institute, Boulder, who has instruments on two CLPS landers. "There's smart people all over the world. So what did they do wrong?"

PHOTO: NASA GSFC SCIENTIFIC VISUALIZATION STUDIO

Schrödinger crater on the Moon's far side is a target for a seismometer and a radio telescope.

From the start, Zurbuchen had pitched CLPS as risky, saying as many as half of the missions could fail given the limited oversight NASA would have over the landers' construction and operation. Even so, one or two early crashes could derail the program, triggering a retreat from the White House or Congress. "The No. 1 priority is let's make sure the first one is successful," says Clive Neal, a lunar scientist at the University of Notre Dame. For Astrobotic, among the first in line, the pressure is intense.

THORNTON GOT THE SPACE ITCH in 1997, when he was in fifth grade. NASA's Pathfinder mission had just landed on the surface of Mars, carrying the Sojourner rover. "I thought that was the coolest thing ever," he says. He built a Lego model of the rover for class.

Later, Thornton studied engineering at Carnegie Mellon University (CMU) in Pittsburgh and worked with William "Red" Whittaker, the famed roboticist. In 2007, as he was nearing graduation, the Google Lunar XPRIZE was announced, promising $30 million in prizes for teams that could land robotic probes on the Moon. Thornton faced a choice: He could go to big aerospace, where he might spend his entire career slowly building a few spacecraft, he says, "or I could take my moonshot."

He joined Astrobotic, which Whittaker led at the time, as an engineer. In those early days, Astrobotic got by with modest NASA tech development grants. The contracts kept growing larger, building the company's credibility, and when Thornton took over in 2013 it began to focus on the business of lunar delivery. Companies were delivering cargo to orbit, he says, and he thought, "Why not the Moon?" But it soon became clear that the $20 million top XPRIZE award, with a final deadline of 2018, was far too small to build a business on. "We saw the writing on the wall," he says.

Ultimately, none of the XPRIZE competitors made it, but Zurbuchen, who was an engineer and space scientist at the University of Michigan, Ann Arbor, before joining NASA in 2016, saw an opportunity. When the incoming Trump administration expressed interest in a return to the Moon, he got on the phone and started asking around to see whether some of the companies that participated in the program were legit. They seemed to be—and if they stayed cheap, he thought, they could give NASA a fast route to the Moon. "It took advantage of that unique opportunity in time," Zurbuchen says. The administration and Congress were sold.

To keep the effort lean, NASA sent out a call for scientific instruments that were, essentially, sitting on the shelf. Some had been built for a previously proposed lunar rover, Resource Prospector, that was ultimately canceled. Others sprung from unflown instruments designed for sounding rockets, orbiting spacecraft, and even a delayed European-Russian lander. As one NASA official told Grimm early in the program, "We're so desperate for payloads, we'll fly rocks back to the Moon."

In May 2019, NASA selected two small U.S. landers to carry the first of these instruments: Peregrine, from Astrobotic, and Nova-C, from Intuitive Machines in Houston. (A third lander selection was rescinded once it became clear its operations were largely in India.) The two landers would aim for easy, flat, and somewhat boring equatorial sites on the Moon's near side: Astrobotic to Lacus Mortis and Nova-C to Oceanus Procellarum. Each would be solar-powered, lasting about half a lunar day—roughly 2 weeks on Earth. And they would be cheap, with the government paying less than

$100 million, when a NASA-built lander might cost $500 million, or likely more.

The pandemic has complicated life for Astrobotic, like everyone else. Two NASA CLPS contracts allowed the company to grow from 18 people less than 3 years ago to more than 150. But distancing and remote work have slowed development time. "We have a lot of employees that we've hired that have never actually been to the office before," Thornton says.

The delay has also complicated Astrobotic's hopes to be the first lander: Both it and Intuitive Machines are slated for 2022 landings now, and the latter is launching on a proven SpaceX Falcon 9 rocket, whereas Peregrine has booked the maiden flight of the next-generation Vulcan rocket from the United Launch Alliance, a joint venture of Lockheed Martin and Boeing.

The pace was starting to pick up by the time of Zurbuchen's June visit. It was a tricky dance for the NASA official. When Zurbuchen visits a typical NASA mission, he likes to reach down deep into the project's guts to find its top three problems. With CLPS, "We're not managing that," he says. "They need to be their own company." And so he pushes instead for higher level details—how the company is testing interfaces for NASA's scientific instruments, for example, or how it is interacting with the agency. "Is NASA helping where they're asking?"

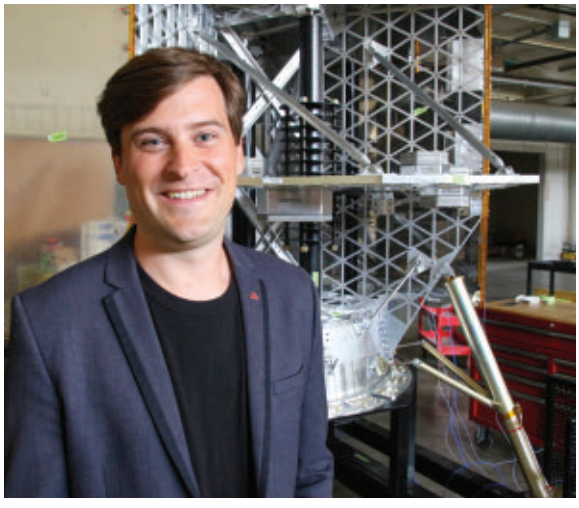
During his visit, a full-scale structural model, which Astrobotic used to test how the lander would handle a rocket launch, sat in the company's high bay, standing roughly 2 meters tall. Peregrine's four stubby legs and squat structure are reminiscent of the Apollo landers. It's an intentionally simple design, based on an aluminum alloy bus with a fixed solar panel on top and no complex moving parts.

Astrobotic once planned to ring the lander with decks that would house its payloads. But the Moon is unforgiving, with no air to carry heat away from sunlit components. To survive on the surface, engineers had to ditch the decks, replacing them with metallic heat shielding and moving around some of the 21 payloads the lander will carry. "It was kind of a push and pull to really find a solution that worked best for us and for them," says John Walker, a payload manager at Astrobotic.

Adapting to this commercial model has been challenging for scientists. "This isn't a mission, it's a bunch of instruments on a lander," says Barbara Cohen, a planetary scientist at NASA Goddard Space Flight Center who leads development of one of the instruments that will fly on Peregrine, the PROSPECT Ion-Trap Mass Spectrometer. Researchers don't have a good sense of the order in which the instruments will be turned on and when their measurements will begin. They also can't dictate the flight path of the lander or which gases, for example, other instruments might use, potentially contaminating their measurements. There's no project scientist for the overall mission to sort it all out.

At the same time, the loose coordination has allowed Cohen and her colleagues to move fast on completing their instrument, which will hunt for signs of water bouncing around just above the lunar surface. "It's been a challenge in an interesting way," Cohen says. "It's very fast."

Sometime next year, the spacecraft will ride its Vulcan rocket to lunar orbit, where it will linger for up to 1 month, awaiting the right moment to land—just after the Sun has risen on Lacus Mortis. It will descend flying on its side, then pivot 100 meters above the surface to a legs-down configuration. A laser-ranging system will help guide it safely to the surface, where it will begin its brief, intense life.



Astrobotic's CEO, John Thornton, with a test model of an engine.

IF THERE'S A DOMINANT THEME to the early CLPS missions, it's water. And for good reason. A series of orbital missions over the past 2 decades has detected tantalizing signs of the stuff across the Moon's surface, defying past notions of a desiccated planetary body. Two NASA missions have turned up evidence of water ice in shadowed areas and near the south pole. And when Chandrayaan-1, India's first lunar mission, orbited the Moon in 2009, a NASA instrument aboard picked up evidence of water all over the surface, the signal strongest during lunar dawn and dusk.

"I call it space dew," says Georgiana Kramer, a lunar scientist at the Planetary Science Institute. The Sun throws hydrogen ions at the lunar surface, in the form of the solar wind, that bombard oxygen-rich rocks, apparently forming small amounts of water. The vapor seems to dance across the surface and may gain enough energy to hop from the equator to shadowed cold traps at the poles, although "we don't really understand its cycle," Cohen says.

Among the water-focused payloads on Peregrine will be three instruments that will form the bass beat of future CLPS missions, reappearing on various landers and ultimately ending up on NASA's Volatiles Investigating Polar Exploration Rover (VIPER), which Astrobotic will carry to the Moon as early as 2023. This VIPER suite will map lunar water and probe its origins with a mass spectrometer that can detect water directly; a neutron counter to measure hydrogen based on the neutrons it absorbs; and a near-infrared spectrometer to detect hydrogen in drill tailings.

Peregrine's mission to Lacus Mortis is unlikely to provide any revelations itself—stuck in one location, it won't be able to sample multiple areas. It will, however, provide baseline measurements of how much the lander contaminates the surface and prove out how these tools work together. "It's risk reduction, big time," for future missions, says Richard Elphic, a space physicist at NASA Ames Research Center and lead on the neutron counter.

Other water-hunting missions will follow close behind. PRIME-1, carried on a lander built by Intuitive Machines, is slated to bring the VIPER mass spectrometer together with a specialized drill to the south pole in November 2022 (see graphic, p. 1191). Another, built by Masten Space Systems in Mojave, California, will launch in 2023. Masten's lander will use powerful cameras to study the texture and composition of the soil surrounding it and wield a dexterous robotic arm to scoop up samples for analysis. It will also include the three VIPER instruments and deploy a small autonomous rover, built by Astrobotic and CMU, that will carry the neutron counter to scan the first meter of the surface for hydrogen spikes indicating buried water ice.

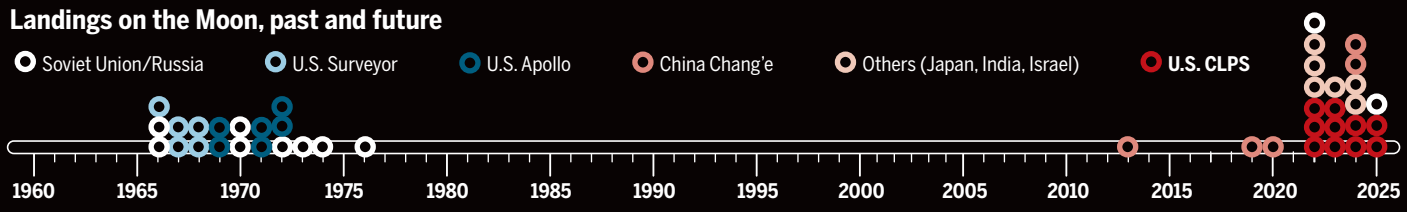
Soon after, in November 2023, it will be VIPER's turn. The $430 million NASA rover will be carried by Astrobotic's Griffin lander, which will be about the size of a sedan. Once it rolls down the ramp, VIPER will be fully independent, running for 100 days and conducting more than 40 drilling operations while constantly counting neutrons. Beyond mapping the sheer amount of water and its

PHOTO: ASTROBOTIC

NASA's new moonshot

Beginning next year, an armada of low-cost robotic landers sponsored by NASA will begin to arrive on the Moon. The Commercial Lunar Payload Services (CLPS) program aims to lower the cost of reaching the Moon, much as NASA has enlisted private space companies to reduce the cost of reaching low-Earth orbit. CLPS may eventually provide infrastructure for astronauts, but its initial payloads will conduct NASA's first surface lunar science in 50 years. *Graphic by* **Chris Bickel**

Landings on the Moon, past and future



1 Peregrine
Provider: Astrobotic
Lacus Mortis
2022

2 Nova-C
Provider: Int. Machines
Oceanus Procellarum
2022

3 Nova-C (PRIME-1)
Provider: Int. Machines
South pole
2022

4 Xelene
Provider: Masten
South pole
2023

5 Blue Ghost
Provider: Firefly
Mare Crisium
2023

6 Griffin (VIPER)
Provider: Astrobotic
South pole
2023

7 PRISM-1A
Provider: TBD
Reiner Gamma
2024

8 PRISM-1B
Provider: TBD
Schrödinger crater
2024

9 PRISM-2A
Provider: TBD
South pole
2025

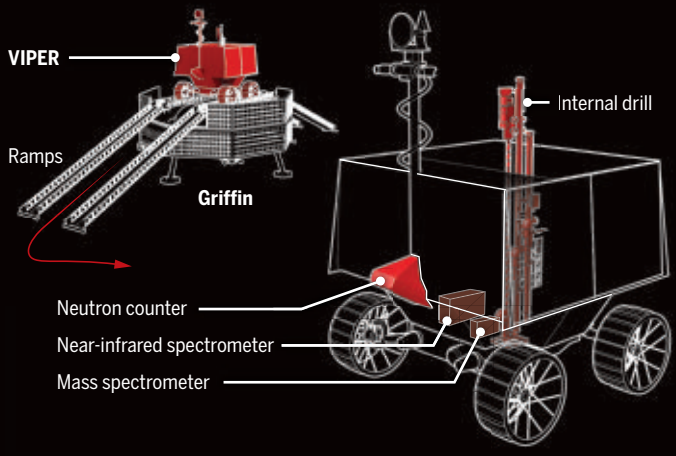
10 PRISM-2B
Provider: TBD
Gruthuisen Domes
2025

Polar rush

Four CLPS landers, along with two Chinese and two Russian missions (not shown), target the south pole, a prime water target. China's Chang'e-6 lander in 2024 will return rock samples from the region, which NASA is also eyeing for human exploration.

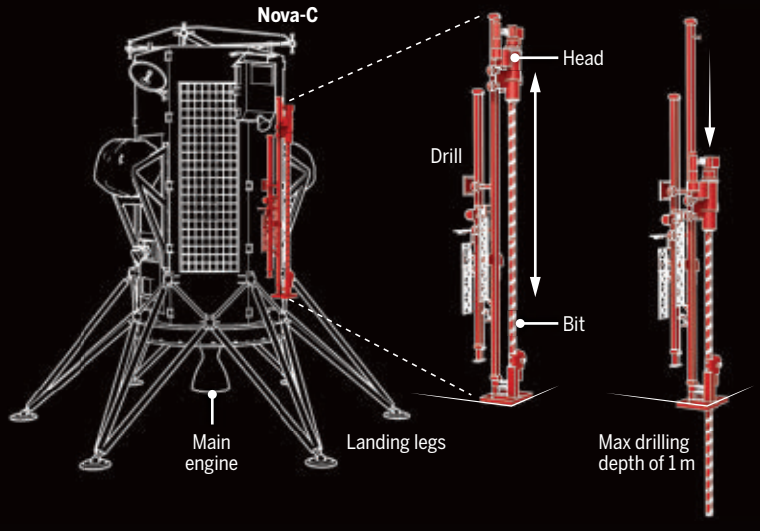
Precious delivery

Astrobotic's landers include Peregrine (not shown) and the larger Griffin, a car-size spacecraft that will carry NASA's $430 million Volatiles Investigating Polar Exploration Rover (VIPER) to the south pole, where it will scan and drill for water.



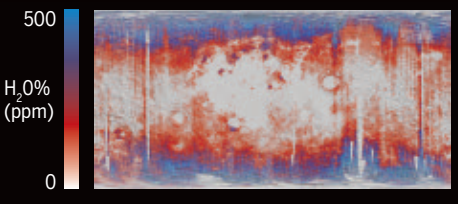
Super 'Novas'

Intuitive Machines's Nova-C landers stand nearly 4 meters tall. They will carry out two of the first three CLPS missions, including the first to the south pole, which will test the fluted drill and a mass spectrometer that will be used on VIPER (above).



KREEPy surface

Most of the Apollo rock samples come from the Procellarum KREEP terrane (PKT), a hot region enriched in thorium. Data from more of the Moon could show whether these rocks give a biased view of its history.



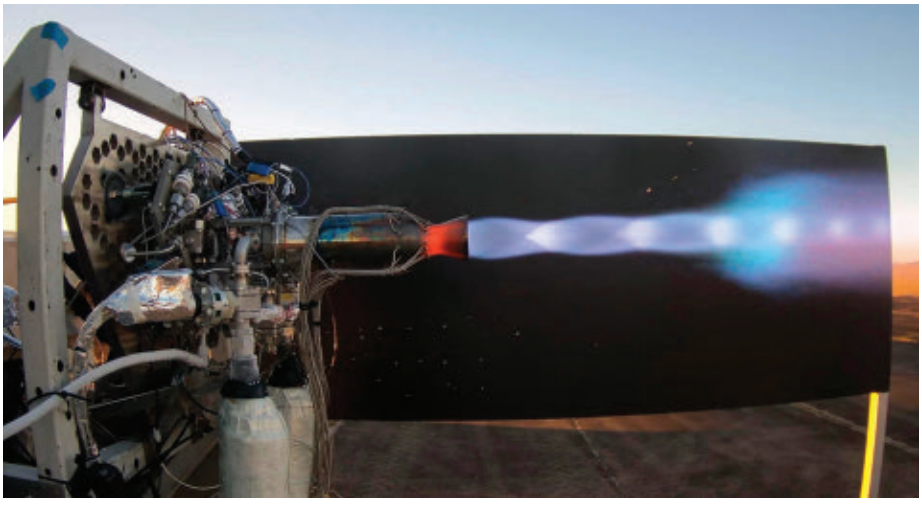
Veil of water

Orbital observations suggest the Moon not only has water at its poles (blue on the map), but a whole hydrological cycle, with water created as hydrogen from the solar wind bombards oxygen-rich rocks.

CREDITS: (GRAPHIC): C. BICKEL/*SCIENCE*; (DATA) R. MILLIKEN/BROWN UNIVERSITY; S. LI/UNIVERSITY OF HAWAII

distribution—important for future use by astronauts—its mass spectrometer will be able to tease out the water's origins by looking at hydrogen isotopes. Whereas some is likely created by solar wind, which contains light hydrogen, other water may be delivered by comets, rich in a heavier isotope. And by sniffing for sulfur, the instrument might detect whether any of this water seeped from the Moon's interior.

VIPER may be the culmination of CLPS's first stage, but it could also be a turning point. If the program is a success, NASA may turn to companies like Astrobotic not just for transportation to the Moon, but also for rovers. "There's a good possibility that the VIPER rover is the first and last NASA-built rover for the Moon," says Anthony Colaprete, the mission's project scientist at Ames. "And I'm perfectly fine with that."



A test run of the engine that will carry Intuitive Machines's Nova-C lander to the Moon in 2022.

ALTHOUGH WATER MAY BE THE MOST pressing question about the Moon, it is far from the only one. "We have a million and one lunar science questions," Cohen says. NASA is now developing instrument packages to address them. In 2023, the first lander built by Firefly Aerospace, for example, will carry a heat probe and an electromagnetic sounder to a nearside lava basin called Mare Crisium, which boasts some of the Moon's thinnest crust. The instruments will provide the first in situ measurements of the interior's heat since Apollo, which found that the Moon's surface was surprisingly warm—and help clarify whether the terrain the astronauts explored was an anomaly and how it might have formed.

Instruments to probe the interior are just part of the cargo, leading Brian Walsh, an astronomer at Boston University, to call Firefly a "beautiful Frankenstein." Walsh's own instrument, the Lunar Environment Heliospheric X-ray Imager, will look back at Earth and take the first ever picture of the planet's magnetic field by detecting solar x-rays careening off it. Although Earth's magnetic field has been measured in many ways, it has never been imaged from the outside. "It's hard to study from the inside," Walsh says. "If you were trying to study a whale, you couldn't do it from its belly."

Still, coordinated science will be a theme going forward, Zurbuchen says. NASA has already selected two new packages of instruments, which CLPS companies are now bidding to take. The first, called Lunar Vertex, will be carried by a rover across a "lunar swirl." Even though the Moon lacks an overall magnetic field, these mysterious, dark areas have a localized field that seems to shield them from the solar wind. The second payload, perhaps CLPS's most scientifically ambitious, will go to the Schrödinger basin, a 316-kilometer-wide impact crater on the Moon's far side that features a peak ring, a central upwelling of rock that is a hallmark of large impacts. Copies of the heat probe and magnetometer flown on Firefly will land here, teasing out the source of the material in this peak ring—does it come from the crust or the mantle? The answer will illuminate how giant impacts have shaped the surfaces of rocky planets and moons throughout the Solar System.

The Schrödinger mission will also carry the first NASA seismometer to run on the Moon since the 1970s—and the first ever on its far side. It will listen for moonquakes originating from the far side, which Apollo never detected—perhaps indicating deeper geological differences between the two hemispheres, says Mark Panning, a seismologist at NASA's Jet Propulsion Laboratory who is leading development of the Schrödinger seismometer. The background seismic data will also reveal the rate at which micrometeorites bombard the surface—essential information for shielding long-term lunar inhabitants.

Finally, the lander will carry a small radio telescope. After the lander shuts off and the Sun dips below the lunar horizon, the telescope's far-side vantage will ensure a quiet unknown on Earth, allowing it to search for cosmological signals of a time before stars, when dark clouds of neutral hydrogen began to slowly pull together. "That's a very challenging measurement," says Bale, the instrument's lead investigator. "I wouldn't claim a priori that we'll go to Stockholm. But it's a goal."

ALL THAT SCIENCE COULD come in the next 5 years—a striking contrast to most NASA science missions, which can take decades to develop, launch, and return data. The demographics of lunar scientists reflect that pace. "Go to a lunar science meeting and the faces look very different than at other planetary meetings," Zurbuchen says. A number of younger researchers are getting involved and writing proposals, drawn by the possibility of getting an instrument to the lunar surface in a matter of a few years. Many haven't been successful yet, but each round of proposals brings valuable experience, Denevi says. "The opportunity is open."

How long it will remain open is not clear. The Biden administration supports CLPS. But its focus is not guaranteed to stay on science; among the landers now vying for CLPS contracts are SpaceX's enormous Starship, a spacecraft that NASA has selected to serve as its lander for astronauts and that might also be used to ferry supplies. Astrobotic also sees a future in providing services for humans on the Moon—delivering permanent solar stations or other equipment. Expanding beyond small science payloads and developing landers that can survive the frigid lunar night will be essential to keeping these companies viable, Neal says. "If you keep CLPS just as small payloads surviving only one [lunar] day, it's a dead end. It's going to get cut."

For now, though, NASA's focus is on those first launches. At the end of Zurbuchen's visit, during an evening talk a short distance from Astrobotic's headquarters, he lauded the company and the innovation it was fostering in Pittsburgh, comparing it to the NASA team that landed Perseverance on Mars. "What happened is an amazing team came together, persevered, and inspired everybody. And John," he said, looking over at Thornton, "that's exactly what I expect from you."

"No pressure," Thornton said afterward. "I'm sweating already." ■

PHOTO: INTUITIVE MACHINES

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

1848 SOCIETY
AAAS

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

INSIGHTS

PERSPECTIVES

ATMOSPHERIC SCIENCE

Convective hydration of the stratosphere

Cirrus plumes above superstorm anvils are visible manifestations of a hydraulic jump

By **Jessica B. Smith**

Strong thunderstorms provide a means for rapidly moving air from the troposphere—the atmospheric layer closest to Earth's surface—to the stratosphere—the next highest layer of the atmosphere, located approximately 10 miles above the surface. Updrafts within the core of the storm can transport pollutants, water vapor, and ice to the stratosphere on time scales that are shorter than those of other troposphere-stratosphere exchange processes. Modeling (*1*–*3*) and observations (*4*, *5*) show that the input of tropospheric air and ice can moisten the normally dry stratosphere and alter its composition and, potentially, its radiative and chemical properties (*6*). Yet, there are many unanswered questions concerning the transport mechanism. On page 1248 of this issue, O'Neill *et al.* (*7*) provide a framework for understanding how some supercell thunderstorms—those that produce above-anvil cirrus plumes—generate an internal hydraulic jump to irreversibly inject air and water from the troposphere to the stratosphere.

Perturbations to the stratosphere can have global consequences. Water vapor is a powerful greenhouse gas and constitutes one of the most important feedbacks within the climate system. The upper troposphere and lower stratosphere, where air is the coldest

PHOTO: SANTIAGO BORJA

Intense thunderstorms can lead to the formation of above-anvil cirrus plumes and transport pollutants and water to the stratosphere.

and driest, is the region that is most sensitive to changes in water vapor because it determines how much outgoing longwave infrared radiation escapes Earth's atmosphere (*8*, *9*). It has also been suggested that decadal-scale variations in stratospheric water vapor may contribute to trends in surface warming (*10*). Moreover, the stratosphere is home to the ozone layer, which is critical for life at Earth's surface because it absorbs the harmful ultraviolet radiation from the Sun. Given the sensitivity of this region to perturbations, it is important to understand both the immediate effects of these thunderstorms and how these effects will change in the future, especially because environmental conditions favorable for severe convection may increase with climate change (*11*, *12*).

Convection arises from the density difference between a parcel of warm air and the cooler air surrounding it. Warm air, which is less dense, rises through the atmospheric column and expands and cools. When the temperature of the rising air parcel has cooled sufficiently, the water vapor it contains will begin to condense and release energy (latent heat). This addition of heat facilitates continued ascent. If the surrounding atmosphere cools more quickly with height than does the rising air parcel, unstable conditions emerge, and the parcel will continue to rise until it is at the same temperature as that of the environment. At this point, the equilibrium level, the parcel becomes neutrally buoyant. This often occurs near the tropopause, the boundary between the troposphere and the stratosphere, where the atmosphere begins to warm with increasing height. The anvil-shaped structure at the top of deep convective storms is indicative of the level where the rising motion has stopped, and air and ice have spread out horizontally.

If air within the core of the thunderstorm has enough upward momentum, however, it can continue to rise past its equilibrium level and overshoot into the stratosphere. The upward motion decelerates in the stable stratosphere, and most of the air descends back into the upper troposphere. A fraction, however, may remain in the stratosphere through turbulent mixing with stratospheric air. The mixing of the convectively lofted air mass with stratospheric air makes it neutrally buoyant at stratospheric altitudes and results in irreversible cross-tropopause transport of tropospheric air and moisture. This has been demonstrated in both modeling and observational studies (*2*, *13*).

The detailed processes responsible for convective transport and mixing between the troposphere and stratosphere have lacked a consistent physical model. Thus, uncertainties remain in estimating the amount of tropospheric air and water injected into the stratosphere by convection, and the impact of these events on large-scale stratospheric composition and chemistry are poorly constrained. There has been substantial progress in the development of observational datasets and models over the past 10 years. For example, a decade-long record of processed Next Generation Weather Radar data provides precise information about the time, location, and morphology of convection, including determination of cross-tropopause overshoot depth over the United States (*14*). Similarly, analysis of infrared imagery acquired by the Geostationary Operational Environmental Satellite has yielded a long-term record of overshooting convection over the North American continent, which has led to the identification and characterization of hundreds of storms that produce above-anvil cirrus plumes (*15*). High-resolution in situ sampling of convective outflow in the lower stratosphere has been used to constrain the amount of water added by isolated storms (*5*).

O'Neill *et al.* propose a "hydraulic jump" model that accounts for multiple observed features of thunderstorms that produce above-anvil cirrus plumes. The model provides a coherent framework that resolves apparent contradictions among other hypothesized mechanisms for this important troposphere-to-stratosphere transport pathway. To strengthen their case, O'Neill *et al.* draw on existing observational data to elucidate their proposed mechanism and demonstrate its likelihood in the real atmosphere.

Improved understanding of the physics of supercell thunderstorms—in conjunction with new observations being gathered by NASA's Dynamics and Chemistry of the Summer Stratosphere mission (over the next two summers, over the US midwest), which is probing outflow in the stratosphere from a broader range of storm types and environmental conditions than that of any prior mission—promises major advancements in addressing the remaining uncertainties in convective troposphere-to-stratosphere transport, quantifying its impact at present, and forecasting how this mechanism will respond to changing climate conditions. ■

Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA.
Email: jsmith@huarp.harvard.edu

REFERENCES AND NOTES

1. P. K. Wang, *J. Geophys. Res.* **108**, D002581 (2003).
2. G. L. Mullendore, D. R. Durran, J. R. Holton, *J. Geophys. Res.* **110**, D005059 (2005).
3. C. R. Homeyer, J. D. McAuliffe, K. M. Bedka, *J. Atmos. Sci.* **74**, 1617 (2017).
4. T. F. Hanisco *et al.*, *Geophys. Res. Lett.* **34**, L04814 (2007).
5. J. B. Smith *et al.*, *J. Geophys. Res. Atmos.* **122**, 9529 (2017).
6. J. G. Anderson, D. M. Wilmouth, J. B. Smith, D. S. Sayres, *Science* **337**, 835 (2012).
7. M. E. O'Neill, L. Orf, G. M. Heymsfield, K. Halbert, *Science* **373**, 1248 (2021).
8. I. M. Held, B. J. Soden, *Annu. Rev. Energy Environ.* **25**, 441 (2000).
9. R. Müller *et al.*, *Earths Futur.* **4**, 25 (2016).
10. S. Solomon *et al.*, *Science* **327**, 1219 (2010).
11. R. J. Trapp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19719 (2007).
12. S. L. Van Klooster, P. J. Roebber, *J. Clim.* **22**, 3317 (2009).
13. W. Frey *et al.*, *Atmos. Chem. Phys.* **15**, 6467 (2015).
14. J. W. Cooney, K. P. Bowman, C. R. Homeyer, T. M. Fenske, *J. Geophys. Res.* **123**, 329 (2018).
15. K. Bedka, E. M. Murillo, C. R. Homeyer, B. Scarino, H. Mersiovsky, *Weather Forecast.* **33**, 1159 (2018).

10.1126/science.abl8740

CHEMISTRY

More versatile synthesis of oligonucleotides

A phosphotriester chemistry may enable sustainable oligonucleotide production

By **Pasi Virta**

Solid-phase phosphoramidite chemistry developed by Beaucage and Caruthers (*1*) couples appropriately protected nucleosides as 3′-phosphoramidites to the 5′-hydroxy group of an immobilized growing oligonucleotide. The phosphite triester intermediate is then either oxidized or sulfurized, the 5′-protection of the connected nucleotide is removed, and the coupling cycle is repeated. The superior coupling efficiency, availability of the reagents, standardized coupling protocols, and established automated instrumentation of this method have limited the use of alternative technologies (such as *H*-phosphonate and phosphotriester coupling chemistry) (*2*–*4*) to small-molecule synthesis and applications that require a special chemical orthogonality. On page 1265 of this issue, Huang *et al.* (*5*) report on a phosphotriester-based technology, a "P(V) platform," that has oligonucleotide-chain elongation efficiency competitive with the phosphoramidite coupling cycle. A unified coupling condition is used for native phosphodiester, either diastereomer of phosphorothioate, and phosphorodithioate linkages.

Therapeutic oligonucleotides (oligomers of ~20 nucleotides) have become an effective drug modality for treating human diseases. To date, 12 oligonucleotide drugs have reached the market, and hundreds of oligonucleotides are in clinical trials. Diagnostic tools and aptamers (*6*) (antibody-like high-affinity binders consisting of nucleotides) are other important applications of oligonucleotides. The increasing demand for such therapeutics has challenged scientists to develop alternative strategies to make oligonucleotide production more sustainable.

In solid-phase synthesis, an excess of reagents is usually needed to drive the reactions to completion, and extensive solvent washing is used to remove them. Some of these reagents are hazardous, toxic to aquatic life, or are produced from nonrenewable petroleum-based sources. Thus, the current standard oligonucleotide synthesis along with its downstream processing present sustainability issues (*7*). Furthermore, real-time monitoring of the individual reaction steps is limited, which hampers optimization of the reactions. Short-term progress has improved the susta www.acsgcipr.org inability of oligonucleotide synthesis (*8*), including minimized reagent excesses, recovery of reagents, and the use of alternative greener solvents.

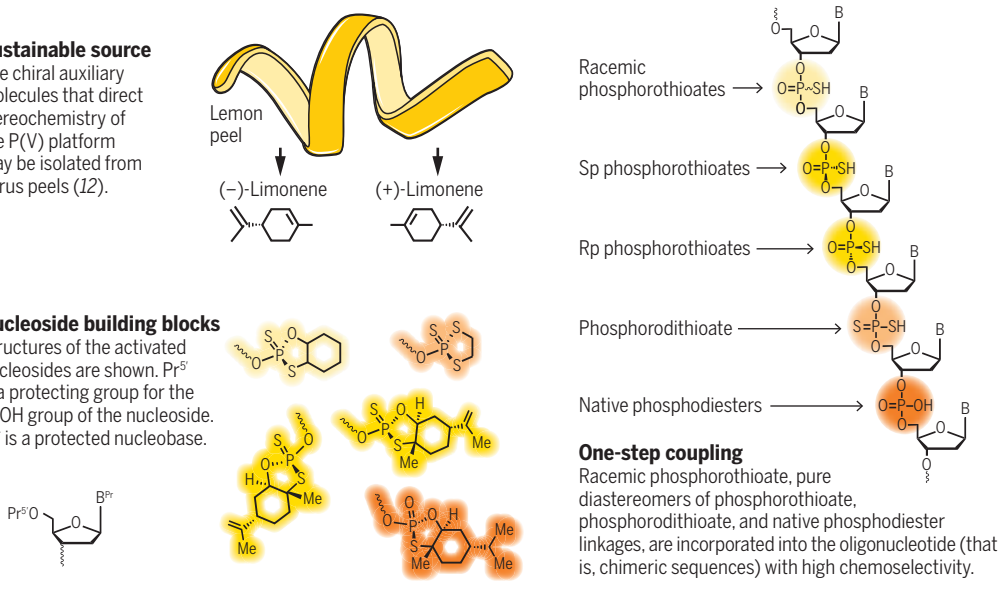
With regard to long-term goals, solution-phase technologies for the biopolymer assemblies (for both peptides and oligonucleotides) have been developed (*9*). In these technologies, the growing oligomers are isolated by membrane filtration or precipitation with appropriate soluble supports (*10*). Relatively short oligonucleotide segments, or blockmers, can be assembled that may be chemically or enzymatically ligated further to gain the full-length oligonucleotide products (*11*). Liquid-phase technologies are not limited to a certain coupling chemistry, but the extra oxidation and sulfurization step dilutes the value of the phosphoramidite coupling cycle.

Huang *et al.* extended their recent studies (*12*) in which limonene-based chiral auxiliaries for the P(V) platform were introduced. The synthetic compatibility of the platform with automated solid-supported synthesis of multiple chimeric 16-nucleotide oligomer oligonucleotides has now been demonstrated. Two equivalents of activated nucleotide building blocks in the presence of an organic base resulted in fast and quantitative coupling to the growing oligonucleotide. This platform can insert the backbone repertoire anywhere in the sequence with high stereo- and chemical selectivity (see the figure). This simple coupling chemistry is especially attractive for liquid-phase technologies.

Another challenge of phosphoramidite chemistry relates to the chirality of the phosphorothioate analogs. In current oligonucleotide drugs, the phosphorothioate linkages are introduced in their racemic form. Such drugs are mixtures of hundreds of thousands of stereoisomers, and each or a certain population of these isomers may

Phosphotriester-based oligonucleotide synthesis

Simple and efficient coupling and the straightforward synthesis of the building blocks in the P(V) oligonucleotide synthesis of Huang *et al.* are an attractive alternative route for the production of oligonucleotides.



GRAPHIC: C. BICKEL/*SCIENCE*

Department of Chemistry, University of Turku, 20500 Turku, Finland. Email: pamavi@utu.fi

have their own contribution to the splice-switching efficacy, ribonuclease H activity, enzymatic stability, toxicity, protein corona formation, biodistribution, and delivery efficency (*13*). Enantiopure oligonucleotide phosphorothioates should also have improved efficacy.

Efficient chiral auxiliaries for the sulfurization step of the phosphoramidite coupling cycle have been developed (*9*), but their greenness and practicability for large-scale manufacturing are outstanding issues. For the P(V) platform, the purified isomers of limonene, a terpene extracted from citrus peels and used to produce chiral auxiliaries, can be prepared on a large scale. This feature, along with operationally simple and efficient coupling chemistry with a high degree of stereocontrol, makes the P(V) platform attractive for the production of enantiopure phosphorothioate oligonucleotides.

The next step for the P(V) platform will likely be to evaluate the scaling up of this process as well as the applicability of the platform for the liquid-phase technologies and compatibility with the alternative protecting group schemes. Although 4,4′-dimethoxytrityl (DMTr) works well as an acid labile 5′-protection for solid-phase assembly, severe problems are frequently encountered in solution because of the reversibility of the detritylation reaction and partial simultaneous acid-catalyzed deadenylation ("depurination") that occurs during deprotection. The simple coupling chemistry in the described P(V) platform combined with an atom-economic alternative protecting group scheme would be an attractive improvement for liquid-phase technologies and the sustainable production of oligonucleotides. ■

REFERENCES AND NOTES

1. S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859 (1981).
2. P. J. Garegg *et al.*, *Tetrahedron Lett.* **27**, 4051 (1986).
3. B. C. Froehler, P. G. Ng, M. D. Matteucci, *Nucleic Acids Res.* **14**, 5399 (1986).
4. C. B. Reese, *Org. Biomol. Chem.* **3**, 3851 (2005).
5. Y. Huang *et al.*, *Science* **373**, 1265 (2021).
6. G. T. Rozenblum, V. G. Lopez, A. D. Vitullo, M. Radrizzani, *Expert Opin. Drug Discov.* **11**, 127 (2016).
7. B. I. Andrews *et al.*, *J. Org. Chem.* **86**, 49 (2021).
8. Y. S. Sanghvi, V. T. Ravikumar, A. N. Scozzari, D. L. Cole, *Pure Appl. Chem.* **73**, 175 (2001).
9. A. Gimenez Molina, Y. S. Sanghvi, *Curr. Protoc. Nucleic Acid Chem.* **77**, e82 (2019).
10. H. Lönnberg, *Beilstein J. Org. Chem.* **13**, 1368 (2017).
11. A. Crameri, D. G. Tew, Novel processes for the production of oligonucleotides. PCT Patent WO2019121500A1 (2019).
12. K. W. Knouse *et al.*, *Science* **361**, 1234 (2018).
13. N. Iwamoto *et al.*, *Nat. Biotechnol.* **35**, 845 (2017).

ACKNOWLEDGMENTS

This article was written with the support of the American Chemical Society (ACS) Green Chemistry Institute Pharmaceutical Roundtable (GCIPR) (www.acsgcipr.org).

10.1126/science.abk3478

NEUROSCIENCE

Increased serotonin prevents compulsion in addiction

Extracellular serotonin decreases cocaine taking in mice

By **Katsuhiko Miyazaki** *and* **Kayoko W. Miyazaki**

Drug addiction is a disease characterized by compulsive drug seeking and consumption despite negative consequences (*1*). About 20% of cocaine users lose self-control and are eventually diagnosed as addicted. Cocaine blocks dopamine, noradrenalin, and serotonin uptake through inhibition of the transporters of these neurotransmitters. The rewarding effect of cocaine is mainly induced by blocking the dopamine transporter (DAT), which increases extracellular dopamine concentrations. Intriguingly, mice that lack DAT still experience rewarding effects from cocaine, and this is abolished when both DAT and serotonin transporter (SERT) are deleted in mice (*2*, *3*). On page 1252 of this issue, Li *et al.* (*4*) show that increased extracellular serotonin antagonizes transition to compulsive cocaine intake by inducing a presynaptic depression in which serotonin inhibits excitatory synaptic transmission to the dorsal striatum (DS) through the serotonin receptor, 5-hydroxytryptamine receptor 1B (5-HTR_{1B}), which is expressed on the presynapse of cortical inputs.

For experimental animal studies in addiction, self-administration (SA) followed by punishment is commonly used (*5*). When cocaine SA subsequently becomes accompanied with punishment (such as foot shock), animals show two types of behavioral response. One is reduced cocaine SA (renouncers, ~80% of mice), and the other is continued cocaine SA (perseverers, ~20% of mice) (*6*). Direct causality between activation of dopamine neurons in the ventral tegmental area (VTA) and addiction was revealed by using optogenetic dopamine neuron self-stimulation (oDASS), which enabled inducible release of dopamine from the VTA (*7*). The use of oDASS results in a bimodal distribution of compulsive (perseverers, >50%) and noncompulsive (renouncers, <50%) individuals. Why is there a larger proportion of perseverers in the oDASS group than in the cocaine SA group?

Neural Computation Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan. Email: miyazaki@oist.jp

Li *et al.* found that cocaine SA in mice that express a SERT that blocks cocaine binding (SertKI mice) (*8*) does not increase extracellular serotonin concentrations, but unexpectedly, this causes increased transition to compulsion compared with that in wild-type (WT) mice. The population of perseverers of SertKI mice in cocaine SA is comparable with that of WT mice in the oDASS group. The authors also found that although oDASS-mediated addiction results in compulsion in more than half of individuals, systemic administration of citalopram, which is a serotonin-selective reuptake inhibitor, during acquisition of addiction through oDASS reduces compulsion transition to the same amount in cocaine SA of WT mice (~20% perseverers). These results indicate that increased extracellular serotonin prevents the transition to compulsion.

What neural and cellular mechanisms contribute to these serotonin effects? In oDASS-mediated addiction, perseverers were associated with increased neural activity in the orbitofrontal cortex (OFC), and chemogenetic inhibition of the OFC reduced transition to compulsion (*7*). Compulsive lever pressing in oDASS-mediated addiction was associated with increased neural activity in the projection terminals from the OFC to the DS, which is related to habitual drug use, and glutamatergic transmission from the OFC to the DS was permanently potentiated in perseverer mice (*9*). Consistently, Li *et al.* revealed that in both cocaine SA and oDASS groups, perseverers show synaptic potentiation of afferents from the OFC to the DS, which did not occur in renouncers. Bath application of serotonin to OFC-DS terminals in brain slices induced a presynaptic depression of excitatory transmission, which was prevented by the 5-HTR_{1B} antagonist NAS181. Moreover, in cocaine SA-mediated addiction, ablation of 5-HTR_{1B} (*10*) increases the proportion of perseverers to >50% because of the lack of the presynaptic depression.

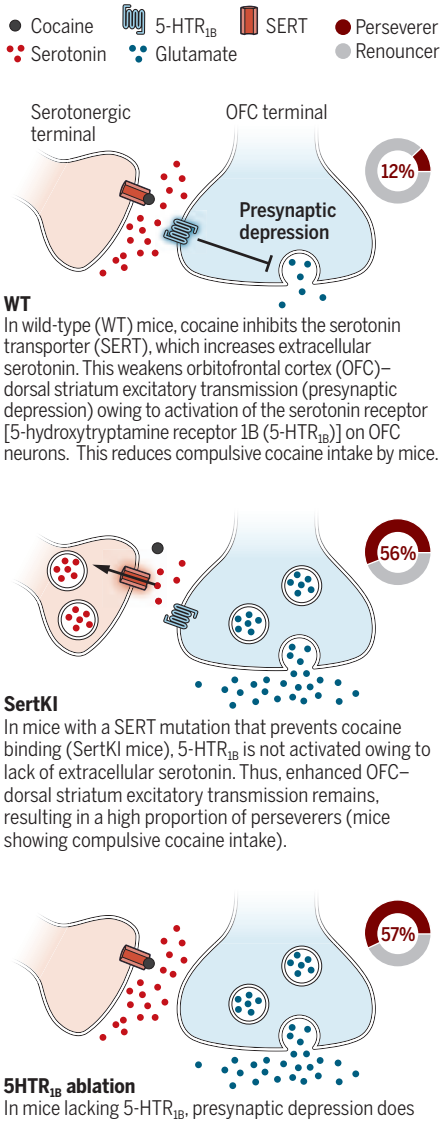
Li *et al.* elegantly show that in WT mice, cocaine binds to SERT to block serotonin reuptake and increases extracellular serotonin concentrations, causing the presynaptic depression (see the figure). When 5-HTR_{1B} at OFC neuron terminals projecting to the DS is activated, synaptic transmission in the

OFC-DS pathway is reduced by the presynaptic depression. Because enhancement of glutamatergic OFC-DS synaptic transmission is causally related to transition to compulsion in addiction, 5-HTR_{1B}-induced presynaptic depression prevents WT mice from transition to compulsion. In SertKI mice, because cocaine cannot bind to SERT, extracellular serotonin concentrations do not change. The OFC-DS synaptic transmission in SertKI mice is more efficient than in WT mice, and this induces increased transition to compulsive cocaine taking. When 5-HTR_{1B} is ablated in mice, although cocaine binds to SERT and extracellular serotonin concentrations increase, the presynaptic depression at OFC-DS synapses does not occur because of lack of 5-HTR_{1B}. Increased OFC-DS synaptic transmission by cocaine SA thus induces increased perseverers in these mice.

The study of Li *et al.* poses further questions regarding mechanisms of compulsion in drug addiction. After initial acquisition before punishment, are mice already shaped as renouncers or perseverers with different strengths of OFC-DS transmission, or do these two behaviors differentiate during relearning with punishment? Does transition to compulsion occur deterministically or stochastically? Does serotonin prevent transition to compulsion in taking other addictive drugs, such as opioids? Addictive drugs commonly increase dopamine in the mesolimbic pathway, and this would play a crucial role in the induction of compulsion in addiction (*11*). Further studies will clarify whether activation of 5-HTR_{1B} during addictive drug intake can prevent transition to compulsion. Compulsive cocaine-taking individuals may show increased neural activity in the OFC, as in the case of oDASS-mediated addiction. In addition to affecting OFC-DS transmission, serotonin may modulate OFC neural activity to shift perseverers to renouncers in rats (*12*). Forebrain serotonin depletion induced compulsive cocaine seeking in rats, and systemic injection of a 5-HTR_{2C} agonist reduced compulsive cocaine seeking (*12*). Serotonergic modulation in the OFC was also reported in a reward waiting task (*13*) and in reversal learning (*14*), indicating that a common mechanism by which serotonin regulates OFC neural activity may exist. Further studies should clarify the neural mechanism underlying serotonin modulation of the transition to compulsion in drug addiction, what agents specific to serotonin receptors can be used, and when these agents can be administered to potentially treat drug addiction. ■

Serotonin inhibits addiction

Extracellular serotonin antagonizes the transition to compulsive cocaine intake in mice by modulating glutamatergic transmission in the dorsal striatum.



REFERENCES AND NOTES

1. G. F. Koob, N. D. Volkow, *Neuropsychopharmacology* **35**, 217 (2010).
2. B. A. Rocha *et al.*, *Nat. Neurosci.* **1**, 132 (1998).
3. I. Sora *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7699 (1998).
4. Y. Li *et al.*, *Science* **373**, 1252 (2021).
5. B. J. Everitt, T. W. Robbins, *Nat. Neurosci.* **8**, 1481 (2005).
6. V. Deroche-Gamonet, D. Belin, P. V. Piazza, *Science* **305**, 1014 (2004).
7. V. Pascoli, J. Terrier, A. Hiver, C. Lüscher, *Neuron* **88**, 1054 (2015).
8. L. D. Simmler, R. D. Blakely, *ACS Chem. Neurosci.* **10**, 3053 (2019).
9. V. Pascoli *et al.*, *Nature* **564**, 366 (2018).
10. K. M. Nautiyal *et al.*, *Neuron* **86**, 813 (2015).
11. C. Lüscher, T. W. Robbins, B. J. Everitt, *Nat. Rev. Neurosci.* **21**, 247 (2020).
12. Y. Pelloux, R. Dilleen, D. Economidou, D. Theobald, B. J. Everitt, *Neuropsychopharmacology* **37**, 2505 (2012).
13. K. Miyazaki *et al.*, *Sci. Adv.* **6**, eabc7246 (2020).
14. J. Alsiö *et al.*, *Cereb. Cortex* **31**, 1090 (2021).

10.1126/science.abl6285

MOLECULAR BIOLOGY

Getting droplets into shape

Protein clusters at interfaces control sizes and properties of biomolecular condensates

By **Wilton T. Snead**[1] *and* **Amy S. Gladfelter**[1,2]

Condensed droplets of protein and nucleic acid regulate a variety of cellular functions, from signaling to RNA processing (*1*). Many of these biomolecular condensates display liquid-like properties and are expected to evolve in size over time (called coarsening) to form a single, large droplet at equilibrium with its surroundings (*2*). However, condensates in cells frequently remain as a dispersed emulsion of droplets that do not increase beyond a certain size and maintain distinct molecular identities despite sharing a common environment. Various mechanisms that restrict coarsening have been proposed, including physical barriers such as chromatin or actin networks that keep droplets apart and prevent fusion (*3*), active processes that suppress material transfer between droplets (called Ostwald ripening) (*4*), and "kinetic arrest" due to protein binding site saturation (*5*). On page 1218 of this issue, Folkmann *et al.* (*6*) find another regulator of droplet coarsening: nanometer-scale protein clusters adhered to droplet interfaces.

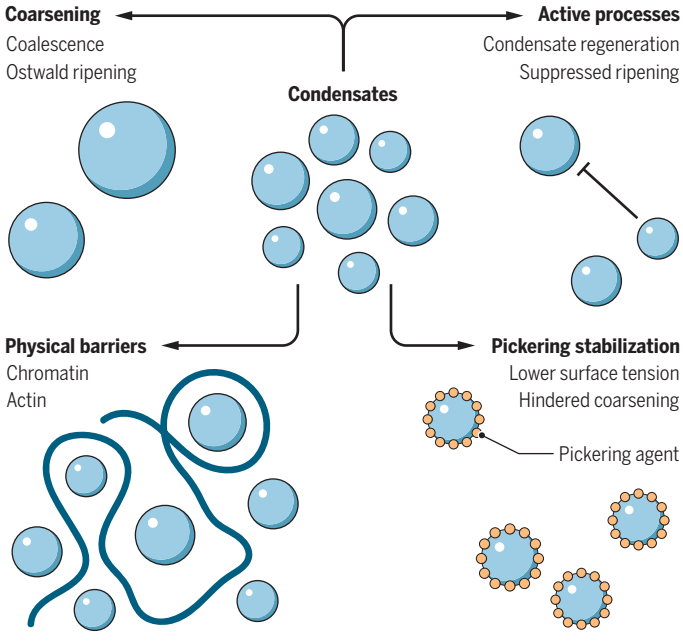
Although interfacial stabilizers are conceptually new in the study of biomolecular condensates, these materials are well established in other fields such as chemical engineering and food science (*7*). Specifically, liquid-liquid emulsions can be stabilized by materials that reduce surface tension, a key parameter that determines how quickly droplets coarsen. A common example is a surfactant or detergent that stabilizes oil-in-water emulsions. Interfacial stabilizers can also include solid particles, collectively called "Pickering agents." These materials were originally described as stabilizers of inorganic emulsions (*8*) and are now used in diverse applications such as drug delivery, in which solid particulates composed of biocompatible polymers stabilize drug-carrying emulsions and assist with the controlled timing of drug release (*7*).

[1]Department of Biology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2]Marine Biological Laboratory, Woods Hole, MA, USA. Email: amyglad@unc.edu

GRAPHIC: KELLIE HOLSKI/*SCIENCE*

Mechanisms of droplet-size regulation in cells

If unrestricted, coarsening due to coalescence or Ostwald ripening leads to larger droplets over time (top left). Examples of mechanisms that restrict droplet coarsening include active processes (top right) or physical barriers (bottom left). Folkmann *et al*. report that gel-like protein clusters adsorbed to interfaces can act as "Pickering agents" that stabilize droplet sizes (bottom right).



Folkmann *et al*. aimed to better understand how condensates in the germline of *Caenorhabditis elegans* called P granules, composed of RNA and the protein PGL-3 (P granule abnormality protein 3), form and dissolve rapidly but resist coarsening during the transition from egg to embryo. Specifically, P granules become asymmetrically distributed in the polarized zygote (fertilized oocyte), but the cellular factors that regulate P granule dynamics and size during this process are poorly understood. The authors were motivated by previous findings that a protein called MEG-3 (maternal-effect germline defective-3) forms a gel-like scaffold around PGL-3 condensates (*9*). This scaffold stabilized droplets under conditions that favor droplet dissolution (*9*), suggesting that MEG-3 may influence the surface tension and coarsening of P granules. Indeed, Folkmann *et al*. reveal that the MEG-3 scaffold is composed of nanometer-scale protein clusters adsorbed to PGL-3 condensate interfaces, which reduce surface tension and stabilize droplet sizes over time. Notably, these clusters do not affect molecular rearrangement within droplets or exchange of materials at the surface. MEG-3 is also asymmetrically distributed within the zygote and is required for the accumulation of P granules on one side of the polarized cell. The authors develop a theoretical model, supported by experiments, to reveal that droplet surface tension is preferentially reduced and that P granules are stabilized on the side of the cell where MEG-3 is enriched, whereas droplets dissolve on the opposite side. Thus, the study describes one of the first examples of a Pickering stabilizer of condensate emulsions in cells (see the figure).

The findings of Folkmann *et al*. raise the question of whether Pickering agents serve a broader function of controlling condensate sizes throughout the cell. Specifically, a variety of condensates display layered, core-shell morphologies (*10*), suggesting that the distinct material properties of the shell may act to limit condensate size through Pickering stabilization. Although the P granule system described by the authors focuses on a protein-based Pickering agent, RNA is a key component of the core-shell structure of many other condensates. For example, spliced RNAs are concentrated at the shells of nuclear speckles (*11*), and an architectural, long noncoding RNA controls the core and shell identities of paraspeckles (*12*). In both examples, RNA sequence and structure define the localization of specific molecules to the shell region. The mechanisms by which shell-localized RNAs recruit specific Pickering agents will be an interesting topic for future work.

Another possibility is that interfacial RNA molecules may regulate condensate size directly. Indeed, recent work showed that an accumulation of negative charge through nucleic acid enrichment at condensate interfaces can lower droplet surface tension, a key driver of coarsening (*13*). Interfacial RNAs may also hinder coarsening by forming solid assemblies at condensate interfaces that act as Pickering agents. In support of this idea, recent work found that some RNAs form stable, gel-like layers at condensate interfaces through intermolecular RNA-RNA interactions (*14*). Notably, an RNA helicase was found to limit both RNA self-assembly and RNA recruitment to condensates (*14*), raising the possibility that active cellular processes may further tune the function of RNAs as Pickering agents.

Proteins involved in biomolecular condensation are frequently classified as "scaffolds" that phase-separate or "clients" that passively partition into droplets (*1*). The findings of Folkmann *et al*. suggest a need for another category: proteins that form solid assemblies that are adhered to droplet interfaces. Although the "molecular grammar" of the driving forces for phase separation are becoming well understood (*15*), the sequence features that lead to stable, gel-like structures at interfaces remain to be explored. Moreover, the mechanisms by which Pickering agents help define droplet identity are still unclear. Specifically, cellular Pickering agents are likely more complex than simple adsorbates but instead engage in specific interactions with the underlying liquid phase to help maintain a distinctive material state. Further study of interfacial assemblies will not only improve our understanding of the regulatory knobs that control phase separation in cells but will also inform the rational design of engineered condensates. In particular, Pickering agents may facilitate tuning of interfacial properties and molecular permeability for applications in biosynthesis and controlled drug delivery. ■

GRAPHIC: A. MASTIN/*SCIENCE*

REFERENCES AND NOTES

1. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285 (2017).
2. J. Berry, C. P. Brangwynne, M. Haataja, *Rep. Prog. Phys.* **81**, 046601 (2018).
3. D. S. W. Lee, N. S. Wingreen, C. P. Brangwynne, *Nat. Phys.* **17**, 531 (2021).
4. D. Zwicker, A. A. Hyman, F. Jülicher, *Phys. Rev. E* **92**, 012317 (2015).
5. S. Ranganathan, E. I. Shakhnovich, *eLife* **9**, e56159 (2020).
6. A. W. Folkmann, A. Putnam, C. F. Lee, G. Seydoux, *Science* **373**, 1218 (2021).
7. Y. Yang *et al*., *Front. Pharmacol.* **8**, 287 (2017).
8. S. U. Pickering, *J. Chem. Soc.* **91**, 2001 (1907).
9. A. Putnam, M. Cassani, J. Smith, G. Seydoux, *Nat. Struct. Mol. Biol.* **26**, 220 (2019).
10. C. M. Fare, A. Villani, L. E. Drake, J. Shorter, *Open Biol.* **11**, 210137 (2021).
11. J. Fei *et al*., *J. Cell Sci.* **130**, 4180 (2017).
12. T. Yamazaki *et al*., *Mol. Cell* **70**, 1038 (2018).
13. I. Alshareedah, G. M. Thurston, P. R. Banerjee, *Biophys. J.* **120**, 1161 (2021).
14. D. Tauber *et al*., *Cell* **180**, 411 (2020).
15. J. Wang *et al*., *Cell* **174**, 688 (2018).

ACKNOWLEDGMENTS

A.S.G. is a scientific advisor for Dewpoint Therapeutics. A.S.G. is supported by the National Institutes of Health (R01-GM081506), the Howard Hughes Medical Institute (HHMI) Faculty Scholars program, and the Air Force Office of Scientific Research (FA9550-20-1-0241). W.T.S. is supported by a Ruth L. Kirschstein National Research Service Award (NRSA) Postdoctoral Fellowship from the National Institutes of Health (F32-GM136055).

10.1126/science.abl6282

GENE THERAPY

Mitochondrial disease: Replace or edit?

Questions over mitochondrial replacement suggest a role for mitochondrial DNA editing

By Eli Y. Adashi[1], Donald S. Rubenstein[2], Jim A. Mossman[3], Eric A. Schon[4], I. Glenn Cohen[5]

Maternal transmission of a wide spectrum of mutant alleles of mitochondrial DNA (mtDNA) gives rise to metabolic diseases of differing penetrance, presentation, and prognosis to offspring (*1*). Only a proportion of these mitochondrial diseases are caused by mtDNA mutations, and when multiple mtDNA haplotypes coexist in cells—called heteroplasmy—the resulting mtDNA composition can cause disease (*1*). An estimated 1000 affected children are born annually in the US, and a prevalence rate of ~1 in 5000 is estimated in the UK (*2*). Effective mitigation includes supportive management of diabetes, deafness, and cardiac involvement. However, recent advances give hope that these diseases might be alleviated by one of two distinct paths of germline remediation (*3, 4*): Oocytes or zygotes at risk for mitochondrial disease could have replacement unaffected mitochondria (*3*), or mtDNA could undergo corrective genome editing (*4*).

First conceptualized ~25 years ago, mitochondrial replacement therapy (MRT) is undergoing preclinical development and early-stage clinical application. In general, MRT involves the transfer of genetic material (pronuclei, maternal spindles, or polar bodies) from at-risk oocytes or zygotes into enucleated donor cells with unaffected mitochondria (*5–9*) (see the figure). Prospective mothers are typically identified as at-risk by a family history of mitochondrial disease. The leading edge of this biotechnology, pronuclear transfer (PNT) and maternal spindle transfer (MST), have been the subject of extensive preclinical interrogation over the past decade (*5–8*). Both PNT and MST techniques have demonstrated proof-of-concept that disease-associated mtDNA can be effectively substituted by an unaffected donor mtDNA haplotype. The more recent polar body transfer (PBT) technique has yet to be explored in comparable detail (*9*). Largely devoid of cytoplasm, polar bodies may be particularly suited to limiting the "carry-over" of mutant maternal mtDNA during transfer (*9*). Moreover, both first polar body (PB1) and second polar body (PB2) transfer techniques gave rise to normal diploid zygotes, suggesting this is a promising approach (*9*).

A human infant has already been born as a result of MRT carried out by MST (*8*). Moreover, MRT has garnered regulatory approval in several jurisdictions. This outcome is at least in part because mtDNA sequences are unperturbed by MRT, unlike mtDNA editing approaches. Notably, in the UK, at least 17 requests for MRT were granted by the Human Fertilisation and Embryology Authority (HFEA) between 31 August 2017 and 31 October 2019. It is unclear how many children have been born through the use of MRT and what their health status is, owing to requests for privacy by parents. The results of the PNT clinical trial (NCT04113447) are still pending.

Common to all forms of MRT is the risk of "carry-over" of perinuclear mtDNA into the reconstituted oocyte or zygote. It is nearly impossible to transfer mitochondria-free nuclear material, and the potential minimization (if not elimination) of carry-over remains the subject of active investigation. Reports of 2.37 to 9.23% mutant mtDNA load in tissues from the child born after MST affirm the reality of the carry-over phenomenon as well as the postnatal persistence thereof (*8*). Carried-over parental mtDNA was documented following MST or PBT with material from nonhuman primates (*10*). What remains poorly understood is whether carried-over mutant maternal mtDNA might outgrow its intact donor counterpart to reestablish the disease phenotype, which has implications for safety and efficacy. Along these lines, "reversion" may all but invalidate the utility of the MRT paradigm.

First noted in nonhuman primate oocytes of female progeny produced by MST, reversion has been repeatedly observed. Studies of adult MST-derived rhesus macaques and their offspring revealed maternal heteroplasmy of mutant alleles (≤16.6%) in several internal organs (*10*). Assessment of subsequent progeny in turn revealed that the paternal mtDNA contribution rose to 33% in select tissues and organs in two individuals (*10*). Moreover, a total of 8 of the 39 human embryonic stem cell lines (derived from MRT blastocysts) studied to date in vitro displayed some reversion of mtDNA (*11*). Yet to be clarified is the possibility that this constitutes a cell culture phenomenon. Additional research is needed to clarify the root cause(s) of the reversion phenomenon.

Enzymes that edit mtDNA, akin to the nuclear counterparts, have long been sought for the treatment of mitochondrial diseases. Regrettably, however, such precision editing has been held back by the inability to deliver guide RNAs for CRISPR-Cas editing into the membrane-bound mitochondrial matrix owing to hydrophobicity and the strong electrochemical potential across the inner mitochondrial membrane. These limitations mean that RNA-free programmable nucleases have been the focus of such efforts. Pathogenic mtDNA of hybrid (mouse–human) mammalian oocytes were targeted by transcription activator–like effector nucleases (TALENs) or zinc finger nucleases (ZFNs) (*12*), which recognize specific DNA sequences and allow targeted editing through endonuclease-induced double-stranded DNA breaks. The mtDNA comprised multiple pathogenic alleles of the mitochondrial diseases Leber's hereditary optic neuropathy (LHON) and neuropathy, ataxia, and retinitis pigmentosa (NARP), and editing was demonstrated to be successful in vitro (*12*).

The prospect of precision germline editing of mutant mtDNA has recently received a boost by the development of a CRISPR-free mitochondrial DNA base editor (*4*) (see the figure). Capitalizing on the distinctive properties of an interbacterial toxin (DddA) of the deaminase superfamily, a DddA-derived cytosine base fusion editor (DdCBE), which can access mtDNA through the addition of a mitochondria-targeting signal, was engineered (*4*). DdCBE catalyzes specifically targeted T•A to C•G transition mutations with a high degree of specificity. This base editor does not induce a double-stranded DNA break (*4*), which would normally destroy the whole molecule because mtDNA lacks a double-strand break repair mechanism, a feature exploited by the TALEN and ZFN approaches (*12*). Off-target activity within mtDNA of human embryonic kidney HEK 293T cells was deemed to be low (0.03 to 0.13% off-target single-nucleotide variants were detected in selected genes), and off-target activity of the nuclear genome was ruled out (*4*). Additional research is required to ensure that editing of mtDNA does not give rise to off-target inser-

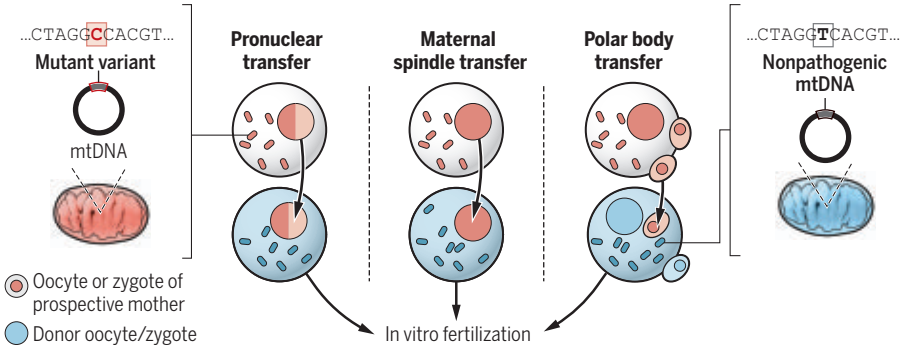
[1]Department of Medical Science, Brown University, Providence, RI, USA. [2]Prisma Health–Upstate, Greenville, SC, USA. [3]Aspen Neuroscience, San Diego, CA, USA. [4]Department of Neurology, and Department of Genetics and Development, Columbia University, New York, NY. [5]Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics, Harvard Law School, Harvard University, Cambridge, MA, USA. Email: eli_adashi@brown.edu

Altering mitochondrial DNA

There are two distinct paths of germline remediation to prevent mitochondrial diseases in offspring.

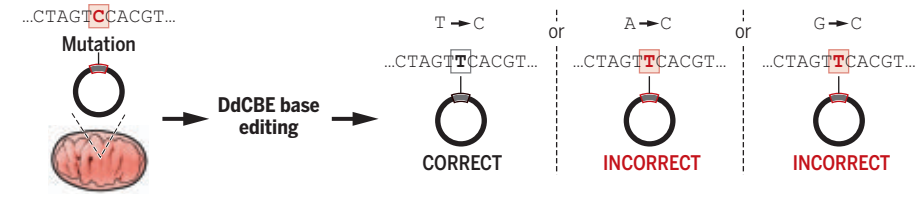
Mitochondrial replacement therapy

Mitochondrial replacement therapy can occur through three methods: pronuclear transfer, maternal spindle transfer, and polar body transfer into enucleated donor oocytes or zygotes. The reconstituted cells contain the unaffected mitochondrial DNA (mtDNA) and can then be used for in vitro fertilization.



Mitochondrial genome editing

Mitochondrial genome editing aims to use base editors to convert the full spectrum of mtDNA substitution mutations. The current DdCBE editor converts C to T when C is preceded by a T and so it corrects transition mutations (T→C), but transversion mutations (A→C or G→C) will also wrongly be edited back to T.



tions, deletions, and rearrangements, as can occur with CRISPR-Cas.

Editing of mtDNA of at-risk oocytes or zygotes could, in time, prevent the transmission of mitochondrial diseases. Realizing the full potential of base editing at the germline level will require the redress of several shortfalls. Topping this list is the reliance of DdCBE on mtDNA replication, when the double-stranded molecule temporarily becomes single stranded and one of the two parental strands is editable. This requirement alone limits the editing efficiency of DdCBE to 50% (*4, 13*). Moreover, mtDNA replication, which occurs in blastocyst-stage embryos ~5 days after fertilization, is unlikely to occur during earlier stages of human embryonic development, which is when remedial germline editing is likely to be applied. It would thus appear that, for now, DdCBE may help reduce, but not eliminate, some deleterious mtDNA alleles that contribute to mitochondrial diseases (*13*). However, a ~50% allele frequency correction may still be efficacious if the mutant allele frequency drops below a mitochondrial disease–promoting threshold.

The editing specificity of DdCBE is limited to C's preceded by a T and thus to the conversion of a mutant CG base pair to TA (*4*). Approximately 10% of the 93 confirmed pathogenic mutations currently listed in MitoMap (www.mitomap.org) may be amenable to correction by DdCBE. This estimate is consistent with the theoretical 6.25% frequency of a TC dinucleotide present randomly in DNA. Notably, DdCBE reverts C back to T, thereby limiting this base editor to correcting a transition mutation (i.e., T→C). However, transversion mutations (e.g., A→C or G→C; which are rarer than transitions) will also be edited back to T, resulting in the wrong sequence. It would thus appear that editing mtDNA of oocytes or zygotes at risk for mitochondrial disease will benefit not only from refinements of the properties of DdCBE but also from further expansion of the editing range by altogether new base editors (*3*). Additionally, mitochondrial base editors will likely prove incapable of repairing pathogenic insertions and deletions, such as m.*3271*delT or m.*5537_5538*insT. Currently, every reasonable preclinical effort should be made to explore the therapeutic potential as well as the safety and efficacy of existing and future base editors of mtDNA. Much could be learned in this context from the accumulated experience in editing the nuclear genome of the human embryo (*14*). But progress should be gradual so as to avoid clinical and regulatory setbacks.

MRT and mtDNA editing have so far been considered as distinct, autonomous approaches to the germline prevention of mitochondrial diseases. In principle, however, subject to the availability of a palette of mtDNA base editors, consideration could be given to the sequential application of these therapeutic modalities. In one potential scenario, MRT-reconstituted oocytes or zygotes could be subjected to mtDNA editing to reduce or eliminate carried-over mutant maternal mtDNA. Such a therapeutic paradigm could also undercut the prospect of reversion. This combination therapy could optimize the germline remediation of mitochondrial diseases in the future. Realization of such a vision, however, will require that the editing efficiencies of future mtDNA base editors substantially increase compared with DdCBE (*4*). Moreover, future mtDNA base editors must not give rise to off-target insertions, deletions, and rearrangements.

Torn between uncertainty and promise, the germline remediation of mitochondrial diseases is in flux. High hopes for the MRT paradigm have given way to a measure of uncertainty over its safety and efficacy: Concerns over the carry-over of mutant maternal mtDNA alleles and postnatal reversion continue to haunt the field. The dependence of MRT on a donor egg or zygote constitutes another challenge replete with ethical, legal, and religious considerations. Using MRT in the US remains infeasible owing to a renewable prohibitory legislation from 2015 (*15*). It is probably just a matter of time before a growing number of mtDNA base editors become available to remedy many of the mutant alleles currently beyond reach. Editing mtDNA could assume a leading (donor-independent) role in the prevention of mitochondrial diseases. In the interim, however, every effort must be made to assure the safety and efficacy of MRT. Further progress in this arena should ensure that relief of the attendant human misery will materialize before long. ■

REFERENCES AND NOTES

1. D. C. Wallace, *Nat. Genet.* **50**, 1642 (2018).
2. G. S. Gorman, J. P. Grady, D. M. Turnbull, *N. Engl. J. Med.* **372**, 885 (2015).
3. P. F. Chinnery, *N. Engl. J. Med.* **382**, 1855 (2020).
4. B. Y. Mok *et al.*, *Nature* **583**, 631 (2020).
5. L. A. Hyslop *et al.*, *Nature* **534**, 383 (2016).
6. M. Tachibana *et al.*, *Nature* **493**, 627 (2013).
7. D. Paull *et al.*, *Nature* **493**, 632 (2013).
8. J. Zhang *et al.*, *Reprod. Biomed. Online* **34**, 361 (2017).
9. T. Wang *et al.*, *Cell* **157**, 1591 (2014).
10. H. Ma *et al.*, *Hum. Reprod.* **36**, 493 (2021).
11. M. Yamada, K. Akashi, R. Ooka, K. Miyado, H. Akutsu, *Int. J. Mol. Sci.* **21**, 5880 (2020).
12. P. Reddy *et al.*, *Cell* **161**, 459 (2015).
13. M. Aushev, M. Herbert, *Nature* **583**, 521 (2020).
14. M. V. Zuccaro *et al.*, *Cell* **183**, 1650 (2020).
15. I. G. Cohen, E. Y. Adashi, *Science* **353**, 545 (2016).

ACKNOWLEDGMENTS

E.A.S. is a member of the Board of Directors of Raft Biosciences, Inc. I.G.C. is a member of the Ethics Advisory Board for Illumina.

10.1126/science.abg0491

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

POLICY FORUM

PUBLIC HEALTH

Digital exposure tools: Design for privacy, efficacy, and equity

Apps can cut transmission of SARS-CoV-2—but how do we ensure that they don't exacerbate public health inequities?

By **Susan Landau**[1,2]

Use of smartphone-based digital contact-tracing apps has shown promise in responding to the COVID-19 pandemic. But such apps can reveal very personal information; thus, their use raises important societal questions, not just during the current pandemic but as we learn and prepare for other inevitable outbreaks ahead. Can privacy-protective versions of such apps work? Are they efficacious? Because the apps influence who is notified of exposure and who gets tested—and possibly treated—we need to consider the apps in the context of health care equity. Exposure-notification apps are predicated on the assumption that if someone is informed of exposure, they will follow instructions to isolate. Such an expectation fails to take into account that isolation—and sometimes even seeking care when ill—is much harder for some populations than others. If apps are to work for all, and not make this worse for disadvantaged populations, there needs to be basic social infrastructure that supports testing, contact tracing, and isolation.

When severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reached Singapore in spring 2020, the nation instituted a centralized contact-tracing app in which identifiers linked to the user were shared with nearby phones running the app. If a person was diagnosed with SARS-CoV-2, the collected identifiers would be communicated to the Ministry of Health, which would use them to do contact tracing (*1*). Such proximity information—who is near whom for how long—is very revelatory. It can expose a journalist's source or the existence of an extramarital affair. Singapore had originally committed to the data being used solely for tracking SARS-CoV-2 exposures, but in January 2021, the government changed its policy, permitting the use of app data in criminal investigations.

Researchers and epidemiologists in Europe, the United States, and Australia developed privacy-protective decentralized apps in which contact information remains on the user's phone even while they can learn of their own exposure. Google and Apple, whose phones run the underlying technology, developed engineering infrastructure for such exposure-notification apps with a focus on privacy. By June 2020, the Google-Apple Exposure Notification (GAEN) infrastructure was available, and SwissCovid, the app from Switzerland's Federal Office of Public Health, was rolled out. Other European apps soon followed.

These exposure-notification apps exchange Bluetooth identifiers when two app users are in close proximity for at least 15 min (the apps also handle more sophisticated situations involving exposure by multiple people for shorter periods of time). To protect privacy, the identifiers are randomized strings that change every 10 to 20 min. If one of the app users later tests positive for SARS-CoV-2, they inform public health and then upload the identifiers the phone sent out during their contagious period. They do not share the identifiers they collected; those remain on their phone. All apps check a public health server at predetermined intervals daily and download newly posted identifiers. If one of these identifiers matches an identifier previously collected by the phone, that means the user was in proximity to someone who has tested positive.

The design choice meant that the infrastructure would not collect personal user information or track users—nor would the apps using the infrastructure be allowed to do so. Keeping proximity information on users' phones and letting the user decide whether to inform public health officials if they receive an exposure notification protect user privacy—but do so at a cost of failing to provide public health officials with data regarding the rate and nature of exposure. Such data would have been particularly useful at the start of the pandemic, when little was known about how SARS-CoV-2 spread.

The apps do not report time or location of exposure; indeed, the apps are designed to prevent anyone, including the user, from finding out such information. Even so, public perception of the apps was often negative, undoubtedly in part because of distrust in how technology companies and the government have stretched uses of peoples' data beyond the original collection purpose. The GAEN protocol's complexity makes it not entirely easy to explain; and the large amounts of misinformation surrounding the pandemic likely played a role as well.

EFFICACY AND OBSTACLES

Exposure-notification apps can be exceedingly useful (*2*). Although the app's anonymized design prevents tracking who received notifications and acted on them, a study provided statistical estimates indicating that between 178,000 and 399,000 cases were averted through use of the UK National Health Service GAEN-based COVID-19 app in October through December 2020 (*2*). Exposure-notification apps speed up notification of exposure and pick up encounters that a user might miss—the people standing nearby during a conference break, the person alongside them on the subway.

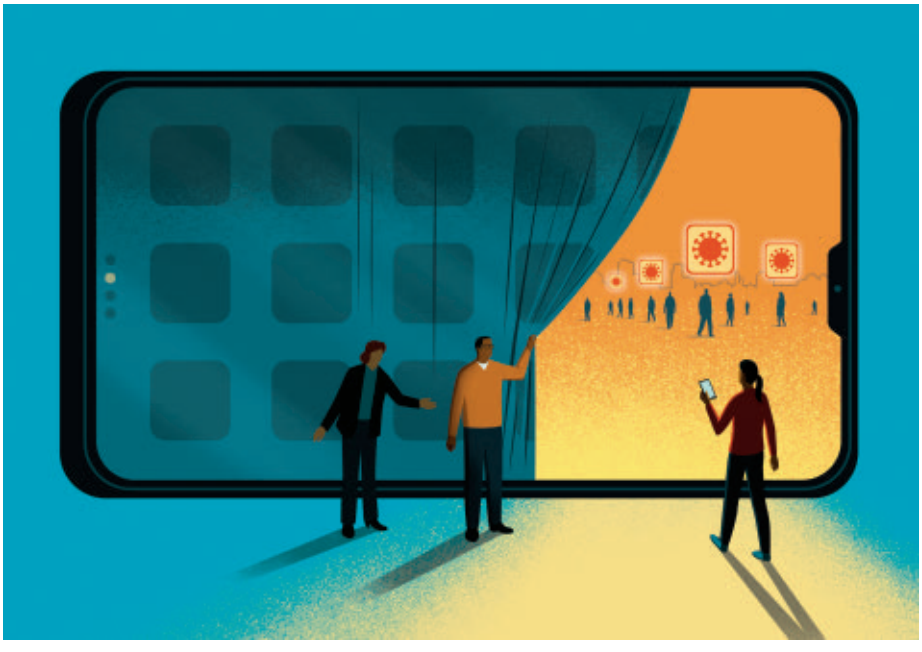
But in this pandemic that has demonstrated a devastatingly disproportionate impact on certain communities, because the apps only run on the most recent iPhone and Android phones, many had anticipated that use of exposure-notification apps would prove problematic to low-income communities and the elderly for economic reasons. Early evidence shows, however, a somewhat more complex set of interactions. A study of the UK app showing decreased incidence of SARS-CoV-2 spread noted that regions of higher app use had "lower levels of poverty, are more rural, and have higher local GDP" (*2*). In those regions, the populations that use the apps are more likely to be white, but also more likely to be elderly.

The reasons behind low app use by marginalized populations involve more concerns than simply access to the right version of smartphone. A recommendation to isolate is markedly different for a low-income worker who has a role that requires physical presence and cannot afford to lose their job than for a middle-class person whose work can be done from home. Research has

[1]The Fletcher School and School of Engineering, Department of Computer Science, Tufts University, Medford, MA 02155, USA. [2]Department of Computer Science, University College London, London, UK. Email: susan.landau@tufts.edu

identified at least three reasons for failure to quarantine and isolate: need to maintain salary, need to buy essentials, and need for social services (*3*).

Reasons for failure to quarantine and isolate will also vary by demographic. For example, researchers found a high correlation between Massachusetts communities with high SARS-CoV-2 infection rates and those with three factors: high immigrant populations who are often reluctant to seek government aid, including health care, even when ill; a high percentage of people working in food services (exposing workers to high numbers of people); and individuals living in multigenerational households (*4*). People who fear involving the government in their affairs and who are unlikely to be able to stay home from work because of a possible SARS-CoV-2 exposure are unlikely to benefit from an exposure-notification app.



Cultural history and socioeconomic challenges are different for distinct population groups. This then plays out in how likely contact-tracing apps are to be used. For example, a history of mistreatment has created great distrust of both the government and public health by many in the US Black community. That distrust is likely to undermine uptake of exposure-notification apps by many in this community. That said, exposure notification is more likely to be adopted by a middle-class person living in an integrated neighborhood than by a low-income individual living in a community with overpolicing (*5*).

The story about the coronavirus is different on the Fort Apache Indian Reservation in eastern Arizona, home to about 12,000 members of the White Mountain Apache Tribe. Yet, despite struggles with poverty and chronic disease and many people on the reservation being infected by SARS-CoV-2, a tailored intervention by the medical team early in the pandemic at the Whiteriver Indian Hospital kept the death rate low (*6*).

Two responses were particularly valuable. First, families on the reservation often live in multigenerational settings, and this highly contagious disease could spread easily between members. So if one person was sick, contact tracers would immediately test oxygen saturation levels for everyone in the household, quickly picking up anyone else who might already be suffering serious signs of the disease. Second, contact tracers, working in the knowledge that children in Apache families often have extended visits with nonresident grandparents, would ask, "Who are your grandparents?"—and then check on their health, even if they did not live in the same household (*6*).

ENABLING EQUITY WITH DIGITAL APPS

Exposure-notification apps provide benefits, but they can also exacerbate health care inequities. In a situation of scarcity—and for much of the pandemic, testing and medical attention have been scarce—testing driven by contact tracing in some communities can divert health care resources from people in other, marginalized communities.

The apps themselves are not contact tracers. Effective use of apps thus starts with contact tracers who are trained to be attuned to community needs. Contact tracers can point to success after success in marginalized communities where developing partnerships with churches, nonprofits, and community organizations—connecting with the connectors of the community—and being hands-on in providing care—access to testing, food, transportation, and treatment—were critical, whether in halting an eruption of syphilis cases in Baltimore in 2004 or responding to an Ebola cluster in Dallas in 2014 (*7*). Actions such as these create the trust that enables people to share crucial information on who might have been exposed to the disease and who might already be ill; it creates the willingness and capability to isolate despite the personal hardship.

Exposure-notification apps assume that users can manage the complexities of exposure and isolation. There are many for whom such an assumption just does not hold true (*8*). Thus, these apps provide neither the emotional support that human contact tracers do nor the encouragement and assistance that lead to patients' trust and willingness to list their contacts (*9*). In that sense, the exposure-notification apps are designed for a user who can handle the complexities that weeks of isolation may create—but not for someone for whom the logistical arrangements and economic consequences may cause household arrangements to capsize.

In light of this, economic support is crucial. The Swiss government financially supports those who receive a public health recommendation to isolate and who cannot work from home. Such aid is not only humane; it can be cost effective. One estimate from Massachusetts suggests that by decreasing the number of infections in the state, government support for quarantine and isolation would actually reduce government costs (*3*).

There is also the support that actually allows someone to isolate safely at home. Contact tracers ask: Are you safe isolating? What do you need? Do you need help with getting food or medication? Then the contact tracers go about ensuring that these needs are met.

The GAEN-based apps were intended to support contact tracing. Indeed, the Republic of Ireland's app, Covid Tracker, provides an option of registering with a phone number; in the case of an exposure, a contact tracer will call the exposed person, check how they are feeling, see if they need support when isolating and whether they are safe in doing so, and trace their contacts (*10*). In September 2020, Google and Apple introduced EN Express, which provides a menu of options enabling states to easily set up an app. But EN Express fails to include an option allowing the user to register

ILLUSTRATION: DAVIDE BONAZZI/SALZMAN ART

their phone number—and that eliminates the possibility of automatically pulling in a contact tracer if a user has been exposed to SARS-CoV-2.

The current set of exposure-notification apps are designed so that an exposed person must reach out to public health officials rather than contact tracers reaching out to the exposed person. Populations with distrust of public health services and government are less likely to take that initial step. To provide people in such communities with the support that enables them to isolate and that creates the trust that allows tracing of their contacts, exposure-notification apps would be most effective if providing a contact phone number was the default.

That means an exposed person would become known to public health officials. The user, however, would have a choice whether to provide a number when downloading the app (and thus preserve their anonymity in the event of exposure). And even if a user does provide a phone number, the exposure-notification app would not be a centralized data collection service; it would provide public health officials with information that someone has been exposed—but not by whom. Information about which users were in someone's proximity would still remain on a user's phone. The only change that would occur by including a phone number would be automatic involvement of contact tracers in the case of exposure.

Providing a contact phone number should always be at the user's choice. An appropriate nudge during app signup indicating that providing a phone number would enable easier access to testing and social services could be used to encourage more users to do so. For privacy's sake, the system should be designed to ensure that opting out from providing a contact phone number is easy and explicit. If a user provides a number but later decides to opt out, doing so should similarly be simple. There should be no penalty for any user who fails to provide a contact phone number or later decides to remove it.

Providing a contact number removes the anonymity the user of an exposure-notification system would otherwise have, so other changes are needed to create trust in the system. Because false positives of exposure can have a disproportionate effect on those who can't easily isolate, app adoption should always be a user's choice *(11–14)*. That right should also be legislatively guaranteed.

Though no one would learn the details of exposure—that is, the where, when, or by whom of exposure—exposures would no longer be anonymous for those who provide a phone number to the app. Thus, to ensure trust in the system—and adoption by people at risk—there would need to be strong policies protecting privacy of the data.

The apps don't distinguish between two people in close proximity in a small room—high risk of contagion—and being in close proximity outdoors, where risk drops substantially. Signal strength—and the corresponding measure of likelihood of exposure—varies depending on whether the phone is in a pocket or a handbag and could also vary by phone model and shape of a room (*15*). Bluetooth low-energy signals travel through building materials, which can lead to other false notifications of exposure, especially for those living in tightly packed housing.

"Public health resources should be adjusted so that those demographics that are 'underusing' the apps do not lose needed resources."

Such false positives of exposure are problematic for everyone. But if false positives of exposure are having an inequitable effect on certain communities because of their inability to isolate and work from home, then there should be protections that ease the impact. These could include easier access to testing (and thus return to work), easier access to social services, and possibly even increased job protections for those who receive recommendations to isolate.

Information regarding exposure should only be used for contact-tracing purposes. Switzerland's Epidemic Act already does this, prohibiting disclosure of SwissCovid data for any other purpose; Australia does the same. Other countries should pass laws that information provided by contact-tracing and exposure-notification apps is to be used only for informing patients of exposure.

Use of the app should also be voluntary. Originally the use of Singapore's TraceTogether was purely at people's discretion. But in May 2021, the government effectively changed policy, requiring app use for entry in many Singaporean establishments, including shopping malls, restaurants, offices, and schools.

The GAEN-based approach is privacy protective, and early results demonstrate efficacy (*2*). The pandemic, however, exposed great health care inequities across different demographics. To understand how use of the app is affecting public health and what adjustments must accordingly be made, apps must be tested before use. Communities experienced the pandemic in various waves; peoples' needs and behaviors have ebbed and flowed with the changes in risk. Thus, the apps must be tested repeatedly during public use, and this must be done across multiple demographic areas (*11–14*). Public health resources should be adjusted so that those demographics that are "underusing" the apps do not lose needed resources.

In every age, we expand our understanding of the causes of pandemics, their transmission mechanisms, and potential techniques for ending them. Exposure-notification apps provide a useful way to slow the spread of SARS-CoV-2. Like any technology, they change the dynamics of the disease's spread. This pandemic will not be the last that humans face. We must use and build tools and supporting health care policy so that the technologies are not only protective of people's rights, health, and safety, but also so that their use enables greater health care equity. Only then will the tools properly address both the medical and social occurrences of the pandemic. ■

REFERENCES AND NOTES

1. Government of Singapore, TraceTogether, Safer Together, www.tracetogether.gov.sg (accessed 15 April 2021).
2. C. Wymant et al., *Nature* **594**, 408 (2021).
3. M. Bourdeaux et al., "Estimating the costs and benefits of supported quarantine and isolation in Massachusetts: The missing link in Covid-19 response," Faculty Research Working Paper Series RWP21-003, Harvard Kennedy School, February 2021.
4. J. F. Figueroa, R. K. Wadhera, D. Lee, R. W. Yeh, B. D. Sommers, *Health Aff. (Millwood)* **39**, 1984 (2020).
5. S. Landau, People Count: Contact-Tracing Apps and Public Health (MIT Press, 2021).
6. R. M. Close, M. J. Stone, *N. Engl. J. Med.* **383**, e15 (2020).
7. C. P. Chaulk, J. Zenilman, R. Bialek, A. Rompalo, *J. Public Health Manag. Pract.* **27**, 322 (2021).
8. L. Smith et al., *BMJ* **372**, (2021).
9. S.-W. Jian, H.-Y. Cheng, X.-T. Huang, D.-P. Liu, *Int. J. Infect. Dis. 101*, **348** (2020).
10. Government of the Republic of Ireland, HSE, "Privacy and How We Use Your Data," https://www2.hse.ie/services/covid-tracker-app/privacy-and-how-we-use-your-data.html (accessed 20 August 2020).
11. S. Landau, C. Lopez, L. Moy, "The importance of equity in contact tracing," Lawfare, 1 May 2020, www.lawfareblog.com/importance-equity-contact-tracing.
12. K. "K. J." Bagchi et al., "Digital tools for COVID-19 contact tracing: Identifying and mitigating the equity, privacy, and civil liberties concerns," Edmond J. Safra Center for Ethics, Open Technology Institute, and New America, 2 July 2020.
13. Ethics Advisory Board, NHSX, Letter to Secretary of State, 24 April 2020, https://nhsbsa-socialtracking.powerappsportals.com/EAB%20Letter%20to%20NHSx.pdf (accessed 18 October 2020).
14. N. S. Guiliani, ACLU, Government Safeguards for Tech-Assisted Contact Tracing, 18 May 2020.
15. D. J. Leith, S. Farrell, *ACM Comput. Commun. Rev.* **50**, 66 (2020).

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grants CNS 1923528 and CNS 1955805.

10.1126/science.abi9852

PHOTO: ARMANDO GONZALES/US NAVY/GETTY IMAGES

The USS *Mason* is resupplied with biofuel made from waste beef fat and inedible vegetable oil in 2016.

MARITIME AFFAIRS

The keepers of the seas

As oceans warm, an author reveals the centrality of maritime activities in the modern world

By **Jason W. Smith**

In 2012, the US Navy launched its Great Green Fleet as part of a multinational training exercise. Steaming for the first time on a combination of diesel and biofuels, the voyage was meant to evoke the Great White Fleet (1907–1909), which marked an international introduction to American sea power. This modern corollary lacked the pageantry associated with its forebear, however, occurring at a time when most Americans seem to know little about maritime affairs, even as the ocean remains vital to the country's interests.

This paradox is at the center of Gregg Easterbrook's *The Blue Age*. Part history, part political science, and part economics lesson, the book aims to reacquaint American readers with how much they take the oceans for granted.

Since 1945, Easterbrook argues, American naval predominance coupled with the flourishing of seaborne commerce unimpeded by large-scale conflict has transformed much of the world, raising standards of living, life expectancy, and many other indicators of progress for most, if not all, people. Free trade under the US Navy's watch ushered in a period of unparalleled global prosperity, he maintains, elegantly dubbing this era "the Blue Age."

In the book's first section, Easterbrook examines the current state of the American fleet and its historical evolution. Citing the navy's lead in supercarriers, amphibious assault ships, and ballistic missile submarines, he notes that the United States possesses a fleet larger than the rest of the world combined. While he offers a historical note of caution about the folly of naval arms races and the perils of misplacing faith in technological paradigms such as the battleship, Easterbrook argues that American naval power must be maintained, particularly in the face of challenges from China and budget cuts.

In part two, Easterbrook documents the transformation of maritime commerce since 1945, the proliferation of flags of convenience and foreign shipbuilding, and the evolution of maritime shipping from break-bulk carriers to containerization. He describes the transformation of ports to accommodate the new intermodal commerce and an economic reorientation that has made Asia the home of 9 of the 10 busiest ports in the world.

In part three, Easterbrook looks to a future defined by a newly navigable Arctic Ocean. He calls for international governance of the seas as the best means to regulate resource extraction and environmental protection, protect sailors' rights, and limit the proliferation of naval arms, thus maintaining the Blue Age in the face of perilous environmental and geostrategic threats.

The Blue Age is a book with big ideas that ranges widely. Its strengths lie in calling more attention to the centrality of maritime and naval activity to our modern world, in centering the evolution of American power and its future on the ocean, and in linking naval and maritime affairs, which in academic scholarship often maintain a stubborn bifurcation.

However, Easterbrook tends to be breezy in his analysis, and his treatment of the historical record is sometimes thin. For example, although his vision of the future rests on international governance, he leaves unexplored the US government's reluctance to sign the United Nations Convention on the Law of the Sea (UNCLOS), which stems primarily from concerns about ceding authority and royalties for deep seabed mining.

At times, Easterbrook sacrifices rigorous historical causation for cliché, for example, with his glib assertion that Margaret Thatcher engaged in the Falklands War because she "wanted to show she was more manly than most men." He also assigns Germany and Japan's defeat in the world wars to the sea power theories of Alfred Thayer Mahan without acknowledging that the United States followed the same principles to global preeminence, and he writes elsewhere that "the US military has almost always been an expeditionary force—intended to operate over the horizon," a claim that mischaracterizes the military's involvement in Native American wars, coast defense, and commerce protection for much of the 19th century.

The Blue Age
Gregg Easterbrook
PublicAffairs, 2021.
304 pp.

The Blue Age, in other words, would have benefited from a deeper dive into the historiography of naval and maritime history. For all its interest in the geostrategic implications of climate change, especially in the Arctic, the book is not primarily concerned with scientific questions, nor does it examine the role of the marine sciences in the navy. Nonetheless, it is provocative with generally sound ideas. For a more nuanced view of American naval dominance and its relationship to ocean science, however, it should remain a starting point for further exploration. ■

The reviewer is at the Department of History, Southern Connecticut State University, New Haven, CT 06515, USA. Email: smithj131@southernct.edu

10.1126/science.abj9481

PSYCHOLOGY

Leveraging the collective mind

Shared social identities can better prepare us to confront societal problems

By **S. Alexander Haslam**

In the early 1990s, Karl Weick and Karlene Roberts conducted extremely influential field research to understand why, on the flight decks of nuclear-powered aircraft carriers where there were "a million accidents waiting to happen," almost none of them did. The answers they provided set the stage for a groundswell of research into what came to be known as high-reliability organizations (HROs).

At the center of their analysis was an observation that the key psychological hallmark of an HRO was something the researchers referred to as collective mind. As they put it, "A well-developed organization mind, capable of reliable performance is thoroughly social. It is built of ongoing interrelating and dense interrelations…As people move toward individualism and fewer interconnections, organization mind is simplified and soon becomes indistinguishable from individual mind. With this change comes heightened vulnerability to accidents."

As HRO research has evolved, many of Weick and Roberts's findings have been confirmed and consolidated. Nevertheless, their central insight has become increasingly obscured as researchers have fallen back on individualized approaches to HRO conceptualization and assessment. This drift toward individualistic models of human behavior mirrors trends in a host of related spheres. However, it is both limited and limiting. In *The Power of Us*, Jay Van Bavel and Dominic Packer explain why this is the case and map out in accessible terms why only an expanded model of self will allow us to tackle the pressing intellectual, social, and political problems of today.

Drawing on a corpus of research dating back to the groundbreaking work of Henri Tajfel and John Turner in the 1970s, they elaborate a model that embraces the productive potential of a collectively defined self wherein people think, feel, and behave in terms of shared social identity ("we" and "us"), not only in terms of personal identity ("I" and "me"). The book clarifies the implications of this model for a wide array of contemporary phenomena—everything from social media use to populist leadership—and is strengthened by its grounding in the authors' own pioneering work on the technological and neuroscientific frontiers of social psychology.

Written specifically for the popular end of the market, the book is also notable for going well beyond other texts that have traversed this terrain in recent years and for avoiding the pitfalls to which many have succumbed (one notable exception being Rutger Bregman's *Humankind*). In particular, when other writers have broached the topic of social identity, they have tended to associate it with ineluctable bias characterized by in-group love and out-group hate. Such interpretations vulgarize the message of Tajfel's original research, in which schoolboys assigned to meaningless groups were found to discriminate in favor of their in-group without obvious reason, implying an alignment with the dark conformist message that commentators customarily derive from psychology's other classic studies. This is seen, for example, in narratives of Latané and Darley's helplessly irresponsible bystanders, Milgram's mindlessly obedient electrocuters, and Zimbardo's heedlessly sadistic prison guards.

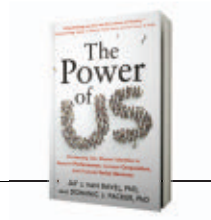
Van Bavel and Packer challenge this dark imagining of group psychology by breathing fresh life into Turner's insistence that the social self is flexible and highly responsive to both social structural reality and normative influence. In particular, this sensibility points to the productive potential of "identity leadership," which crafts and embeds particular versions of the collective self in ways that can render "us" a source of resistance and emancipation.

In the process of advancing this argument, *The Power of Us* compellingly debunks many of the myths that have arisen not only around social identity research but also around social psychology as a whole. Of these, the most enduring are those that paint groups as inevitable sites for corruption of self and loss of reason. The triumph of this book is how it shows that this framing is not only wholly wrong but also dangerously misleading. As Weick and Roberts recognized, if we are to have functional and resilient organizations, institutions, and societies, we must understand and draw upon the power of collective mind. ■

10.1126/science.abl5269

The reviewer is a professor of psychology and Australian Laureate Fellow at the University of Queensland, St Lucia, QLD 4072, Australia, and the coeditor of *Together Apart: The Psychology of COVID-19* (Sage, 2020). Email: a.haslam@uq.edu.au

The Power of Us
Jay J. Van Bavel and Dominic J. Packer
Little, Brown Spark, 2021. 320 pp.



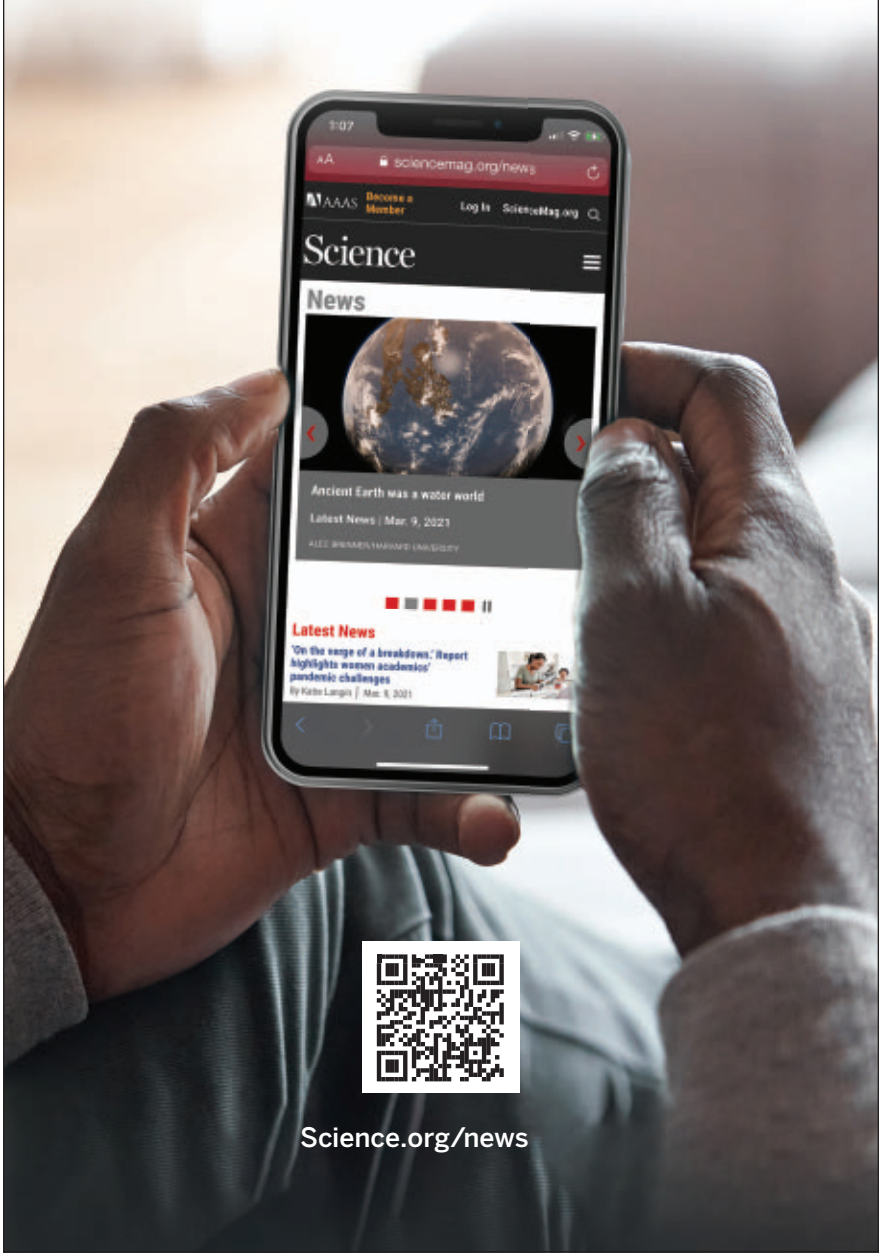
A man signs a petition in 1969, joining like-minded opponents of the Vietnam War.

PHOTO: BEV GRANT/GETTY IMAGES

NEWS FROM Science

Up-to-the-minute research and policy news you won't find in print

Visit us online to read all the news coverage that there just wasn't enough room to print in this issue.



Science.org/news

COULD YOU BE THE NEXT LEADER IN IMMUNOLOGY?

APPLY NOW FOR THE $30,000 GRAND PRIZE

APPLY TODAY:
www.science.org/michelson

MICHELSON & Science PRIZE FOR IMMUNOLOGY

A recently approved mining project threatens the endemic species and fragile ecosystems of northern Chile, including those on Tilgo Island.

PHOTO: TUI DE ROY/MINDEN PICTURES

Edited by **Jennifer Sills**

Extractivist policies hurt Chile's ecosystems

On 11 August, the Chilean government approved the construction and operation of Dominga (*1*), a high-grade iron and copper open-pit mine and port at the heart of the Humboldt Archipelago ecosystem (seven islands near northern Chile and the adjacent Chilean coast). The scientific community, National Forest Service, and Ministry of the Environment had previously rejected this project, which will cause huge and irreversible impacts on both land and sea to an ecosystem globally recognized as a biodiversity "hope spot" (*2*). The Humboldt Archipelago ecosystem harbors a rich and unique biota, and 60% of the total area of the reserve is classified as very fragile (*3*). An estimated 60 to 88% of the ecosystem's species are endemic (*3*), and many are included in the International Union for Conservation of Nature Red List. Given the high environmental costs of Dominga, the project should be halted.

Dominga will disrupt the Humboldt Archipelago ecosystem and coastal land. Persistent upwelling will transfer large volumes of desalination concentrate from the submarine pipeline, affecting pelagic and benthic food webs that support seabirds and marine mammals of global importance. About 25 seabird species forage around the islands, which provide nesting sites for at least 14 species, 9 of which are endemic (*4*). Marine mammals such as the endangered marine otter (*4*), resident bottlenose dolphin (*5*), and 11 cetacean species feed, breed, or both, in the Humboldt Archipelago ecosystem (*4*). Port infrastructure, risk of oil spills, noise and light contamination, and increased vessel traffic will directly threaten seabirds and cetaceans (*6*). Mining will also compromise habitats of endangered cacti (*7*) and populations of birds (*8*), reptiles (*9*), scorpions (*10*), and insects (*11*).

Extractivist policies in Chile have a long history despite a track record of disastrous consequences for the environment, communities, and human rights (*12*). In line with the Intergovernmental Panel on Climate Change's sixth assessment report to mitigate climate change, Chile should implement cleaner environmental and socioeconomic policies, preserving both richness and species composition of the Humboldt Archipelago ecosystem. Chile is currently at a crossroads as elected constituents draft a new constitution: Either we continue on the path of traditional extractivism mortgaging our biodiversity or we adapt to a greener future, promoting a sustainable economy that balances progress with current social imperatives for environmental justice.

Mauricio A. Urbina[1,2], Pablo C. Guerrero[3], Viviane Jerez[1], Fulgencio Lisón[1], Guillermo Luna-Jorquera[4], Camilo Matus-Olivares[5], Juan C. Ortiz[1], Guido Pavez[6,7,8], María J. Pérez-Alvarez[7,9,10], Ramiro Riquelme-Bugueño[1,2], Macarena Santos-Carvallo[6,7], Maritza Sepúlveda[6,8], Pedro F. Victoriano[1], Daniel Gomez-Uchida[1,8]*

[1]Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile. [2]Instituto Milenio de Oceanografía, Universidad de Concepción, Concepción, Chile. [3]Departamento de Botánica, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile. [4]Núcleo Milenio Ecología Manejo Sustentable Islas Oceánicas, Departamento de Biología Marina, Universidad Católica del Norte, Coquimbo, Chile. [5]Departamento de Ciencias Forestales y Medioambiente, Facultad de Ciencias Agropecuarias y Forestales, Universidad de La Frontera, Temuco, Chile. [6]Centro de Investigación y Gestión de Recursos Naturales, Facultad de Ciencias, Universidad de Valparaíso, Chile. [7]Centro de Investigación Eutropia, Santiago, Chile. [8]Núcleo Milenio Salmónidos Invasores, Concepción, Chile. [9]Escuela de Medicina Veterinaria, Facultad de Ciencias, Universidad Mayor, Santiago, Chile. [10]Instituto de Ecología y Biodiversidad, Santiago, Chile.
*Corresponding author. Email: dgomezu@udec.cl

REFERENCES AND NOTES

1. Mining[dot]com, "Chile's $2.5 billion Dominga copper-iron project approved by regulators" (2021); www.mining.com/web/chiles-2-5bn-dominga-copper-iron-mine-project-approved-by-regulators/.
2. *BBC News*, "Humboldt Archipelago: The 'place of hope' of the oceans in Chile" (2018).
3. Corporación Nacional Forestal, "Plan de manejo Reserva Nacional Pinguino de Humboldt, Part A" (2018) [in Spanish].
4. M. Sepúlveda *et al.*, "Determinación del estado

poblacional en las Reservas Marinas isla Chañaral e islas Choros y Damas, de las especies delfín nariz de botella, chungungo, pingüino de Humboldt y cetáceos" (Informe Final Proyecto FIPA 2018-43, 2021) [in Spanish].
5. M. J. Pérez-Alvarez *et al.*, *Anim. Behav.* **139**, 81 (2018).
6. R. Quispe *et al.*, *Mar. Ornithol.* **48**, 205 (2020).
7. P. C. Guerrero *et al.*, *Taxon* **68**, 557 (2019).
8. J. F. Masello *et al.*, *Front. Zool.* **8**, 16 (2011).
9. P. Valladares *et al.*, *Gayana* **75**, 81 (2011) [in Spanish].
10. J. Pizarro-Araya, A. A. Ojanguren-Affilastro, *Gayana* **82**, 8 (2018) [in Spanish].
11. V. Jerez, J. Pizarro-Araya, *Gayana* **82**, 101 (2020) [in Spanish].
12. S. Smart, *Crit. Plan.* **23**, 59 (2017).

10.1126/science.abm0157

Preventing urban soil loss in China

China has urbanized at an unprecedented pace, expanding not only its large cities but also smaller towns (*1*). China's urbanization rate has increased from 10.6% in 1949 to 63.9% in 2020 and is expected to rise to 75% by 2030 (*2*). The Chinese government monitors urban air and drinking water quality (*3*) but pays less attention to loss and deterioration of urban soil. The weakening of urban soil's ecosystem functions may have pronounced impacts on human welfare, particularly during accelerating urbanization (*4*).

Urban soil loss is largely anthropogenic and has adverse geological consequences. Urban infrastructure construction and land reclamation projects change city layouts and damage pristine soil systems (*5*). Rainfall cannot infiltrate into the urban soil because it is sealed and compacted by hard surfaces and buildings, resulting in urban flooding (*6*). The overexploitation of underground water in northern Chinese cities can lead to land subsidence (*7*), increasing the risk of geological and flooding disasters.

The loss in urban soil decreases the service functions of urban and suburban ecosystems. In addition to increasing the risks of floods and landslides, soil loss can change ground heat absorption and radiation, exacerbating urban heat island effects (*8*). The dust generated by fine soil particles will increase the total suspended particles in the ambient air; in turn, urban air pollution will increase the risk of soil pollution (*9*). The sediments eroded from soil by stormwater enter the nearby water conservancy infrastructures and rivers, raising the riverbed and occupying reservoir capacity. Urban soil particles may carry and transfer harmful chemicals that may accumulate in grain, vegetables, livestock, or fish, posing health risks to humans (*10*).

China is taking measures to optimize urban land management. The Fifth Plenary Session of the 19th Central Committee of the Communist Party of China proposed strategies to guide smarter urbanization and development (*11*). The Ministry of Natural Resources has proposed the regular analysis and evaluation of the effects of urban land spatial planning (*12*). However, much work remains to be done. The land administration department should establish monitoring networks for urban soil erosion, including nutrients and pollutants. Sealed or compacted areas should be integrated with vegetated areas to reduce the likelihood of flooding. Regional soil erosion data and remote sensing technology should be applied to estimate local urban soil losses. The laws and regulations on urban soil protection should be updated, and urban construction should take ecosystem health into account. Collectively, these measures will help protect urban soil from loss in both area and ecosystem functions.

San'an Nie[1], Yaping Fang[2], Shun Li[3,4]*
[1]College of Resources and Environment, Hunan Agricultural University, Changsha, 410128, China. [2]School of Humanities and Foreign Languages, Hunan Agricultural University, Changsha, 410128, China. [3]Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China. [4]Ningbo Urban Environment Observation and Research Station, Chinese Academy of Sciences, Ningbo, 315800, China.
*Corresponding author. Email: snli@iue.ac.cn

REFERENCES AND NOTES

1. W. H. Kuang, *Sci. Chin. Earth Sci.* **62**, 1591 (2019).
2. Morgan Stanley, "Rise of China's supercities: New era of urbanization" (2019); www.morganstanley.com.au/ideas/2019-supercities.
3. A. Lin *et al.*, *Science* **351**, 674 (2016).
4. D. L. Herrmann, L. A. Schifman, W. D. Shuster, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6751 (2018).
5. J. D. Li *et al.*, *Sustain. Basel* **7**, 7137 (2015).
6. Ministry of Water Resources of the People's Republic of China, "China flood and drought disaster prevention bulletin" (2019); www.mwr.gov.cn/sj/tjgb/zgshzhgb/202104/t20210409_1513262.html [in Chinese].
7. C. F. Zhou *et al.*, *Geomorphology* **335**, 48 (2019).
8. C. S. Tang *et al.*, *Energ. Build.* **43**, 3090 (2011).
9. D. Y. Huang *et al.*, *Environ. Pollut.* **157**, 35 (2009).
10. Y-G. Zhu *et al.*, *Environ. Sci. Technol.* **45**, 5099 (2011).
11. Ministry of Natural Resources of the People's Republic of China, Standard for Land Administration Industry of the People's Republic of China (TD/T 1063-2021), "Code of practice for city examination & evaluation in spatial planning" (2021) [in Chinese].
12. Xinhua News, "Draft outline of China's 14th Five-Year Plan: Optimizing the new type of urbanization strategy and improving the quality of urbanization development" (2021); www.xinhuanet.com/politics/2021lh/2021-03/05/c_1127173066.htm [in Chinese].

10.1126/science.abl8918

China's carbon emissions in Brazil

In her In Depth story "Dire warming report triggers calls for more action from China" (20 August, p. 842), L. Pike describes how China is under pressure to reduce its carbon emissions—particularly its use of coal—in the wake of the Intergovernmental Panel on Climate

"THE SUPPORT OF MICHELSON PHILANTHROPIES CAME AT A CRITICAL TIME IN MY CAREER. IT HELPS ME APPLY CUTTING-EDGE TECHNOLOGIES TO DEVELOP A NOVEL APPROACH TO VACCINE DESIGN."

DANIKA HILL, PH.D.
2020 Michelson Prize Winner
Monash University
Australia, Melbourne

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY

Change's sixth assessment report. However, domestic emissions are only a part of the problem; China's imports and investments have a massive impact on emissions worldwide (*1*), including Amazon deforestation. China imports a vast amount of commodities from the Amazon (*2*), and it plays a key role in financing Amazonian roads, railways, and other infrastructure projects that drive deforestation (*3*).

Discussions about decreasing Amazon deforestation by reducing international exports of Brazil's commodities often focus on Europe [e.g., (*4*)]. European countries make good targets for persuasion given their concern for the environment and sense of responsibility for the impacts they cause beyond their borders. However, 79% of Brazil's exported soy goes to China (*5*) compared with 9.5% to Europe (*6*), and 52% of exported beef goes to China (*7*) compared with 5.3% to Europe (*8*).

Brazil is the world's largest exporter of these commodities (*6*). Much more beef and soy are produced than can be consumed in the country, and every hectare of Amazon forest cleared to produce them is for export rather than to feed the Brazilian people. Cattle pasture and soy are by far the largest replacements for Amazon forest (*9*), making deforestation a direct result of international demand.

Curtailing deforestation in the Amazon would reduce emissions and protect the forest's biodiversity, water recycling function, and Indigenous peoples (*10*). If China were to impose environmental conditions on its imports from Brazil, the effect on Amazon deforestation would be immediate, and it would help reduce China's responsibility for global warming.

Philip M. Fearnside
National Institute for Research in Amazonia (INPA), 69067-375, Manaus, Amazonas, Brazil.
Email: pmfearn@inpa.gov.br

REFERENCES AND NOTES

1. F. Ascensão *et al.*, *Nat. Sustain.* **1**, 206 (2018).
2. P. M. Fearnside *et al.*, *Environ. Dev. Sustain.* **15**, 325 (2013).
3. P. M. Fearnside, A. M. R. Figueiredo, in *China and Sustainable Development in Latin America: The Social and Environmental Dimension*, R. Ray *et al.*, Eds. (Anthem Press, New York, 2016), pp. 229–265.
4. L. Kehoe *et al.*, *Science* **364**, 341 (2019).
5. COMEX, "Exportação de Soja: Entenda melhor" (2021); www.fazcomex.com.br/blog/exportacao-de-soja/ [in Portuguese].
6. ABIOVE, "Brasil é o maior exportador de soja do mundo" (2021); https://abiove.org.br/wp-content/uploads/2019/07/Abiove_Exportação.pdf [in Portuguese].
7. COMEX, "Exportação no Brasil: Quais os principais produtos exportados?" (2021); www.fazcomex.com.br/blog/quais-principais-produtos-exportados-brasil/ [in Portuguese].
8. COMEX, "Exportação de Carne Bovina" (2021); www.fazcomex.com.br/blog/exportacao-de-carne-bovina/ [in Portuguese].
9. E. Barona *et al.*, *Environ. Res. Lett.* **5**, 024002 (2010).
10. C. A. Nobre *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10759 (2016).

10.1126/science.abl9962

Where Science Gets Social.

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

REVEALING THE NEXT BREAKTHROUGH IN IMMUNOLOGY

APPLY NOW FOR THE $30,000 GRAND PRIZE

APPLICATIONS DUE: OCTOBER 1, 2021

The Michelson Philanthropies & Science Prize for Immunology recognizes research that will have a significant impact on vaccine and immunotherapy development.
Together with AAAS/Science, Michelson Philanthropies awards breakthrough research from the past three years. The applications are judged on innovation and potential impact.

With these prizes, scientists (35 and younger) have the opportunity to advance their research projects, are provided with a renowned platform in Science to publish their work and connect them with other innovators through the Michelson Prizes alumni network.

Young researchers from a wide range of disciplines are encouraged to apply.

APPLY TODAY:
www.science.org/michelson

MICHELSON PHILANTHROPIES & Science
PRIZE FOR IMMUNOLOGY

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PAIN

Opioid attenuation by astrocytes

The chronic use of morphine and other opioids can lead to tolerance and a paradoxical increase in pain in patients. Hu *et al*. found that these effects are mediated in part by interleukin-33 (IL-33) signaling in astrocytes and oligodendrocytes in the spine. Treating mice with morphine increased the levels of IL-33 and its receptor, ST2, which activated the production and release of the chemokine CXCL12 from astrocytes. Suppressing this pathway specifically in astrocytes or pharmacologically blocking the chemokine receptor CXCR4 dampened morphine-enhanced pain in mice. —LKF *Sci. Signal.* **14**, eabe3773 (2021).

Artist's rendition of an astrocyte, a type of glial cell that plays important roles in pain perception and enhancement

MUSCULAR DYSTROPHY

Protecting muscles with antibodies

Latent transforming growth factor β-binding protein 4 (LTBP4) has previously been shown to be a modifier of muscular dystrophy in mice. Demonbreun *et al.* developed a monoclonal antibody to human LTBP4 that binds to the protein's hinge region, preventing proteolytic cleavage. In dystrophic mice expressing the human LTBP4 protein, anti-LTBP4 treatment protected muscle function, with enhanced force production and reduced fibrosis. Twenty-four weeks of treatment also improved diaphragm muscle function. The combination of anti-LTBP4 treatment with prednisone substantially enhanced muscle function and protected against injury in dystrophic mice, suggesting that anti-LTBP4 antibodies may be useful in treating muscular dystrophies in humans. —MN

Sci. Transl. Med. **13**, eabf0376 (2021).

PHYSICS

Setting bounds on a fifth force

Some extensions to the Standard Model of particle physics posit the existence of a fifth force to complement the existing four fundamental forces. To set bounds on the strength of such an interaction, experiments on vastly different length scales have been performed. Heacock *et al*. used an unusual method called Pendellösung interferometry to measure the neutron structure factors of silicon. The momentum dependence of the structure factors enabled the researchers to put more stringent bounds on the strength of a type of fifth force called the Yukawa force, as well as measure the charge radius of the neutron. —JS

Science, abc2794, this issue p. 1239

RESPIRATORY ENZYMES

Breaking down oxygen

Molecular oxygen (O_2) is the terminal oxidant for respiration in mitochondria and many bacteria. Within membrane-bound heme–copper oxidases, a controlled, four-electron reduction of O_2 to water is coupled to pumping of protons across the membrane that can be used, among other outcomes, to generate adenosine triphosphate. Studying cytochrome bo_3 ubiquinol oxidase, Jose *et al*. investigated the key P_M intermediate, which forms after O–O bond cleavage and precedes proton pumping, using magnetic circular dichroism spectroscopy. The authors observed features demonstrating that P_M is a three-spin system, which is consistent with

ILLUSTRATION: KATERYNA KON/SCIENCE SOURCE

a consensus model including an iron(IV)-oxo species, copper(II) ion, and tyrosyl radical. These results provide an important validation of the O–O cleavage mechanism and open the door to understanding the proton-pumping step. —MAF

Science, abh3209, this issue p. 1225

MATERIALS SCIENCE

A match made clear

The fabrication of strong and tough composites is of interest in many technologies, such as the combination of mechanical performance with transparency for robust display systems. Amini *et al*. combined and centrifuged glass flakes with poly(methyl methacrylate) (PMMA) to make a transparent composite. By doping the glass flakes, it was possible to alter the refractive index of PMMA to maximize optical clarity. Such composites show good strength and toughness and could have a wide range of potential applications as an alternative to current glass composites. —MSL

Science, abf0277, this issue p. 1229

PHASE-CHANGE MEMORY

Flexing computer memory

Phase change materials leverage changes in structure into differences in electrical resistance that are attractive for computer memory and processing applications. Khan *et al*. developed a flexible phase change memory device with layers of antimony telluride and germanium telluride deposited directly on a flexible polyimide substrate. The device shows multilevel operation with a low switching current density. The combination of phase change and mechanical properties is attractive for the large number of emerging applications for flexible electronics. —BG

Science, abj1261, this issue p. 1243

A flexible electronic device made with phase change materials

CREDITS: (LEFT TO RIGHT) KHAN *ET AL*.; ISTOCK.COM/HNJJAR007

2D MATERIALS

Fermi nesting

Correlated states have been shown to emerge in bilayer and trilayer graphene with the two-dimensional layers at just the right angle with respect to each other. Key to the enhanced importance of interactions are the so-called moiré electronic bands that form in such systems. Rickhaus *et al*. explored a related system of two graphene bilayers twisted with respect to each other. The twist angle was set so that the layer coupling was strong enough to form moiré bands but weak enough for the carrier concentration in the top and bottom bilayer to be controlled separately. Doping the top bilayer with electrons and the bottom bilayer with holes, the researchers created a correlated state with nested Fermi surfaces. —JS

Science, abc3534, this issue p. 1257

OPTICS

Absorbingly exceptional

Most oscillating systems have a resonance or multiple resonances at which they ring out and are most sensitive to excitation. In non-Hermitian systems, open systems with gain and loss, the resonances have been found to coalesce into an exceptional point when the gain and loss can be engineered. Complementing these resonant exceptional points, Wang *et al*. show that controlling the absorption of a coupled microresonator system can produce a new kind of absorbing exceptional point. They also show how these exceptional points are distinct and how systems can be engineered to exhibit new scattering behavior. —ISO

Science, abj1028, this issue p. 1261

IN OTHER JOURNALS

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

Dogs, unlike wolves, are highly responsive to human voice, gestures, and eye contact.

CANINE BEHAVIOR

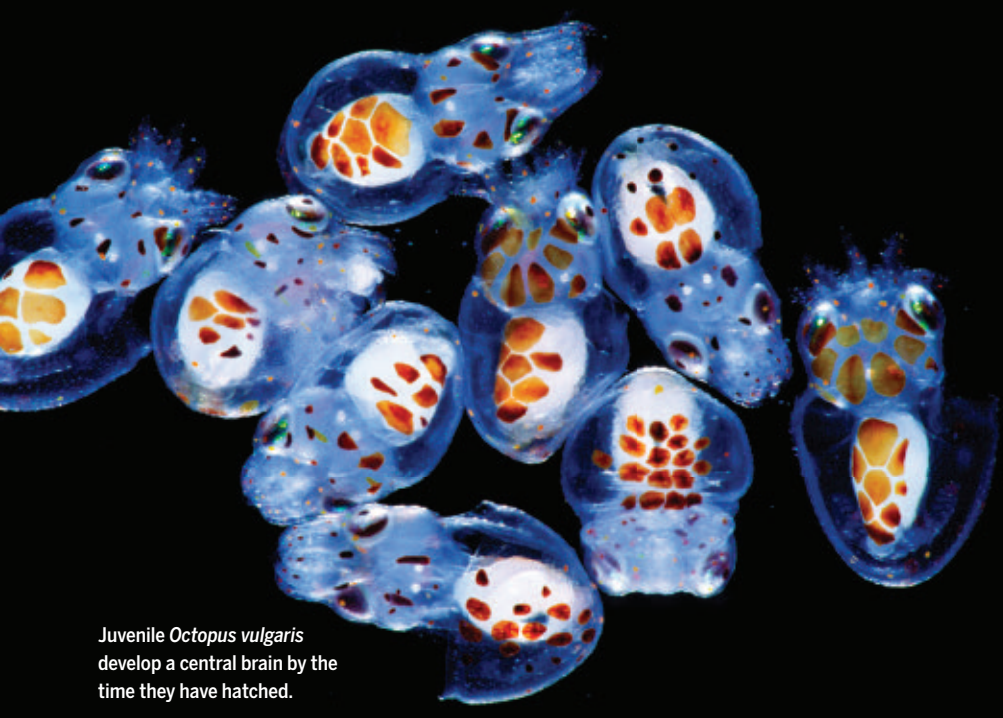
Friends from the start

The closest relative to dogs, "man's best friend," is the wolf, a wily predator that generally avoids human interaction. For decades, researchers and dog owners have wondered how the leap to domestication occurred. The main hypothesis invoked very early selection for wolves that "liked"—or least tolerated—humans, and the connection strengthened from there. However, there is still some debate about whether the degree to which dogs interact and communicate with humans is a learned trait. Two recent studies appear to close the book on this learning hypothesis. Bray *et al.* looked at about 400 puppies and found that at this young age and without much human interaction, they were adept at following human gestures and positively responded to high-pitched "puppy talk." Further, there was variation in these responses with an association between relatedness and social communication skills, which supports a genetic driver. Salomons *et al.* compared dog and wolf puppies and found no difference in general cognitive responses, but much greater responsiveness to human gestures and eye contact, in dog puppies. Importantly, this happened even though the dog pups had received less actual human interaction than did the wolf pups. These studies confirm that dogs' interest in communication with humans is an evolved trait unique to their lineage. —SNV *Curr. Biol.* **31**, 3132, 3137 (2021).

SUPERCONDUCTIVITY

Peeking into vortex cores

When placed in a magnetic field, some superconductors develop a lattice of vortices, tiny regions in which supercurrent circulates around a non-superconducting central region called the vortex core. In *d*-wave superconductors, such as the cuprates, the density of electronic states right at the



Juvenile *Octopus vulgaris* develop a central brain by the time they have hatched.

NEURODEVELOPMENT

Brainy octopus development

It is surprising that a relatively short-lived relative of snails should have evolved extraordinary intelligence. The octopus brain accounts for about one-fourth of its larval body size at hatching. Deryckere *et al.* compared the development of this charismatic invertebrate mollusk's brain with that of the vertebrate (mammalian) brain. By the time they hatch from their transparent eggs at about 40 days of development and despite their tiny size, *Octopus vulgaris* have developed a central brain that surrounds the esophagus. Tracking expression of genes characteristic of differentiating neurons identified the lateral lips as a transient embryonic neurogenic zone. From there, postmitotic neurons migrate to the embryonic central brain following pathways suggestive of guidance cues. The lateral lips shrink as development draws to a close. —PJH *eLife* **10**, e69161 (2021).

vortex core has been predicted to have a peak near zero energy that splits in two away from the core. Gazdić *et al.* performed scanning tunneling spectroscopy measurements in the material $Bi_2Sr_2CaCu_2O_{8+\delta}$ at low magnetic fields and at large hole doping, where cuprate physics are less complicated by competing phases than elsewhere in the phase diagram. These conditions led to reduced interactions between vortices, enabling the researchers to verify the long-standing prediction. —JS

Phys. Rev. X **11**, 031040 (2021).

ELECTROCHEMISTRY

Volta's reduction

Among the most groundbreaking discoveries in science was Alessandro Volta's "pile" of silver and zinc, which we now call the first battery. The observation proved so profoundly useful that zinc is still the anode in disposable batteries today. The cathode has undergone much more tweaking, and, surprisingly, some questions still linger about precisely which reduction reaction took place in Volta's original device. Clarke *et al.* marshal experimental evidence for oxygen (rather than proton) reduction as the key contributor. They further outline a protocol for students to recreate and explore the pile with pennies. —JSY

J. Chem. Educ. 10.1021/acs.jchemed.1c00016 (2021).

IMMUNOLOGY

Tissue cues steer heterogeneity

The CD8 memory T cell pool contains both circulating and noncirculating subsets. This latter population of cells, known as tissue-resident memory T (T_{RM}) cells, occupy tissues and act as rapid responders to infection. Christo *et al.* investigated mouse T_{RM} cell heterogeneity across multiple organs and found distinct tissue microenvironments that drove considerable differences in T_{RM} cell form and function. Variable sensitivity to the cytokine transforming growth factor β (TGF-β) was a major impetus for these differences. TGF-β–responsive CD103$^+$ T_{RM} cells in epithelial tissues such as the skin showed reduced proliferation, impaired function, and restrained differentiation capacity but potentially more durable responses compared with TGF-β–unresponsive CD103$^-$ T_{RM} cells in the liver. —STS

Nat. Immunol. **22**, 1140 (2021).

BIOCHEMISTRY

From light to darkness with vitamin A

Vitamin A plays an essential biochemical role in vision as part of the chromophore in light receptors. Zhang *et al.* studied how vitamin A and its metabolic by-products change during aging and how these changes might affect how receptors adapt to dark conditions. In human eyes, the amount of by-product increases linearly with age. Experiments in mice engineered with defects in vitamin A processing suggest detrimental effects of by-products on receptor recovery after moving from light to dark. Although there may be other mechanisms by which dark adaptation is impaired during aging, treatments that reduce vitamin A by-product formation are worth investigating. —MAF

J. Biol. Chem. 10.1016/j.jbc.2021.101074 (2021).

DEVELOPMENT

Brain organoids with eye features

The complex workings of organ systems are being elucidated by analyzing three-dimensional organ-like structures (organoids) generated from stem cells. Starting with induced pluripotent stem cells, Gabriel *et al.* derived human brain organoids that spontaneously developed primordial eye fields around day 30. Pigmented structures invaginated to assemble optic vesicle–like structures with characteristic cell compositions and the transcriptomic signature for brain, early retina, and eye-related cell populations. The optic vesicle brain organoids were bilaterally symmetrical and contained a lens-like structure with a light-sensitive, corneal epithelium–like layer. —BAP

Cell Stem Cell 10.1016/j.stem.2021.07.010 (2021).

PHOTO: YASUMASA KOBAYASHI/MINDEN PICTURES

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

MARINE CONSERVATION

Consistency in conservation

Marine protected areas (MPAs) are now well established globally as tools for conservation, for enhancing marine biodiversity, and for promoting sustainable fisheries. That said, which regions are labeled as MPAs varies substantially, from those that full protect marine species and prohibit human extraction to those that permit everything from intensive fishing to mining. This inconsistency can in some cases inhibit both conservation and quantifying the proportion of the marine environment that is truly protected. Grorud-Colvert *et al*. review the consistency of MPAs and propose a framework by which levels of protection can be evaluated and improved. —SNV

Science, abf0861, this issue p. 1215

GENE THERAPY

To replace or edit?

Mitochondrial diseases are a complex set of maladies caused by mutations in mitochondrial DNA (mtDNA). Mitochondrial replacement techniques, in which maternal nuclear DNA is transferred into enucleated donor egg cells or zygotes, are showing promise for preventing the transmission of these diseases from at-risk mothers to offspring. Indeed, at least one child has been born using this technique. However, questions have arisen about the efficacy and safety of mitochondrial replacement. In a Perspective, Adashi *et al.* discuss the potential issues with mitochondrial replacement and the alternative approach of editing mtDNA using targeted and specific nucleases. Although much more development of the mtDNA-editing approach is needed, the authors suggest that this may be a powerful approach that might even be combined with mitochondrial replacement to optimize mitigation of mitochondrial diseases. —GKA

Science, abg0491, this issue p. 1200

RIBOSOME

Making ribosomes in human cells

The assembly of the human ribosome requires a vast number of assembly factors that first form a giant precursor of the ribosomal small subunit (SSU), the SSU processome. Singh *et al*. provide new insights into human ribosome assembly by using cryo–electron microscopy, x-ray crystallography, and functional studies to reveal the structures of the human SSU processome as it matures in the nucleolus. This study reveals the tightly controlled molecular choreography by which an ensemble of assembly factors controls early irreversible maturation steps that bring about the formation of the human SSU. —DJ

Science, abj5338, this issue p. 1216

NEUROSCIENCE

Neuroinflammation after brain injury

Traumatic brain injury affects millions of people every year and is a major cause of disability worldwide. Most of the maladaptive outcomes develop months or years later and are thought to be caused by secondary injuries that are indirect and long-term effects after the initial impact. Holden *et al*. found that secondary and chronic neuroinflammation and neurodegeneration are caused by the C1q molecule, a mediator of the complement pathway. C1q is responsible for chronic inflammation and secondary neuronal loss specifically in the cortico-thalamo-cortical circuit. Traumatic brain injury also leads to altered brain states that are caused by the C1q complement pathway. —PRS

Science, abj2685, this issue p. 1217

MOLECULAR BIOLOGY

A Pickering-stabilized intracellular emulsion

Pickering emulsions, droplet suspensions stabilized by solid particles, were discovered more than 100 years ago and are well studied in foods, oils, cosmetics, and pharmaceuticals. The particles adsorb to the droplet interface and prevent the emulsion from coarsening. Folkmann *et al*. report that P granules, biomolecular condensates in *Caenorhabditis elegans*, are an example of an intracellular Pickering emulsion (see the Perspective by Snead and Gladfelter). Biomolecular condensates are cellular compartments that form without traditional lipid membranes. This work raises the possibility that Pickering agents fulfill the role of membranes in biomolecular condensates. —DJ

Science, abg7071, this issue p. 1218; see also abl6282, p. 1198

SUPERCONDUCTIVITY

Exploring cuprate chains

Superconductivity in cuprates takes place in their two-dimensional (2D) layers, but solving even the simplest model of interacting fermions in 2D is a challenge. The theory problem simplifies in 1D, with experiment becoming the tricky part. Chen *et al*. synthesized a cuprate that consists of parallel chains and behaves like a 1D system. Crucially, the material could be doped over a wide range of hole concentrations. The researchers showed that including a near-neighbor attractive interaction in a 1D model of interacting fermions was necessary to explain their photoemission measurements. —JS

Science, abf5174, this issue p. 1235

ATMOSPHERIC DYNAMICS

Water up

Above-anvil cirrus plumes are stratospheric cloud formations that form downwind from the tops of some very strong thunderstorms. Despite their common occurrence, an adequate physical model explaining many of their features and effects does not exist. O'Neill *et al*. show that the storm supercell that extends into the stratosphere functions like a physical barrier, deflecting wind streams there like a topographic obstacle and driving a hydraulic jump downstream at the tropopause (see the Perspective by Smith). This feature triggers the intense injection of water vapor deep into the stratosphere, leading to a substantial increase in stratospheric humidity. —HJS

Science, abh3857, this issue p. 1248; see also abl8740, p. 1194

NEUROSCIENCE

Prevention of compulsive cocaine taking

Over time, about 20% of chronic cocaine users lose control and become addicted. There are indications that the differential efficacy of the brain serotonin (5-HT) system may be involved in the vulnerability to drug addiction. However, the relevant circuits and underlying cellular processes remain elusive. Li *et al*. discovered a synaptic mechanism in mice that underlies the modulatory role of 5-HT in reducing the likelihood of transition to compulsion and eventually addiction (see the Perspective by Miyazaki and Miyazaki). Cocaine binds to 5-HT transporters to block 5-HT reuptake. The elevated extracellular 5-HT activates 5-HT_{1B} receptors and causes presynaptic depression of a projection from the orbitofrontal cortex to the dorsal striatum. These changes reduce the likelihood of inducing postsynaptic potentiation at these synapses, which ultimately drives compulsion. —PRS

Science, abi9086, this issue p. 1252; see also abl6285, p. 1197

OLIGONUCLEOTIDES

Platform for the synthesis of diverse oligos

DNA is primarily viewed as a carrier of information encoded in the sequence of bases, but the chemistry of the phosphodiester backbone is crucial to oligonucleotide stability and structure. Building on previous work in synthetic P(V) phosphorothioate coupling chemistry, Huang *et al*. developed two new reagents for making phosphorodithioate- and phosphate-based linkages (see the Perspective by Virta). The authors incorporated all of these reagents into a unified P(V)-based synthesis platform capable of running at high efficiency on a commercial automated solid-phase synthesizer. They demonstrate the flexibility of this system by producing oligonucleotides with all three linkage types in specific positions. Access to such precisely constructed molecules opens new approaches to therapeutic oligonucleotide design. —MAF

Science, abi9727, this issue p. 1265; see also abk3478, p. 1196

HIV

PD-1 blockade improves SIV vaccination

Effective HIV therapies must induce functional $CD8^+$ T cells and clear latent viral reservoirs during antiretroviral therapy (ART). Using immune checkpoint blockade (ICB) alone during ART does not substantially boost T cell responses. Combining vaccines with ICB improves $CD8^+$ T cell function in other infections and thus might be an effective strategy for HIV. Rahman *et al.* used a chronic simian immunodeficiency virus (SIV) rhesus macaque model to test whether the addition of anti–PD-1 antibody to an SIV vaccine would induce SIV-specific T cells and control virus after ART interruption. This strategy increased cytotoxic $CD8^+$ T cells in the lymph nodes and was correlated with control of virus and prolonged survival compared with the vaccine alone. These results suggest that the combination of ICB with an HIV vaccine is a promising therapeutic strategy. —IRW

Sci. Immunol. **6**, eabh3034 (2021).

REVIEW SUMMARY

MARINE CONSERVATION

The MPA Guide: A framework to achieve global goals for the ocean

Kirsten Grorud-Colvert*, Jenna Sullivan-Stack, Callum Roberts, Vanessa Constant, Barbara Horta e Costa, Elizabeth P. Pike, Naomi Kingston, Dan Laffoley, Enric Sala, Joachim Claudet, Alan M. Friedlander, David A. Gill, Sarah E. Lester, Jon C. Day, Emanuel J. Gonçalves, Gabby N. Ahmadia, Matt Rand, Angelo Villagomez, Natalie C. Ban, Georgina G. Gurney, Ana K. Spalding, Nathan J. Bennett, Johnny Briggs, Lance E. Morgan, Russell Moffitt, Marine Deguignet, Ellen K. Pikitch, Emily S. Darling, Sabine Jessen, Sarah O. Hameed, Giuseppe Di Carlo, Paolo Guidetti, Jean M. Harris, Jorge Torre, Zafer Kizilkaya, Tundi Agardy, Philippe Cury, Nirmal J. Shah, Karen Sack, Ling Cao, Miriam Fernandez, Jane Lubchenco

BACKGROUND: Marine Protected Areas (MPAs) are places in the ocean that receive protection to safeguard biodiversity from abatable threats. Confusion exists about the definition of "protection" and likely MPA outcomes. This is because not all MPAs are the same. They range from full to minimal protection; some exist only on paper, not in practice. The resulting, understandably divergent outcomes can lead to controversies about effectiveness, undermine confidence in MPAs, and jeopardize conservation goals, including those of the Convention on Biological Diversity and the United Nations (UN) Sustainable Development Agenda. We integrated decades of research to clarify these issues.

ADVANCES: We propose a science-based, policy-relevant framework—The MPA Guide—to categorize, evaluate, and plan MPAs. It complements the well-known International Union for Conservation of Nature (IUCN) Protected Area Categories for management objectives and governance types. Together, these tools enable a comprehensive picture of any MPA.

The guide consists of four elements that define types of MPAs and activities, conditions for success, and likely outcomes. First, the four STAGES of establishment of an MPA are (i) Proposed/Committed, by a governing or other organizing body; (ii) Designated, by law or other authoritative rulemaking; (iii) Implemented, with activated regulations; and (iv) Actively Managed, with ongoing monitoring and adaptive management.

Second, the four LEVELS of protection from abatable activities within an MPA (or MPA zone), based on allowed activities, are (i) Fully Protected—no impact from extractive or destructive activities; (ii) Highly Protected—minimal impact; (iii) Lightly Protected—moderate impact; and (iv) Minimally Protected—high total impact, although still an MPA by IUCN criteria.

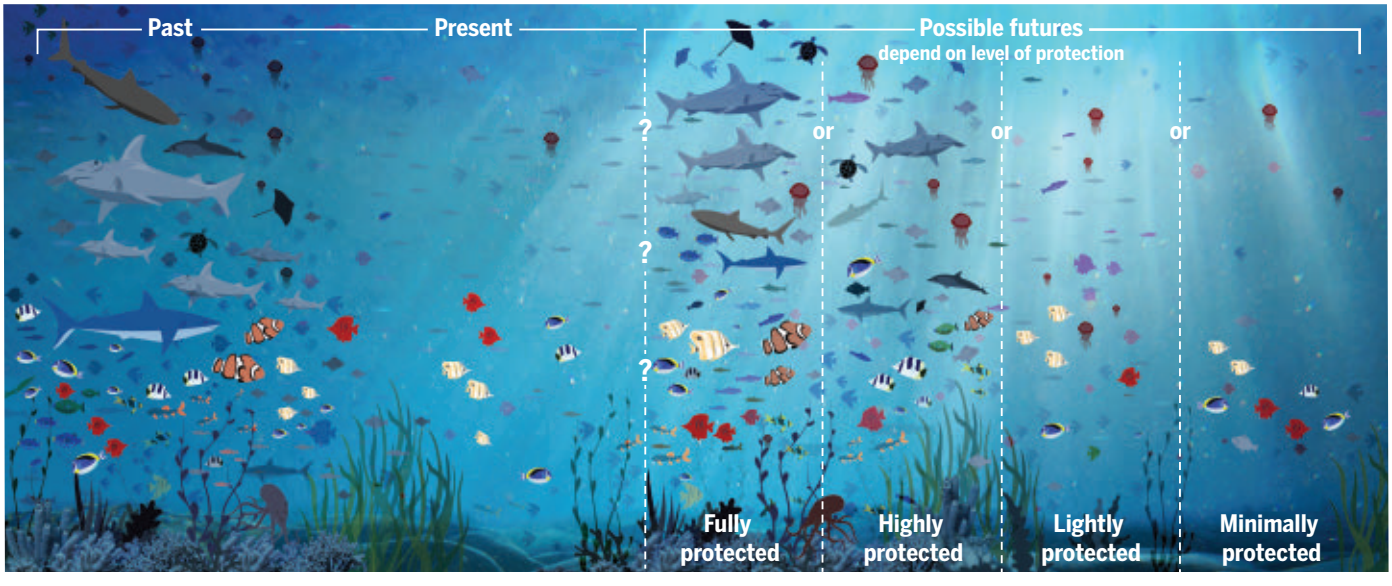
Third, to succeed, an MPA should be established and sustained through the enabling CONDITIONS for effective and equitable MPA planning, design, governance, and management.

Fourth, the likely OUTCOMES of an MPA depend directly on STAGE, LEVEL, and CONDITIONS to succeed.

OUTLOOK: The MPA Guide enables smart planning, design, and evaluation of new or existing MPAs by informing decisions about scientific, societal, and policy priorities and facilitates evaluating progress on international conservation targets. The guide draws attention to quality, not just quantity, of MPAs. It points to fully or highly protected areas as having the greatest likelihood of achieving biodiverse and healthy ecosystems, once the MPA is implemented or actively managed, if enabling CONDITIONS are in place. Last, our synthesis also identifies research priorities, including examining MPAs' effectiveness across LEVEL of protection for climate mitigation and adaptation, social change, and comprehensive marine spatial planning. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: grorudck@oregonstate.edu
Cite this article as Grorud-Colvert *et al.*, *Science* **373**, eabf0861 (2021). DOI: 10.1126/science.abf0861

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf0861



The level of protection, and therefore the effectiveness of MPAs, will greatly influence the future state of the ocean. Past ocean ecosystems were abundant and diverse in species and habitats. Over time, expanded and intensified human activities depleted and disrupted ocean ecosystems and reduced their services. MPAs, in conjunction with climate mitigation strategies and more sustainable uses of the ocean, can conserve and restore biodiversity and the resilient ecosystems needed for human well-being. Different levels of protection will result in different outcomes, if enabling conditions are satisfied.

REVIEW

MARINE CONSERVATION

The MPA Guide: A framework to achieve global goals for the ocean

Kirsten Grorud-Colvert[1]*, Jenna Sullivan-Stack[1], Callum Roberts[2], Vanessa Constant[3]†, Barbara Horta e Costa[4], Elizabeth P. Pike[5], Naomi Kingston[6], Dan Laffoley[7], Enric Sala[8], Joachim Claudet[9], Alan M. Friedlander[10,11], David A. Gill[12], Sarah E. Lester[13], Jon C. Day[14], Emanuel J. Gonçalves[15,16], Gabby N. Ahmadia[17], Matt Rand[18], Angelo Villagomez[18], Natalie C. Ban[19], Georgina G. Gurney[20], Ana K. Spalding[21,22], Nathan J. Bennett[23], Johnny Briggs[18], Lance E. Morgan[24], Russell Moffitt[5], Marine Deguignet[6], Ellen K. Pikitch[25], Emily S. Darling[26], Sabine Jessen[27], Sarah O. Hameed[28], Giuseppe Di Carlo[29], Paolo Guidetti[30,31], Jean M. Harris[32], Jorge Torre[33], Zafer Kizilkaya[34], Tundi Agardy[35], Philippe Cury[36], Nirmal J. Shah[37], Karen Sack[38], Ling Cao[39], Miriam Fernandez[40], Jane Lubchenco[1]

Marine Protected Areas (MPAs) are conservation tools intended to protect biodiversity, promote healthy and resilient marine ecosystems, and provide societal benefits. Despite codification of MPAs in international agreements, MPA effectiveness is currently undermined by confusion about the many MPA types and consequent wildly differing outcomes. We present a clarifying science-driven framework—The MPA Guide—to aid design and evaluation. The guide categorizes MPAs by stage of establishment and level of protection, specifies the resulting direct and indirect outcomes for biodiversity and human well-being, and describes the key conditions necessary for positive outcomes. Use of this MPA Guide by scientists, managers, policymakers, and communities can improve effective design, implementation, assessment, and tracking of existing and future MPAs to achieve conservation goals by using scientifically grounded practices.

Marine Protected Areas (MPAs) are one of many tools that policy-makers, managers, and communities use to stem the loss of biodiversity, disruption of ocean ecosystems, and the decline of the many benefits provided to people by healthy ocean ecosystems (*1*, *2*). Although most of the ocean used to be a de facto MPA because of limited access, technology has enabled exploitation of almost all of the ocean (*3*). In addition, although there are numerous examples of successful traditional resource management, customary marine governance including the use of closed areas has been eroded in many countries as a result of processes such as colonization and market expansion (*4*). Because degradation, pollution, and exploitation have substantially affected the open ocean, the coast, and adjoining lands (*1*), integrated efforts are urgently needed to make extractive uses sustainable, minimize impact of destructive activities, and expand effective protection of species, habitats, and ecosystem functioning (*5*, *6*).

MPAs by definition prioritize the conservation of nature [International Union for Conservation of Nature (IUCN); (*7*)] and are the primary area-based tool for marine biodiversity conservation. In this Review, we focus only on MPAs because of their prevalence and extensive scientific underpinnings based on decades of tracking and evaluation (*8*, *9*). Other area-based management tools for which biodiversity conservation is not the primary goal are not MPAs, although they may confer some conservation benefit. For example, Locally Managed Marine Areas (*10*) or Fisheries Management Areas (*11*) have different management priorities. If the protection they provide effectively conserves biodiversity, they may qualify as Other Effective area-based Conservation Measures (OECMs) (*12*, *13*). International governing bodies have set global targets for MPAs and OECMs—for example, to protect 10% of the ocean by 2020 [Convention on Biological Diversity's (CBD's) Aichi Target 11 (*14*) and United Nations Sustainable Development Goal 14.5 (*15*)]. Calls are increasing for the more ambitious target of effectively protecting at least 30% of the ocean by 2030 (*16*, *17*).

However, confusion and disagreements pervade many discussions of MPAs and detract from conservation efforts. Quantifying how much of the ocean's biodiversity is effectively protected is challenging. Substantial discrepancies exist over what "protected" means, when to "count" an area as protected, and which types of MPAs achieve the intended conservation goals (*8*, *9*, *18*, *19*). At present, global databases document that a relatively small proportion of the ocean is protected in MPAs. Specifically, at the time of this writing 7.7% of the ocean is self-reported by countries as existing in some type of designated MPA (*20*), but only 5.9% is in MPAs that have been implemented, with likely much less actively managed (*21*). Additionally, not all of the tallied areas in those percentages meet the IUCN definition of an MPA (*7*). The race to simply protect a certain percentage of the ocean could detract from the importance of MPA quality, leading to perverse outcomes from establishing MPAs that are insufficiently protected or not adequately designed to achieve conservation goals (*22*).

Removing confusion around MPAs

We posit that much of this confusion can be resolved by addressing three critical questions.

[1]Department of Integrative Biology, Oregon State University, 3029 Cordley Hall, Corvallis, OR, USA. [2]Department of Environment and Geography, University of York, York YO10 5DD, UK. [3]Department of Integrative Biology, Oregon State University, 3029 Cordley Hall, Corvallis, OR, USA. [4]Center of Marine Sciences, CCMAR, University of Algarve, Campus de Gambelas, Faro, 8005-139, Portugal. [5]Marine Protection Atlas, Marine Conservation Institute, Seattle, WA, 98103-9090, USA. [6]UN Environment Programme World Conservation Monitoring Centre, Cambridge, UK. [7]IUCN World Commission on Protected Areas, International Union for Conservation of Nature (IUCN), CH-1196 Gland, Switzerland. [8]National Geographic Society, Washington, DC, USA. [9]National Center for Scientific Research, PSL Université Paris, CRIOBE, USR 3278 CNRS-EPHE-UPVD, Maison des Océans, 75005 Paris, France. [10]Hawai'i Institute of Marine Biology, University of Hawai'i, Kāne'ohe, HI 96744, USA. [11]Pristine Seas, National Geography Society, Washington, DC 20036, USA. [12]Duke University Marine Laboratory, Nicholas School of the Environment, Duke University, Beaufort, NC 28516, USA. [13]Department of Geography, Florida State University, Tallahassee, FL 32306-2190, USA. [14]ARC Centre of Excellence in Coral Reef Studies, James Cook University, Townsville QLD 4811, Australia. [15]Marine and Environmental Sciences Centre (MARE), ISPA–Instituto Universitário, 1149-041 Lisbon, Portugal. [16]Oceano Azul Foundation, Oceanário de Lisboa, Esplanada D. Carlos I,1990-005 Lisbon, Portugal. [17]Ocean Conservation, World Wildlife Fund, Washington, DC 20037, USA. [18]Pew Bertarelli Ocean Legacy Project, The Pew Charitable Trusts, Washington, DC 20004-2008, USA. [19]School of Environmental Studies, University of Victoria, Victoria, BC V8W 2Y2, Canada. [20]Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia. [21]School of Public Policy, Oregon State University, Corvallis, OR 97330, USA. [22]Smithsonian Tropical Research Institute, Panama City, Panama; Coiba Scientific Station (Coiba AIP), Panama City, Panama. [23]The Peopled Seas Initiative, Vancouver, BC, Canada. [24]Marine Conservation Institute, Seattle, WA 98103, USA. [25]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY 11794, USA. [26]Wildlife Conservation Society, 2300 Southern Blvd, Bronx, NY 10460, USA. [27]National Ocean Program, Canadian Parks and Wilderness Society, Ottawa, ON K2P 0A4, Canada. [28]Blue Parks Program, Marine Conservation Institute, Seattle, WA 98103, USA. [29]World Wide Fund For Nature (WWF)–Mediterranean, Rome 00198, Italy. [30]Department of Integrative Marine Ecology (EMI), Stazione Zoologica A. Dohrn–National Institute of Marine Biology, Ecology and Biotechnology, Villa Comunale, 80121 Naples, Italy. [31]National Research Council, Institute for the Study of Anthropic Impact and Sustainability in the Marine Environment (CNR-IAS), V16149 Genoa, Italy. [32]Institute for Coastal and Marine Research (CMR), Nelson Mandela University, Gomeroy Avenue, Summerstrand, Port Elizabeth 6031, South Africa. [33]Comunidad y Biodiversidad, A.C. Isla del Peruano 215, Col. Lomas de Miramar, Guaymas, Sonora, 85454, Mexico. [34]Mediterranean Conservation Society, Bornova, Izmir 35100 Turkey. [35]Sound Seas, Colrain, MA 01340, USA. [36]MARBEC, Montpellier University, CNRS, IRD, IFREMER, Sète, France. [37]Nature Seychelles, Centre for Environment and Education, Sanctuary at Roche Caiman, Mahe, Seychelles. [38]Ocean Unite, Washington, DC 20007, USA. [39]School of Oceanography, Shanghai Jiao Tong University, Shanghai 230000, China. [40]Estación Costera de Investigaciones Marinas de Las Cruces and Departmento de Ecología, Pontificia Universidad Católica de Chile, Santiago, Chile.

***Corresponding author. Email: grorudck@oregonstate.edu** †Present address: The National Academies of Sciences, Engineering, and Medicine, 500 5th St., NW, Washington, DC 20001, USA.

(i) What does "protected" mean for biodiversity conservation? Even under the IUCN definition, the "protected" in "marine protected area" encompasses numerous levels of protection with an almost endless variety and combination of activities that are allowed or not allowed and, consequently, lead to a wide range of impacts on biodiversity (*8*, *23*). As a result, the ecological and social outcomes expected from MPAs, or a zone within an MPA, vary widely [for example, (*18*)]. Clarifying why MPAs differ from one another and which types will deliver specific desired outcomes is essential to help evaluate whether any given MPA is set up with the appropriate protection to achieve its aims.

(ii) When should an MPA count as effectively protected? There are many steps in the process to create an MPA. Global tallies differ from one another in part because they use different criteria to count MPAs [for example, (*20*, *21*)]—for example, when it is proclaimed in law versus when it is implemented in the water. This disparity becomes problematic when some MPAs are counted as achievements toward global targets but no real protection is in place in the water (*24*). There is a need to track all stages of MPAs and clarify that biodiversity protection is not expected to begin until the MPA rules and regulations are in place and active.

(iii) What is needed to achieve effective ocean protection? To prevent overestimation of how much ocean is actually protected (*9*, *19*, *25*, *26*), knowledge of the total MPA coverage across different levels of protection is needed at the global scale. This requires assessment of the number, area, and impact of MPAs to ensure these are sufficient to achieve local, national, or international goals for healthy, productive, and resilient ocean ecosystems that support biodiversity and sustainable use (*27*).

A new framework to understand protected areas in the ocean

With input from diverse global collaborators, we reviewed MPA science and its implications for global biodiversity conservation targets to develop a multidisciplinary, collaborative scientific synthesis that addresses the above three questions. We present our findings as a new framework called The MPA Guide. This guide organizes MPAs according to stage of establishment (STAGE) and level of protection (LEVEL), defined below. We then link these MPA types to measured outcomes (OUTCOMES), on the basis of the enabling social and ecological conditions (CONDITIONS) that research shows are key to an MPA successfully achieving its goals.

This guide strategically complements the IUCN Protected Area Categories, an existing framework that categorizes areas by their management objectives and governance types [IUCN (*28*)], but not by STAGE or LEVEL of protection. Together, the MPA Guide and the IUCN Categories provide a comprehensive picture of an MPA. This guide helps to consolidate and advance the reporting framework of United Nations Environment Programme (UNEP)–World Conservation Monitoring Centre (WCMC) World Database on Protected Areas (WDPA) and the IUCN MPA Standards (*7*), which summarize and distill approved motions by the global conservation community in past World Conservation Congresses. As long as an MPA (or zone within a multizone MPA) meets the IUCN definition of an MPA (*7*), it will fit into one STAGE of establishment and one LEVEL of protection at any point in time.

Stage of establishment (STAGE) and when to count an MPA

MPA establishment generally occurs as a series of steps by governing or other authorities on the basis of their local and national context. This guide specifies minimum criteria for an MPA to achieve each STAGE and provides guidelines for best practices (STAGES Expanded Guidance) (fig. S1). In some cases, it may take several years between an announcement of intent to create an MPA to the time when in situ protection and management occurs. In other situations, an MPA may be designated and implemented simultaneously if the announcement has legal authority and a management plan. Below, we describe each STAGE and provide examples.

At STAGE Proposed/Committed, the intent to create an MPA is made public. An MPA must be announced in some formal (although nonbinding) manner by means of a statement by a government, community, conservation organization, or other organizing group—for example, through an international meeting, a press release, or online. The MPA site must be identified, ideally with clear goals and informed by stakeholder and rights-holder participation, and that of Indigenous or other local peoples, and scientific knowledge of the social-ecological context. At the time of this writing, two examples of proposed/committed MPAs are in the East Antarctic (*29*) and in the Weddell Sea (*30*), where potential MPAs are currently under consideration by the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR).

At STAGE Designated, the MPA is established or recognized through legal means or other authoritative rulemaking. A designated MPA must satisfy three minimum criteria: (i) defined boundaries, (ii) legal gazetting or equivalent Indigenous or traditional authorization or customary recognition, and (iii) clearly stated goals and process to define allowed uses and associated regulations or rules to control impact. MPA boundaries (including zones within the MPA) are ideally published, unambiguous, and known to local users. A designated MPA should have a database ID number in the WDPA that signifies official recognition of the MPA. The MPA should be long term; for example, it should not have a sunset clause or review process that allows for rescinding protection in less than 25 years (*7*). As an example, Seychelles recently legally designated 30% of its ocean territory as an MPA network, which is currently in the process of being implemented (*31*).

MPAs that are proposed/committed or designated are not yet implemented with changes in activities and thus will not accrue biodiversity conservation benefits. Protection does not begin until implementation. MPAs that are designated for an extended period of time without being implemented are often referred to as "paper parks." These situations may reflect a lack of capacity and support (*24*).

At STAGE Implemented, the MPA has transitioned from existence "on paper" to being operational "in the water," with management plans activated. Biodiversity conservation benefits begin to accrue at this stage, not before. Resource users are aware of the rules, and mechanisms to promote compliance and enforcement exist. Plans for regulating MPA activities are in place. Stakeholders are engaged, users are aware of regulations, financial and human resource management systems are established, and performance measures are part of a plan to evaluate and monitor the MPA. Ideally, governance and administrative structures for management, implementation, and sustainable financing are specified (such as in management plans). Zones and their goals should be described, if applicable [for example, (*32*)]. A management body should exist to implement and review plans. For example, Niue's Moana Mahu MPA is implemented (*33*) and includes 40% of the exclusive economic zone (EEZ) as fully protected, with enforcement activities underway, partnerships in place, and ongoing stakeholder engagement.

At STAGE Actively Managed, MPA management is ongoing, including monitoring, periodic review, and adjustments made as needed to achieve biodiversity conservation and other ecological and social goals. All necessary MPA management activities for sustained functioning and achievement of goals continue. The MPA management authority documents, monitors, and evaluates MPA outcomes. Adaptive management will lead to adjustments in plans and activities as needed to ensure good compliance, stakeholder and rights-holder collaboration, and achievement of MPA goals. Comprehensive systems exist to evaluate actively managed MPAs, such as the IUCN Green List (*34*) and the Marine Conservation Institute's Blue Parks Program (*35*). Periodic reviews of actively managed areas are based on evaluations of MPA management function such as sustainable financing, staffing, and outreach

as well as data collected frequently inside and outside the MPA. These should involve a social-ecological systems approach (*36*) and ideally the participation of local communities and stakeholders (*37*). For example, the California network of MPAs established by the Marine Life Protection Act is actively managed and undergoes a systematic and comprehensive periodic 5-year review by using monitoring data to evaluate current management and inform future decisions (*38*).

Level of protection (LEVEL) for biodiversity conservation

By IUCN's definition, an MPA's primary goal is the conservation of nature (*7*). Thus, this guide focuses on evaluating protection on the basis of the biodiversity outcomes that different activities at different scales are expected to produce.

Extensive peer-reviewed research shows that MPAs, or specific MPA zones, effectively protect biodiversity if they adequately prohibit extractive and destructive uses [for example, (*39*–*42*); a list of others is provided in the supplementary materials, LEVELS Expanded Guidance, and fig. S1] and if key factors for positive desired outcomes are in place (CONDITIONS). It is possible to conserve biodiversity while also balancing sustainable uses (*43*); assuming full compliance with rules, some extractive and destructive activities may be allowed in an MPA, albeit with conservation outcomes that are likely more limited [for example, (*18*, *42*, *44*)].

This MPA Guide describes four LEVELS of protection based on the impact of allowed activities. It incorporates guidance from the Regulations-Based Classification System for MPAs (*23*) and IUCN's guidelines (*7*). Impact is determined by activity type, intensity, scale, duration, and frequency relative to biodiversity conservation goals and is described as "none," "low," "moderate," "high/large," or "incompatible with biodiversity conservation" (Fig. 1 and fig. S2). Using this impact scale, a LEVEL of protection can be assigned for any given MPA or zone regardless of location, species, or circumstances. Impacts of certain activities may scale differently considering specific features of an MPA or zone, such as size; for example, distribution of an activity across areas of different sizes may render it high impact in a smaller MPA but moderate impact in a larger MPA. Incompatible activities include industrial extraction such as industrial fishing (for example, vessels > 12 m using towed or dragged gears), oil and gas exploration, mining, or other extremely impactful activities such as fishing with dynamite or poison (supplementary materials, LEVELS Expanded Guidance,

Fig. 1. Level of protection based on maximum allowed impact of seven potential activities in MPAs. An MPA or MPA zone can be categorized into one of four LEVELS of protection: Fully, Highly, Lightly, or Minimally, on the basis of seven types of activities and their impacts (a decision tree approach is available in fig. S2). Dials indicate the scale of impact that may be occurring at a given protection level: none, minimal, low, moderate, or high/large. If impacts are high/large, the site must still provide some conservation benefit to meet the definition of an MPA. If the impact of any of these activities is greater than high, the MPA is incompatible with the conservation of nature (fig. S2). For example, some activities such as mining and mineral and oil prospecting have such a high impact that they are incompatible with biodiversity conservation and should not occur in any MPA; here, the allowed impact of mining is scored as "none" across all four LEVELS.

and fig. S1). Any activity that may be conducted for scientific research purposes in an MPA or zone is subject to the review and approval of the MPA management authority based on its impact.

At LEVEL Fully Protected, no extractive or destructive activities are allowed; all abatable impacts are minimized. Minimizing impacts requires attention to the scale of the protected area and the scale of the activity. MPAs cannot abate or prevent some impacts (such as climate change, coastal urbanization, or pollution), although in certain circumstances, they can enhance ecosystem resistance and resilience (both the ability of the ecosystem to resist impacts of disturbance and to return to a healthy state after disturbance) to some of these threats (*45*). The meaning of other, similar terms such as "strong" or "strict" protection, "marine reserves," or "no-take" areas varies considerably from user to user (*46*). We use and clearly define the term "fully protected" because it encompasses more than just extractive activities and emphasizes the positive intent of the action (compared with "no-take," which emphasizes what is prohibited). Nonextractive low-impact tourism or low-impact cultural activities may be compatible with fully protected areas, provided collective impact is low (Fig. 1 and fig. S2). Potentially impactful activities such as aquaculture are only allowed for restoration purposes and not extraction. Examples include small-scale, decades-old community comanaged MPAs in the Philippines (*47*), large-scale MPAs such as the Palau National Marine Sanctuary [which covers 80% of the country's EEZ (*48*)], or zones within multizone MPAs (*19*).

At LEVEL Highly Protected, only light extractive activities with low total impact are allowed, with all other abatable impacts minimized. Some allow a small amount of subsistence or small-scale fishing with minimal impact, depending on the number of fishers and gear types [up to five or fewer low-impact gears; for example, use by few fishers of highly selective gear such as hand lines or collection by freedivers may be compatible with highly protected status (*23*)]. Allowed activities include low-impact tourism and low-density, unfed aquaculture. Highly protected areas may allow low-impact cultural and traditional activities such as sustainable fishing by Indigenous communities [for example, (*49*)], which are supported by clear property rights affording local stakeholders and rights-holders the authority to govern areas, including restricting exploitation by nonlocal actors (*50*). The 2016 expansion zone of the United states's Papahānaumokuākea Marine National Monument, which allows only low-frequency and low-impact activities, is highly protected (*51*).

At LEVEL Lightly Protected, some protection of biodiversity exists, but moderate to substantial extraction and other impacts are allowed. These MPAs can achieve some protection of biodiversity for certain species or habitats, but the number and impacts of activities allowed are greater than for highly protected areas. A larger number of fishing gear types might be used [10 or fewer (*23*)], or fishing occurs with less selective gear types (such as gill, trammel, or small-scale drift nets). Tourism could have moderate impacts on habitats and species, such as damage caused by high-intensity recreational diving. Aquaculture may occur by means of semi-intensive, unfed methods or small-scale and low-density fed methods. The vast majority of MPAs worldwide are lightly protected or minimally protected (*9*, *19*, *21*) and often attempt to balance biodiversity conservation goals with resource use and development goals. For example, Habitat Protection Zones in Australia's Great Barrier Reef Marine Park are lightly protected because they allow multiple types of fishing (*52*).

At LEVEL Minimally Protected, extensive extraction and other impacts are allowed, but the site still provides some conservation benefit in the area. Extensive extraction and other impacts occur in a minimally protected area, but the area still achieves sufficient biodiversity conservation to satisfy the IUCN definition of an MPA. For example, the area must not allow industrial fishing (*53*). Nonetheless, minimally protected areas are unlikely to deliver substantial biodiversity conservation benefits for nature and people. A recent analysis showed that more than 10 fishing gear types used in an MPA either recreationally or commercially likely leads to large-scale impacts (*18*, *23*). Minimally protected MPAs often allow many or high-impact gear types for extraction and may include medium- to high-density aquaculture and/or large-impact anchoring or infrastructure. For example, the US Hawaiian Islands Humpback Whale National Marine Sanctuary is minimally protected because it allows extensive fishing and anchoring (*54*).

LEVELs of protection are designed to harmonize with and build on, but not replicate, the information provided by the IUCN Protected Area Categories. For this reason, LEVEL in The MPA Guide does not map directly to an IUCN Category. The zones in the Great Barrier Reef (GBR) Marine Park provide useful examples. Some fully protected zones correspond with IUCN Category Ia; for example, the GBR Preservation Zones are "no-go" areas with all extractive and destructive activities prohibited. Other Category Ia areas, such as the GBR Scientific Research Zones, are highly protected with low-impact extractive research and traditional resource use allowed. The GBR Conservation Park and Buffer zones are both IUCN Category IV MPAs, but Buffer Zones are highly protected, whereas Conservation Park zones are lightly protected because of the range of fishing gears allowed.

Within The MPA Guide, LEVEL of protection for any particular MPA, or zone within a multizone MPA, depends on activity types that are explicitly permitted or prohibited by the MPA rules or are based on overlapping regulations for the surrounding area (Fig. 1). Some activity types or impact levels are not explicitly stated in MPA rules and regulations, often because they are not within the management jurisdiction of the MPA authority. In these circumstances, knowledge of whether or not that activity occurs may be used. Because it is the current activities that influence the degree to which an MPA is protecting biodiversity at a given point in time, the assessment of MPA LEVEL should reflect activities actually occurring in the site at the time of reporting, whether or not they are explicitly stated in the management plans.

Seven main types of activities determine LEVEL: (i) mining/oil and gas extraction; (ii) dredging and dumping; (iii) anchoring; (iv) infrastructure; (v) aquaculture; (vi) fishing, whether it is subsistence, professional, or recreational fishing—this activity encompasses extraction of wild fish and other marine species and includes gleaning; and (vii) nonextractive activities, including recreational, traditional, and cultural (supplementary materials, LEVELS Expanded Guidance, and fig. S1). The compatibility of each activity with conservation goals was evaluated through multiple, iterative workshops that used peer-reviewed literature, scientific judgment, expert opinion, and IUCN resolutions and protected area guidance [for example, (*23*)].

This guide does not include every possible activity but provides best practices wherever possible. For example, shipping is not explicitly addressed because the right of innocent passage is mandated under international law and regulated by International Maritime Organization treaties. As a result, it is challenging for an MPA managing authority to restrict shipping movement. Nonetheless, it is recommended that ships with dangerous goods or toxic antifouling chemicals not transit MPAs, and that shipping activity be restricted to shipping lanes to minimize noise pollution and other negative impacts, such as collisions with marine life (supplementary materials, LEVELS Expanded Guidance, and fig. S1) (*55*). Guidance is intended to evolve with new knowledge, activities, and technology. Emerging threats due to electromagnetic fields, excessive or persistent noise, high-energy active sonar, or other technologies not explicitly addressed here are subject to the burden of proof [for example, (*55*, *56*)], meaning that management bodies should receive evidence of their expected impacts before allowing their use and should monitor to assess and actively manage

their actual impacts. Impacts should not exceed those associated with a given LEVEL.

Enabling conditions (CONDITIONS) for effective MPAs

MPAs cannot achieve their goals unless key CONDITIONS are in place. These are the conditions by which an MPA is effectively planned, designed, implemented, governed, and managed to achieve desired ecological outcomes and the direct and indirect human well-being outcomes that result. These CONDITIONS may vary in their importance during the process of achieving each of the four STAGES [for example, (*57*–*59*)] (Table 1), but aspects of each apply when moving from proposed/committed to designated [for example, (*60*–*62*)], to implemented [for example, (*34*, *63*)], and to actively managed [for example, (*34*, *63*, *64*)]. They will also vary according to local challenges, opportunities, and resources, requiring engagement in a prioritization process that is specific to each context.

The beneficial governance practices that these CONDITIONS span—such as inclusivity, transparency, and accountability—increase legitimacy, ownership, support, and overall effectiveness of conservation (*65*, *66*). These practices give voice to those who often disproportionately bear the costs of degradation or conservation and identify livelihood support or other strategies to help mitigate impacts and increase benefits. For example, MPAs in the Mediterranean received greater support from community members with transparent decision-making that recognized and strengthened the rights of local resource users (*66*). MPAs that exclude resource users from decision-making and ignore their rights and livelihood dependencies can erode their well-being and undermine compliance. In Mnazi-Bay, Tanzania, exclusion of resource users from

Table 1. Enabling conditions for effective MPAs. These CONDITIONS may vary in their importance during the process of achieving each of the four STAGES.

Enabling conditions across all stages of establishment	• Clearly defined vision and objectives • Long-term political will and commitment • Sustainable financing • Public participation with contextual and procedural fairness • Evidence-based decision-making • Knowledge integration, e.g., across academic disciplines, local, Indigenous, practitioner domains • Coordination with related governance institutions • Collaboration across jurisdictions	• Transparency and communication • Upward and downward accountability to legal mandate and to stakeholders • Recognition and support of existing governance by Indigenous peoples and local rights-holders, including sovereignty, self-determination, and rights of access, use, and management • Conflict resolution mechanisms
Enabling conditions from Proposed/ Committed to Designated	*All the Enabling Conditions above, plus:* Ecological design principles: • Viability based on MPA location, size, spacing, shape, and permanence • Representativeness and replication of habitats • Incorporation of habitats and species of unique conservation value • Design for connectivity and resilience • Precautionary approach considering current and emerging threats • Consideration of existing threats and mitigation	Social design principles: • Inclusion of social objectives for multi-dimensional human well-being • Recognition of pre-existing rights, tenure, uses: extractive and non-extractive • Consideration of pre-existing resource use and socio-economic status • Accounting for unequal costs and benefits to different social groups • Impact- and benefit-sharing with distributional fairness
Enabling conditions from Designated to Implemented	*All the Enabling Conditions above, plus:* • Sufficient and properly organized staffing and funding • Appropriate and adequate administrative structures and processes • Stakeholder engagement plan	• Compliance and enforcement (including graduated sanctioning) • Education and outreach initiatives • Clarity of rules, rights, and boundaries
Enabling conditions from Implemented to Actively Managed	*All the Enabling Conditions above, plus:* • Ongoing monitoring, evaluation, and knowledge sharing • Adaptive management • Support for livelihoods, e.g. development programs, capacity building, hiring • Effective management of broader seascape and external pressures	• Ongoing efforts to build trust, strong local leadership, partnerships with local users • Local collaboration in monitoring, enforcement, and management • Ongoing consideration of cultural values, traditions, and activities in site management

the MPA process led to negative social outcomes, including increased food insecurity, violent conflict, and lower educational outcomes (*67*).

Linking MPA goals to measurable outcomes: Achieving ocean protection

We integrate peer-reviewed scientific literature and expert working group products to link STAGE and LEVEL with the ecological and social OUTCOMES expected from different types of MPAs. If biodiversity is conserved, an MPA would be considered successful at meeting its primary goal. However, this does not preclude other outcomes from also occurring and producing benefits, including those for human well-being. Once an MPA is implemented with CONDITIONS in place, it can lead to interrelated ecological and social outcomes based on LEVEL of protection.

Ecological outcomes of MPAs

Thousands of MPA studies document the ecological effects of MPAs across almost all ocean regions and seas, demonstrating that MPAs are an effective tool to conserve biodiversity and improve ecosystem functioning (Table 2; expanded references are provided in table S1). Outside their borders, MPAs can also enhance fish stocks through egg and larval export and spillover of juveniles and adults to areas outside the MPA boundaries (*68*). Interconnected networks of MPAs are expected to deliver scaled benefits (*69*). Highly mobile species and those with very large home ranges may receive lower benefit levels from MPA protection than that of more sedentary species, unless MPAs are larger or dynamic with mobile boundaries (*70*, *71*) or they protect critical life stages [such as spawning aggregations, nursery or feeding grounds, or migration bottlenecks (*72*, *73*)]. Long-ranging species require well-designed MPA networks and effective management outside MPAs (*74*).

Research is often biased toward ecological and fisheries responses to MPA protection [for example, (*75*)] because these are related to the biodiversity conservation goals of MPAs (*7*) and the main impact that MPAs abate (fishing). However, other benefits are possible (Table 2 and table S1). Water quality can

Table 2. Ecological OUTCOMES of MPAs as a result of LEVEL of protection. The outcomes discussed here assume that best practices in CONDITIONS have been met and that the system has had time to progress from a degraded state to one with relatively few fluctuations. Not all OUTCOMES can be expected from all MPAs because they vary by habitat type, oceanographic conditions, and previous state of degradation. Levels of confidence are indicated with shaded circles; the darker the circle, the higher the confidence, either high, moderate, or low confidence. Confidence level represents expert judgments based on the quantity and quality of research available. Citations are available in table S1.

	Outcome	Level of Protection: Fully	Highly	Lightly	Minimally
Biodiversity conservation	Abundance	●	●	●	●
	Population age structure	●	●	●	●
	Biomass	●	●	●	●
	Species richness (no. of species)	●	●	●	●
	Reproductive output and replenishment	●	●	●	●
	Connectivity of populations	●	●	●	●
	Rare and endangered species protected	●	●	●	●
	Genetic diversity	●	●	●	●
	Habitats	●	●	●	●
	Ecosystem functioning	●	●	●	●
	Ecosystem resilience (ability to recover after disturbance)	●	●	●	●
Effects on exploited species	Spillover	●	●	●	●
	Larval export	●	●	●	●
	Insurance against management failure or stock collapse	●	●	●	●
	Protection of vulnerable life stages	●	●	●	●
Water quality	Eutrophication .	●	●	●	●
	Pathogens and pollutants	●	●	●	●
	Suspended sediment	●	●	●	●
Climate resilience	Carbon	●	●	●	●
	Acidification	●	●	●	●
	Productivity	●	●	●	●
	Coastal protection	●	●	●	●

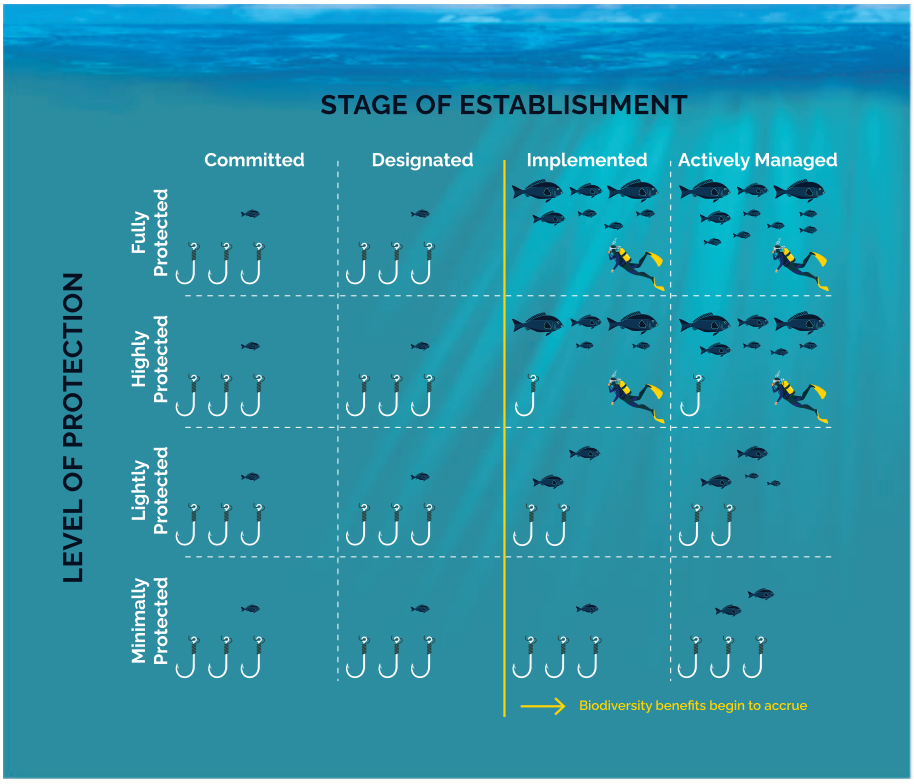


Fig. 2. Matrix based on LEVEL of protection and STAGE of establishment of MPAs. Any MPA or MPA zone sits in one of the 16 cells in this matrix according to its LEVEL and STAGE, and global area of ocean protected in MPAs can also be tallied by each matrix cell. Hooks indicate extractive use; divers indicate recreational, traditional, and cultural use; and fish indicate biodiversity outcomes. As long as CONDITIONS are in place, the OUTCOMES of an MPA will depend primarily on its protection LEVEL and STAGE, as depicted (other factors such as state of ecosystem degradation before establishment of the MPA may also enhance or reduce outcomes). Protection does not begin until an MPA is implemented or actively managed. The most effective biodiversity conservation OUTCOMES from an MPA are likely in the top right quadrant of this matrix, where MPAs are fully or highly protected and implemented or actively managed. In considering the global area protected, a larger percentage in the top right quadrant would indicate more effective protection than that of a larger percentage in the bottom left quadrant.

improve if MPAs restore and recover vegetated habitats and filter-feeding bottom communities (*76*). Evidence is accumulating that MPAs can enhance mitigation, adaptation, and resilience to climate change (*45*, *77*, *78*). Protecting "blue carbon" habitats can preserve their ability to provide carbon sequestration and coastal protection, particularly if supplemented by restoration (*79*).

In well-managed MPAs, ecological benefits relative to surrounding unprotected areas are more prominent where species have previously been depleted, particularly by factors that can be managed or excluded. Substantial prior habitat damage or human impacts outside MPAs can slow recovery (*65*). In more intact areas, protection can guard against future losses (*80*). Threats that cannot be abated by protection may reduce benefits, especially in the short term. However, protection may partially mitigate some of these impacts by protecting functioning ecosystems, boosting resilience, and hastening recovery (*45*, *81*, *82*). Extractive activities displaced by protection may lead to impacts outside MPAs, underscoring the need to integrate MPAs into comprehensive marine spatial planning to ensure that damaging activities are not displaced onto more sensitive habitats or the ranges of more vulnerable species.

Whereas some benefits occur quickly after protection, others can take decades. Species respond to protection at different rates depending on factors such as life history characteristics, behavior, depletion at the time of protection (such as for fished species), and other human impacts (*40*). Early results often include increases in species already common within the MPA, but as time passes, such benefits also include increases in rare and vulnerable species, reestablishment of natural population age structure (especially for long-lived species), and recovery of degraded structural ecosystem elements and habitats (*83*). The Outcomes in Table 2 assume that adequate protection has been in place long enough for effects to develop.

Recovery is more likely, faster, and more complete at the higher LEVELS of protection: Positive ecological outcomes are more substantial and less variable in fully and highly protected areas than in lightly and minimally protected areas (Table 2) [for example, (*18*, *23*, *84*)], with greater potential for ecosystem restoration when areas are fully protected (*85*). In protection levels with more activities occurring, management often addresses competing or conflicting uses of an area and may advantage certain groups of users (such as small-scale or recreational over larger commercial fishers). Decisions about the appropriate protection LEVEL will depend on conservation and management goals, social context, and CONDITIONS, which enable OUTCOMES (Table 1). For example, poorly designed, managed, and resourced MPAs, with low compliance and staff, will deliver fewer benefits (*57*, *86*), and a highly protected area could produce better outcomes than those of a fully protected area if it has stronger enabling conditions.

Social outcomes of MPAs

MPAs can directly and indirectly affect all aspects of human well-being [social, health, culture, economic, and governance (*87*)] for different rights-holders and stakeholders (such as Indigenous peoples, fishers, tourism operators, and coastal residents). When key CONDITIONS are in place, positive benefits of MPAs can be enhanced, and negative impacts can be minimized. A recent comprehensive review found that about half of all documented human well-being outcomes of MPAs were positive and about one-third were negative, with the remaining showing no change or change that was not attributable as positive or negative (*88*). Common positive outcomes were community involvement, increased catch per unit of fishing effort (CPUE), and higher income, whereas negative outcomes commonly manifested through increasing costs of activities (fishing) and conflict (*89*). Both positive and negative impacts can occur at the same time [for example, (*90*, *91*)]. Four MPAs in Indonesia had positive effects on material wealth and scientific environmental knowledge but negative effects on perceived well-being, fish catch, and marine resource control (*92*).

Direct effects of MPAs on human well-being can be immediate owing to changes in access or decision-making (*93*). For example, discussions about whether to have MPAs, where to place them, and what management measures to include can directly affect levels of conflict, perceptions of procedural fairness, access to resources and incomes, and sense of agency in resource management, either negatively or positively (*94*, *95*). Indirect effects also occur

through subsequent management actions and ecosystem changes, including altered catches, CPUE, and income from resource extraction or nonextractive activities (*65*, *96*, *97*). Such effects are the most common positive MPA outcomes for human well-being; the timing of these outcomes varies according to ecological recovery rates. Both direct and indirect effects may shift over time [for example, (*92*)]; negative impacts on the fish catch for certain commercial fishers increased over 4 years in two MPAs in the Gulf of Mexico (*98*).

The effects of MPAs can vary substantially across and within stakeholder groups depending on previous rights, dependence, and uses (*99*, *100*) and across the broader social-ecological context (*100*, *101*). Differential MPA effects have been examined most commonly for fishers, particularly by fishing method (such as commercial or artisanal and use of different gear types) (*88*, *90*). This variability can depend on level of resource dependency [such as dietary dependency or livelihood diversity (*102*, *103*)], ability to adapt to changes [such as fishing areas or jobs (*94*, *103*, *104*)], involvement in MPA establishment processes [for example, (*65*)], and other sociocultural characteristics that structure society [such as age, gender, and ethnicity (*92*)].

The direction and strength of MPA impacts on different societal groups can also change temporally [for example, (*100*)] and can affect power dynamics within coastal communities as some members of the community benefit and others are excluded [for example, (*99*)]. Individuals from marginalized groups with high resource dependency and low adaptive capacity often bear disproportionate costs (*67*, *105*), particularly when excluded from decision-making processes (*104*). Alternatively, if protection strengthens local community property rights and excludes outside users, and/or provides economic benefits (such as from tourism), an MPA may benefit local communities. Achieving more positive outcomes requires attention to the MPA goals and the CONDITIONS during all STAGES to support stakeholders and rights-holders; the contribution of marine ecosystems to their wellbeing, including livelihoods; and long-term MPA functioning (Table 1).

Protection LEVEL influences all indirect social impacts but only some direct impacts. A higher LEVEL of protection can generate greater recovery of socially, culturally, and economically important species or habitats, especially over the longer term (an indirect impact). Such protection could also increase the likelihood of conflict resulting from fishers being displaced but may not change other direct effects (such as empowerment in decision-making). In some cases, lightly or minimally protected areas may meet the needs of the local community, at least in the short term. Overall, when key CONDITIONS are met (such as long-term protection or high levels of compliance) (Table 1), fully and highly protected areas are associated with more positive outcomes (*88*), aligning with the positive outcomes found in ecological studies [for example, (*106*)].

Moving forward with clarity and transparency

Existing international targets highlight the key role of MPAs in conserving biodiversity and supporting a sustainable ocean economy—the blue economy. Achieving these goals has become even more important because of escalating threats to ocean biodiversity and ecosystem functioning (*1*, *5*) and the disproportionately large impacts of the COVID-19 pandemic on near-shore communities (*107*). We argue that three key actions using this framework would strengthen MPA understanding and use at local, national, and global scales.

(i) Incorporate STAGE of establishment and LEVEL of protection into global reporting on progress toward international targets. The WDPA (*20*) reports on progress toward the 10% Aichi Target 11 and will report on the subsequent Post-2020 Target yet to be adopted by the CBD at the time of this writing. The WDPA is mandated to report designated MPAs (combined with implemented and actively managed areas) in their tally of total protected area worldwide and does not track LEVEL of protection. Tracking STAGE and LEVEL—for example, by using them as indicators of progress toward the future CBD Target—provides a full matrix evaluation (Fig. 2) of MPA quality and thereby moves global assessment of protection beyond a single percentage metric. This matrix approach may be similarly useful for OECMs and terrestrial protected areas.

(ii) Use this framework to identify immediate opportunities to strengthen existing or to create new MPAs. An urgent need to recover ocean health and concomitant benefits to people means that high-priority and ample pay-off opportunities exist to create new MPAs and to strengthen the level of protection and compliance for existing MPAs. Doing so was one of the Ocean Panel's five immediate action opportunities for COVID-19 recovery (*107*). The MPA Guide can help local, regional, and national bodies develop, implement, and manage new and existing MPAs.

Recognizing the distinct STAGE of an MPA can help MPA agencies and those working in civil society to progress an MPA to the next step—for example, by thinking through and addressing capacity constraints such as lack of financial, social, and scientific capital. When developing a new MPA, decision-makers and managers can also assess different protection LEVELS and their expected outcomes when deciding which activities to allow. A review of the rules and regulations in existing MPAs and how they map to protection LEVELS can help to determine whether these activities are consistent with desired ecological and social outcomes. Indonesia recently underwent an evaluation of their MPAs by STAGE and LEVEL, highlighting the impressive resource and capacity the country has invested toward active management while also identifying MPAs that may require increased protection to achieve their goals (*108*). At a regional level, we can track how much of an ecosystem or habitat type is in each LEVEL of protection and identify sites in need of increased protection for biodiversity (*19*). Identifying how much ocean is still in proposed/committed or designated MPAs shows what has been promised for protection but is still in need of further action to implement [for example, (*109*)].

(iii) Develop research agendas to link MPA protection LEVEL, CONDITIONS, and OUTCOMES. Although some types of MPAs have been studied for decades, two paths forward are required in a new era of MPA research. First, datasets should be organized around the protection provided by different MPAs in different LEVELS. Most existing ecological research lumps MPAs into fully protected areas and "partially protected" areas, the latter of which combines highly, lightly, and minimally protected [for example, (*42*, *44*, *84*)]. Combining these levels limits our ability to understand and predict OUTCOMES (Table 2) and to assess trade-offs to biodiversity conservation and trade-offs among different stakeholder groups. Explicit research across these three levels of protection is now possible by using this framework.

More research is also needed to better understand MPA effects on specific social outcomes, across different societal groups (such as gender, age, and ethnic groups), and over time (*88*, *90*). Research should expand geographically to assess how MPAs affect the multiple dimensions of human well-being in diverse contexts (*36*, *88*, *90*) and should use an impact-evaluation lens, including rigorous counterfactual study designs (*37*, *92*) in qualitative as well as quantitative studies (*88*). Further research is also needed to better understand the CONDITIONS as they relate to an MPA's STAGE and LEVEL of protection, and the specific aspects of MPA planning, governance, and management that produce positive or negative outcomes for equity (*110*) and other dimensions of human well-being.

Conclusion

The stakes have never been higher for connecting MPA science to policy and action. Development of the new CBD and other MPA goals and targets requires improved clarity and harmonization to be effective from local to global scales. Use of The MPA Guide would shift the conversation from arguments about what MPAs can deliver to answering questions

such as, "What level of protection is needed for an MPA to produce the desired outcomes for biodiversity and human well-being?" and "What is the global tally of MPAs by stage of establishment and level of protection, and what does this tell us about progress toward ocean conservation goals?" This scientific synthesis and guide offers a framework, language, and detailed guidance toward doing so.

REFERENCES AND NOTES

1. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services," E. S. Brondizio, J. Settele, S. Díaz, H. T. Ngo, Eds. (IPBES secretariat, 2019).
2. United Nations, Ed., *The First Global Integrated Marine Assessment: World Ocean Assessment I* (Cambridge Univ. Press, 2017).
3. J. Lubchenco, S. D. Gaines, A new narrative for the ocean. *Science* **364**, 911–911 (2019). doi: 10.1126/science.aay2241; pmid: 31171668
4. A. M. Friedlander, Marine conservation in Oceania: Past, present, and future. *Mar. Pollut. Bull.* **135**, 139–149 (2018). doi: 10.1016/j.marpolbul.2018.05.064; pmid: 30301023
5. Intergovernmental Panel on Climate Change (IPCC), "IPCC Special Report on the Ocean and Cryosphere in a Changing Climate," H.-O. Pörtner et al., Eds. (IPCC, 2019).
6. E. Sala *et al*., Protecting the global ocean for biodiversity, food and climate. *Nature* **592**, 397–402 (2021). doi: 10.1038/s41586-021-03371-z; pmid: 33731930
7. IUCN, WCPA, "Applying IUCN's Global Conservation Standards to Marine Protected Areas (MPAs). Delivering effective conservation action through MPAs, to secure ocean health & sustainable development. Version 1.0." (IUCN, 2018).
8. J. Lubchenco, K. Grorud-Colvert, OCEAN. Making waves: The science and politics of ocean protection. *Science* **350**, 382–383 (2015). doi: 10.1126/science.aad5443; pmid: 26472764
9. E. Sala *et al*., Assessing real progress towards effective ocean protection. *Mar. Policy* **91**, 11–13 (2018). doi: 10.1016/j.marpol.2018.02.004
10. J. A. Kawaka *et al*., Developing locally managed marine areas: Lessons learnt from Kenya. *Ocean Coast. Manage.* **135**, 1–10 (2017). doi: 10.1016/j.ocecoaman.2016.10.013
11. D. Vilas *et al*., Current and potential contributions of the Gulf of Lion Fisheries Restricted Area to fisheries sustainability in the NW Mediterranean Sea. *Mar. Policy* **123**, 104296 (2020). doi: 10.1016/j.marpol.2020.104296
12. D. Laffoley *et al*., An introduction to 'other effective area-based conservation measures' under Aichi Target 11 of the Convention on Biological Diversity: Origin, interpretation and emerging ocean issues. *Aquat. Conserv.* **27**, 130–137 (2017). doi: 10.1002/aqc.2783
13. J. M. Reimer, R. Devillers, J. Claudet, Benefits and gaps in area-based management tools for the ocean Sustainable Development Goal. *Nat. Sustain.* **4**, 349–357 (2020). doi: 10.1038/s41893-020-00659-2
14. S. Woodley *et al*., Meeting Aichi Target 11: What does success look like for Protected Area systems? *Parks* **18**, 1 (2012).
15. E. S. Nocito, C. M. Brooks, A. L. Strong, Gazing at the crystal ball: Predicting the future of marine protected areas through voluntary commitments. *Front. Mar. Sci.* **6**, 835 (2020). doi: 10.3389/fmars.2019.00835
16. B. C. O'Leary *et al*., Effective coverage targets for ocean protection. *Conserv. Lett.* **9**, 398–404 (2016). doi: 10.1111/conl.12247
17. K. R. Jones *et al*., Area requirements to safeguard Earth's marine species. *One Earth* **2**, 188–196 (2020). doi: 10.1016/j.oneear.2020.01.010
18. M. Zupan *et al*., Marine partially protected areas: Drivers of ecological effectiveness. *Front. Ecol. Environ.* **16**, 381–387 (2018). doi: 10.1002/fee.1934
19. J. Claudet, C. Loiseau, M. Sostres, M. Zupan, Underprotected marine protected areas in a global biodiversity hotspot. *One Earth* **2**, 380–384 (2020). doi: 10.1016/j.oneear.2020.03.008
20. UNEP-WCMC, World Database on Protected Areas. *Protected Planet* (2020); www.protectedplanet.net/marine.
21. Marine Conservation Institute, The Atlas of Marine Protection, (2020); http://mpatlas.org.
22. M. D. Barnes, L. Glew, C. Wyborn, I. D. Craigie, Prevent perverse outcomes from global protected area policy. *Nat. Ecol. Evol.* **2**, 759–762 (2018). doi: 10.1038/s41559-018-0501-y; pmid: 29556080
23. B. Horta e Costa *et al*., A regulation-based classification system for Marine Protected Areas (MPAs). *Mar. Policy* **72**, 192–198 (2016). doi: 10.1016/j.marpol.2016.06.021
24. M. Pieraccini, S. Coppa, G. A. De Lucia, Beyond marine paper parks? Regulation theory to assess and address environmental non-compliance. *Aquat. Conserv.* **27**, 177–196 (2017). doi: 10.1002/aqc.2632
25. B. Horta e Costa *et al*., Categorizing ocean conservation targets to avoid a potential false sense of protection to society: Portugal as a case-study. *Mar. Policy* **108**, 103553 (2019). doi: 10.1016/j.marpol.2019.103553
26. J. Claudet, C. Loiseau, A. Pebayle, Critical gaps in the protection of the second largest exclusive economic zone in the world. *Mar. Policy* **124**, 104379 (2021). doi: 10.1016/j.marpol.2020.104379
27. CBD, "Decisions Adopted by the Conference of the Parties to the Convention on Biological Diversity at its Eighth Meeting (Decision VIII/15, Annex IV)," Convention on Biological Diversity, Curitiba, Brazil, 2006.
28. IUCN, "Guidelines for applying the IUCN protected area management categories to marine protected areas" (IUCN, ed. 2, 2019); www.iucn.org/content/guidelines-applying-iucn-protected-area-management-categories-marine-protected-areas-0.
29. CCAMLR, "Proposal to establish an East Antarctic Marine Protected Area," CCAMLR-38/21 (2019); www.ccamlr.org/en/ccamlr-38/21.
30. CCAMLR, "Proposal to establish a Marine Protected Area across the Weddell Sea region (Phase 1)," CCAMLR-38/23 (2019); www.ccamlr.org/en/ccamlr-38/23.
31. D. Faure, Speech on the occasion of 30% of Seychelles' EEZ Designated as Marine Protected Area (2020); www.statehouse.gov.sc/speeches/4786/speech-by-president-danny-faure-on-the-occasion-of-30-of-seychelles-eez-designated-as-marine-protected-area.
32. J. C. Day, R. A. Kenchington, J. M. Tanzer, D. S. Cameron, Marine zoning revisited: How decades of zoning the Great Barrier Reef has evolved as an effective spatial planning approach for marine ecosystem-based management. *Aquat. Conserv.* **29** (S2), 9–32 (2019). doi: 10.1002/aqc.3115
33. Department of Agriculture, Forestry, and Fisheries, "Niue Moana Mahu Marine Protected Area Regulations 2020," no. 2020/04 (2020); https://old.mpatlas.org/media/filer_public/bc/95/bc959065-13b7-42d7-97dd-507503fc4b01/reg_2020-04_niue_moana_mahu_marine_protected_area_regulations_1.pdf.
34. S. Wells *et al*., Using the IUCN Green List of Protected and Conserved Areas to promote conservation impact through marine protected areas. *Aquat. Conserv.* **26**, 24–44 (2016). doi: 10.1002/aqc.2679
35. Marine Conservation Institute, BlueParks (2020); https://blueparks.org.
36. G. G. Gurney *et al*., Implementing a social-ecological systems framework for conservation monitoring: Lessons from a multi-country coral reef program. *Biol. Conserv.* **240**, 108298 (2019). doi: 10.1016/j.biocon.2019.108298
37. G. N. Ahmadia *et al*., Integrating impact evaluation in the design and implementation of monitoring marine protected areas. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140275 (2015). doi: 10.1098/rstb.2014.0275; pmid: 26460128
38. S. Murray, T. T. Hee, A rising tide: California's ongoing commitment to monitoring, managing and enforcing its marine protected areas. *Ocean Coast. Manage.* **182**, 104920 (2019). doi: 10.1016/j.ocecoaman.2019.104920
39. S. E. Lester *et al*., Biological effects within no-take marine reserves: A global synthesis. *Mar. Ecol. Prog. Ser.* **384**, 33–46 (2009). doi: 10.3354/meps08029
40. J. Claudet *et al*., Marine reserves: Size and age do matter. *Ecol. Lett.* **11**, 481–489 (2008). doi: 10.1111/j.1461-0248.2008.01166.x; pmid: 18294212
41. P. P. Molloy, I. B. McLean, I. M. Côté, Effects of marine reserve age on fish populations: A global meta-analysis. *J. Appl. Ecol.* **46**, 743–751 (2009). doi: 10.1111/j.1365-2664.2009.01662.x
42. S. Giakoumi *et al*., Ecological effects of full and partial protection in the crowded Mediterranean Sea: A regional meta-analysis. *Sci. Rep.* **7**, 8940 (2017). doi: 10.1038/s41598-017-08850-w; pmid: 28827603
43. M. Zupan *et al*., How good is your marine protected area at curbing threats? *Biol. Conserv.* **221**, 237–245 (2018). doi: 10.1016/j.biocon.2018.03.013
44. M. Sciberras *et al*., Evaluating the relative conservation value of fully and partially protected marine areas. *Fish Fish.* **16**, 58–77 (2015). doi: 10.1111/faf.12044
45. C. M. Roberts *et al*., Marine reserves can mitigate and promote adaptation to climate change. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6167–6175 (2017). doi: 10.1073/pnas.1701262114; pmid: 28584096
46. J. Claudet, France must impose strict levels of marine protection. *Nature* **570**, 36–36 (2019). doi: 10.1038/d41586-019-01750-1; pmid: 31164767
47. A. C. Alcala, G. R. Russ, No-take marine reserves and reef fisheries management in the Philippines: A new people power revolution. *Ambio* **35**, 245–254 (2006). doi: 10.1579/05-A-054R1.1; pmid: 16989509
48. Palau International Coral Reef Center-PICRC, "Palau National Marine Sanctuary Act Overview"; https://picrc.org/picrcpage/wp-content/uploads/2019/12/PNMS-Act-Overview-2019AUG20.pdf.
49. K. Kikiloi, A. M. Friedlander, 'Aulani Wilhelm, N. Lewis, K. Quiocho, W. 'Āila, S. Kaho'ohalahala, Papahānaumokuākea: Integrating culture in the design and management of one of the world's largest marine protected areas. *Coast. Manage.* **45**, 436–451 (2017).
50. M. B. Mascia, C. A. Claus, A property rights approach to understanding human displacement from protected areas: The case of marine protected areas. *Conserv. Biol.* **23**, 16–23 (2009). doi: 10.1111/j.1523-1739.2008.01050.x; pmid: 18847441
51. National Oceanic and Atmospheric Administration, United States Fish and Wildlife Service, Hawai'i Department of Land and Natural Resources, "Papahānaumokuākea Marine National Monument Management Plan" (2008); https://nmspapahanaumokuakea.blob.core.windows.net/papahanaumokuakea-prod/media/archive/management/mp/vol1_mmp08.pdf.
52. Great Barrier Reef Marine Park Authority, Interpreting zones; www.gbrmpa.gov.au/access-and-use/zoning/interpreting-zones.
53. International Union for the Conservation of Nature, "Motion 066–Guidance to identify industrial fishing incompatible with protected areas" (2020); www.iucncongress2020.org/motion/066.
54. Office of National Marine Sanctuaries, National Oceanic and Atmospheric Administration, United States Department of Commerce, State of Hawai'i, Hawai'i Department of Land and Natural Resources, "Hawai'ian Islands Humpback Whale National Marine Sanctuary Management Plan" (2020); https://nmshawaiihumpbackwhale.blob.core.windows.net/hawaiihumpbackwhale-prod/media/docs/2020-hihwnms-management-plan.pdf.
55. L. M. Petruny, A. J. Wright, C. E. Smith, Getting it right for the North Atlantic right whale (Eubalaenaglacialis): A last opportunity for effective marine spatial planning? *Mar. Pollut. Bull.* **85**, 24–32 (2014). doi: 10.1016/j.marpolbul.2014.06.004; pmid: 24998798
56. C. Bracciali, D. Campobello, C. Giacoma, G. Sarà, Effects of nautical traffic and noise on foraging patterns of Mediterranean damselfish (*Chromis chromis*). *PLOS ONE* **7**, e40582 (2012). doi: 10.1371/journal.pone.0040582; pmid: 22792375
57. D. A. Gill *et al*., Capacity shortfalls hinder the performance of marine protected areas globally. *Nature* **543**, 665–669 (2017). doi: 10.1038/nature21708; pmid: 28329771
58. M. Lockwood, Good governance for terrestrial protected areas: A framework, principles and performance outcomes. *J. Environ. Manage.* **91**, 754–766 (2010). doi: 10.1016/j.jenvman.2009.10.005; pmid: 19896262
59. N. J. Bennett, T. Satterfield, Environmental governance: A practical framework to guide design, evaluation, and analysis. *Conserv. Lett.* **11**, e12600 (2018). doi: 10.1111/conl.12600
60. IUCN-WCPA, "Establishing marine protected area networks–Making it happen" (IUCN World Commission on Protected Areas, National Oceanic and Atmospheric Administration, and The Nature Conservancy, 2008), p. 118.
61. R. L. Pressey, M. C. Bottrill, Approaches to landscape- and seascape-scale conservation planning: Convergence, contrasts and challenges. *Oryx* **43**, 464–475 (2009). doi: 10.1017/S0030605309990500
62. H. E. Fox *et al*., Reexamining the science of marine protected areas: Linking knowledge to action. *Conserv. Lett.* **5**, 1–10 (2012). doi: 10.1111/j.1755-263X.2011.00207.x

63. M. Hockings, S. Stolton, F. Leverington, N. Dudley, J. Courrau, *Evaluating Effectiveness: A Framework for Assessing Management Effectiveness of Protected Areas*, Best Practice Protected Area Guidelines Series (IUCN, 2006).
64. B. Spergel, M. Moye, *Financing Marine Conservation a Menu of Options* (World Wildlife Fund, Center for Conservation Finance, 2004).
65. P. Guidetti, J. Claudet, Comanagement practices enhance fisheries in marine protected areas. *Conserv. Biol.* **24**, 312–318 (2010). doi: 10.1111/j.1523-1739.2009.01358.x; pmid: 19906064
66. N. J. Bennett *et al*., Local support for conservation is associated with perceptions of good governance, social impacts, and ecological effectiveness. *Conserv. Lett.* **12**, e12640 (2019). doi: 10.1111/conl.12640
67. V. R. Kamat, "The ocean is our farm": Marine conservation, food insecurity, and social suffering in southeastern Tanzania. *Hum. Organ.* **73**, 289–298 (2014). doi: 10.17730/humo.73.3.f43k115544761g0v
68. H. B. Harrison *et al*., Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Curr. Biol.* **22**, 1023–1028 (2012). doi: 10.1016/j.cub.2012.04.008; pmid: 22633811
69. K. Grorud-Colvert *et al*., Marine protected area networks: Assessing whether the whole is greater than the sum of its parts. *PLOS ONE* **9**, e102298 (2014). doi: 10.1371/journal.pone.0102298; pmid: 25084458
70. A. Di Franco *et al*., Linking home ranges to protected area size: The case study of the Mediterranean Sea. *Conserv. Biol.* **221**, 175–181 (2018). doi: 10.1016/j.biocon.2018.03.012
71. S. M. Maxwell, K. M. Gjerde, M. G. Conners, L. B. Crowder, Mobile protected areas for biodiversity on the high seas. *Science* **367**, 252–254 (2020). doi: 10.1126/science.aaz9327; pmid: 31949070
72. B. Erisman *et al*., Fish spawning aggregations: Where well-placed management actions can yield big benefits for fisheries and conservation. *Fish Fish.* **18**, 128–144 (2017). doi: 10.1111/faf.12132
73. C. M. Hernández *et al*., Evidence and patterns of tuna spawning inside a large no-take Marine Protected Area. *Sci. Rep.* **9**, 10772 (2019). doi: 10.1038/s41598-019-47161-0; pmid: 31341251
74. M. H. Carr *et al*., The central importance of ecological spatial connectivity to effective coastal marine protected areas and to meeting the challenges of climate change in the marine environment. *Aquat. Conserv.* **27**, 6–29 (2017). doi: 10.1002/aqc.2800
75. M. Di Lorenzo, J. Claudet, P. Guidetti, Spillover from marine protected areas to adjacent fisheries has an ecological and a fishery component. *J. Nat. Conserv.* **32**, 62–66 (2016). doi: 10.1016/j.jnc.2016.04.004
76. "Reef 2050 Water Quality Improvement Plan 2017-2020" (State of Queensland, 2018), p. 63.
77. J. E. Duffy, J. S. Lefcheck, R. D. Stuart-Smith, S. A. Navarrete, G. J. Edgar, Biodiversity enhances reef fish biomass and resistance to climate change. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6230–6235 (2016). doi: 10.1073/pnas.1524465113; pmid: 27185921
78. S. D. Gaines *et al*., Improved fisheries management could offset many negative effects of climate change. *Sci. Adv.* **4**, eaao1378 (2018). doi: 10.1126/sciadv.aao1378; pmid: 30167455
79. L. Sheehan, E. T. Sherwood, R. P. Moyer, K. R. Radabaugh, S. Simpson, Blue carbon: An additional driver for restoring and preserving ecological services of coastal wetlands in Tampa Bay (Florida, USA). *Wetlands* **39**, 1317–1328 (2019). doi: 10.1007/s13157-019-01137-y
80. G. W. Allison, S. D. Gaines, J. Lubchenco, H. P. Possingham, Ensuring persistence of marine reserves: Catastrophes require adopting an insurance factor. *Ecol. Appl.* **13** (sp1), 8–24 (2003). doi: 10.1890/1051-0761(2003)013[0008:EPOMRC]2.0.CO;2
81. C. Mellin, M. Aaron MacNeil, A. J. Cheal, M. J. Emslie, M. Julian Caley, Marine protected areas increase resilience among coral reef communities. *Ecol. Lett.* **19**, 629–637 (2016). doi: 10.1111/ele.12598; pmid: 27038889
82. M. J. Emslie *et al*., Expectations and outcomes of reserve network performance following re-zoning of the Great Barrier Reef Marine Park. *Curr. Biol.* **25**, 983–992 (2015). doi: 10.1016/j.cub.2015.01.073; pmid: 25819564
83. R. C. Babcock *et al*., Decadal trends in marine reserves reveal differential rates of change in direct and indirect effects. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18256–18261 (2010). doi: 10.1073/pnas.0908012107; pmid: 20176941
84. A. M. Friedlander, M. K. Donovan, H. Koike, P. Murakawa, W. Goodell, Characteristics of effective marine protected areas in Hawai'i. *Aquat. Conserv.* **29** (S2), 103–117 (2019). doi: 10.1002/aqc.3043
85. O. Aburto-Oropeza *et al*., Large recovery of fish biomass in a no-take marine reserve. *PLOS ONE* **6**, e23601 (2011). doi: 10.1371/journal.pone.0023601; pmid: 21858183
86. P. Guidetti *et al*., Italian marine reserve effectiveness: Does enforcement matter? *Conserv. Biol.* **141**, 699–709 (2008). doi: 10.1016/j.biocon.2007.12.013
87. M. Kaplan-Hallam, N. J. Bennett, Adaptive social impact management for conservation and environmental management. *Conserv. Biol.* **32**, 304–314 (2018). doi: 10.1111/cobi.12985; pmid: 29063710
88. N. C. Ban *et al*., Well-being outcomes of marine protected areas. *Nat. Sustain.* **2**, 524–532 (2019). doi: 10.1038/s41893-019-0306-2
89. D. A. Gill *et al*., Social synergies, tradeoffs, and equity in marine conservation impacts. *Annu. Rev. Environ. Resour.* **44**, 347–372 (2019). doi: 10.1146/annurev-environ-110718-032344
90. D. A. Gill, H. A. Oxenford, P. W. Schuhmann, in *Viability and Sustainability of Small-Scale Fisheries in Latin America and The Caribbean*, S. Salas, M. J. Barragán-Paladines, R. Chuenpagdee, Eds. (Springer, 2019), pp. 295–328. doi: 10.1007/978-3-319-76078-0_13
91. M. Schratzberger, S. Neville, S. Painting, K. Weston, L. Paltriguera, Ecological and socio-economic effects of Highly Protected Marine Areas (HPMAs) in temperate waters. *Front. Mar. Sci.* **6**, 749 (2019). doi: 10.3389/fmars.2019.00749
92. G. G. Gurney *et al*., Poverty and protected areas: An evaluation of a marine integrated conservation and development project in Indonesia. *Glob. Environ. Change-Hum. Pol. Dimens.* **26**, 98–107 (2014). doi: 10.1016/j.gloenvcha.2014.04.003
93. C. E. Hattam, S. C. Mangi, S. C. Gall, L. D. Rodwell, Social impacts of a temperate fisheries closure: Understanding stakeholders' views. *Mar. Policy* **45**, 269–278 (2014). doi: 10.1016/j.marpol.2013.09.005
94. R. Chuenpagdee *et al*., Marine protected areas: Re-thinking their inception. *Mar. Policy* **39**, 234–240 (2013). doi: 10.1016/j.marpol.2012.10.016
95. H. S. Macedo, R. P. Medeiros, P. McConney, Are multiple-use marine protected areas meeting fishers' proposals? Strengths and constraints in fisheries' management in Brazil. *Mar. Policy* **99**, 351–358 (2019). doi: 10.1016/j.marpol.2018.11.007
96. C. Leisher, P. van Beukering, L. Scherl, "Nature's investment bank: How marine protected areas contribute to poverty reduction" (The Nature Conservancy, 2007).
97. E. Sala *et al*., A general business model for marine reserves. *PLOS ONE* **8**, e58799 (2013). doi: 10.1371/journal.pone.0058799; pmid: 23573192
98. M. D. Smith, J. Zhang, F. Coleman, Effectiveness of marine reserves for large-scale fisheries management. *Can. J. Fish. Aquat. Sci.* **63**, 153–164 (2006). doi: 10.1139/f05-205
99. K. Hogg, T. Gray, P. Noguera-Méndez, M. Semitiel-García, S. Young, Interpretations of MPA winners and losers: A case study of the Cabo De Palos- Islas Hormigas Fisheries Reserve. *Marit. Stud.* **18**, 159–171 (2019). doi: 10.1007/s40152-019-00134-5
100. G. G. Gurney, R. L. Pressey, J. E. Cinner, R. Pollnac, S. J. Campbell, Integrated conservation and development: Evaluating a community-based marine protected area project for equality of socioeconomic impacts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140277 (2015). doi: 10.1098/rstb.2014.0277; pmid: 26460130

ACKNOWLEDGMENTS

The authors thank the many ocean experts from more than 40 countries who generously shared their knowledge about MPAs during this process. The MPA Guide is facilitated by the founding partners: The MPA Project at Oregon State University, IUCN, UNEP-WCMC, the World Database on Protected Areas and Protected Planet, National Geographic Society Pristine Seas, and Marine Conservation Institute's Marine Protection Atlas. We thank the following partner organizations for in-kind support: Pew Bertarelli Ocean Legacy, Centro de Ciências do Mar (CCMAR), Fundação para a Ciência e Tecnologia (FCT), Centre National de la Recherche Scientifique (CNRS), Prince Albert I of Monaco Foundation's Institut Océanographique de Paris, the Prince Albert II of Monaco Foundation, the Ocean Sanctuary Alliance, and the Oceano Azul Foundation. Graphics were developed in collaboration with Communications, Inc. **Funding:** This project was funded by Oceans 5, the Kingfisher Foundation, the David and Lucile Packard Foundation, and JL's Ocean Science Innovation Fund at Oregon State University. J.C. was supported by BiodivERsA. E.J.G. was supported by FCT–Foundation for Science and Technology through UID/MAR/04292/2018/2019, MARE/UIDB/MAR/04292/2020, and MARE/UIDP/MAR/04292/2020, and funds from the Oceano Azul Foundation. **Author contributions:** K.G-C., J.L., and J.S-S. conceived the research; all authors contributed to the analysis of the literature and the concepts and writing of the manuscript. All authors approved the submission. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abf0861
Figs. S1 and S2
Table S1
References (*101*–*213*)

10.1126/science.abf0861

RESEARCH ARTICLE SUMMARY

RIBOSOME

Nucleolar maturation of the human small subunit processome

Sameer Singh†, Arnaud Vanden Broeck†*, Linamarie Miller, Malik Chaker-Margot, Sebastian Klinge*

INTRODUCTION: The ribosome catalyzes protein synthesis in all cells by coupling the decoding of messenger RNA by the small ribosomal subunit with peptide bond formation by the large ribosomal subunit. As both subunits are composed of ribosomal RNA and ribosomal proteins, the assembly of these molecular machines is tightly controlled. In human cells, more than 200 ribosome assembly factors catalyze the maturation of both ribosomal subunits. Ribosome assembly is initiated in the nucleolus, a biomolecular condensate formed through multivalent protein–nucleic acid interactions. Inside this biomolecular condensate, the first stable eukaryotic ribosome assembly intermediate of the small subunit, the small subunit (SSU) processome, is formed. Within the SSU processome, more than 70 proteins and an RNA chaperone, the small nucleolar RNA (snoRNA) U3, work in concert to bring about RNA folding, modifications, rearrangements, and cleavage as well as targeted degradation of preribosomal RNA by the RNA exosome. The essential nature of this process is highlighted by mutations in ribosomal proteins and ribosome assembly factors that have been associated with human diseases.

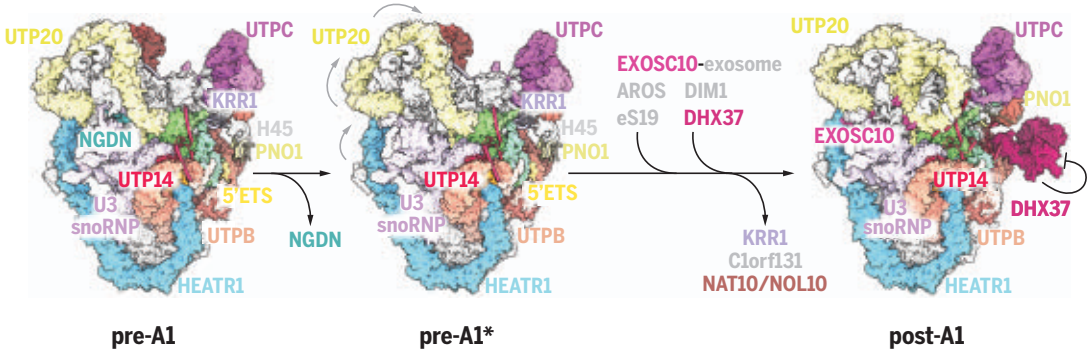
RATIONALE: Little is known about the structure, function, and maturation of the human SSU processome in the nucleolus as it transitions through different states to assemble the small ribosomal subunit. The retention of preribosomal particles within the biomolecular condensate of the nucleolus has thus far prevented their biochemical and structural characterization. To elucidate the nucleolar maturation of the human SSU processome, the isolation of these preribosomal particles from the nucleolus was therefore a key hurdle. To this end, we have developed a human genome editing platform to biallelically affinity tag the endogenous SSU processome and have established biochemical procedures for its isolation from the nucleolus for structural studies.

RESULTS: The cryo–electron microscopy (cryo-EM) structures of three distinct states of the human SSU processome (denoted pre-A1, pre-A1*, and post-A1) were obtained at resolutions of 3.6, 3.9, and 2.7 Å, respectively. Together these structures reveal the molecular mechanisms responsible for the progression during nucleolar SSU processome maturation. Intricate layers of regulation are exerted through the concerted interplay of ribosome assembly factors that control irreversible steps of ribosome assembly, such as RNA processing by the RNA exosome, RNA unwinding by the DEAH-box helicase DHX37 (Dhr1 in yeast), and RNA cleavage by the endonuclease UTP24. The earliest state, pre-A1, highlights how only a small portion of a large preribosomal RNA segment, the 5′ external transcribed spacer (5′ ETS) is recognized by the SSU processome. We have generated a synthetic human ribosomal DNA (rDNA) locus that enables the production of recombinant ribosomes in human cells. By derivatizing this rDNA locus, we have identified a minimal 5′ ETS that can function to generate mature small ribosomal subunits in human cells. As the SSU processome transitions from states pre-A1 and pre-A1* to post-A1, a specific binding site for the RNA exosome-associated exoribonuclease EXOSC10 (Rrp6 in yeast) is generated. In the context of the post-A1 state, the DEAH-box helicase DHX37 is recruited, and its helicase activity must be tightly regulated to prevent the premature unwinding of preribosomal RNA from the RNA chaperone U3. By combining x-ray crystallography with cryo-EM, we show that DHX37 is initially autoinhibited and tightly controlled on the SSU processome through a precisely choreographed interplay that dictates the availability of binding sites and the accessibility of an activating peptide of the ribosome assembly factor UTP14. Within the core of the SSU processome, we have identified a network of proteins that guide cleavage at site A1, thereby generating the mature end of the small ribosomal subunit RNA. Lastly, the structure of the human SSU processome in state post-A1 provides further insight into human diseases, showing how the mammalian-specific ribosome assembly factor AROS stabilizes ribosomal protein eS19, which is most abundantly mutated in Diamond-Blackfan anemia.

CONCLUSION: The cryo-EM structures of three human SSU processome complexes presented here reveal the structural plasticity of the human SSU processome, which endows this particle with the ability to coordinate conformational changes with distinct stages of ribosome assembly. The combined cryo-EM, x-ray crystallography, and functional data serve as a framework for the mechanistic understanding of the human SSU processome. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: klinge@rockefeller.edu (S.K.); avanden@rockefeller.edu (A.V.B.)
†These authors contributed equally to this work.
Cite this article as S. Singh *et al.*, *Science* **373**, eabj5338 (2021). DOI: 10.1126/science.abj5338

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj5338



Schematic model of nucleolar maturation of the human SSU processome. The dissociation of neuroguidin (NGDN) signifies the transition from state pre-A1 to state pre-A1*, which is followed by large-scale conformational changes within the SSU processome (gray arrows) as well as the dissociation of pre-A1–specific factors and the association of the exosome and post-A1–specific factors. The RNA helicase DHX37 is autoinhibited in state post-A1. Protein complexes are color coded, and gray labels indicate proteins that are not shown.

RESEARCH ARTICLE

RIBOSOME

Nucleolar maturation of the human small subunit processome

Sameer Singh[1]†, Arnaud Vanden Broeck[1]†*, Linamarie Miller[1,2], Malik Chaker-Margot[1,2]‡, Sebastian Klinge[1]*

The human small subunit processome mediates early maturation of the small ribosomal subunit by coupling RNA folding to subsequent RNA cleavage and processing steps. We report the high-resolution cryo–electron microscopy structures of maturing human small subunit (SSU) processomes at resolutions of 2.7 to 3.9 angstroms. These structures reveal the molecular mechanisms that enable crucial progressions during SSU processome maturation. RNA folding states within these particles are communicated to and coordinated with key enzymes that drive irreversible steps such as targeted exosome-mediated RNA degradation, protein-guided site-specific endonucleolytic RNA cleavage, and tightly controlled RNA unwinding. These conserved mechanisms highlight the SSU processome's impressive structural plasticity, which endows this 4.5-megadalton nucleolar assembly with the distinctive ability to mature the small ribosomal subunit from within.

The biogenesis of human ribosomes requires more than 200 ribosome assembly factors, which facilitate transcription, RNA folding, and processing and modification of preribosomal RNA as well as the incorporation of 80 ribosomal proteins to form the small (40*S*) and large (60*S*) ribosomal subunits (*1*). This process starts in the human nucleolus, a tripartite biomolecular condensate formed through multivalent protein–nucleic acid interactions (*2*). Here, a 47*S* precursor transcript is formed, which contains coding segments for the small ribosomal subunit (18*S*) and large ribosomal subunit (28*S* and 5.8*S*) as well as key regulatory regions including the 5′ external transcribed spacer (5′ ETS) (*3*). The 5′ ETS plays a pivotal role during the initial stages of ribosome assembly, where it coordinates the recruitment of assembly factors and provides a structural blueprint for early assembly intermediates of the small subunit (SSU) (*4*). In contrast to lower eukaryotes such as *Saccharomyces cerevisiae*, human preribosomal RNA contains drastically larger spacer regions, so that the 5′ ETS is almost twice the size of the entire 18*S* rRNA. The trend toward larger 5′ ETS segments in mammals suggests that in addition to an immediate function in ribosome assembly, these 5′ ETS segments may have acquired additional macroscale functions, such as participation in biomolecular condensate formation in mammalian nucleoli (*5*).

After the transcription of preribosomal RNA, the first stable nucleolar ribosome assembly intermediate is the SSU processome, a giant precursor of the small ribosomal subunit (*6*). In yeast, the temporal order in which more than 50 ribosome assembly factors form the SSU processome has been elucidated (*7*, *8*). Cryo–electron microscopy (cryo-EM) structures of yeast SSU processomes have been determined before cleavage at site A1, which separates the 5′ ETS and 18*S* (*9*–*12*), and during transition states toward a pre-40*S* particle that involve the RNA degradation machinery, the RNA exosome (*13*–*15*). However, despite these advances, limited resolution near 4 Å has so far only allowed a partial interpretation of these molecular snapshots. As a result, the molecular logic that drives SSU processome maturation at a mechanistic level has remained obscure, and the mechanisms and control of key enzymes, including the yeast RNA helicase Dhr1 (DHX37 in humans) within the SSU processome, or substrate recognition by the RNA exosome are still poorly understood.

Insights into human small ribosomal subunit assembly are even more modest and currently limited to late nuclear and cytoplasmic stages (*16*, *17*), as the human nucleolus exhibits distinctive biochemical properties in the form of a biomolecular condensate whose contents have thus far proved impenetrable for biochemical and structural studies. An inability to purify intact human nucleolar ribosome assembly intermediates such as the human SSU processome has therefore limited our knowledge of their composition to known homologs of yeast ribosome assembly factors and large-scale genetic screens (*18*–*20*). The lack of structural information has further hindered our understanding of the roles and functions of about one-quarter of human ribosome assembly factors, precluding a mechanistic understanding of severe human diseases in which mutations of ribosomal proteins result in dysfunctional ribosome assembly at the level of the SSU processome (*21*).

In this study, we developed a human genome editing platform and biochemical methodologies to permeabilize the biomolecular condensate of the human nucleolus and determined the high-resolution cryo-EM structures of maturing human SSU processomes at resolutions of 2.7 to 3.9 Å. These structures reveal the exquisite levels of control that the SSU processome provides by enforcing a strict chronology and molecular choreography for irreversible steps of ribosome assembly, such as RNA degradation, site-specific RNA cleavage at site A1, and RNA unwinding.

Cryo-EM structures of maturing human SSU processomes

A critical hurdle for the study of endogenous human ribosome assembly intermediates is the ability to access all endogenous components. To overcome this challenge, we developed a human genome editing platform that couples the genome editing activity of the CRISPR RNA-guided endonuclease Cas9 with cell surface display to rapidly isolate biallelically edited cells by flow cytometry. In this platform, two DNA repair templates are used, one for each allele of the same gene. While both DNA repair templates code for an identical C-terminal tag of the gene of interest, they contain two different cell surface epitopes, which are produced as separate polypeptides. Thus, cells in which two alleles of a given gene have been repaired with two different DNA repair templates can be selected using flow cytometry (fig. S1, A to E). Using this editing tool, the endogenous human ribosome assembly factor NOC4L, which is present in human SSU processomes, was biallelically tagged and used as bait for subsequent purifications (fig. S1, F to H). To permeabilize the human nucleolus, we developed a sequential extraction procedure in which human nucleolar assembly intermediates containing tagged NOC4L are released in the last step of the extractions, thereby providing the starting material for subsequent nanobody-based purification and structural analysis by cryo-EM. The purified particles include pre-rRNA species that are uncut (30*S* and 26*S*) or cut (21*S*) at site A1 (fig. S2, B and C). A total of 84,904 cryo–electron micrographs were collected on a Titan Krios microscope equipped with a K3 detector, yielding 9.3 million autopicked particles (fig. S3). Extensive three-dimensional (3D) classification revealed three classes that

[1]Laboratory of Protein and Nucleic Acid Chemistry, The Rockefeller University, New York, NY 10065, USA.
[2]Tri-Institutional Training Program in Chemical Biology, The Rockefeller University, New York, NY 10065, USA.
*Corresponding author. Email: klinge@rockefeller.edu (S.K.); avanden@rockefeller.edu (A.V.B.)
†These authors contributed equally to this work.
‡Present address: Biozentrum, University of Basel, 4056 Basel, Switzerland.

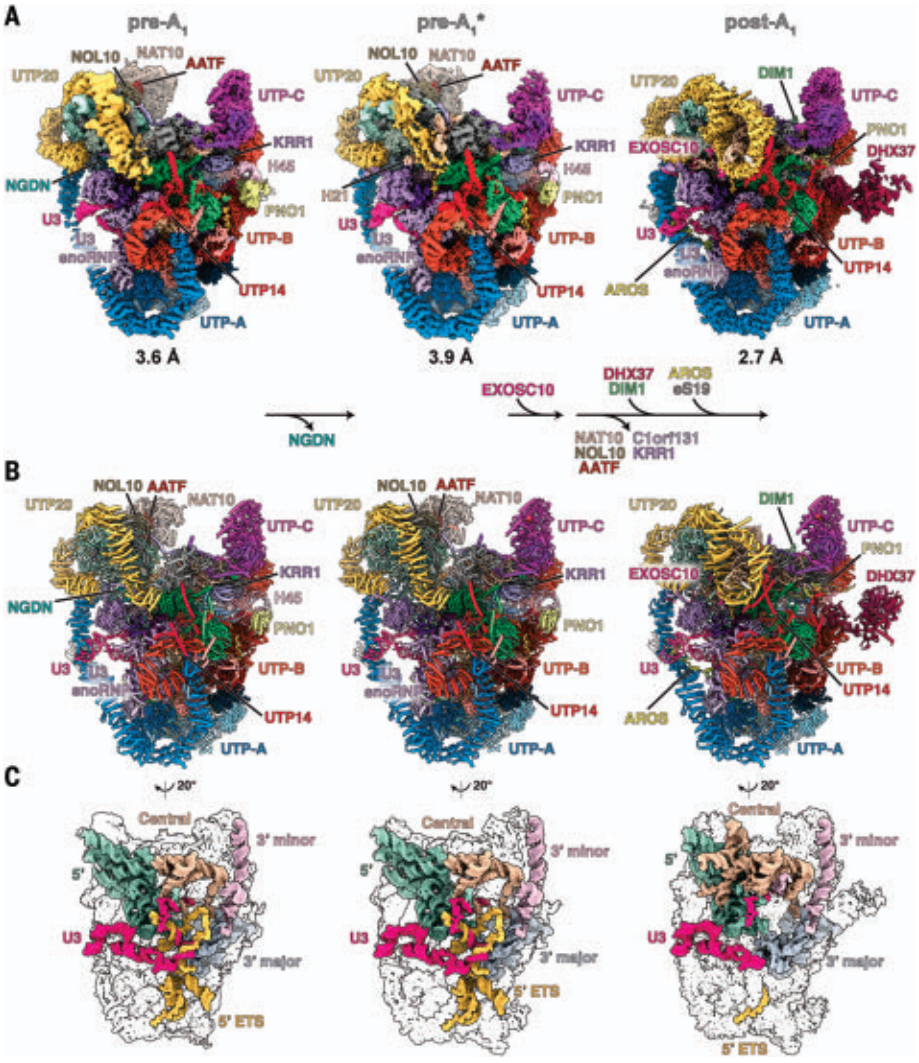


Fig. 1. Nucleolar maturation of human SSU processome. (**A**) Cryo-EM density maps of nucleolar states pre-A1, pre-A1*, and post-A1 at 3.6, 3.9, and 2.7 Å, respectively. Modules, assembly factors, and RNA are labeled, and compositional changes are indicated at the bottom. (**B**) Atomic models of the three states displayed and labeled as in (A). (**C**) Structures of RNA elements in each state. 5′ domain, green; central domain, beige; 3′ major domain, gray; 3′ minor domain, light pink; 5′ ETS, yellow; U3 snoRNA, pink.

represent distinct maturation states of the human SSU processome, here referred to as pre-A1, pre-A1*, and post-A1, which were resolved to resolutions of 3.5, 3.9, and 2.7 Å, respectively (Fig. 1A and fig. S3). As was done for the first complete yeast SSU processome structure (*11*), the use of focused classifications and refinements of subregions of particles belonging to states pre-A1 and post-A1 facilitated atomic model building and yielded high-resolution composite maps (figs. S4 to S9 and table S1). These reconstructions enabled the assignment and precise model building of human preribosomal RNA (5′ ETS and 18*S* rRNA), U3 small nucleolar RNA (snoRNA), 21 ribosomal proteins, and 50 human ribosome assembly factors (Fig. 1B, figs. S10 and S11, tables S2 to S4, and movies S1 to S4). For an in-depth view of these states at the atomic coordinate level, PyMOL sessions and morphs highlighting the differences between states pre-A1, pre-A1*, and post-A1 are available (data S1 and movies S5 and S6). For the post-A1 state, posttranscriptionally modified nucleotides could be built that are in agreement with prior mass spectrometry and structural data (*22*, *23*). The three states are characterized by the maturation of preribosomal RNA as well as state-specific protein components. Whereas states pre-A1 and pre-A1* are characterized by the presence of the 5′ ETS, the post-A1 state only contains a minimal ordered segment of the 5′ ETS near the interacting UtpA and UtpB complexes (Fig. 1C and fig. S12). In addition to vertebrate-specific extension of the UtpA and UtpB complexes, the human UtpA complex lacks a Utp9 homolog and instead contains two copies of WDR43 (Utp5 in yeast), like what was previously observed in *Chaetomium thermophilum* (*24*). Similarly, we do not observe a human homolog of Utp30, a ribosome assembly factor that is replaced by the small ribosomal subunit protein eS19 during the assembly of the yeast SSU processome. The loss of neuroguidin (NGDN) signifies the transition from pre-A1 to pre-A1*, whereas a range of pre-A1–specific factors (NAT10, NOL10, AATF, C1orf131, and KRR1) are removed during the subsequent transition toward post-A1 in which DHX37, DIM1, the RNA exosome subunit EXOSC10, and eS19 and its chaperone AROS are recruited.

Recognition and processing of the human 5′ ETS

A distinctive feature of mammalian preribosomal RNAs is the drastically enlarged 5′ ETS (fig. S13), for which both structure and function have remained obscure. The critical structured elements of the human 5′ ETS can be visualized in states pre-A1 and pre-A1* (Fig. 2, A to D). Although the structured parts of the human 5′ ETS occupy a similar region as first reported in yeast (*10*), the overall topology of the human 5′ ETS differs from that observed in yeast (fig. S14). Recognition of the human 5′ ETS is mediated by extensive protein–RNA interactions with nearby assembly factors as well as base pairing with U3 snoRNA. Downstream of the cotranscriptional A′ cleavage site, the evolutionarily conserved motif (ECM) within the 5′ ETS (*25*) is specifically recognized by a composite binding site formed by UTP4 and WDR75 (Fig. 2C). Following a set of seven ordered helical segments, the 5′ ETS is first recognized by the U3 snoRNA via its 3′ hinge. Notably, the 5′ ETS segment paired with the 3′ hinge is located several hundred nucleotides away from the previously postulated position (*26*). After a largely single-stranded segment, the 5′ ETS protrudes into the solvent with a long disordered helical segment, which is followed by an ordered sequence that pairs with the U3 5′ hinge before once again extending into the solvent. As in yeast (*11*, *27*), U3 snoRNA forms a central architectural anchor point to coordinate both the 5′ ETS via 3′ and 5′ hinges and the 18*S* via boxes A and A′ (Fig. 2B and figs. S10 and S14).

The small fraction of the human 5′ ETS that we could visualize within the human SSU processome prompted us to analyze the functional significance of the other elements that make up this 3.6-kb segment of human preribosomal RNA. We generated a synthetic human ribosomal DNA (rDNA) locus under a native promoter on a plasmid for transfection into human cells. This synthetic rDNA contains distinctive sequences within the 18*S* and 28*S* sequences that allow mature ribosomal subunits originating specifically from this recombinant

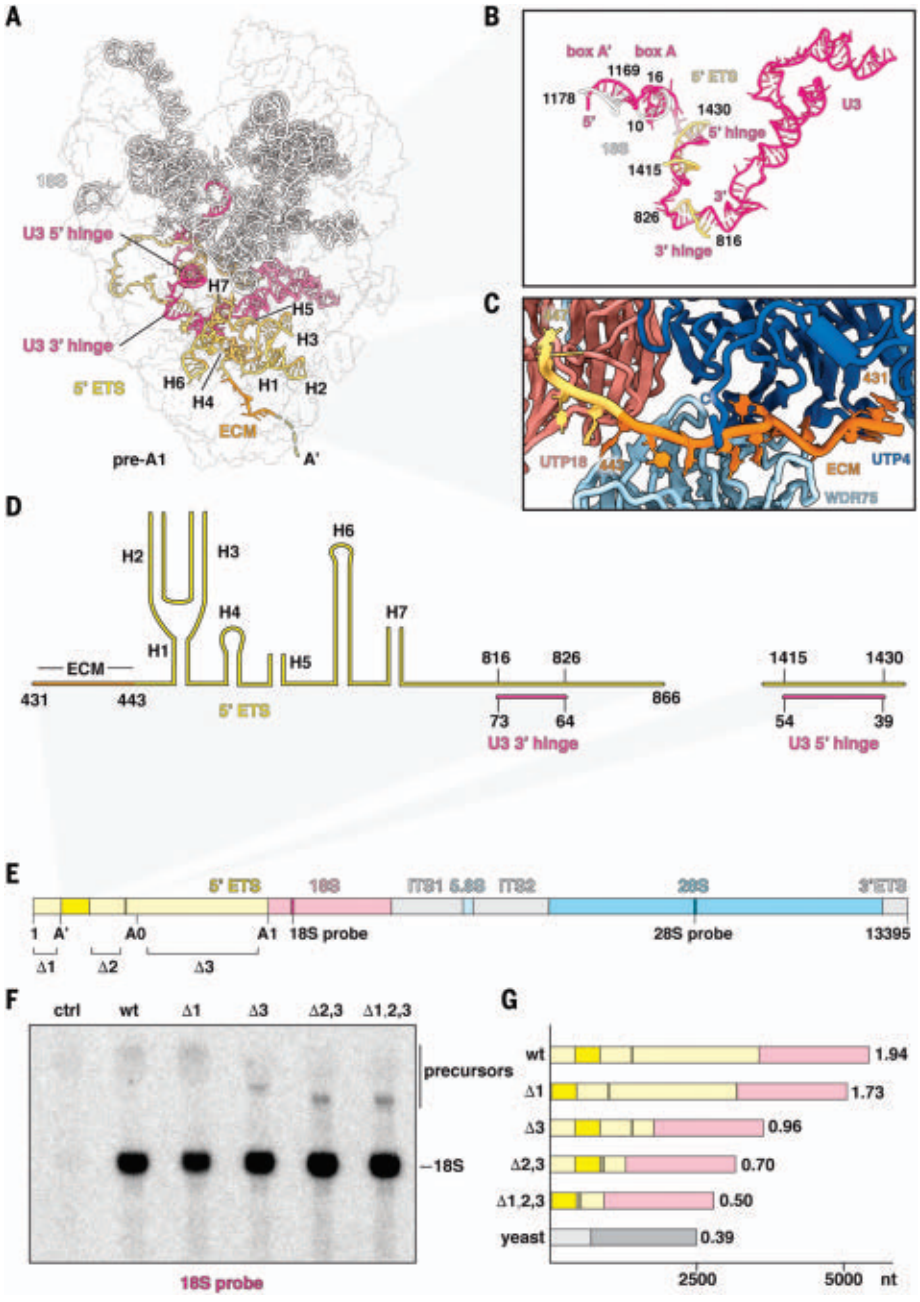


Fig. 2. Recognition and processing of the human 5′ ETS. (**A**) Architectural view of state pre-A1 with pre-18*S* (white), 5′ ETS (yellow), and U3 snoRNA (pink). The ECM (orange) and U3 hinges are indicated. (**B**) Detailed view of U3 snoRNA and its interactions with pre-18*S* and 5′ ETS RNAs. (**C**) Recognition of the ECM by UTP4 (dark blue) and WDR75 (light blue). (**D**) Secondary structure of the observed 5′ ETS in state pre-A1. (**E**) Schematic representation of a synthetic human rDNA locus containing distinct sequences for the small and large ribosomal subunits (18*S* and 28*S* probe, respectively). Ordered regions of the 5′ ETS are indicated in dark yellow, and segments for truncations are indicated with Δ1, Δ2, and Δ3. (**F**) Northern blot analysis of cells transfected without plasmid (control, ctrl), wild-type (wt), and truncated synthetic human rDNA loci (Δ1; Δ3; Δ2,3; and Δ1,2,3). Precursors and mature 18*S* rRNA originating from the synthetic template are indicated. (**G**) Schematic illustrating the ratios between 5′ ETS and 18S of wild-type and truncated rDNAs. Ratios between 5′ ETS and 18*S* are indicated at the end of each bar, with yeast (bottom bar) serving as reference. nt, nucleotides.

rDNA locus to be detected during Northern blotting analysis. On the basis of our secondary structure model and the built residues of the human 5′ ETS observed in states pre-A1 and pre-A1*, this system was derivatized to generate internal deletions of 5′ ETS regions that were not visualized in our cryo-EM reconstructions. These deletions included sequences upstream of the A′ cleavage site (Δ1), the long disordered helical element upstream of the 5′ hinge (Δ2), or a large helical segment between sites A0 and A1 (Δ3) (Fig. 2E and fig. S14).

Unexpectedly, single, double, and triple truncations of these sequences within the human 5′ ETS gave rise to mature small ribosomal subunits, as evidenced by the presence of a specific 18*S* band (Fig. 2F and fig. S15). Just as combined truncations in the 5′ ETS reduce the size of the transcript, a corresponding reduction in the size of the precursor species was observed, further supporting the specific assembly of mature 18*S* particles from the synthetic truncated rDNA locus. These data highlight that the human 5′ ETS can be separated into two functional units: the minimal functional core of the 5′ ETS that is required for SSU processome assembly (~25% of the 5′ ETS) and a large fraction of the 5′ ETS that is not required for SSU processome formation and may therefore function in biomolecular condensate formation (~75% of the 5′ ETS). The 75% reduction in size of the minimal 5′ ETS more closely matches the 5′ ETS:18*S* ratio observed in yeast (Fig. 2G).

Licensing for exosome-mediated remodeling

A critical function of the SSU processome is to signal the stage of assembly of the small ribosomal subunit that triggers subsequent maturation events of preribosomal RNA. One of these key events is the exosome-mediated remodeling of the 5′ ETS. In yeast SSU processomes, large conformational changes within Utp20 have been observed that correlate with the dissociation of the assembly factor

Lcp5 (the yeast homolog of NGDN) as well as the dissociation of the Kre33 module (yeast homolog of NAT10, NOL10, and AATF) (*13*). However, owing to limited resolution, the molecular significance and consequences of these events for eukaryotic ribosome assembly have remained elusive. The structures of the human SSU processomes reveal how conformational changes within the particles are used as a direct molecular readout for the recruitment of the RNA exosome component EXOSC10 (yeast Rrp6) (Fig. 3). In the pre-A1 state, the helical repeat proteins HEATR1 and UTP20 wrap around most of the SSU processome, with UTP20 lining both sides of NGDN, which is sandwiched between eS4 and uS4 (Fig. 3A).

The transition from states pre-A1 to pre-A1* is associated with the dissociation of NGDN and the conformational switch of the U3IP2 N terminus to a post-A1 conformation, thereby creating a largely vacant space in which helix 21 of the 18*S* rRNA is disordered (Fig. 3B). A key event occurring during the transition between pre-A1* and post-A1 states is the dissociation of the NAT10-NOL10-AATF module upon which large-scale movements of the 5′ domain of the 18*S* rRNA are transmitted via UTP20 and HEATR1. While HEATR1 moves upward and approaches the U3 snoRNP, UTP20 adopts a drastically different conformation, with its C-terminal region looping back toward the middle of the protein. In concert with these large conformational changes, eS4 adopts a different conformation to stabilize helix 21 in a new environment (movie S6). Unexpectedly, we observe the so-called "lasso" region of the exosome-associated nuclease EXOSC10 (*28*) in the post-A1 state, which is sensitive to these conformational changes and thus only binds to state post-A1 of the human SSU processome (Fig. 3C). Specifically, four peptide motifs of EXOSC10 probe this area for UTP20 (motif I), the formation of a UTP20-eS4 interface (motif II), the availability of eS24 for β sheet formation (motif III), and the proximity of HEATR1 (motif IV). Multiple sequence alignments of these motifs show that motifs I to III are present from yeast to humans whereas motif IV is a mammalian-specific extension (fig. S16). This mechanism of substrate recognition by EXOSC10 (Rrp6) upon conformational changes within the SSU processome is universal, as we have identified corresponding elements in exosome-associated SSU processomes from yeast in which these regions could not be interpreted previously (fig. S16, A to D) (*13*–*15*). These analyses highlight that there are two links between the SSU processome and the exosome: The exosome recognizes the SSU processome by way of the C terminus of EXOSC10, and the SSU processome recruits Mtr4 by way of arch-interacting motifs (AIMs) present in either Utp18 [in yeast (*29*)] or presumably TDIF2 (human), which contains a similar sequence motif in the solvent-exposed region (fig. S16E).

Regulation of the DEAH-box helicase DHX37

A key RNA helicase involved in the progression of later SSU processome intermediates is the RNA helicase DHX37, an enzyme of the DEAH-box family whose conformational state in the context of maturing SSU processomes remains the subject of controversy. Indeed, although high-resolution crystal structures of substrate-bound DHX37 (*30*) and a partially proteolyzed apo structure of the yeast homolog Dhr1 have been determined (*31*), a radically different configuration of the subdomains of Dhr1 was observed in yeast SSU processomes at lower resolution, a discrepancy originally attributed to a different nucleotide state (*13*). To elucidate the precise atomic basis of this altered state, we determined the 2.2-Å crystal structure of the yeast Dhr1 core enzyme (Fig. 4A and table S5). This structure reveals the presence of a loop within the RecA2 domain blocking the substrate channel. A comparison with the RNA-bound DHX37 shows that this RecA2 loop occupies the same region

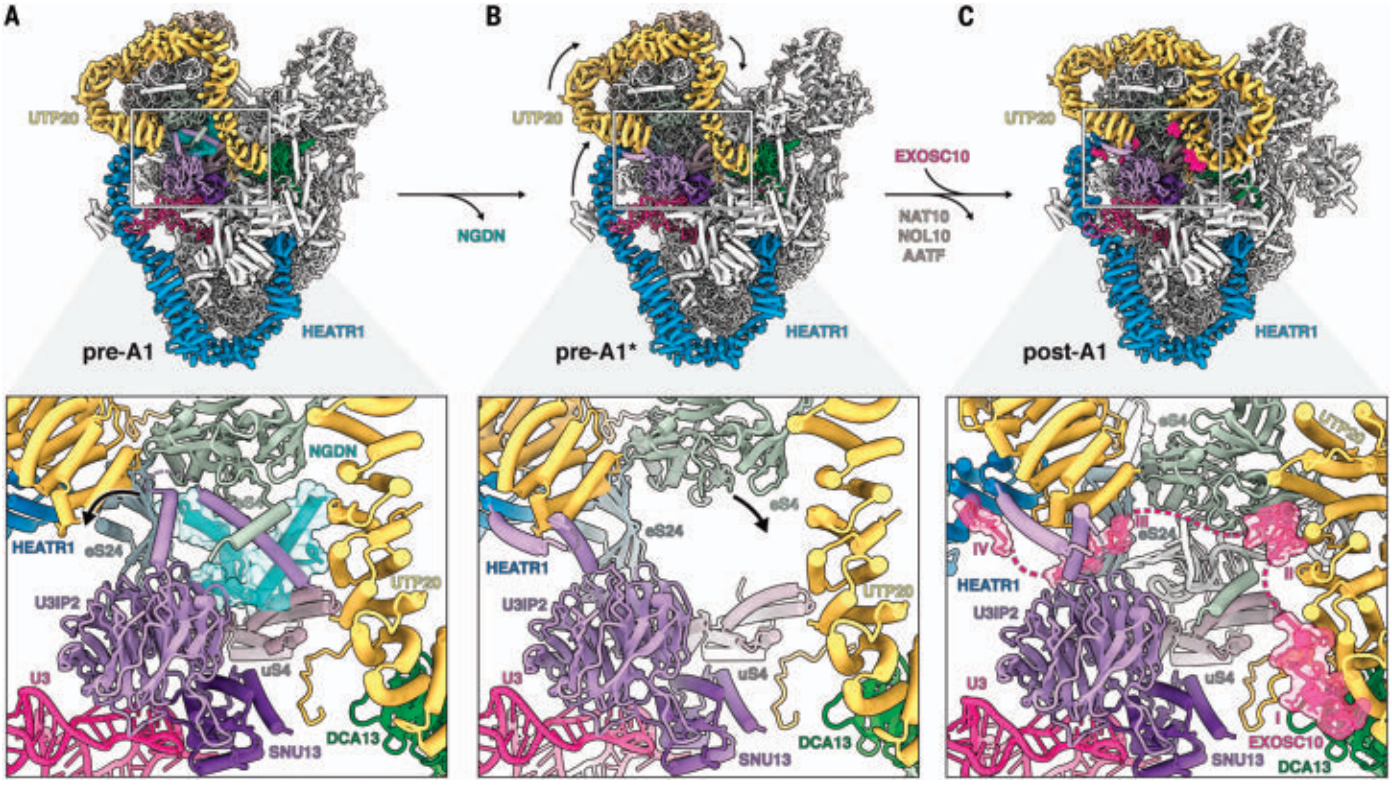


Fig. 3. Structural remodeling facilitates licensing for exosome-mediated maturation. (**A**) (Top) Architecture of state pre-A1. (Bottom) Zoomed view highlighting the NGDN (cyan, transparent surface) binding site near UTP20 (yellow), U3IP2 (purple), and ribosomal proteins eS4, eS24, and uS4. (**B**) (Top) Architecture of state pre-A1*. (Bottom) Zoomed view highlighting the loss of NGDN (cyan) and rearrangement of the U3IP2 N terminus. (**C**) (Top) Architecture of state post-A1. (Bottom) Zoomed view highlighting the presence of the EXOSC10 (pink, transparent surface) lasso with four peptide epitopes (I to IV) near the moved eS4.

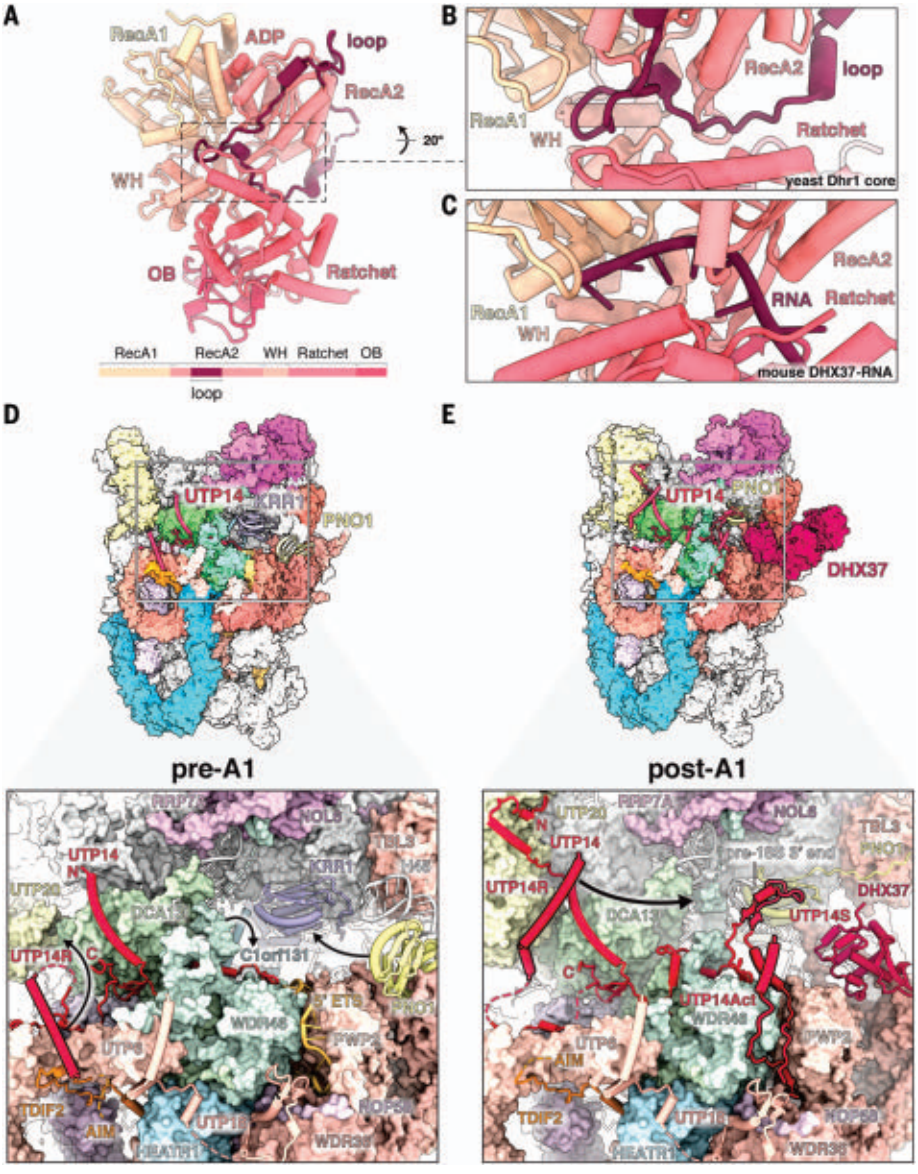


Fig. 4. Regulation of the DEAH-box helicase DHX37. (**A**) Crystal structure of the yeast Dhr1 core enzyme with the RecA2-associated autoinhibitory loop. A schematic of the yeast Dhr1 domain organization is shown at the bottom. ADP, adenosine diphosphate; WH, winged helix domain; OB, oligonucleotide/oligosaccharide binding-fold domain. (**B** and **C**) Detailed views of the Dhr1 autoinhibitory loop in the substrate channel (B) compared with the substrate RNA in the mouse Dhx37 structure (PDB ID 6O16) (C). (**D**) (Top) Structure of human pre-A1 state highlighting UTP14 proximal proteins. (Bottom) Detailed view of UTP14 and the recruitment domain (UTP14R). Black arrows depict subsequent movements of UTP14R, WDR46, and PNO1 from pre-A1 to post-A1. (**E**) (Top) Structure of human post-A1 state highlighting UTP14 proximal proteins. (Bottom) Detailed view of UTP14 and its recruitment, activation, and stabilizing domains (UTP14R, UTP14Act, and UTP14S, respectively). Positioning of these domains depends on movement of UTP20 (UTP14R), processing of the 5′ ETS (UTP14Act), and movement of Pno1 (UTP14S). The black arrow depicts the putative path of UTP14R toward its binding site, as seen in the yeast state Dis-C. The putative AIM of TDIF2 is indicated.

as the RNA substrate, suggesting that autoinhibition modulates enzymatic activity (Fig. 4, B and C, and fig. S17). This autoinhibited state of the enzyme can be docked into all yeast SSU processome states in which Dhr1 is not engaging its substrate, confirming the physiologically relevant role of the proposed mechanism (fig. S18, A to I) (*13*, *15*). Moreover, the principle of DHX37 autoinhibition appears to be conserved in the human pathway as well, because the domain orientations observed at relatively low resolution in the human post-A1 state display a related overall topology (fig. S18, J to M).

The autoinhibited DHX37 is recruited to the human SSU processome in the post-A1 state via its C-terminal domain, which is structurally distinct from the yeast homolog but occupies the same location between PWP2 and WDR36 of the UtpB complex (Fig. 4E). A prerequisite for this localization is the transition of the assembly factor PNO1 with helix 45 of the 18*S* rRNA into the interior of the particle to replace a site that was previously occupied by the assembly factor KRR1 (Fig. 4D). The recruitment site of DHX37 is close to UTP14, a protein that had been previously implicated in DHX37 recruitment and activation (*30*, *32*, *33*). As recruitment and activation of DHX37 need to be tightly controlled, UTP14 elements involved in either of these processes are themselves highly regulated and respond to the overall processing state of the human SSU processome.

In late states of SSU processome disassembly in yeast (Dis-C), an N-terminal helical segment of UTP14 (here referred to as UTP14R) recruits DHX37 close to its substrate U3 snoRNA near the 18*S* 3′ end (*13*). Positioning of UTP14R and access to its final binding site therefore require extensive control (Fig. 4 and fig. S19). In the pre-A1 state, UTP14R is sequestered on the UtpB subunit UTP6, and its final binding site is occupied by C1orf131, the human homolog of Faf1 that we have identified in the present study (Fig. 4D and fig. S20). In the post-A1 state, UTP14R is targeted to a different binding site near the rearranged UTP20 C-terminal segment. In this state, the final UTP14R binding site is now occupied by the C terminus of WDR46 (Utp7 in yeast) that has replaced C1orf131 to block premature binding of UTP14R (Fig. 4E). Upon dissociation of WDR46 during later stages of SSU processome disassembly, UTP14R can translocate to its final binding site to recruit DHX37. The ability to recruit DHX37 to its final location near its substrate therefore depends on a perfectly choreographed movement of the UTP14R segment and access to its binding site that is

guided by large conformational changes within the SSU processome that signal distinct maturation stages (Fig. 4 and fig. S19).

A second layer of control is provided by the activation of DHX37 helicase activity. An activating segment of UTP14 (here referred to as UTP14Act) has been shown to stimulate the adenosine triphosphatase activity of DHX37 in vitro (*30*), yet how premature activation of DHX37 on the SSU processome is prevented has remained unclear. In the pre-A1 state, the 5′ ETS RNA threads through a narrow groove between PWP2 and WDR46 (Fig. 4D), and upon processing of the 5′ ETS from pre-A1 to post-A1, this same location is used to sequester UTP14Act. At the same time, a stabilizing segment of UTP14 (here referred to as UTP14S) stabilizes both PNO1 and the 18*S* 3′ end to help create half of the binding site that will later be used by UTP14R and the C-terminal domain of DHX37 during substrate engagement. The release of WDR46 during later stages of SSU processome disassembly therefore not only vacates the UTP14R binding site to enable DHX37 recruitment but also releases UTP14Act for subsequent activation of DHX37.

Together, our crystallographic and cryo-EM analyses reveal elegant layers of control that are used to keep DHX37 in an inactive state and regulate both its localization and activity as a function of larger conformational changes that are structurally read out by UTP14 (movie S6). Lastly, we observe that the presence of DHX37 in the post-A1 state is also necessary for the recruitment of the DIM1 methyltransferase, explaining the previously observed co-dependence of both proteins in depletion experiments of either protein in human cells (fig. S21) (*33*). The timing of DHX37 and DIM1 incorporation into the SSU processome is further coordinated with the dissociation of the NAT10-NOL10-AATF complex, as overlapping binding sites near eS6 ensure a mutually exclusive presence of these enzymes (fig. S21).

Protein-guided mechanism of cleavage at site A1

A key, irreversible event during the maturation of the small ribosomal subunit is the cleavage between the 5′ ETS and 18*S* rRNA that generates the mature 5′ end of the 18*S* rRNA. Within the core of the human SSU processome, cleavage at site A1 is orchestrated through protein-guided RNA movements that reposition the substrate RNA within the PIN domain nuclease UTP24 (*34*). Here, a set of helical elements originating from NOP14, IMP4, and the guanosine triphosphatase BMS1 undergo radical conformational changes to stabilize boxes A and A′ in the pre-A1 and post-A1 states (Fig. 5). In the pre-A1 state, the C terminus of BMS1 reaches across U3 snoRNA between boxes A and A′ to make contact with a protein cluster containing IMP4, IMP3, and C1orf131, while the C-terminal segment of NOP14 stabilizes box A′ in a lower position. The site of RNA cleavage, the 5′ end of the 18*S* rRNA, is positioned far away from the nuclease UTP24, which is additionally occluded by C1orf131 (Fig. 5A). In the post-A1 state, C1orf131 has disappeared and KRR1 has been replaced by PNO1. Box A′ is moved upward and stabilized by the NOP14 C-terminal segment that has changed its secondary structure from a previously continuous helix to a partially disrupted helix. Box A has been relocated closer to UTP24, and nucleotide 2 of the 18*S* rRNA is now seen in close proximity to the nuclease active site (Fig. 5B). The configuration of boxes A and A' is further stabilized by long helices of IMP4 and BMS1, both of which have undergone substantial conformational changes (movie S5).

During the transition from the pre-A1 state to the post-A1 state, the N-terminal segment of UTP24 reads out conformational changes that directly depend on the destabilization of the 5′ ETS (Fig. 5, C and D). In the pre-A1 state, the UTP24 N terminus is wedged between the 5′ hinge, the IMP3 N terminus, PWP2, and WDR46 (Fig. 5C). With the processing of the 5′ ETS in the post-A1 state, IMP3, PWP2, and WDR46 undergo conformational changes that prevent the UTP24 N terminus from binding in the same region. Instead, the N terminus of UTP24 now forms a lid on top of the nuclease active site, presumably to stabilize the incoming substrate 18*S* rRNA (Fig. 5D). The combined insights obtained from human SSU processomes in states pre-A1 and post-A1 highlight that upon exosome-mediated processing of the 5′ ETS, the movement of boxes A and A′

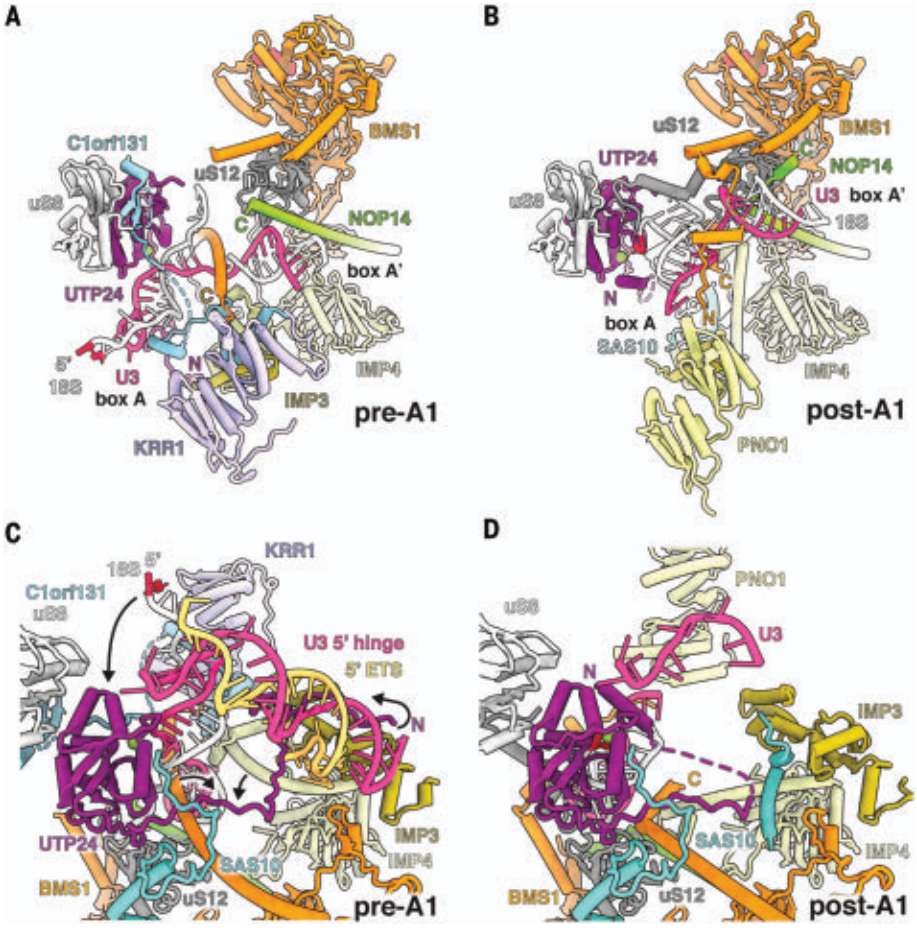


Fig. 5. Protein-guided mechanism of cleavage at site A1. (**A**) Catalytic core of state pre-A1 with U3 boxes A and A′ (pink) and base-pairing 18*S* (white). C1orf131 (light blue) occludes the nuclease UTP24 (purple), and nucleotide 2 of the 18*S* rRNA (red) is positioned far away. (**B**) In state post-A1, C1orf131 has disappeared and KRR1 has been replaced by PNO1 (yellow). Rearrangements in U3 box A and A′ allow nucleotide 2 (red) of the 18*S* rRNA to be in close proximity to the UTP24 nuclease active site (magnesium ion, green circle). (**C**) In state pre-A1, the UTP24 N terminus (purple) is sequestered via interactions with the 5′ hinge (pink), 5′ ETS (yellow), and the IMP3 N terminus (olive). Black arrows highlight structural changes within the N terminus of UTP24, the 5′ end of 18*S* rRNA, and other proteins that are required for A1 cleavage. (**D**) Destabilization of the 5′ ETS and accompanying changes in the post-A1 state allow the rearranged UTP24 N terminus to stabilize incoming substrate 18*S* rRNA.

is guided by the coordinated structural adaptation of a number of proximal protein segments. In addition, the activation of UTP24 depends on the release of the inhibitory C1orf131 and the release of the UTP24 N terminus to act as part of the enzyme-substrate complex for cleavage at site A1.

Human diseases of the SSU processome

Mutations in both ribosomal proteins and assembly factors present in the human SSU processome have been associated with human diseases, so-called ribosomopathies (*35*), such as Diamond-Blackfan anemia, and several other childhood diseases, which can now be visualized in the context of the human SSU processome (Fig. 6). In >25% of patients suffering from the ribosomopathy Diamond-Blackfan anemia (*36*), mutations in eS19 have been identified, which are linked to the accumulation of a 21*S* preribosomal RNA precursor species (*37*, *38*). In the human SSU processome, we have identified AROS (also known as RPS19BP1) as a chaperone that specifically mediates the integration of eS19 in state post-A1 (Fig. 6, A, B, and D, and fig. S22). The significance of AROS in the post-A1 state is further highlighted by previous biochemical data showing that small interfering RNA–mediated knockdown of AROS phenocopies the loss of eS19 (*39*) and that the AROS C-terminal domain is required for eS19 binding (*40*). The binding site for the AROS N terminus is created in state post-A1 upon the conformational changes in UTP20 during the transition from pre-A1 to post-A1, which are communicated through HEATR1 near U3 snoRNA. These changes result in the creation of a composite binding site between HEATR1 and U3, allowing AROS to thread along a similar trajectory as the unrelated protein Rrt14 in yeast (Fig. 6B, fig. S22, and movie S6). As eS19 is recognized by a conserved small C-terminal helix of AROS, any mutations within eS19 likely abrogate this interaction and thus the incorporation of eS19 into the SSU processome (Fig. 6, B and D). The structure of the human SSU processome further provides physiological context for the effects of proximal mutations found in uS12 and BMS1, which give rise to different pathologies (Fig. 6C) and illustrates how mutations found in Diamond-Blackfan anemia [ribosomal protein S7 (RPS7), RPS19, RPS24, RPS27, and RPS28], 5q- syndrome (RPS14), RPS23-related ribosomopathies, aplasia cutis congenita (BMS1), Bowen-Conradi syndrome (NEP1), and North American Indian childhood cirrhosis (UTP4) (*36*, *41*–*45*) may affect SSU processome formation and maturation (Fig. 6).

Discussion

The recognition and controlled processing of large preribosomal RNA segments are the key functions of the human SSU processome. Here we provide high-resolution insights into the nucleolar maturation of human SSU processomes. These structural insights, together with complementary functional data, elucidate the mechanisms by which 50 ribosome assembly factors work in concert to precisely tune a strict chronology of assembly events in the nucleolus.

A key principle that emerges from this study is how changes in RNA folding are transmitted through conformational changes in assembly factors and how these conformational changes are subsequently recognized by downstream factors involved in ensuing pre-rRNA maturation steps. Two prominent examples of this principle are the conformation-specific recognition of human SSU processomes by the RNA exosome (Fig. 3) and the complex interplay of different placeholder proteins that precisely orchestrate localization and activation of the DEAH-box helicase DHX37 (Fig. 4). These concepts are broadly applicable to large assemblies

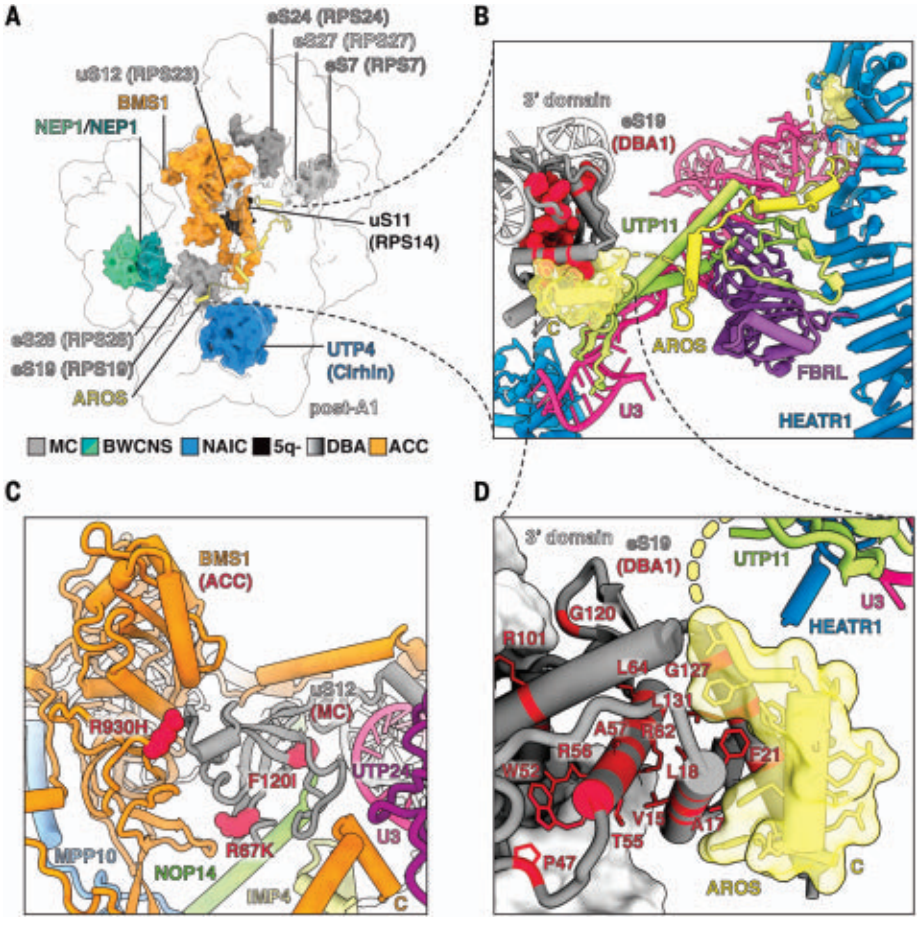


Fig. 6. Human diseases of the SSU processome. (**A**) SSU processome factors with disease-causing mutations are highlighted within state post-A1. MC, microcephaly; BWCNS, Bowen-Conradi syndrome; NAIC, North American Indian childhood cirrhosis; 5q-, 5q- syndrome; DBA, Diamond-Blackfan anemia; ACC, aplasia cutis congenita. (**B**) Detailed view of AROS (yellow, transparent surface) and its interacting proteins. DBA1-causing mutations in eS19 are colored red. (**C**) Detailed view of BMS1 (orange), uS12 (gray), and neighboring proteins. Disease-causing mutations in BMS1 (ACC) and uS12 (MC) are labeled and shown as red spheres. (**D**) A close-up view of the interface between eS19 (gray) and AROS (yellow, transparent surface). DBA1-causing mutations in eS19 are labeled and shown as red sticks.

such as maturing preribosomal particles or assembling spliceosomes.

Eukaryotic ribosome assembly is a very tightly controlled process, and our discovery that the exosome component EXOSC10 (Rrp6) lasso directly recognizes a particular state of the SSU processome highlights redundant pathways that intimately link the exosome with one of its key substrates, the SSU processome. At least two redundant links exist between the exosome and the SSU processome: First, the exosome-associated helicase Mtr4 can interact with SSU processome factors containing AIM motifs [Utp18 in yeast (*29*) and presumably TDIF2 in humans]. Second, the lasso of EXOSC10 (Rrp6) can directly bind to SSU processomes that have already lost NGDN (Lcp5) (Fig. 3). Whereas a loss of either of these links should be tolerated, the removal of both links should be deleterious, and this is indeed observed in yeast where Utp18 AIM mutants are synthetically lethal with an Rrp6 deletion (*29*).

On the basis of our data, a model emerges in which biomolecular condensate formation within the nucleolus enables human SSU processome maturation by concentrating ribosome assembly factors (Fig. 7A). The recognition of the SSU processome for exosome-mediated maturation is triggered by the removal of NGDN (Lcp5) (Fig. 7, B to D). With dual connections between the exosome and the SSU processome, the exosome uses its 3′-5′ exonuclease activity to remove the 5′ ETS from the 3′ end. With this trajectory, it is likely that the exosome triggers the major rearrangements that we observe in the SSU processome core. The exosome-mediated removal of the 5′ ETS RNA within the interior of the SSU processome will further drive drastic conformational changes of those assembly factors that are in direct contact with the 5′ ETS in state pre-A1 (WDR46, PWP2, UTP24, IMP4, BMS1, and NOP14, among others). These conformational changes include the assembly factors involved in the protein-guided cleavage at site A1 within the core of the SSU processome as well as conformational changes within the periphery of the SSU processome that contribute to the control of the helicase DHX37 before it engages U3 snoRNA during subsequent maturation steps.

Our functional studies illuminate the evolution of eukaryotic ribosome assembly and the expansion of the 5′ ETS sequences in higher eukaryotes, which may explain the observed changes in morphology of the nucleolus toward a tripartite architecture in humans (fig. S13). As mature human small ribosomal subunits can be assembled from a minimal pre-rRNA with a greatly reduced 5′ ETS (Fig. 2 and fig. S14), these data suggest that there are two distinct functions of the human 5′ ETS: First, a minimal 5′ ETS is sufficient for ribosome assembly in cis, as a human SSU processome can form on this pre-rRNA and mature 40*S* ribosomes are generated (Fig. 2F). Second, the large extensions found in human 5′ ETS sequences likely drive ribosome assembly in trans by forming a biomolecular condensate with intrinsically disordered regions (IDRs) of highly concentrated ribosome assembly factors, the dense fibrillar component (DFC) (Fig. 7A). Therefore, the targeted degradation of the 5′ ETS by the RNA exosome eliminates large interaction partners within the DFC, which could liberate late human SSU processomes from the biomolecular condensate that they initiate.

Lastly, the structures of maturing human SSU processomes further provide physiological context for a range of disease-causing mutations in ribosomal proteins and assembly factors. The observation that AROS is directly involved in the recruitment of eS19 into the SSU processome suggests that, in the future, mutations within AROS may be found that give rise to a similar pathology as mutations within eS19 that are currently observed in patients suffering from Diamond-Blackfan anemia.

Materials and methods

Generation of a biallelically tagged NOC4L-GFP cell line

For the swift generation of a biallelic tagged NOC4L-GFP cell line, we developed an iterative genome editing platform called SNEAK PEEC (surface engineered fluorescence assisted kit with protein epitope enhanced capture), which combines CRISPR-Cas9 genome editing with cell-surface display. SNEAK PEEC links each precisely targeted knock-in with the expression of a distinct exogenous protein epitope on the cell surface, which allows clonal identification using fluorescent binders of each epitope. The system uses two DNA repair templates, each encoding a 3C protease cleavable green fluorescent protein (GFP) tag followed by a surface display sequence encoding a distinct protein epitope. For this study, the btuF vitamin B12 binding protein in *Escherichia coli* [Protein Data Bank (PDB) ID 5OVW] and capsid protein p24 from human immunodeficiency virus 1 (PDB ID 5O2U) were used as the surface display epitopes.

Both DNA repair template plasmids contained homology arms (left and right) that covered ~600 base pairs (bp) in either direction of the Cas9 cut site in the last exon of the *NOC4L* gene. Homology arms flanking a multiple

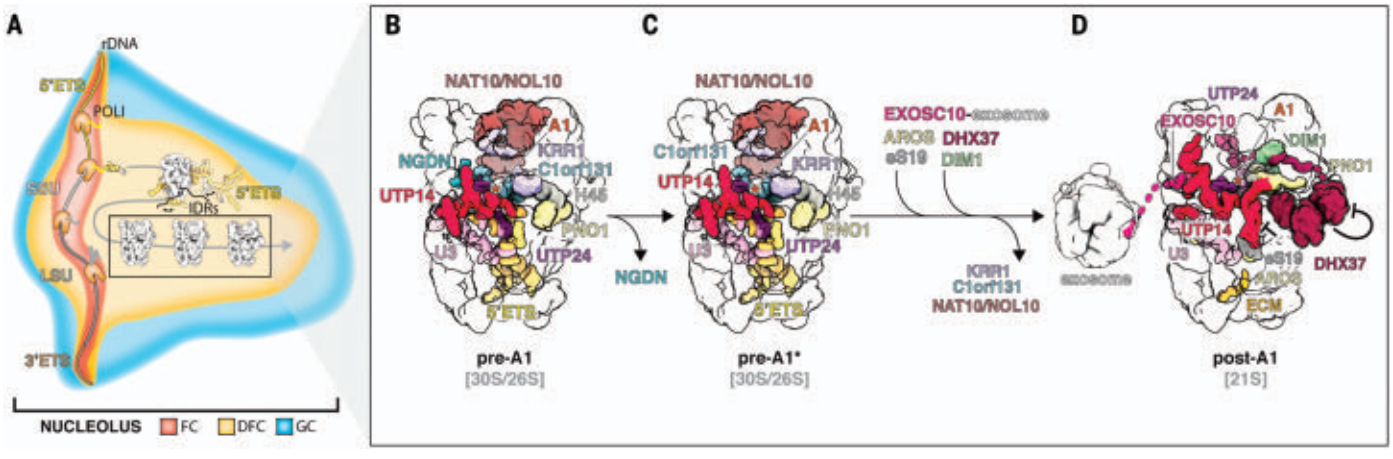


Fig. 7. Model for nucleolar maturation of the human ribosomal small subunit. (**A**) Illustration depicting the assembly of human SSU processomes within the nucleolus, with fibrillar center (FC, red), dense fibrillar component (DFC, yellow), and granular component (GC, blue). The expanded human 5′ ETS (yellow) is shown interacting with IDRs of nucleolar proteins. Processing of pre-18*S* rRNA progresses via different states of the SSU processome (black box). (**B** to **D**) Detailed views of three states indicating the chronology of events with components present in states pre-A1 (B), pre-A1* (C), and post-A1 (D). Components leaving and joining the particles are indicated, and inhibitory mechanisms are highlighted.

cloning site (MCS) were first cloned into pUC57 vectors (GenScript, Piscataway, NJ). The entire knock-in sequence (3C-GFP-2A-display) was then introduced using NotI/PacI sites in the MCS. A plasmid expressing a single guide RNA (sgRNA) together with a high specificity *Streptococcus pyogenes* Cas9 variant [eSpCas9(1.1)] was a gift from F. Zhang (*46*) (Addgene plasmid # 71814). The 20-bp sgRNA target sequence (5′-AAACTCCAGTGGCACCGGCT-3′, PAM: CGG) was cloned into this vector using a pair of BbsI sites. The sgRNA sequences were selected using an online resource, crispor.telfor.net (*47*).

Transfections were carried out on HEK 293-F cells (ThermoFisher Scientific, R79007) grown in 24-well plates (Falcon, 353047). Each well received 1 μg of total DNA, split in equimolar concentrations of each plasmid (Repair template 1, Repair template 2, Cas9 +sgRNA). All transfections were carried out in 500 μl of Opti-MEM medium/well (ThermoFisher Scientific, 31985070) using the commercial transfection reagent Lipofectamine 2000 (ThermoFisher Scientific, 11668019). Cells were transfected at 70 to 90% confluency, and 14 to 16 hours after transfection, cells were washed and resuspended in Freestyle 293 Expression Medium (ThermoFisher Scientific, 12338026) supplemented with 2% heat inactivated fetal bovine serum. Cells were allowed to recover, after which they were expanded to six-well plates (VWR, 10062-892) for at least 7 days.

For flow cytometry, cells were harvested from six-well plates after a minimum of 7 days after transfection. Cells were first detached by gentle aspiration and washed 1× with 1X phosphate-buffered saline (PBS), 0.1% bovine serum albumin (BSA). They were then resuspended in 1X PBS plus 0.1% BSA at a concentration of 1×10^6 to 10×10^6 cells/ml. For cell surface staining, two fluorescently labeled nanobodies, each of which specifically binds to one of the surface display epitopes, were added to 200-μl cell suspensions at a final concentration of 10 nM. Cells were labeled on ice in the dark for 30 min, washed two times with 1X PBS plus 0.1% BSA and resuspended in 200 μl of the same buffer. Samples were filtered before cell sorting to remove clumps (BD Falcon, 352235). DAPI (4′,6-diamidino-2-phenylindole) was used to stain dead cells. For the identification of a biallelically tagged NOC4L-GFP cell line, sorting was used to first select cells expressing GFP, followed by selection of cells within this population that stained positive for both btuF and p24 surface display epitopes. Cell sorting was carried out using a BD FACSAria cell sorter (BD Biosciences) using FACSDiva Software (BD Biosciences). All necessary compensation was performed using single color controls before measurement and analysis of the data. FlowJo (FlowJo, LLC) was used for analysis of the data.

Sorted single cell clones were expanded for 2 weeks, after which they were screened via polymerase chain reaction (PCR) to confirm biallelic editing. Extraction of human genomic DNA was carried out using the QuickExtract DNA Extraction Solution (Lucigen, QE09050), per the manufacturer's protocol. Thirty microliters of extracted solution was used per screening PCR reaction (50 μl final volume). The following PCR primers were used in this study (fig. S1H): PCR1, fwd (CCTTTCATCTGTAACCTGCTGCGCC) and rev (GACTCCACGGGGCCAACTGTCTCAAGG); PCR2, fwd (CCTTTCATCTGTAACCTGCTGCGCC) and rev (TGCTGTCATCATTTCCTCGAGCGTAGCACC).

Purification of human SSU processomes

Purification of the human SSU processome was carried out by means of sequential lysis. NOC4L-GFP cells were grown as suspension cultures, reaching a cell concentration between 4×10^6 and 5×10^6 cells/ml. Cell pellets were washed twice with cold 1X PBS and frozen in liquid nitrogen for storage at −80°C until use. Frozen cell pellets were resuspended in basic buffer [25 mM Hepes pH 7.6, 65 mM NaCl, 65 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, 0.05% Triton, 1 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM pepstatin, 1 mM E-64 protease inhibitor] and the resulting lysate centrifuged, and the supernatant (cytoplasmic fraction) discarded. Pellets were washed with basic buffer, the supernatants discarded, and the pellet was resuspended in active buffer [25 mM Hepes pH 7.6, 10 mM KCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM EDTA, 1 mM spermidine, 5% glycerol, 0.1% Triton, 100 mM arginine, 25 mM adenosine triphosphate, 1 mM DTT, 0.5 mM PMSF, 1 mM pepstatin, 1 mM E-64 protease inhibitor, 800U RNase-free DnaseI (Worthington)]. The insoluble fraction was removed by centrifugation at 4°C, 40,000*g* for 20 min. The supernatant was incubated with anti-GFP nanobody beads (Chromotek) for 4 hours at 4°C, and the beads were washed 2× with wash buffer 1 (25 mM Hepes pH 7.6, 75 mM NaCl, 75 mM KCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 2.5% glycerol, 0.1% Triton, 100 mM arginine, 1 mM DTT, 0.5 mM PMSF, 1 mM pepstatin, 1 mM E-64). This was followed by 1× wash with wash buffer 2 [25 mM Hepes pH 7.6, 75 mM NaCl, 75 mM KCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 2.5% glycerol, 0.05% $C_{12}E_8$ (Anatrace, O330), 100 mM arginine]. The complex was eluted using 3C-protease cleavage, the integrity of particles confirmed using negative stain EM, and immediately followed by preparation of cryo-EM grids.

Cryo-EM grid preparation and data acquisition

Cryo-EM grids were prepared using a Vitrobot Mark IV robot (FEI Company) maintained at 12°C and 90% humidity. Three and a half microliters of the sample was applied to glow discharged Quantifoil R 2/2 grids coated with a layer of 2-nm ultrathin carbon (Au 400 2nm C, Electron Microscopy Sciences). After 70-s incubation, the sample was manually blotted while still in the Vitrobot chamber, and 3.5 μl of fresh sample was reapplied to the same grid. After 70-s incubation, the sample was blotted for 8 s using a blot force of 8, followed by plunging into liquid ethane.

Imaging was carried out on a Titan Krios electron microscope (FEI) equipped with an energy filter (slit width: 20 eV) and a K3 Summit detector (Gatan) operating at 300 kV with a nominal magnification of 64,000×. Using SerialEM (*48*), six datasets totaling 84,904 micrographs were collected with a defocus range of −0.7 to −2.7 μm and a super-resolution pixel size of 0.54 Å. Images with 45 subframes were collected using a total dose of 30.1 electrons per pixel per second (1.08 Å pixel size at the specimen) with an exposure time of 2.25 s and a total dose of 58 $e^-/Å^2$. A multishot imaging strategy was used to record seven micrographs per hole. The beam tilt was restricted to a single hole.

Cryo-EM data processing

Upon data collection, images were gain corrected, dose weighted, aligned, and binned to a pixel size of 1.08 Å using RELION's implementation of a MotionCor2-like algorithm (*49*, *50*). The defocus value of each micrograph was estimated using Gctf (*51*). A total of 9,297,626 particles were picked from six datasets comprising 84,904 micrographs using CrYOLO (*52*). The particles from each dataset were extracted at a pixel size of 4.32 Å and were separately subjected to a global search 3D classification in RELION 3.1 (*53*). An ab initio model generated in RELION 3.1 or a well-resolved 3D class from dataset 1 were used as initial references for 3D classification of dataset 1 and datasets 2 to 6, respectively (fig. S3). For each dataset, the particles from the best class were re-centered and reextracted at a pixel size of 1.08 Å and subjected to three rounds of CTF refinement and Bayesian polishing in RELION 3.1. The polished particles from each dataset were merged into a consensus particle stack of 824,089 particles. This particle stack was subjected to a 3D refinement resulting in a reconstruction of the human SSU processome at a global resolution of 2.68 Å. To separate different states present in the consensus reconstruction, three parallel 3D classifications were performed: one local search within eight classes with an angular search of 15°, and two classifications within 12 and 16 classes without alignment. A pre-A1 state containing the NAT10-NOL10 module and a post-A1 state containing DHX37 were discovered (fig. S3). Particles belonging to the pre-A1 or post-A1 state were merged separately and duplicate particles were removed. To further distinguish other possible states from

the pre-A1 state, a focused classification on NGDN was performed and resulted in the separation of pre-A1 and pre-A1* states, characterized by the presence or absence of NGDN. States pre-A1, pre-A1*, and post-A1 were refined in RELION 3.1 to global resolutions of 3.51, 3.87, and 2.65 Å, respectively (fig. S3).

To facilitate accurate atomic model building of the entire human SSU processome, high-resolution and high-quality cryo-EM reconstructions of the different states were essential. Focused 3D classifications followed by focused 3D refinements on 15 different regions of the pre-A1 and post-A1 states were performed in RELION 3.1. The focused maps were further improved using phenix.resolve_cryo_em (*54*), which helped considerably with the model building (figs. S4 to S6). Ultimately, the 15 focused maps for each state were postprocessed in RELION 3.1 and combined into composite maps using phenix.combine_focused_maps (table S1) (*55*). The same procedure was used to generate composite half-maps that were used to calculate the final overall resolution of the composite pre-A1 (3.60 Å) and post-A1 (2.69 Å) maps. This strategy was not pursued for the pre-A1* state owing to its low number of particles.

All reported resolutions are based on the gold standard FSC-0.143 criterion (*56*), and FSC-curves were corrected using high-resolution noise substitution methods (*57*) in RELION 3.1. All reconstructions were corrected for the modulation transfer function of the detector and sharpened by applying a negative B-factor that was estimated using automated procedures in RELION 3.1 (*53*). Local resolution for the overall, focused, and composite maps was calculated using Blocres (*58*) within cryoSPARC v3 (*59*) (figs. S7 to S9).

Model building and refinement

A combination of homology modeling using Modeller (*60*), a few existing x-ray structures, artificial intelligence (AI)–based neural network methods (*61*), and de novo model building was used to build the entire human SSU processome. Model building was first done for the highest resolution map, the post-A1 state solved at 2.65 Å. The core of the map reaches local resolution up to 2.5 Å, and only a few parts have a resolution lower than 3.5 Å (figs. S8 and S9). Most of the human ribosomal protein atomic coordinates were extracted from the SARS-CoV-2-Nsp1-40S EM structure (PDB ID 6ZOJ), rigid-body docked in our EM density and manually adjusted in COOT (*62*). X-ray structures of some proteins contained within the human U3 snoRNP and of very few assembly factors were available and also docked into our EM map with manual adjustment in COOT (*62*). For the UtpA and UtpB subcomplexes and the rest of the assembly factors, homology models were generated with MODELLER (*60*) using templates from yeast SSU processome structures (PDB ID 5WLC or 6ZQD) or de novo using trRosetta (*63*). These homology models, together with AI-based models generated with Deep Tracer (*61*), served as a starting point for manual adjustments and manual building in COOT (*62*). The post-A1 final model was then used to build the pre-A1 and pre-A1* models. Proteins specific to the post-A1 were removed, and assembly factors specific to the pre-A1 and pre-A1* states were added and manually curated with the same procedure as for the post-A1 state. Conformational changes between the different states of many assembly factors were also observed and manually adjusted in COOT (*62*).

A complete list of templates and crystal structures used to build the final pre-A1, pre-A1*, and post-A1 models can be found in tables S2 and S3. The final models for the three states were real-space refined using phenix.real_space_refine (*64*) with secondary structure restraints for proteins and RNAs. The model refinement statistics can be found in table S4. The maps and models were visualized in either Chimera (*65*), ChimeraX (*66*), or PyMOL (*67*). The figures were generated using ChimeraX (*66*).

Expression and purification of yeast Dhr1 for crystallization

Crystallization of both human DHX37 and yeast Dhr1 was attempted, but only yeast Dhr1 yielded crystals. Yeast Dhr1 (residues 379 to 1185) was PCR amplified from *Saccharomyces cerevisiae* genomic DNA and cloned into an RSF-Duet1 vector containing an N-terminal His-14 SUMO tag (H14-SUMO) using BglII and XhoI restriction sites. This vector was transformed into RIL cells for expression. Large-scale growth was done in 2x Yeast-Tryptone (2xYT) medium at 37°C until an optical density (OD) of 1. At this point, expression was induced by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), and growth was continued at 20°C. Twenty hours after induction, cells were harvested by centrifugation at 8000*g* and resuspended in buffer A (20 mM Tris/HCl pH 7.6, 500 mM NaCl, 0.1% Triton X-100) supplemented with protease inhibitors (E64, pepstatin, and PMSF) and DNase/RNase. Cells were lysed at 4°C by passing the cell suspension once through a Constant Systems cell disruptor at 30 kPSI. Lysates were centrifuged at 40,000*g* for 45 min, and clarified lysates were applied to a His-Trap HP column (GE Lifesciences). The column was washed with wash buffer (20 mM Tris/HCl pH 7.6, 30 mM imidazole, 500 mM NaCl) and then cleaved on-column with SUMO protease buffer (20 mM Tris/HCl pH 7.6, 500 mM NaCl, 15 μg/ml SUMO protease). Elutions were diluted to 125 mM NaCl with 20 mM Tris/HCl pH 7.6 and then loaded onto a HiTrap Heparin HP column (GE Lifesciences). The Heparin column was washed with two column volumes of 20 mM Tris/HCl pH 7.6 and 100 mM NaCl before being eluted with a salt gradient from 20 mM Tris/HCl pH 7.6, 100 mM NaCl to 20 mM Tris/HCl pH 7.6, 1 M NaCl. Fractions containing Dhr1 were pooled and concentrated before size exclusion chromatography on a HiLoad 16/600 Superdex 200 equilibrated with 20 mM Tris/HCl pH 7.6, 250 mM NaCl. Size exclusion fractions were concentrated to 8 mg/ml using a 30-kDa molecular weight cut-off Macrosep Advance centrifugal device (Pall Corporation).

Crystallization of yeast Dhr1 (residues 379 to 1185)

Purified yeast Dhr1 (residues 379 to 1185) at 8 mg/ml was supplemented with 1 mM adenosine 5′-diphosphate monopotassium salt and 1 mM magnesium chloride. Crystals were grown using the sitting drop vapor-diffusion method, where 1 μl of Dhr1 was mixed with 2 μl well solution containing 10.5% PEG 6000 and 0.1 M HEPES pH 7.0 at 4°C. Crystals were cryoprotected with a mix of 75% well solution and 25% PEG 400 before flash freezing in liquid nitrogen.

X-ray data collection and structure determination

X-ray data was collected from beam line 24-ID-E of the Advanced Photon Source and Argonne National Laboratories (Chicago, IL) at a wavelength of 0.97919 Å. The Dhr1 crystal was solved using molecular replacement with Prp43 subdomains (PDB ID 2XAU) as search models using PHENIX (*68*).

Model building and refinement

Model building was performed with COOT (*69*), and the structure was refined using PHENIX (*68*). Model statistics for the Dhr1 core are provided in table S5.

Recombinant human rDNA engineering

A full copy of the human rDNA sequence present on a BAC plasmid (clone JH4) (*70*) was extracted and cloned into an in-house modified pUC57 plasmid. The copy includes the Pol I promoter, 5′ ETS, 18*S*, ITS1, 5.8*S*, ITS2, 28*S*, 3′ ETS, and a shorter terminator comprising only five termination signals. By subcloning fusion-PCR amplicons into the rDNA, we were able to introduce distinct sequences into the expansion segments of the small (ES3S) and large subunits (ES15L). These sequences are not recognized within the human transcriptome and can be used to monitor the production of recombinant ribosomes in human cells transfected with the rDNA plasmid by Northern blot. To assess the role and requirement of different regions of the human 5′ ETS in the maturation of the small ribosomal subunit, we performed structure-based truncations into

specific regions of the 5′ ETS that are not immediately stabilized within the SSU processome (Fig. 2). We first removed the 5′ region from the Pol I promoter to the A′ site, just before the ECM (Δ1: 39 to 422). Then, the region between the two base-paring sites (U3 3′ and 5′ hinges) was removed (Δ2: 893 to 1382). Finally, the remaining the large helical segment between sites A0 and A1 was removed (Δ3: 1722 to 3546). We also engineered combinations of truncations (Δ1-Δ2, Δ2-Δ3, and Δ1-Δ2-Δ3). All the plasmids were fully sequenced, and no mutations were detected.

Cell transfection for functional studies

Three micrograms of each plasmid harboring the different truncations, as well as the wild-type rDNA plasmid, were reverse transfected into human HEK 293-F cells using Lipofectamine 3000 reagent (1 M cells per transfection per well in six-well plates). The cells were incubated at 37°C, 8% CO_2. After 48 hours, the cells were washed once with cold PBS, gently detached from the plate with cold PBS, and then pelleted at 100 g for 5 min. The pellets were flash frozen into liquid nitrogen and stored at −80°C until usage.

RNA extraction and Northern blot analysis

The RNA was extracted from the cells using Trizol/chloroform and isopropanol precipitation following standard procedures. Ten micrograms of total RNA from each sample was resolved on a 1.2% denaturing agarose gel. After complete migration, the RNA was transferred to a positively charged nylon membrane and fixed by ultraviolet cross-linking. The following 5′-^{32}P-labeled oligonucleotide probes were used for Northern blot analysis: ITS1 probe (5′-GTGAGCACGACGTCACCACATCGATCGAAG-ATC-3′), 18*S* probe (5′- CGAGGATCGAGGC-TTT -3′), and 28*S* probe (5′-AGGTACACTCGA-GAGCTTCA-3′). The Millenium RNA marker (ThermoFisher Scientific) was used as a molecular weight ladder.

Sequence alignments and secondary structure prediction

Sequence alignments were performed using Clustal Omega (*71*) and analyzed using Consurf (*72*). Secondary structure predictions were performed using PSIPRED (*73*).

REFERENCES AND NOTES

1. K. E. Bohnsack, M. T. Bohnsack, Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J.* **38**, e100278 (2019). doi: 10.15252/embj.2018100278; pmid: 31268599
2. D. L. J. Lafontaine, J. A. Riback, R. Bascetin, C. P. Brangwynne, The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021). doi: 10.1038/s41580-020-0272-6; pmid: 32873929
3. B. Michot, J. P. Bachellerie, Secondary structure of the 5′ external transcribed spacer of vertebrate pre-rRNA. Presence of phylogenetically conserved features. *Eur. J. Biochem.* **195**, 601–609 (1991). doi: 10.1111/j.1432-1033.1991.tb15743.x; pmid: 1999184
4. M. Hunziker *et al*., Conformational switches control early maturation of the eukaryotic small ribosomal subunit. *eLife* **8**, e45185 (2019). doi: 10.7554/eLife.45185; pmid: 31206356
5. R.-W. Yao *et al*., Nascent pre-rRNA sorting via phase separation drives the assembly of dense fibrillar components in the human nucleolus. *Mol. Cell* **76**, 767–783.e11 (2019). doi: 10.1016/j.molcel.2019.08.014; pmid: 31540874
6. F. Dragon *et al*., A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002). doi: 10.1038/nature00769; pmid: 12068309
7. M. Chaker-Margot, M. Hunziker, J. Barandun, B. D. Dill, S. Klinge, Stage-specific assembly events of the 6-MDa small-subunit processome initiate eukaryotic ribosome biogenesis. *Nat. Struct. Mol. Biol.* **22**, 920–923 (2015). doi: 10.1038/nsmb.3111; pmid: 26479197
8. L. Zhang, C. Wu, G. Cai, S. Chen, K. Ye, Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast. *Genes Dev.* **30**, 718–732 (2016). doi: 10.1101/gad.274688.115; pmid: 26980190
9. M. Kornprobst *et al*., Architecture of the 90S pre-ribosome: A structural view on the birth of the eukaryotic ribosome. *Cell* **166**, 380–393 (2016). doi: 10.1016/j.cell.2016.06.014; pmid: 27419870
10. M. Chaker-Margot, J. Barandun, M. Hunziker, S. Klinge, Architecture of the yeast small subunit processome. *Science* **355**, eaal1880 (2017). doi: 10.1126/science.aal1880; pmid: 27980088
11. J. Barandun *et al*., The complete structure of the small-subunit processome. *Nat. Struct. Mol. Biol.* **24**, 944–953 (2017). doi: 10.1038/nsmb.3472; pmid: 28945246
12. J. Cheng, N. Kellner, O. Berninghausen, E. Hurt, R. Beckmann, 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nat. Struct. Mol. Biol.* **24**, 954–964 (2017). doi: 10.1038/nsmb.3476; pmid: 28967883
13. J. Cheng *et al*., 90*S* pre-ribosome transformation into the primordial 40*S* subunit. *Science* **369**, 1470–1476 (2020). doi: 10.1126/science.abb4119; pmid: 32943521
14. Y. Du *et al*., Cryo-EM structure of 90S small ribosomal subunit precursors in transition states. *Science* **369**, 1477–1481 (2020). doi: 10.1126/science.aba9690; pmid: 32943522
15. B. Lau *et al*., Structure of the maturing 90S pre-ribosome in association with the RNA exosome. *Mol. Cell* **81**, 293–303.e4 (2021). doi: 10.1016/j.molcel.2020.11.009; pmid: 33326748
16. M. Ameismeier, J. Cheng, O. Berninghausen, R. Beckmann, Visualizing late states of human 40S ribosomal subunit maturation. *Nature* **558**, 249–253 (2018). doi: 10.1038/s41586-018-0193-0; pmid: 29875412
17. M. Ameismeier *et al*., Structural basis for the final steps of human 40S ribosome maturation. *Nature* **587**, 683–687 (2020). doi: 10.1038/s41586-020-2929-x; pmid: 33208940
18. L. Tafforeau *et al*., The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol. Cell* **51**, 539–551 (2013). doi: 10.1016/j.molcel.2013.08.011; pmid: 23973377
19. L. Badertscher *et al*., Genome-wide RNAi screening identifies protein modules required for 40S subunit synthesis in human cells. *Cell Rep.* **13**, 2879–2891 (2015). doi: 10.1016/j.celrep.2015.11.061; pmid: 26711351
20. K. I. Farley-Barnes *et al*., Diverse regulators of human ribosome biogenesis discovered by changes in nucleolar number. *Cell Rep.* **22**, 1923–1934 (2018). doi: 10.1016/j.celrep.2018.01.056; pmid: 29444442
21. S. B. Sondalle, S. J. Baserga, Human diseases of the SSU processome. *Biochim. Biophys. Acta* **1842**, 758–764 (2014). doi: 10.1016/j.bbadis.2013.11.004; pmid: 24240090
22. S. K. Natchiar, A. G. Myasnikov, H. Kratzat, I. Hazemann, B. P. Klaholz, Visualization of chemical modifications in the human 80S ribosome structure. *Nature* **551**, 472–477 (2017). doi: 10.1038/nature24482; pmid: 29143818
23. M. Taoka *et al*., Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46**, 9289–9298 (2018). doi: 10.1093/nar/gky811 pmid: 30202881
24. J. Cheng *et al*., Thermophile 90S pre-ribosome structures reveal the reverse order of co-transcriptional 18S rRNA subdomain integration. *Mol. Cell* **75**, 1256–1269.e7 (2019). doi: 10.1016/j.molcel.2019.06.032; pmid: 31378463
25. S. Kass, N. Craig, B. Sollner-Webb, Primary processing of mammalian rRNA involves two adjacent cleavages and is not species specific. *Mol. Cell. Biol.* **7**, 2891–2898 (1987). doi: 10.1128/MCB.7.8.2891; pmid: 3670298
26. N. J. Watkins, M. T. Bohnsack, The box C/D and H/ACA snoRNPs: Key players in the modification, processing and the dynamic folding of ribosomal RNA. *WIREs RNA* **3**, 397–414 (2012). doi: 10.1002/wrna.117; pmid: 22065625
27. Q. Sun *et al*., Molecular architecture of the 90S small subunit pre-ribosome. *eLife* **6**, e22086 (2017). doi: 10.7554/eLife.22086; pmid: 28244370
28. E. V. Wasmuth, C. D. Lima, The Rrp6 C-terminal domain binds RNA and activates the nuclear RNA exosome. *Nucleic Acids Res.* **45**, 846–860 (2017). doi: 10.1093/nar/gkw1152; pmid: 27899565
29. M. Thoms *et al*., The exosome is recruited to RNA substrates through specific adaptor proteins. *Cell* **162**, 1029–1038 (2015). doi: 10.1016/j.cell.2015.07.060; pmid: 26317469
30. F. M. Boneberg *et al*., Molecular mechanism of the RNA helicase DHX37 and its activation by UTP14A in ribosome biogenesis. *RNA* **25**, 685–701 (2019). doi: 10.1261/rna.069609.118; pmid: 30910870
31. A. Roychowdhury *et al*., The DEAH-box RNA helicase Dhr1 contains a remarkable carboxyl terminal domain essential for small ribosomal subunit biogenesis. *Nucleic Acids Res.* **47**, 7548–7563 (2019). doi: 10.1093/nar/gkz529; pmid: 31188444
32. J. Zhu, X. Liu, M. Anjos, C. C. Correll, A. W. Johnson, Utp14 recruits and activates the RNA helicase Dhr1 to undock U3 snoRNA from the preribosome. *Mol. Cell. Biol.* **36**, 965–978 (2016). doi: 10.1128/MCB.00773-15; pmid: 26729466
33. P. Choudhury, P. Hackert, I. Memet, K. E. Sloan, M. T. Bohnsack, The human RNA helicase DHX37 is required for release of the U3 snoRNP from pre-ribosomal particles. *RNA Biol.* **16**, 54–68 (2019). doi: 10.1080/15476286.2018.1556149; pmid: 30582406
34. G. R. Wells *et al*., The PIN domain endonuclease Utp24 cleaves pre-ribosomal RNA at two coupled sites in yeast and humans. *Nucleic Acids Res.* **44**, 5399–5409 (2016). doi: 10.1093/nar/gkw213; pmid: 27034467
35. K. I. Farley-Barnes, L. M. Ogawa, S. J. Baserga, Ribosomopathies: Old concepts, new controversies. *Trends Genet.* **35**, 754–767 (2019). doi: 10.1016/j.tig.2019.07.004; pmid: 31376929
36. J. C. Ulirsch *et al*., The genetic landscape of Diamond-Blackfan anemia. *Am. J. Hum. Genet.* **103**, 930–947 (2018). doi: 10.1016/j.ajhg.2018.10.027; pmid: 30503522
37. N. Draptchinskaia *et al*., The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat. Genet.* **21**, 169–175 (1999). doi: 10.1038/5951; pmid: 9988267
38. J. Flygare *et al*., Human *RPS19*, the gene mutated in Diamond-Blackfan anemia, encodes a ribosomal protein required for the maturation of 40S ribosomal subunits. *Blood* **109**, 980–986 (2007). doi: 10.1182/blood-2006-07-038232; pmid: 16990592
39. J. R. P. Knight, A. E. Willis, J. Milner, Active regulator of SIRT1 is required for ribosome biogenesis and function. *Nucleic Acids Res.* **41**, 4185–4197 (2013). doi: 10.1093/nar/gkt129; pmid: 23462953
40. N. Maeda, S. Toku, N. Kenmochi, T. Tanaka, A novel nucleolar protein interacts with ribosomal protein S19. *Biochem. Biophys. Res. Commun.* **339**, 41–46 (2006). doi: 10.1016/j.bbrc.2005.10.184; pmid: 16289379
41. B. L. Ebert *et al*., Identification of *RPS14* as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008). doi: 10.1038/nature06494; pmid: 18202658
42. N. A. Paolini *et al*., A ribosomopathy reveals decoding defective ribosomes driving human dysmorphism. *Am. J. Hum. Genet.* **100**, 506–522 (2017). doi: 10.1016/j.ajhg.2017.01.034; pmid: 28257692
43. A. G. Marneros, BMS1 is mutated in aplasia cutis congenita. *PLOS Genet.* **9**, e1003573 (2013). doi: 10.1371/journal.pgen.1003573; pmid: 23785305
44. J. Armistead *et al*., Mutation of a gene essential for ribosome biogenesis, *EMG1*, causes Bowen-Conradi syndrome. *Am. J. Hum. Genet.* **84**, 728–739 (2009). doi: 10.1016/j.ajhg.2009.04.017; pmid: 19463982
45. P. Chagnon *et al*., A missense mutation (R565W) in *Cirhin* (FLJ14728) in North American Indian childhood cirrhosis. *Am. J. Hum. Genet.* **71**, 1443–1449 (2002). doi: 10.1086/344580; pmid: 12417987
46. I. M. Slaymaker *et al*., Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016). doi: 10.1126/science.aad5227; pmid: 26628643
47. J.-P. Concordet, M. Haeussler, CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018). doi: 10.1093/nar/gky354; pmid: 29762716
48. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). doi: 10.1016/j.jsb.2005.07.007; pmid: 16182563

49. J. Zivanov *et al*., New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). doi: 10.7554/eLife.42166; pmid: 30412051
50. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
51. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
52.. T. Wagner *et al*., SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019). doi: 10.1038/s42003-019-0437-z; pmid: 31240256
53. J. Zivanov, T. Nakane, S. H. W. Scheres, Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in *RELION*-3.1. *IUCrJ* **7**, 253–267 (2020). doi: 10.1107/S2052252520000081; pmid: 32148853
54. T. C. Terwilliger, S. J. Ludtke, R. J. Read, P. D. Adams, P. V. Afonine, Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020). doi: 10.1038/s41592-020-0914-9; pmid: 32807957
55. D. Liebschner *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in *Phenix*. *Acta Crystallogr. D* **75**, 861–877 (2019). doi: 10.1107/S2059798319011471; pmid: 31588918
56. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003). doi: 10.1016/j.jmb.2003.07.013; pmid: 14568533
57. S. Chen *et al*., High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013). doi: 10.1016/j.ultramic.2013.06.004; pmid: 23872039
58. J. B. Heymann, Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. *Protein Sci.* **27**, 159–171 (2018). doi: 10.1002/pro.3293; pmid: 28891250
59. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
60. B. Webb, A. Sali, Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2199**, 239–255 (2021). doi: 10.1007/978-1-0716-0892-0_14; pmid: 33125654
61. J. Pfab, N. M. Phan, D. Si, DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017525118 (2021). doi: 10.1073/pnas.2017525118; pmid: 33361332
62. A. Casañal, B. Lohkamp, P. Emsley, Current developments in *Coot* for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020). doi: 10.1002/pro.3791; pmid: 31730249
63. J. Yang *et al*., Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020). doi: 10.1073/pnas.1914677117; pmid: 31896580
64. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **74**, 531–544 (2018). doi: 10.1107/S2059798318006551; pmid: 29872004
65. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
66. E. F. Pettersen *et al*., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021). doi: 10.1002/pro.3943; pmid: 32881101
67. The PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC, 2017).
68. P. D. Adams *et al*., R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
69. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
70. J.-H. Kim *et al*., Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. *Nucleic Acids Res.* **46**, 6712–6725 (2018). doi: 10.1093/nar/gky442; pmid: 29788454
71. F. Sievers *et al*., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011). doi: 10.1038/msb.2011.75; pmid: 21988835
72. H. Ashkenazy *et al*., ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016). doi: 10.1093/nar/gkw408; pmid: 27166375
73. D. W. A. Buchan, D. T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019). doi: 10.1093/nar/gkz297; pmid: 31251384

ACKNOWLEDGMENTS

We thank M. Ebrahim, J. Sotiris, and H. Ng for help with grid screening and data collection at the Evelyn Gruss Lipper Cryo-Electron Microscopy Resource Center at The Rockefeller University. We thank V. L. Larionov at the Center for Cancer Research (NCI-NIH) for sharing the BAC containing the human rDNA sequence. Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. We thank members of the Klinge laboratory for critical reading of this manuscript. **Funding:** A.V.B. is supported by an EMBO long-term fellowship (ALTF 711-2019). S.K. is supported by a NIH New Innovator Award (1DP2GM123459). **Author contributions:** S.S. and S.K. conceived of the study. S.S. and S.K. developed the human genome editing platform, S.S., A.V.B., and S.K. designed the experiments and analyzed the data. S.S. performed genome editing and prepared the samples of human SSU processomes for structural analyses with help from A.V.B. A.V.B. performed the cryo-EM analyses and model building of the human SSU processome with help from S.K. and S.S. A.V.B. and S.K. engineered the rDNA plasmids, and A.V.B. performed the functional assays. L.M. and M.C.-M. crystallized, determined the structure, and built an atomic model of the yeast Dhr1 core enzyme with help from S.K. All authors wrote and edited the manuscript. **Competing interests:** The Rockefeller University has filed a patent related to the described human genome editing platform, on which S.S. and S.K. are inventors. **Data and materials availability:** The raw unaligned multiframe movies and aligned micrographs have been deposited in the Electron Microscopy Public Image Archive (EMPIAR-10781).The cryo-EM maps and atomic models have been deposited in the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB): state pre-A1 (EMD-23936, EMD-23939, and PDB ID 7MQ8), state pre-A1* (EMD-23937 and PDB ID 7MQ9), and state post-A1 (EMD-23938, EMD-23940, and PDB ID 7MQA). All focused refinement maps have also been deposited in the EMDB (table S1). The atomic coordinates for the yeast Dhr1 core have been deposited to the Protein Data Bank under ID 7MQJ. Materials in the main text or the supplement are available from S.K. upon reasonable request.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abj5338
Figs. S1 to S22
Tables S1 to S5
References (*74*–*76*)
MDAR Reproducibility Checklist
Movies S1 to S6
Data S1

19 May 2021; accepted 6 August 2021
10.1126/science.abj5338

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Complement factor C1q mediates sleep spindle loss and epileptic spikes after mild brain injury

Stephanie S. Holden, Fiorella C. Grandi, Oumaima Aboubakr, Bryan Higashikubo, Frances S. Cho, Andrew H. Chang, Alejandro Osorio Forero, Allison R. Morningstar, Vidhu Mathur, Logan J. Kuhn, Poojan Suri, Sethu Sankaranarayanan, Yaisa Andrews-Zwilling, Andrea J. Tenner, Anita Luthi, Eleonora Aronica, M. Ryan Corces, Ted Yednock, Jeanne T. Paz*

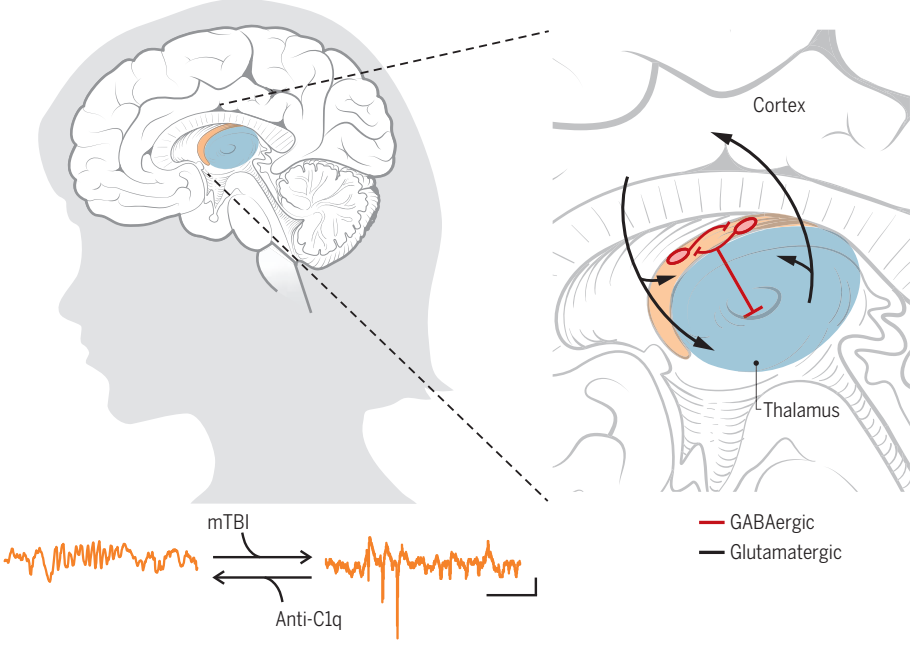
INTRODUCTION: Traumatic brain injury (TBI) is a leading cause of disability in children and adults. TBI affects 69 million people worldwide yearly and can lead to cognitive dysfunction, difficulty with sensory processing, sleep disruption, and epilepsy. Most of these adverse outcomes manifest months or years after TBI and are caused by indirect secondary injuries that develop as consequences of the initial impact. Because the primary injury is essentially irreversible, understanding where, when, and how secondary injuries develop is crucial for preventing or treating disability after TBI.

RATIONALE: Although the cortex is the primary site of injury, the thalamus experiences secondary injury after a TBI, presumably because of its intimate long-range reciprocal connections with the cerebral cortex. These secondary injuries manifest as structural changes that have been implicated in a number of long-term TBI-related health outcomes and as chronic neurodegeneration and inflammation.

Chronic neuroinflammation is common at secondary injury sites. However, most attempts to improve post-TBI cognitive outcomes with broad anti-inflammatory agents have failed, perhaps because inflammation leads to both damage and repair of neuronal circuits. A trigger of inflammation often observed at injury sites is the immune molecule C1q, the initiator of the classical complement cascade. We investigated the role of C1q in post-TBI impairment of the corticothalamic system's function, with particular emphasis on the timing and location of C1q expression. We used a mouse model of mild TBI, which does not acutely affect subcortical structures. Mild TBI cases are ~10 times more common than severe TBI cases and can still lead to chronic health issues such as impaired cognition and sleep. By using a TBI model that only directly affects the cortex, we can also attribute effects on subcortical structures to chronic injury processes rather than the initial impact, which provides a greater window of opportunity for treatment. We monitored neurophysiological changes in various regions of the cortex and thalamus using cellular electrophysiology and wireless cortical recordings in freely behaving mice. An antibody specifically directed against the functional head groups of C1q gave us the opportunity to test the impact of C1q on the structure and neurophysiological functions of the thalamus after TBI.

RESULTS: We found a chronic increase in C1q expression specifically in the cortex surrounding the injury site, and in the functionally connected subcortical thalamus which was not directly affected by the injury. C1q did not appear to originate from the thalamic neurons. Instead, single-nucleus RNA sequencing demonstrated that microglia are a source of thalamic C1q. Similarly, C4, which acts downstream of C1 in the classical complement pathway, was specifically up-regulated in thalamic oligodendrocytes. Increased C1q and C4 expression colocalized with areas of neuron loss and chronic inflammation, and correlated with disruption of sleep spindles, the electroencephalographic hallmarks of early stages of sleep, and the development of epileptic activities. Blocking C1q counteracted all of these outcomes. Neurons in the GABAergic reticular thalamus appeared uniquely sensitive to secondary damage linked to C1q, unlike cortical neurons or the excitatory thalamocortical neurons, which recovered at a chronic time point without C1q inhibition.

CONCLUSION: The GABAergic reticular thalamus is a vulnerable node of circuit dysfunction after mild TBI. The immune molecule C1q marks the secondary injury and is a disease modifier that could be targeted, at the proper time and location, to prevent secondary neuronal loss and the transformation of sleep spindles into epileptic spikes that occur after a mild TBI. Because manipulations of the complement pathway can also improve the outcomes of severe TBI, the complement pathway could become a therapy target for both types of trauma. ■



Complement factor C1q and reticular thalamus mediate chronic brain dysfunction after mild TBI. Left: A mild cortical TBI induces secondary damage to the thalamus (red and blue). Bottom: An anti-C1q antibody reverses the transformation of sleep spindles into epileptic spikes in a mouse model of mild TBI. Scale bar, 150 μV, 0.5 s. Right: The GABAergic reticular thalamus (red) is the vulnerable node of the cortico-thalamo-cortical loop.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jeanne.paz@gladstone.ucsf.edu
Cite this article as S. S. Holden *et al.*, *Science* 373, eabj2685 (2021). DOI: 10.1126/science.abj2685

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj2685

RESEARCH ARTICLE

NEUROSCIENCE

Complement factor C1q mediates sleep spindle loss and epileptic spikes after mild brain injury

Stephanie S. Holden[1,2,3], Fiorella C. Grandi[2,3], Oumaima Aboubakr[2], Bryan Higashikubo[2], Frances S. Cho[1,2,3], Andrew H. Chang[2], Alejandro Osorio Forero[4], Allison R. Morningstar[2], Vidhu Mathur[5], Logan J. Kuhn[5], Poojan Suri[5], Sethu Sankaranarayanan[5], Yaisa Andrews-Zwilling[5], Andrea J. Tenner[6], Anita Luthi[4], Eleonora Aronica[7,8], M. Ryan Corces[2,3], Ted Yednock[5], Jeanne T. Paz[1,2,3,9]*

Although traumatic brain injury (TBI) acutely disrupts the cortex, most TBI-related disabilities reflect secondary injuries that accrue over time. The thalamus is a likely site of secondary damage because of its reciprocal connections with the cortex. Using a mouse model of mild TBI (mTBI), we found a chronic increase in C1q expression specifically in the corticothalamic system. Increased C1q expression colocalized with neuron loss and chronic inflammation and correlated with disruption in sleep spindles and emergence of epileptic activities. Blocking C1q counteracted these outcomes, suggesting that C1q is a disease modifier in mTBI. Single-nucleus RNA sequencing demonstrated that microglia are a source of thalamic C1q. The corticothalamic circuit could thus be a new target for treating TBI-related disabilities.

Traumatic brain injury (TBI) is a leading cause of disability in children and adults (*1*). TBI affects 69 million people worldwide yearly (*2*) and can lead to cognitive dysfunction, difficulty with sensory processing, sleep disruption, and epilepsy. Most of these adverse outcomes manifest months or years after TBI and are caused by indirect secondary injuries that develop as consequences of the initial impact (*3*). Because the primary injury is essentially irreversible, understanding where, when, and how secondary injuries develop is crucial for preventing or treating disability after TBI.

The cortex is often the site of primary injury because it sits directly beneath the skull. However, the cortex belongs to many larger circuits, in particular the cortico-thalamo-cortical loop. This circuit is important for sensory processing, attention, cognition, and sleep (*3*–*5*). The thalamus itself, although not acutely injured in TBI, experiences secondary injury, presumably because of its long-range reciprocal connections with the cerebral cortex (*6*–*10*). Structural changes in the thalamus have been implicated in a number of long-term TBI-related health outcomes (*11*, *12*), and patients with TBI display secondary and chronic neurodegeneration and inflammation in thalamic nuclei (*13*, *14*).

Chronic neuroinflammation is common at secondary injury sites (*15*), but most attempts to improve post-TBI cognitive outcomes with broad anti-inflammatory agents have failed (*15*, *16*). One potential mediator of post-TBI inflammation and injury is the complement pathway, which is activated in the peri-injury area of brain lesions (*17*–*19*). Complement activation contributes to inflammation and neurotoxicity in central nervous system (CNS) injury and is increased in human brains afflicted with injury, epilepsy, and Alzheimer's disease (*19*–*23*). Aberrant activation of C1, the initiating molecule of the classical complement cascade, can trigger elimination of synapses and contribute to the progression of neurodegenerative disease (*24*). Conversely, C1/C1q is involved in normal synapse pruning during development (*25*), and the complement system plays an important part in brain homeostasis by clearing cellular debris and protecting the CNS from infection (*20*).

We investigated the role of C1q in post-TBI impairment of the corticothalamic system's function, with particular emphasis on the timing and location of C1q expression. We used a mouse model of mild TBI (mTBI), which does not acutely affect subcortical structures, and monitored neurophysiological changes in various regions of the cortex and thalamus using cellular electrophysiology and wireless cortical recordings in freely behaving mice.

[1]Neurosciences Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA. [2]Gladstone Institute of Neurological Disease, Gladstone Institutes, San Francisco, CA 94158, USA. [3]Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, USA. [4]Department of Fundamental Neurosciences, University of Lausanne, CH-1005 Lausanne, Switzerland. [5]Annexon Biosciences, South San Francisco, CA 94080, USA. [6]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697, USA. [7]Department of Neuropathology, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, Netherlands. [8]Stichting Epilepsie Instellingen Nederland (SEIN), 2103 SW Heemstede, Netherlands. [9]The Kavli Institute for Fundamental Neuroscience, and The Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA.
*Corresponding author. Email: jeanne.paz@gladstone.ucsf.edu

Secondary C1q expression coincides with chronic inflammation, neurodegeneration, and synaptic dysfunction in the thalamus

We induced a mild impact injury to the right primary somatosensory cortex (S1) of adult mice (Fig. 1A) and assessed its effect on the brain 3 weeks later. This period corresponds to the latent phase in humans, when the brain is undergoing adaptive and maladaptive changes after injury (*26*). We stained coronal brain sections with markers of neurons (NeuN) and of glial inflammation (C1q, classical complement pathway; GFAP, astrocytes; IBA1, microglia/macrophages) (Fig. 1, C to E). Three weeks after surgery, mTBI mice had significantly higher GFAP, C1q, and IBA1 expression in the peri-TBI S1 cortex and the functionally connected ventrobasal thalamus (VB) and reticular thalamic nucleus (nRT) than did sham mice (Fig. 1, B to E). We also saw increased expression of similar inflammatory markers in thalamic tissue from human TBI patients, confirming that thalamic inflammation is also a consequence of TBI in humans (fig. S1).

Glial inflammation was associated with significant neuronal loss in the thalamic region, particularly in the nRT (Figs. 1, D and E, and 2A). The nRT of mTBI mice had significantly fewer neurons than the nRT of sham mice, particularly in the "body" region, which receives most of its excitatory inputs from the injured somatosensory cortex (*27*–*31*) (Fig. 2, B and C).

To test whether C1q might mark functional damage in the cortex and thalamus, we performed whole-cell patch-clamp recordings in brain slices obtained 3 to 6 weeks after injury. We recorded layer 5 pyramidal neurons, a layer with neurons known to project to both relay thalamus and nRT (*32*), and fast-spiking GABAergic interneurons in the peri-TBI S1 cortex, glutamatergic neurons in the relay VB thalamus, and GABAergic neurons in the nRT. The neurons' intrinsic membrane electrical properties and the spontaneous excitatory and inhibitory postsynaptic current (sEPSC and sIPSC, respectively) properties were similar between sham and mTBI mice in both the peri-TBI cortex and the VB thalamus (table S1). However, in the nRT, mTBI led to a reduction in the frequency of sIPSCs (Fig. 2, D and E). Furthermore, nRT sEPSCs were smaller in amplitude and trended toward a lower frequency (Fig. 2, F and G). Immunofluorescence staining for green fluorescent protein in mice expressing Thy1-GCaMP6f, a marker of neuronal calcium levels in corticothalamic neurons, revealed reduced fluorescence in the thalamus after mTBI (Fig. 2, H and I).

We conclude that the major long-term effect of mTBI involves disruption of synaptic transmission in the nRT, which coincides with increased C1q expression, reduced cortical inputs, and local neuronal loss. By contrast, neurons in

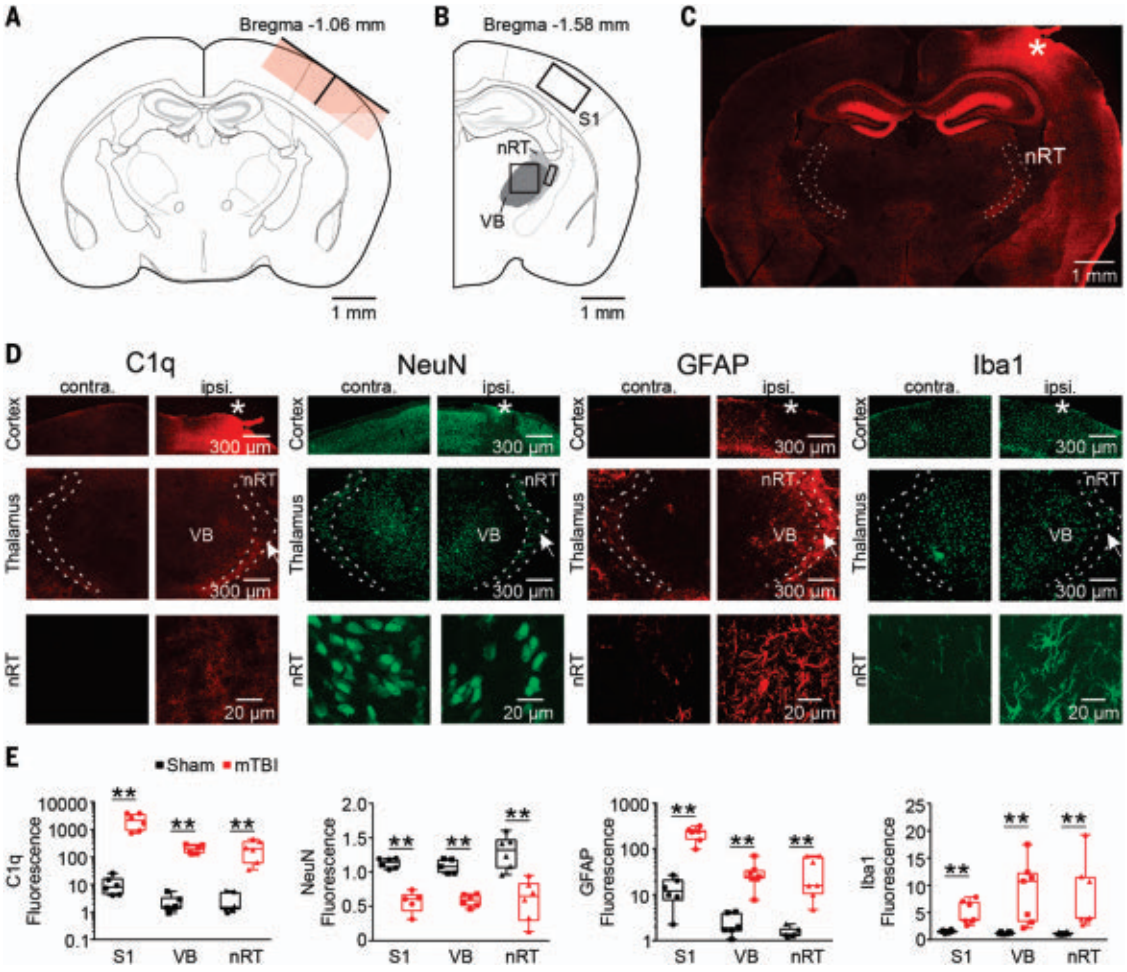


Fig. 1. The injured cortex and functionally connected thalamus show chronic inflammation and neuron loss 3 weeks after mTBI. (**A** and **B**) Schematic of the controlled cortical impact (A) and of the S1 cortex and nRT and VB thalamic regions (B). (**C**) Coronal brain section from a mTBI mouse stained for C1q. Note that bilateral C1q expression is typical for the hippocampus. (**D**) Close-up images of S1 (top), VB and nRT (middle), and confocal images of nRT (bottom) stained for C1q, NeuN, GFAP, and Iba1. Injury site in the right S1 cortex is marked by an asterisk. Arrow in nRT indicates location of confocal image. (**E**) Fluorescence ratios between ipsilateral and contralateral regions in sham and mTBI mice. Data represent all points from minimum to maximum analyzed with a Mann-Whitney test; α = 0.05 (*P < 0.05, **P < 0.01). Analysis includes between five and seven mice per group (three sections per mouse, one image per region).

the peri-TBI cortex and the glutamatergic relay VB thalamus appear normal at chronic stages after mTBI (table S1), suggesting that neuronal vulnerability is specific to the GABAergic nRT.

Microglia are a source of C1q in the thalamus

To determine the cellular origin of C1q in the thalamus, we microdissected nRT and VB thalamic tissue 3 weeks after injury (Fig. 3A) and performed single-nucleus RNA sequencing (snRNA-seq) on 6228 nuclei from sham mice and 5220 nuclei from mTBI mice. After correcting for ambient RNA (*33*) and removing potential doublets (*34*) or nuclei with a high percentage of mitochondrial reads, clustering analysis identified the expected cell types, including microglia (*Cx3cr1*, *P2ry12*), astrocytes (*Cldn10*, *Fgfr3*), oligodendrocytes (*Mobp*, *Olig1*), oligodendrocyte progenitors (*Sox8*, *Pdgfra*), GABAergic neurons (*Gad1*, *Gad2*), and glutamatergic neurons (*Slc17a6*, *Slc17a7*), which originated from adjacent thalamocortical relay nuclei (Fig. 3B and fig. S2A). The cellular composition was similar between sham and mTBI samples (fig. S2, B and C), and we detected a few markers of microglial (*Apoe*, *Cst3*) and astrocytic (*Apoe*, *Clu*) activation after mTBI (fig. S3A).

Microglia expressed high levels of *C1qa*, *C1qb*, and *C1qc*, the three genes that together encode the 18 subunits of C1q (*35*) (Fig. 3C and fig. S3B). In contrast to our findings at the protein level, the expression of the *C1q* genes in nuclear RNA was not significantly different between mTBI and sham samples (Fig. 3D and fig. S3B). Because this was consistent with previous reports that microglia activation is encoded in cytoplasmic rather than nuclear RNA (*36*), we examined *C1qa* mRNA in the bulk cytoplasmic fractions of our nuclei preparations using quantitative reverse transcription polymerase chain reaction (qRT-PCR). This analysis showed a trend toward increased *C1qa* mRNA expression after mTBI in both the thalamus and the cortex (fig. S3C). Similarly, mature oligodendrocytes and astrocytes in both sham and mTBI mice expressed *C4b*, which acts downstream of C1 in the classical complement pathway (Fig. 3E and fig. S3C). *C4b* expression in nuclear RNA increased 5.2-fold in one subcluster of oligodendrocytes after mTBI (Fig. 3F and fig. S3, E to G) but did not significantly increase in astrocytes. Cytoplasmic bulk qRT-PCR confirmed a trend toward increased *C4b* and *C2* expression (fig. S3G), whereas transcripts for other components of the complement pathway, such as *C1ra*, *C1s1*, and *Hc* (C5), were not detected by snRNA-seq or cytoplasmic qRT-PCR (fig. S3, B and I). These data show that microglia are a source of C1q in the thalamus, although we cannot exclude the possibility that other cell types such as astrocytes

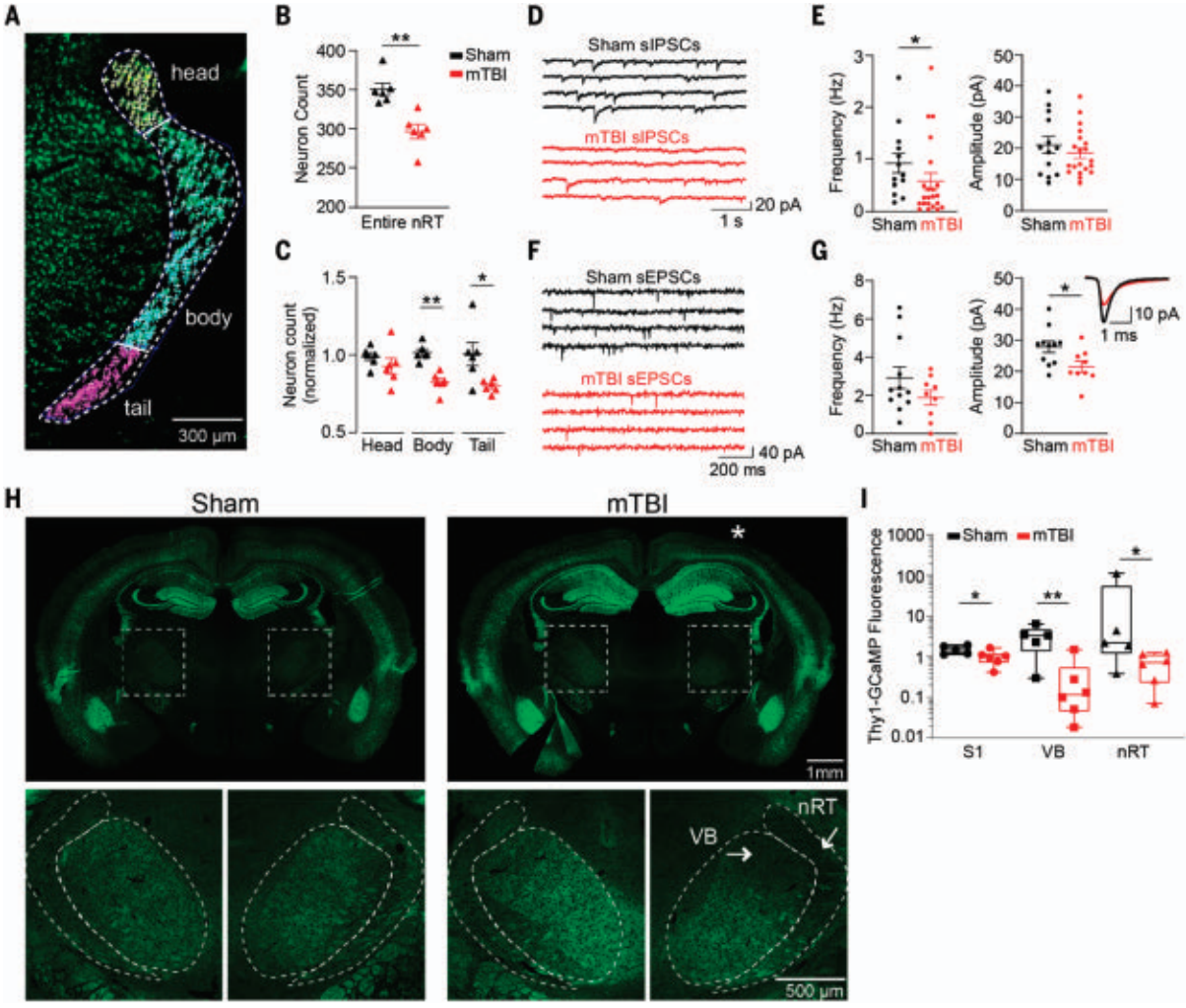


Fig. 2. The nRT ipsilateral to the injured cortex shows neuron loss and altered sIPSC and sEPSC properties 3 weeks after mTBI. (**A** to **C**) High-magnification coronal image of the nRT (A) and neuron counts across the entire ipsilateral nRT (B) or per subdivision, normalized to the median value from the sham group (C). Six mice per group (three sections per mouse, averaged) were studied. (**D** and **E**) sIPSC recordings from nRT neurons (D) and frequency and amplitude distributions (E) in 13 neurons from four sham mice and 22 neurons from six mTBI mice. (**F** and **G**) sEPSC recordings from nRT neurons (F) and frequency and amplitude distributions (G) in 11 neurons from six sham mice and nine neurons from seven mTBI mice. Inset: averaged EPSC traces from single nRT neurons. (**H**) Coronal brain sections from Thy1-GCaMP6f mice with sham surgery and mTBI (asterisk). Bottom: projection terminals from the cortex. (**I**) Thy1-GCaMP fluorescence ratios between ipsilateral and contralateral regions. Data represent all points from minimum to maximum, mean ± SEM analyzed with a Mann-Whitney test; $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$). Analysis includes five sham mice and six mTBI mice (three sections per mouse, one image per region).

and neurons could be additional sources (*37*–*39*). None of the GABAergic neurons showed differential expression of components of the complement pathway between mTBI and sham (fig. S4). Rather, these neurons up-regulated several genes related to mitochondrial function and oxidative phosphorylation, including *Cox6c* and *Cox5a* after mTBI (fig. S4G and table S2).

Blocking C1q function reduces chronic glial inflammation and neuron loss

Increased C1q expression was chronic (fig. S5, A and B) and might therefore explain the long-term effects of mTBI. To test this hypothesis, we used an antibody that specifically binds to C1q and blocks its downstream activity (*40*). Mice were given intraperitoneal injections of the C1q antibody or a mouse IgG1 isotype control 24 hours after mTBI or sham surgery, followed by twice-weekly treatments for 3 weeks. mTBI mice treated with the anti-C1q antibody showed a strong reduction in inflammation and reduced neuronal loss (Fig. 4, A to C) relative to control-treated mTBI mice, as monitored by immunofluorescent staining. On average, they had the same number of nRT neurons as antibody-treated sham mice (Fig. 4C), whereas mTBI mice treated with the control IgG still showed inflammation and neuron loss 3 weeks after mTBI. As an alternative approach to the antibody treatment, we repeated the study using $C1q^{-/-}$ mice and found that they too exhibited reduced chronic inflammation and a nonsignificant NeuN loss in the nRT after TBI (fig. S6). Measurements of anti-C1q and C1q in the brain and the plasma confirmed that the antibody exerted its effect in the brain rather than peripherally (fig. S7).

mTBI leads to loss of sleep spindles and increased epileptic spikes that are prevented by anti-C1q treatment

We investigated the impact of mTBI in vivo using brain rhythms as a readout of cortico-thalamic circuit function. We implanted chronic wireless electrocorticographic (ECoG) devices into sham and mTBI mice during the mTBI induction surgery and analyzed changes in ECoG rhythms within a 12-hour window 3 weeks after surgery. We focused our analysis on sleep

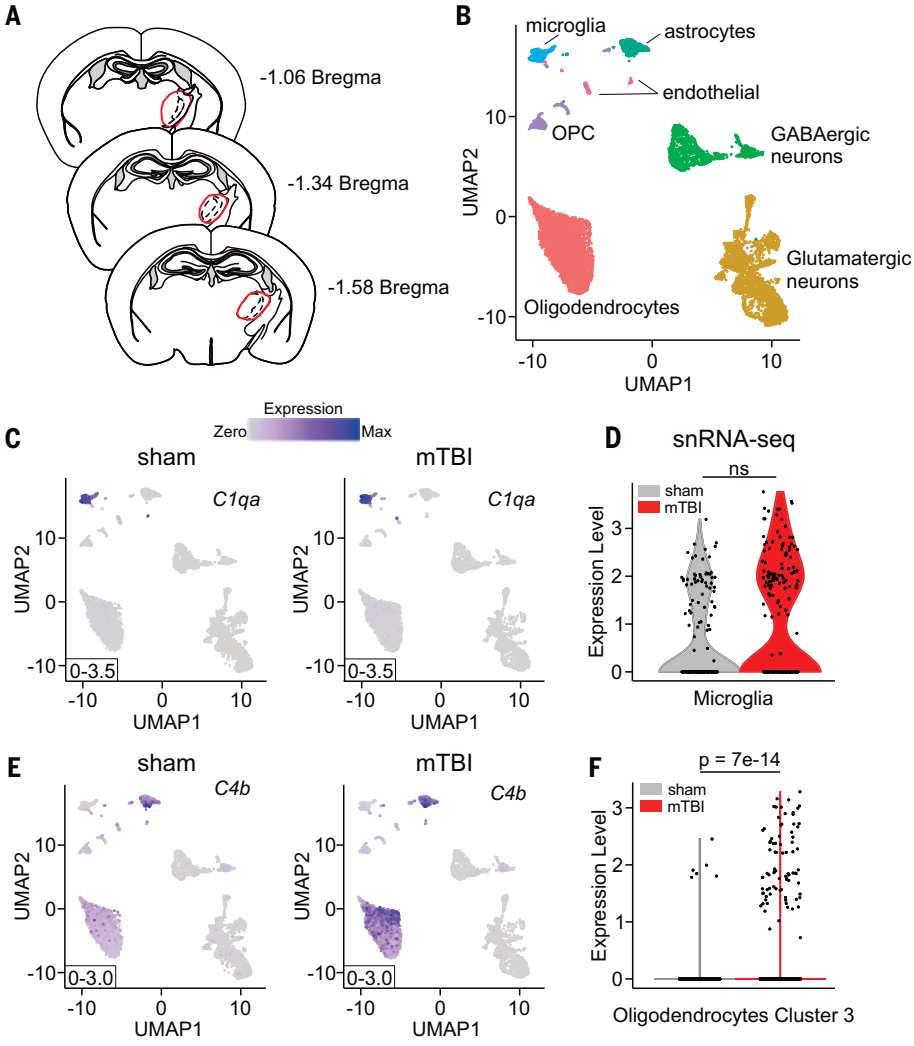


Fig. 3. Microglia are a source of C1q in the thalamus 3 weeks after mTBI. (**A**) Schematic of coronal brain sections showing the location of thalamic tissue dissection. (**B**, **C**, and **E**) Uniform manifold approximation and projection (UMAP) representation of snRNA-seq data (n = 4908 sham nuclei, n = 4338 mTBI nuclei, after data cleaning) colored by cell type or lineage (B), and for nuclear *C1qa* (C) or *C4b* (E) expression. Lineage markers are described in fig. S2A. Normalized expression scale shown above, zero to maximum, with a maximum value for each panel. (**D** and **F**) Violin plots of *C1qa* expression in microglial nuclei (D) and of *C4b* expression in oligodendrocyte nuclei (F) from cluster 3 (Oligo 3; fig. S3E) from sham and mTBI mice analyzed with a Wilcoxon rank sum test. ns, not significant. Analysis combines both technical replicates, collectively representing nine sham mice and 10 mTBI mice. Each dot represents a single nucleus.

spindles, which originate from the nRT during non–rapid eye movement sleep in mice (*41*). Sham mice had similar numbers of sleep spindles in the left and right sensory cortices. In mTBI mice, however, the cortex ipsilateral to injury had fewer sleep spindles than the contralateral cortex (Fig. 5, A to D). mTBI mice also had focal epileptic spikes ipsilateral to the injury (Fig. 6, A to D). After 3 weeks of antibody treatment, mTBI mice receiving the anti-C1q antibody had normal numbers of sleep spindles (Fig. 5, B, C, and D) and fewer epileptic spikes than mTBI mice treated with the isotype control (Fig. 6, B to E).

Discussion

Using electrophysiological approaches at the cellular and circuit levels, we found that mTBI in mice altered the synaptic properties of nRT neurons specifically and was associated with C1q accumulation. The ensuing sleep disruption and development of epileptic spikes were prevented by blocking the C1q pathway with an antibody that has translational potential.

Previous observations of severe head injury show neurodegeneration in the human nRT (*42*). Our studies showed that even mild cortical injury can lead to neuronal loss in the nRT 3 weeks after the injury. The loss of neurons in the nRT could explain some of the synaptic changes that we observed in this area. For instance, the reduced frequency of IPSCs 3 weeks after mTBI could reflect a deficit in GABAergic neurons. Loss of GABAergic inhibition in the nRT is known to result in corticothalamic circuit hyperexcitability (*43*, *44*), and intra-nRT GABAergic connections are important for coordinating inhibitory output to the excitatory thalamic nuclei and for controlling oscillatory thalamic activity (*45*). We also observed deficits in nRT EPSCs, particularly a lower amplitude, which is consistent with a loss of the excitatory cortical inputs. A similar alteration was found in a mouse model of epilepsy that lacks GluA4 AMPA receptors at the cortico-nRT glutamatergic synapse (*46*). This defect results in loss of feed-forward inhibition in the thalamus and epileptic activities (*46*). We therefore propose that alterations to the nRT EPSCs also contribute to corticothalamic circuit hypersynchrony and seizures.

Unlike nRT neurons, cortical neurons, such as excitatory layer 5 pyramidal neurons and GABAergic fast-spiking interneurons, and the

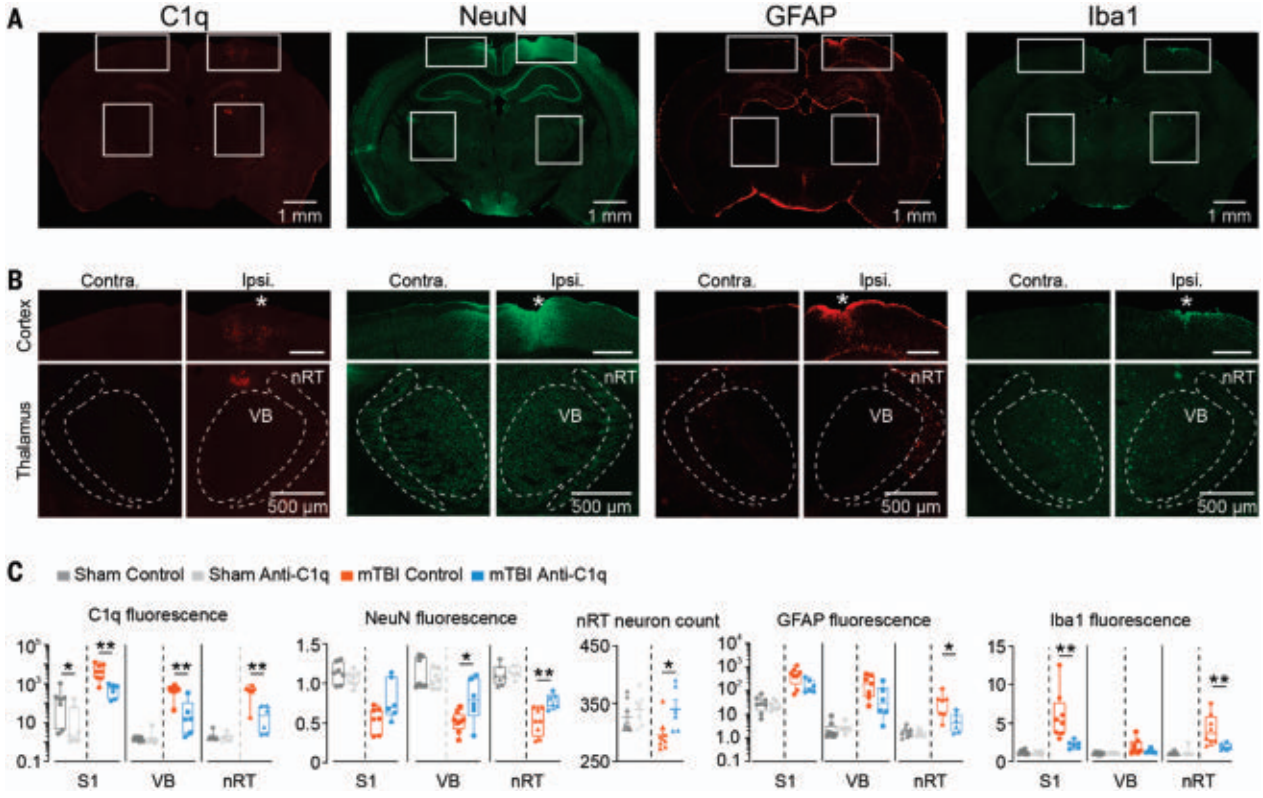


Fig. 4. Anti-C1q antibody reduces chronic inflammation and neuron loss 3 weeks after mTBI. (**A** and **B**) Coronal brain sections (A) and close-ups (B) of S1 (top) and VB and nRT (bottom) from mTBI mice treated with anti-C1q antibody and stained for C1q, NeuN, GFAP, and Iba1. Asterisk marks injury site. (**C**) Quantification of nRT neuron counts and fluorescence ratios between ipsilateral and contralateral regions in sham and mTBI mice treated with the anti-C1q antibody or the isotype control. Data represent all points from minimum to maximum analyzed with a Mann-Whitney test; $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$). Analysis includes between six and eight mice per group (three sections per mouse, one image per region).

excitatory relay thalamic neurons were not altered by mTBI at chronic time points. These observations suggest the presence of homeostatic mechanisms that reduce chronic hyperexcitability after mTBI in the cortex and relay thalamus. They also confirm that at least some long-term outcomes of mTBI must result from nRT dysfunction rather than simply from damage to the cortex. Nevertheless, ECoG did reveal some local cortical deficits in sleep spindles and epileptic spikes. The disruption in sleep spindles has been reported in rats and humans with severe TBI (*5*, *47*) and is more frequent in humans with mTBI than in those with severe TBI (*48*).

Given the emerging role of the nRT in generating local sleep spindles in the cortex (*41*), we speculate that the local loss of sleep spindles in the cortex results from the secondary damage to the nRT.

C1q has well-documented roles in normal brain function, such as synaptic pruning during development (*25*), forgetting memories in adulthood (*49*), and in several neurological disorders, including severe TBI (*17*–*24*, *50*). We found that even after mTBI, C1q expression was highly increased in the thalamus for up to 4 months after injury. Because our mice with mTBI did not develop chronic generalized tonic-clonic seizures, we could not determine whether blocking C1q could prevent full-blown seizures. However, we observed many other protective effects of the anti-C1q antibody, including reduced inflammation and neurodegeneration and protection against sleep spindle disruption and epileptic spikes. A maladaptive role of C1q, including aberrant synapse loss, has been reported in an in vivo mouse model of Alzheimer's disease (*22*), although other studies have shown that C1q can also play a "beneficial" role in Alzheimer's disease by preventing the neurotoxicity caused by amyloid-β (*51*, *52*). C1q's maladaptive role in mTBI may seem paradoxical in light of its beneficial role in neuronal development and adaptive synapse pruning (*25*, *49*). However, there is evidence that different inflammatory types of astrocytes or microglia can be protective or harmful in the brain (*53*, *54*), and this concept could extend to C1q as well. Indeed, C1q levels also increased in the cortex and relay thalamus but did not appear to have a damaging role at these sites because, unlike in the nRT, their neuronal physiology returned to normal at chronic time points.

Our snRNA-seq results suggest that microglia, rather than the neurons themselves, are a source of C1q, whereas astrocytes and oligodendrocytes of C4. C4 appears to mediate injury after severe TBI, as shown by reduced motor deficits in *C4*$^{-/-}$ mice (*18*), and is a genetic risk factor in schizophrenia, in which sleep spindle loss is a major electrophysiological feature (*55*, *56*). C4 overexpression in mice increases microglia engulfment of synapses (*57*), which provides a potential mechanism linking C1q/C4 to nRT dysfunction. Thus, activation of C1q/C4 in thalamic glial cells could underlie the nRT vulnerability leading to the transformation of sleep spindles into epileptic spikes. Lack of C5 in our model suggests that nRT dysfunction is not caused by C5a-C5aR1 microglial polarization or direct local (nRT) neuronal damage caused by C5a or the membrane attack complex, mechanisms that require C5 expression.

The fact that GABAergic nRT neurons overexpress *Cox6c* and *Cox5a*, which are mitochondrial indicators of disrupted neuronal homeostasis and cell death (*58*, *59*), 3 weeks after injury indicates that nRT neurons are at least dysfunctional and perhaps still dying at that chronic stage. Because anti-C1q treatment

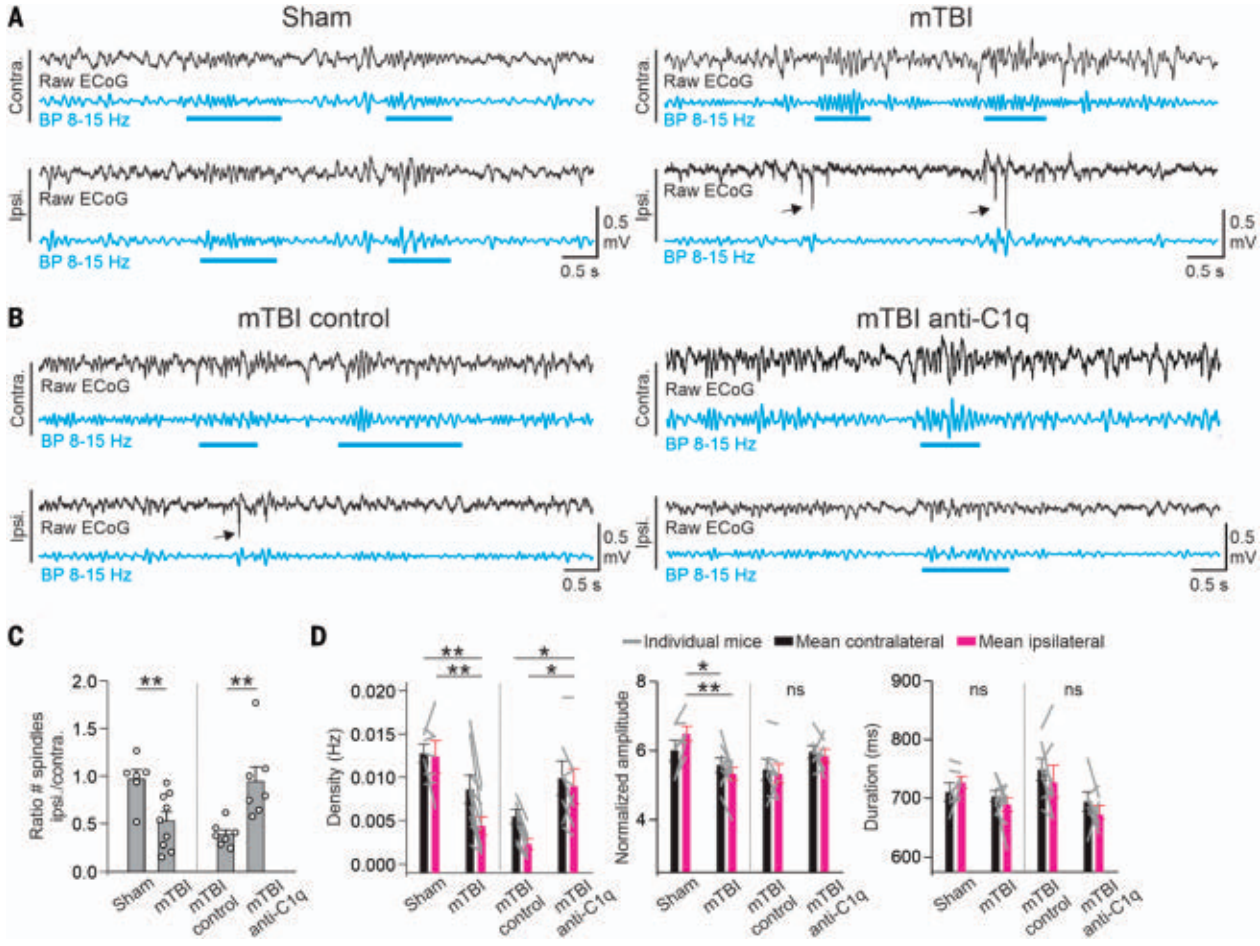


Fig. 5. Anti-C1q antibody reverts chronic sleep spindle reduction after mTBI. (**A**) ECoG recordings from a sham and mTBI mouse 3 weeks after mTBI. Blue traces represent the bandpass-filtered (BP, 8 to 15 Hz) ECoG. Horizontal blue lines show the detected spindles. Arrows indicate epileptic spikes. (**B**) Same as (A) from mTBI mice treated with an isotype control or the anti-C1q antibody. (**C**) Ratio of sleep spindle counts in ipsilateral ECoG to sleep spindle counts in contralateral ECoG detected within a 12-hour window. Data represent mean ± SEM analyzed with a Mann-Whitney rank sum test; $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$). Analysis included six sham mice and nine mTBI mice (left) and seven control-treated mTBI mice and six antibody-treated mTBI mice (right). (**D**) Density, normalized amplitude, and duration of sleep spindles in contralateral and ipsilateral ECoG from the mice in (C). Data represent mean ± SEM analyzed with a Kruskal-Wallis one-way analysis of variance on ranks, all pairwise multiple-comparison procedures (Holm-Sidak method); $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$). Gray lines connect contralateral and ipsilateral data points for each mouse.

reduces nRT neuronal loss, we speculate that sometime after injury, C1q initiates a cycle of complement activation, including C4 production by oligodendrocytes, inflammation, and neuronal loss. This cycle adds to the damage presumably caused by the initial loss of cortical input and maintains the nRT's dysfunctional state, leading to the loss of sleep spindles and a hyperactive thalamic output to the cortex. What makes the nRT uniquely vulnerable to this cycle remains unclear.

In summary, our data point to the complement cascade as a disease modifier that could be targeted after injury (in this study, starting 24 hours later) to avoid the devastating long-term outcomes of mTBI. This prospect is a possibility because anti-C1q treatment is already being used in a clinical trial for another neurological disorder (*40*, *60*), and there is an expanding pursuit of complement therapeutics for CNS disorders (*61*).

Genetic or pharmacological manipulations of the complement pathway can also improve the outcomes of severe TBI, specifically motor function and memory deficits in mice (*17*, *18*, *62*, *63*). The complement pathway could therefore become a therapy target for both types of trauma.

Materials and Methods

Animals

All of our protocols were approved by the institutional animal care and use committees at the University of California, San Francisco, and Gladstone Institutes. Precautions were taken to minimize stress and the number of animals used in each set of experiments. Mice were separately housed after surgical implants.

Controlled cortical impact

We anesthetized mice with 2 to 5% isoflurane and placed them in a stereotaxic frame. We performed a 3-mm craniotomy over the right S1 centered at −1 mm posterior from bregma, +3 mm lateral from the midline. TBI was performed with a CCI device (Impact One Stereotaxic Impactor for CCI, Leica Microsystems) equipped with a metal piston using the following parameters: 3 mm tip diameter, 15° angle, 0.8 mm depth from the dura, 3 m/s velocity, and 100 ms dwell time. Sham animals received identical anesthesia and craniotomy, but the injury was not delivered.

Patch-clamp electrophysiology

Recordings were performed as previously described (*31*, *46*). We visually identified S1, nRT, and VB neurons by differential contrast optics with an Olympus microscope and an infrared video camera. Recording electrodes made of borosilicate glass had a resistance of 2.5 to 4 MΩ when filled with intracellular solution. Access resistance was monitored in all recordings, and cells were included for analysis only if the

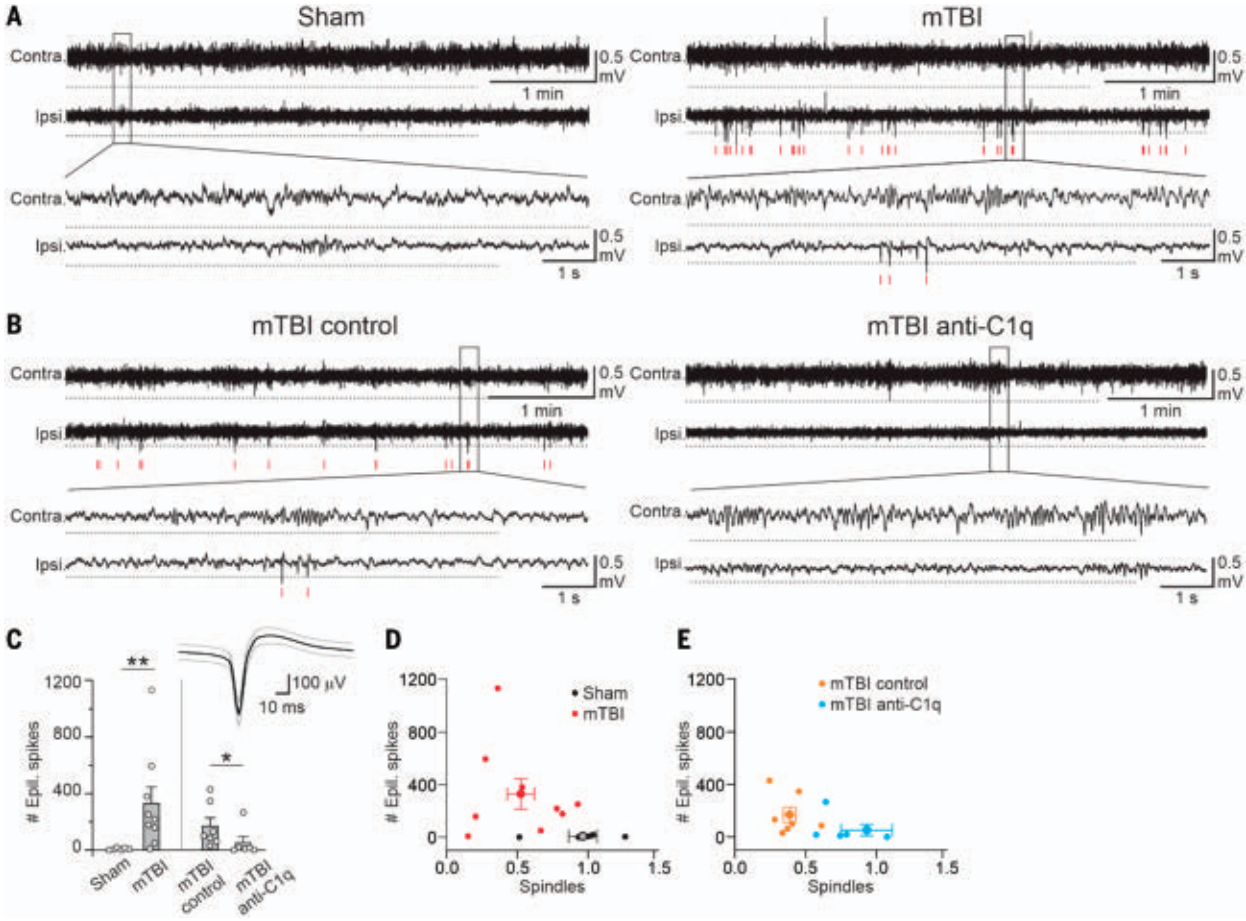


Fig. 6. Anti-C1q antibody reduces focal epileptic spikes that develop 3 weeks after mTBI. (**A**) ECoG recordings from a sham and mTBI mouse 3 weeks after mTBI. Horizontal dashed lines represent the spike detection threshold. Vertical red lines indicate detected spikes. (**B**) Same as (A) from mTBI mice treated with an isotype control or the anti-C1q antibody. Traces in (A) and (B) are from episodes of NREM sleep. (**C**) Number of epileptic spikes detected within a 12-hour window. Data represent mean ± SEM analyzed with a Mann-Whitney rank sum test; $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$). Inset: an average epileptic spike from the mTBI mouse shown in (B) [$n = 592$ spikes; mean (black) ± SD (gray)]. Analysis includes six sham mice and nine mTBI mice (left) and seven control-treated mTBI mice and six antibody-treated mTBI mice (right). (**D** and **E**) Number of epileptic spikes as a function of the ratio of sleep spindles in the sham versus mTBI (D) and anti-C1q versus control mTBI (E) mice from (C). Individual points represent each mouse, and error bars represent mean ± SEM across both axes.

access resistance was <25 MΩ. Intrinsic and bursting properties and sEPSCs were recorded in the presence of picrotoxin (50 µM, Sigma-Aldrich), and the internal solution contained 120 mM potassium gluconate, 11 mM EGTA, 11 mM KCl, 10 mM HEPES, 1 mM $CaCl_2$, and 1 mM $MgCl_2$, pH adjusted to 7.4 with KOH (290 mOsm). We corrected the potentials for −15 mV liquid junction potential.

sIPSCs were recorded in the presence of kynurenic acid (2 mM, Sigma-Aldrich), and the internal solution contained 135 mM CsCl, 10 mM EGTA, 10 mM HEPES, 5 mM Qx-314 (lidocaine *N*-ethyl bromide), and 2 mM $MgCl_2$, pH adjusted to 7.3 with CsOH (290 mOsm).

snRNA library construction and sequencing

snRNA-seq libraries were processed using the Chromium Next GEM Single Cell 3′v3 library kit with dual indexing (10x Genomics) according to the manufacturer's specifications. For every sample, nuclei were diluted to 1000 nuclei/µl in nuclei dilution buffer, and 9900 nuclei were loaded onto the Chromium, with a targeted recovery of 6000 nuclei. Replicate 1 and 2 nuclei were processed on different Chromium runs. Libraries were pooled on the basis of their molar concentrations and sequenced on an Illumina NovaSeq 6000 system using an S1 flow cell and a v1 300-cycle Reagent Kit with 28 cycles for read 1, 90 cycles for read 2, 10 cycles for index i7, and 10 cycles for index i5. Cell Ranger (4.0.0) (10x Genomics) was used to perform sample de-multiplexing, barcode processing, and single-nucleus gene–unique molecular identifier counting. Reads were mapped to mm10 (GENCODE vM23/Ensembl 98, from 10x). From replicate 1, we recovered 2337 nuclei from sham mice with a mean of 92,533 reads per nucleus and 650 nuclei from mTBI mice with a mean of 162,363 reads per nucleus. From Replicate 2, we recovered 3891 nuclei from sham mice with a mean of 47,592 reads per nucleus and 4575 nuclei from TBI mice with a mean of 41,658 reads per nucleus. Across all samples and replicates, the median number of genes per nucleus was 2200. Additional details are provided in the supplementary materials and methods.

Statistical analyses

All numerical values are given as means and error bars are standard error of the mean (SEM) unless stated otherwise. Parametric and non-parametric tests were chosen as appropriate and are reported in the figure legends. Data analysis was performed with MATLAB (SCR_001622), GraphPad Prism 7/8 (SCR_002798), ImageJ (SCR_003070), Ponemah/NeuroScore (SCR_017107), pClamp (SCR_011323), and Spike2 (SCR_000903) software.

Analysis of electrophysiological properties

The input resistance (R_{in}) and membrane time constant (τ_m) were measured from the membrane hyperpolarization in response to low-intensity current steps (−20 to −60 pA). The reported

rheobase averages and SEMs were calculated on the basis of the current that first caused at least one action potential during the stimulus per recording. All data shown in table S1 were analyzed using a Mann-Whitney test with $\alpha = 0.05$ (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$) using GraphPad Prism 7 (SCR_002798).

Cumulative probability distributions were generated in MATLAB (SCR_001622) from 11 sham nRT neurons and nine TBI neurons using 200 randomly selected events from each cell.

Spindle and epileptic spike event detection in ECoG

Data analysis was performed using Spike2 (version 7.20, Cambridge Electronic Design, Cambridge, UK) and Python 3.7 (Python Software Foundation). Epochs were assigned as NREM sleep if the ratio of delta power (δ, 1.5 to 4 Hz) to total power (1.5 to 80 Hz) for ECoG was higher than the threshold value with no locomotor activity. Sleep spindle analysis was performed for a period of 12 hours (7:00 A.M. to 7:00 P.M.) at day 20 or 21 after mTBI or sham surgery. Epileptic spikes were analyzed during the same time frame. Recordings were not analyzed during locomotion because it was challenging to reliably distinguish movement artifacts from epileptic spikes (Fig. 6).

We used the Morlet wavelet function to detect spindles in the 8- to 15-Hz frequency range. We applied a threshold of [1 × the mean + 1.5 × the standard deviation (SD)] of the ECoG power and detected all events above this threshold that lasted at least 0.5 s (Fig. 5 and fig. S8). All detected events were visually validated by a scientist blinded to the groups. The onset and offset times of a spindle event were extended to the closest cycle at zero crossing before and after the threshold. Amplitudes of spindles were computed from the average amplitude of the spindle (8 to 15 Hz) power between onset and offset time, divided by root mean square of the raw ECoG signal, and averaged per mouse. False-positive events that contained epileptic spikes (defined as events that exceeded the threshold of 1× the mean of the baseline + 7 × SD; Figs. 5 and 6) were rejected after visual inspection of a scientist blinded to the groups.

Anti-C1q antibody treatment

Anti-C1q antibody (ANX-M1, Annexon Biosciences) (*22*, *40*, *64*) and control (mouse isotype IgG1 antibody, MOPC, Clone BE0083, from BioXCell) were administered at 100 mg/kg body weight, a nontoxic dose in rodents, first at 24 hours after TBI and then every 3 days for 3 weeks. Detailed treatment paradigm and pharmacological assays are described in the supplementary materials and methods.

REFERENCES AND NOTES

1. World Health Organization, *Neurological Disorders: Public Health Challenges* (WHO, 2006).
2. M. C. Dewan *et al*., Estimating the global incidence of traumatic brain injury. *J. Neurosurg.* 10.3171/2017.10.JNS17352 (2018). pmid: 29701556
3. A. C. Mckee, D. H. Daneshvar, The neuropathology of traumatic brain injury. *Handb. Clin. Neurol.* **127**, 45–66 (2015). doi: 10.1016/B978-0-444-52892-6.00004-0; pmid: 25702209
4. F. Briggs, W. M. Usrey, Emerging views of corticothalamic function. *Curr. Opin. Neurobiol.* **18**, 403–407 (2008). doi: 10.1016/j.conb.2008.09.002; pmid: 18805486
5. D. K. Sandsmark, J. E. Elliott, M. M. Lim, Sleep-wake disturbances after traumatic brain injury: Synthesis of human and animal studies. *Sleep* **40**, (2017). doi: 10.1093/sleep/zsx044; pmid: 28329120
6. C. Auladell, P. Pérez-Sust, H. Supèr, E. Soriano, The early development of thalamocortical and corticothalamic projections in the mouse. *Anat. Embryol.* **201**, 169–179 (2000). doi: 10.1007/PL00008238; pmid: 10664178
7. C. Blakemore, Z. Molnár, Factors involved in the establishment of specific interconnections between thalamus and cerebral cortex. *Cold Spring Harb. Symp. Quant. Biol.* **55**, 491–504 (1990). doi: 10.1101/SQB.1990.055.01.048; pmid: 2132833
8. N. Antón-Bolaños, A. Espinosa, G. López-Bendito, Developmental interactions between thalamus and cortex: A true love reciprocal story. *Curr. Opin. Neurobiol.* **52**, 33–41 (2018). doi: 10.1016/j.conb.2018.04.018; pmid: 29704748
9. M. P. Beenhakker, J. R. Huguenard, Neurons that fire together also conspire together: Is normal sleep circuitry hijacked to generate epilepsy? *Neuron* **62**, 612–632 (2009). doi: 10.1016/j.neuron.2009.05.015; pmid: 19524522
10. J. T. Paz, J. R. Huguenard, Microcircuits and their interactions in epilepsy: Is the focus out of focus? *Nat. Neurosci.* **18**, 351–359 (2015). doi: 10.1038/nn.3950; pmid: 25710837
11. E. J. Grossman *et al*., Thalamus and cognitive impairment in mild traumatic brain injury: A diffusional kurtosis imaging study. *J. Neurotrauma* **29**, 2318–2327 (2012). doi: 10.1089/neu.2011.1763; pmid: 21639753
12. E. J. Grossman, M. Inglese, The role of thalamic damage in mild traumatic brain injury. *J. Neurotrauma* **33**, 163–167 (2016). doi: 10.1089/neu.2015.3965; pmid: 26054745
13. D. T. Ross, F. F. Ebner, Thalamic retrograde degeneration following cortical injury: An excitotoxic process? *Neuroscience* **35**, 525–550 (1990). doi: 10.1016/0306-4522(90)90327-Z; pmid: 2166245
14. G. Scott *et al*., Thalamic inflammation after brain trauma is associated with thalamo-cortical white matter damage. *J. Neuroinflammation* **12**, 224 (2015). doi: 10.1186/s12974-015-0445-y; pmid: 26627199
15. D. W. Simon *et al*., The far-reaching scope of neuroinflammation after traumatic brain injury. *Nat. Rev. Neurol.* **13**, 171–191 (2017). doi: 10.1038/nrneurol.2017.13; pmid: 28186177
16. M. V. Russo, D. B. McGavern, Inflammatory neuroprotection following traumatic brain injury. *Science* **353**, 783–785 (2016). doi: 10.1126/science.aaf6260; pmid: 27540166
17. K. Krukowski *et al*., Traumatic brain injury in aged mice induces chronic microglia activation, synapse loss, and complement-dependent memory deficits. *Int. J. Mol. Sci.* **19**, 3753 (2018). doi: 10.3390/ijms19123753; pmid: 30486287
18. Z. You *et al*., Reduced tissue damage and improved recovery of motor function after traumatic brain injury in mice deficient in complement component C4. *J. Cereb. Blood Flow Metab.* **27**, 1954–1964 (2007). doi: 10.1038/sj.jcbfm.9600497; pmid: 17457366
19. B.-M. Bellander, S. K. Singhrao, M. Ohlsson, P. Mattsson, M. Svensson, Complement activation in the human brain after traumatic head injury. *J. Neurotrauma* **18**, 1295–1311 (2001). doi: 10.1089/08977150152725605; pmid: 11780861
20. A. Hammad, L. Westacott, M. Zaben, The role of the complement system in traumatic brain injury: A review. *J. Neuroinflammation* **15**, 24 (2018). doi: 10.1186/s12974-018-1066-z; pmid: 29357880
21. S. K. Wyatt, T. Witt, N. M. Barbaro, A. A. Cohen-Gadol, A. L. Brewster, Enhanced classical complement pathway activation and altered phagocytosis signaling molecules in human epilepsy. *Exp. Neurol.* **295**, 184–193 (2017). doi: 10.1016/j.expneurol.2017.06.009; pmid: 28601603
22. S. Hong *et al*., Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* **352**, 712–716 (2016). doi: 10.1126/science.aad8373; pmid: 27033548
23. H. Lui *et al*., Progranulin deficiency promotes circuit-specific synaptic pruning by microglia via complement activation. *Cell* **165**, 921–935 (2016). doi: 10.1016/j.cell.2016.04.001; pmid: 27114033
24. K. Cho, Emerging roles of complement protein C1q in neurodegeneration. *Aging Dis.* **10**, 652–663 (2019). doi: 10.14336/AD.2019.0118; pmid: 31165008
25. B. Stevens *et al*., The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164–1178 (2007). doi: 10.1016/j.cell.2007.10.036; pmid: 18083105
26. K. Ding, P. K. Gupta, R. Diaz-Arrastia, "Epilepsy after traumatic brain injury," in *Translational Research in Traumatic Brain Injury*, D. Laskowitz, G. Grant, Eds. (CRC Press/Taylor and Francis, 2016); https://www.ncbi.nlm.nih.gov/books/NBK326716/.
27. A. Destexhe, D. Contreras, M. Steriade, Mechanisms underlying the synchronizing action of corticothalamic feedback through inhibition of thalamic relay cells. *J. Neurophysiol.* **79**, 999–1016 (1998). doi: 10.1152/jn.1998.79.2.999; pmid: 9463458
28. P. Golshani, X. B. Liu, E. G. Jones, Differences in quantal amplitude reflect GluR4- subunit number at corticothalamic synapses on two populations of thalamic neurons. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4172–4177 (2001). doi: 10.1073/pnas.061013698; pmid: 11274440
29. J. Bourassa, D. Pinault, M. Deschênes, Corticothalamic projections from the cortical barrel field to the somatosensory thalamus in rats: A single-fibre study using biocytin as an anterograde tracer. *Eur. J. Neurosci.* **7**, 19–30 (1995). doi: 10.1111/j.1460-9568.1995.tb01016.x; pmid: 7711933
30. Y.-W. Lam, S. M. Sherman, Functional organization of the thalamic input to the thalamic reticular nucleus. *J. Neurosci.* **31**, 6791–6799 (2011). doi: 10.1523/JNEUROSCI.3073-10.2011; pmid: 21543609
31. A. Clemente-Perez *et al*., Distinct thalamic reticular cell types differentially modulate normal and pathological cortical rhythms. *Cell Rep.* **19**, 2130–2142 (2017). doi: 10.1016/j.celrep.2017.05.044; pmid: 28591583
32. M. Deschênes, P. Veinante, Z. W. Zhang, The organization of corticothalamic projections: Reciprocity versus parity. *Brain Res. Brain Res. Rev.* **28**, 286–308 (1998). doi: 10.1016/S0165-0173(98)00017-4; pmid: 9858751
33. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020). doi: 10.1093/gigascience/giaa151; pmid: 33367645
34. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019). doi: 10.1016/j.cels.2019.03.003; pmid: 30954475
35. I. H. Lepow, G. B. Naff, E. W. Todd, J. Pensky, C. F. Hinz Jr., Chromatographic resolution of the first component of human complement into three activities. *J. Exp. Med.* **117**, 983–1008 (1963). doi: 10.1084/jem.117.6.983; pmid: 13929797
36. N. Thrupp *et al*., Single-nucleus RNA-seq is not suitable for detection of microglial activation genes in humans. *Cell Rep.* **32**, 108189 (2020). doi: 10.1016/j.celrep.2020.108189; pmid: 32997994
37. A. R. Bialas, B. Stevens, TGF-β signaling regulates neuronal C1q expression and developmental synaptic refinement. *Nat. Neurosci.* **16**, 1773–1782 (2013). doi: 10.1038/nn.3560; pmid: 24162655
38. N. D. Schartz, A. J. Tenner, The good, the bad, and the opportunities of the complement system in neurodegenerative disease. *J. Neuroinflammation* **17**, 354 (2020). doi: 10.1186/s12974-020-02024-8; pmid: 33239010
39. T. Wu *et al*., Complement C3 is activated in human AD brain and is required for neurodegeneration in mouse models of amyloidosis and tauopathy. *Cell Rep.* **28**, 2111–2123.e6 (2019). doi: 10.1016/j.celrep.2019.07.060; pmid: 31433986
40. J. A. Lansita *et al*., Nonclinical development of ANX005: A humanized anti-C1q antibody for treatment of autoimmune and neurodegenerative diseases. *Int. J. Toxicol.* **36**, 449–462 (2017). doi: 10.1177/1091581817740873; pmid: 29202623
41. L. M. Fernandez *et al*., Thalamic reticular control of local sleep in mouse sensory cortex. *eLife* **7**, e39111 (2018). doi: 10.7554/eLife.39111; pmid: 30583750
42. D. T. Ross, D. I. Graham, J. H. Adams, Selective loss of neurons from the thalamic reticular nucleus following severe human head injury. *J. Neurotrauma* **10**, 151–165 (1993). doi: 10.1089/neu.1993.10.151; pmid: 8411218
43. M. M. Huntsman, D. M. Porcello, G. E. Homanics, T. M. DeLorey, J. R. Huguenard, Reciprocal inhibitory connections and network synchrony in the mammalian

thalamus. *Science* **283**, 541–543 (1999). doi: 10.1126/science.283.5401.541; pmid: 9915702
44. V. S. Sohal, J. R. Huguenard, Inhibitory interconnections control burst pattern and emergent network synchrony in reticular thalamus. *J. Neurosci.* **23**, 8978–8988 (2003). doi: 10.1523/JNEUROSCI.23-26-08978.2003; pmid: 14523100
45. V. S. Sohal, M. M. Huntsman, J. R. Huguenard, Reciprocal inhibitory connections regulate the spatiotemporal properties of intrathalamic oscillations. *J. Neurosci.* **20**, 1735–1745 (2000). doi: 10.1523/JNEUROSCI.20-05-01735.2000; pmid: 10684875
46. J. T. Paz *et al*., A new mode of corticothalamic transmission revealed in the Gria4(-/-) model of absence epilepsy. *Nat. Neurosci.* **14**, 1167–1173 (2011). doi: 10.1038/nn.2896; pmid: 21857658
47. P. Andrade, J. Nissinen, A. Pitkänen, Generalized seizures after experimental traumatic brain injury occur at the transition from slow-wave to rapid eye movement sleep. *J. Neurotrauma* **34**, 1482–1487 (2017). doi: 10.1089/neu.2016.4675; pmid: 27707084
48. O. Mahmood, L. J. Rapport, R. A. Hanks, N. L. Fichtenberg, Neuropsychological performance and sleep disturbance following traumatic brain injury. *J. Head Trauma Rehabil.* **19**, 378–390 (2004). doi: 10.1097/00001199-200409000-00003; pmid: 15597029
49. C. Wang *et al*., Microglia mediate forgetting via complement-dependent synaptic elimination. *Science* **367**, 688–694 (2020). doi: 10.1126/science.aaz2288; pmid: 32029629
50. A. Vukojicic *et al*., The classical complement pathway mediates microglia-dependent remodeling of spinal motor circuits during development and in SMA. *Cell Rep.* **29**, 3087–3100.e7 (2019). doi: 10.1016/j.celrep.2019.11.013; pmid: 31801075
51. M. E. Benoit *et al*., C1q-induced LRP1B and GPR6 proteins expressed early in Alzheimer disease mouse models, are essential for the C1q-mediated protection against amyloid-β neurotoxicity. *J. Biol. Chem.* **288**, 654–665 (2013). doi: 10.1074/jbc.M112.400168; pmid: 23150673
52. K. Pisalyaput, A. J. Tenner, Complement component C1q inhibits beta-amyloid- and serum amyloid P-induced neurotoxicity via caspase- and calpain-independent mechanisms. *J. Neurochem.* **104**, 696–707 (2008). pmid: 17986223
53. S. A. Liddelow *et al*., Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017). doi: 10.1038/nature21029; pmid: 28099414
54. C. D. Mills, K. Kincaid, J. M. Alt, M. J. Heilman, A. M. Hill, M-1/M-2 macrophages and the Th1/Th2 paradigm. *J. Immunol.* **164**, 6166–6173 (2000). doi: 10.4049/jimmunol.164.12.6166; pmid: 10843666
55. A. Sekar *et al*., Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016). doi: 10.1038/nature16549; pmid: 26814963
56. D. S. Manoach, J. Q. Pan, S. M. Purcell, R. Stickgold, Reduced sleep spindles in schizophrenia: A treatable endophenotype that links risk genes to impaired cognition? *Biol. Psychiatry* **80**, 599–608 (2016). doi: 10.1016/j.biopsych.2015.10.003; pmid: 26602589
57. M. Yilmaz *et al*., Overexpression of schizophrenia susceptibility factor human complement C4A promotes excessive synaptic loss and behavioral changes in mice. *Nat. Neurosci.* **24**, 214–224 (2021). doi: 10.1038/s41593-020-00763-8; pmid: 33353966
58. L. K. Harris *et al*., Traumatic brain injury-induced changes in gene expression and functional activity of mitochondrial cytochrome C oxidase. *J. Neurotrauma* **18**, 993–1009 (2001). doi: 10.1089/08977150152693692; pmid: 11686499
59. W. O. Opii *et al*., Proteomic identification of oxidized mitochondrial proteins following experimental traumatic brain injury. *J. Neurotrauma* **24**, 772–789 (2007). doi: 10.1089/neu.2006.0229; pmid: 17518533
60. BioSpace. "Annexon announces initiation of phase 2/3 Trial of ANX005 in patients with Guillain-Barré syndrome" (BioSpace, 2021); https://www.biospace.com/article/annexon-announces-initiation-of-phase-2-3-trial-of-anx005-in-patients-with-guillain-barre-syndrome/).
61. W. M. Zelek, L. Xie, B. P. Morgan, C. L. Harris, Compendium of current complement therapeutics. *Mol. Immunol.* **114**, 341–352 (2019). doi: 10.1016/j.molimm.2019.07.030; pmid: 31446305
62. A. Alawieh *et al*., Complement drives synaptic degeneration and progressive cognitive decline in the chronic phase after traumatic brain injury. *J. Neurosci.* **41**, 1830–1843 (2021). doi: 10.1523/JNEUROSCI.1734-20.2020; pmid: 33446516
63. K. Mallah *et al*., Complement mediates neuroinflammation and cognitive decline at extended chronic time points after traumatic brain injury. *Acta Neuropathol. Commun.* **9**, 72 (2021). doi: 10.1186/s40478-021-01179-6; pmid: 33879257
64. R. McGonigal *et al*., C1q-targeted inhibition of the classical complement pathway prevents injury in a novel mouse model of acute motor axonal neuropathy. *Acta Neuropathol. Commun.* **4**, 23 (2016). doi: 10.1186/s40478-016-0291-x; pmid: 26936605

ACKNOWLEDGMENTS

We thank B. Barres for inspiration and scientific guidance at the early stage of this project, J. Anink for help with histology on human tissue, I. Lew for mouse colony management, F. Chanut for editorial support, and M. Calvert and the Gladstone Histology & Light Microscopy Core for help with confocal microscopy and ROI analysis. Sequencing was performed at the Center for Advanced Technology (CAT) at UCSF. **Funding:** This study was funded mainly by Department of Defense grant EP150038 to J.T.P., who also had support from the National Institutes of Health (NIH grant R01 NS078118), the National Science Foundation (NSF grant 1608236), Gladstone Institutes, the Michael Prize, the Vilcek Prize, and the Kavli Institute for Fundamental Neuroscience. S.S.H. was supported by an Achievement Rewards for College Scientists Scholarship, a Ford Foundation Dissertation Fellowship, the NIH (grant T32-GM007449), and the Weill Foundation. B.H. was supported by an American Epilepsy Society Postdoctoral Research Fellowship. F.S.C. was supported by the National Institute of Neurological Disorders and Stroke (NINDS grant F31 NS111819-01A1), the NSF (Graduate Research Fellowship 1144247), and by a UCSF Discovery Fellowship. E.A. was supported by Epilepsiefonds project 2020-02. A.L. was supported by the Swiss National Science Foundation (grant 310030-184759). A.O.F. was supported by the Faculty of Biology and Medicine–University of Lausanne doctoral fellowship. A.J.T. was supported by the NIH (grant R01AG060148). **Author contributions:** Conceptualization: J.T.P. and S.S.H.; Data curation: S.S.H., J.T.P., F.S.C., E.A., F.C.G., M.R.C.; Formal analysis: S.S.H., F.S.C., F.C.G., M.R.C., J.T.P.; Funding acquisition: J.T.P.; Investigation: J.T.P., S.S.H., F.C.G., O.A., B.H., F.S.C., A.H.C., A.R.M., V.M., L.J.K., P.S., S.S., Y.A.-Z., E.A., M.R.C.; Methodology: J.T.P., S.S.H., F.C.G., B.H., S.S., M.R.C., A.O.F., A.L.; Project administration: J.T.P.; Resources: J.T.P., M.R.C., T.Y.; Supervision: J.T.P., M.R.C., A.L., Y.A.-Z., T.Y.; Visualization: S.S.H., F.C.G., O.A., M.R.C., J.T.P.; Writing – original draft: J.T.P. and S.S.H.; Writing – review and editing: all authors. **Competing interests:** V.M., L.J.K., P.S., S.S., Y.A.-Z., and T.Y. are employees of Annexon Inc., a publicly traded biotechnology company. The remaining authors declare no competing interests. **Data and materials availability:** Raw and processed data have been deposited to the Gene Expression Omnibus (GEO) under accession number GSE179541. All data are available in the manuscript or the supplementary materials. The anti-C1q antibody was obtained under a materials transfer agreement with Annexon.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abj2685
Materials and Methods
Figs. S1 to S8
Tables S1 and S2
References (*65*–*80*)
MDAR Reproducibility Checklist

1 May 2021; accepted 3 August 2021
10.1126/science.abj2685

Science Webinars

AAAS

Science Webinars help you keep pace with emerging scientific fields!

Stay informed about scientific breakthroughs and discoveries.

Gain insights into current research from top scientists.

Take the opportunity to ask questions during live broadcasts.

Get alerts about upcoming free webinars.

Sign up at: webinar.science.org/stayinformed

RESEARCH ARTICLES

MOLECULAR BIOLOGY

Regulation of biomolecular condensates by interfacial protein clusters

Andrew W. Folkmann[1,2]†, Andrea Putnam[1,2]†, Chiu Fan Lee[3], Geraldine Seydoux[1,2]*

Biomolecular condensates are cellular compartments that can form by phase separation in the absence of limiting membranes. Studying the P granules of *Caenorhabditis elegans*, we find that condensate dynamics are regulated by protein clusters that adsorb to the condensate interface. Using in vitro reconstitution, live observations, and theory, we demonstrate that localized assembly of P granules is controlled by MEG-3, an intrinsically disordered protein that forms low dynamic assemblies on P granules. Following classic Pickering emulsion theory, MEG-3 clusters lower surface tension and slow down coarsening. During zygote polarization, MEG-3 recruits the DYRK family kinase MBK-2 to accelerate spatially regulated growth of the P granule emulsion. By tuning condensate-cytoplasm exchange, interfacial clusters regulate the structural integrity of biomolecular condensates, reminiscent of the role of lipid bilayers in membrane-bound organelles.

Liquid-liquid phase separation has emerged as a new principle for cellular organization (*1*). Phase separation of proteins, frequently with RNA, can create dense condensates that are visible by microscopy as micron-scale assemblies containing dozens or more protein and RNA species. Condensates have been reconstituted in vitro using purified proteins that interact by using multivalent, low-affinity binding sites to generate large, interconnected networks. Some proteins form condensates that exhibit fast internal dynamics and rapid exchange with the dilute phase on a time scale of seconds or minutes (*1*). Other proteins form more viscous condensates with slower internal and exchange dynamics, in some cases becoming glass-like with time (*2*, *3*). All condensate systems (emulsions) are predicted to evolve (coarsen) toward a single-condensate equilibrium at a rate that correlates positively with their internal and exchange dynamics (*4*–*8*) (see supplementary text). Like condensates assembled in vitro, condensates in cells exhibit a range of dynamic behaviors, but these do not always fit theoretical predictions (*9*, *10*). For example, some condensates exhibit little coarsening over hour time scales, yet maintain low viscosity, maintain fast exchange dynamics, and dissolve within minutes in response to changes in the cellular environment (*11*). Several hypotheses have been put forward to explain the lack of coarsening of cellular condensates with fast internal dynamics, including physical barriers that keep condensates away from each other (*5*, *12*, *13*), active mechanisms that continuously regenerate small condensates (*14*), and chemical reactions and protein gradients that suppress Ostwald ripening (*15*–*18*) (see supplementary text). In this study, we investigate the mechanisms that control the coarsening and dynamics of P granules, condensates in the *Caenorhabditis elegans* germline. We find that P granule coarsening is controlled by nanoscale protein clusters that adsorb to the condensate interface, a phenomenon first described by Ramsden (in 1904) and Pickering (in 1907) for inorganic emulsions (*19*, *20*).

P granules were the first cellular condensates proposed to form by liquid-liquid phase separation (*21*). At the core of P granules is a liquid-like phase assembled by PGL proteins, paralogs PGL-1 and PGL-3 (*21*–*23*). During most of the *C. elegans* life cycle, PGL condensates associate stably with the cytoplasmic face of nuclei (*24*). During the oocyte-to-zygote transition, PGL condensates redistribute to the cytoplasm and undergo two rapid cycles of dissolution and condensation. The first cycle occurs during oocyte maturation when most PGL condensates dissolve before reassembling after fertilization in the zygote. The second cycle occurs during zygote polarization when PGL condensates dissolve in anterior cytoplasm and assemble in posterior cytoplasm. Both cycles are completed within minutes without substantial coarsening. Factors that regulate P granule dynamics during oocyte maturation have not yet been identified. Factors that regulate dynamics during polarization include MEX-5, an RNA-binding protein; MEG proteins MEG-3 and MEG-4, two paralogous intrinsically disordered proteins; and MBK-2, a DYRK family kinase that interacts physically and genetically with MEG-3 (*22*, *25*–*28*). During polarization, MEX-5 becomes enriched in the anterior cytoplasm, where it promotes P granule dissolution, possibly by competing with P granule proteins for RNA (*22*, *27*). In zygotes lacking *mbk-2* or *meg-3* and *meg-4* activity, P granules do not dissolve in the anterior cytoplasm, despite a normal MEX-5 gradient (*28*). In this study, we investigate how the MEGs and MBK-2 collaborate with MEX-5 to regulate P granule dynamics.

MEG-3 forms low-dynamic clusters that adsorb to the PGL-3 interface

MEG-3 has been reported to form assemblies on the surface of PGL-3 condensates in newly fertilized zygotes (*28*, *29*) and in PGL-3 and MEG-3 co-condensates reconstituted in vitro (*29*). Superresolution three-dimensional confocal microscopy (see methods) confirmed that MEG-3 forms diffraction-limited clusters (<160 nm) at the PGL-3 interface in vivo and in vitro (Fig. 1A and fig. S1). Consistent with MEG-3 clusters adsorbing to the interface, MEG-3 modifies the wetting behavior of PGL-3 condensates assembled in vitro, reducing the extent to which PGL-3 droplets wet the surface of untreated glass slides (Fig. 1, B and C).

MEG-3 clusters are resistant to dilution, high temperature, and salt treatment; by contrast, PGL-3 condensates readily dissolve in dilute conditions and at increased temperatures (*29*). MEG-3 condensates exchange more slowly than PGL-3 condensates, as measured by fluorescence recovery after photobleaching (FRAP) in vitro and in vivo (*29*). Using a single-molecule method adapted from Wu *et al.* (*30*), we measured the dynamics of MEG-3 and PGL-3 molecules in P granules in vivo (Fig. 1, D to F, and movies S1 and S2). Most PGL-3 molecules exhibited short-lived trajectories in P granules with an average apparent diffusion coefficient of $D = 0.056\ \mu m^2/s$ (Fig. 1, E and F). By contrast, all MEG-3 molecules exhibited restricted long-lived trajectories with an average apparent diffusion coefficient of $D = 0.0018\ \mu m^2/s$ (Fig. 1, E and F). In three of three cases where we captured the trajectories of three labeled MEG-3 molecules in the same P granule, their relative position remained fixed over time (Fig. 1D and movie S2). Together these observations confirm that PGL-3 molecules exist primarily in a dynamic liquid-like phase (albeit highly viscous), whereas MEG-3 molecules experience much slower dynamics, resembling solid clusters within our experimental time scales.

MEG-3 reduces the surface tension of PGL-3 condensates without changing viscosity

Solid particulates that adsorb to liquid surfaces reduce surface tension without affecting internal dynamics (viscosity) (*31*). Previous studies have shown that PGL-3 condensates

[1]Department of Molecular Biology and Genetics, Johns Hopkins University, Baltimore, MD 21205, USA. [2]Howard Hughes Medical Institute, Johns Hopkins University, Baltimore, MD 21205, USA. [3]Department of Bioengineering, Imperial College London, London SW7 2AZ, UK.
*Corresponding author. Email: gseydoux@jhmi.edu
†These authors contributed equally to this work.

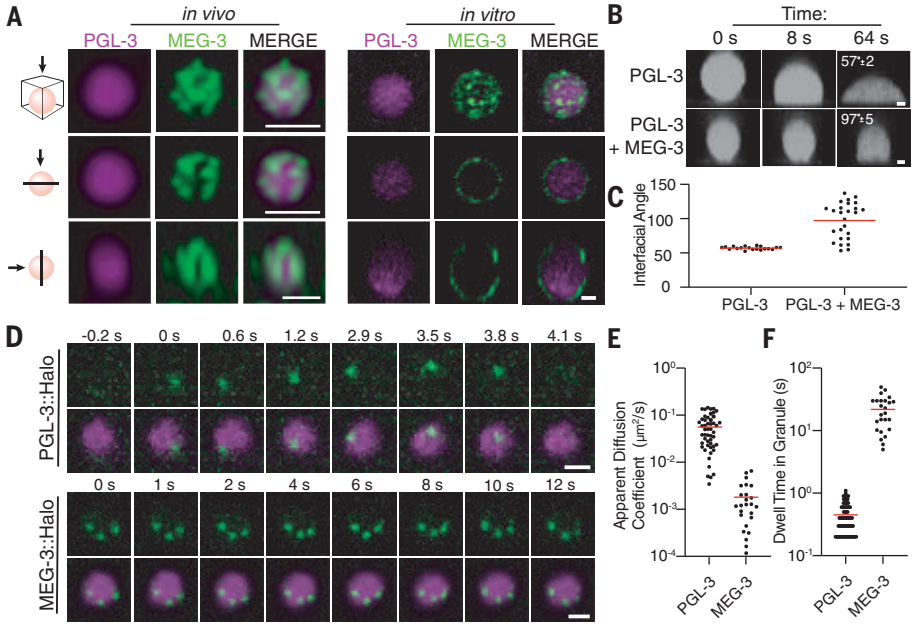


Fig. 1. MEG-3 forms low-dynamic clusters that adsorb to the surface of PGL-3 condensates. (**A**) Photomicrographs of a P granule in vivo labeled with PGL-3::mCherry and MEG-3::meGFP (GFP, green fluorescent protein) and of a P granule reconstituted in vitro with purified PGL-3 and MEG-3 trace-labeled with Dylight 488 and Alexa 647, respectively. Scale bars are 500 nm (in vivo, each scale bar applies to all images in the same row) and 3 μm (in vitro, scale bar applies to all images in the set). The top panels are a maximum projection of a *z*-stack through the granule. The middle panels are a single *x*-*y* plane through the middle of the same granule. The lower panels are a single *z*-*x* plane through the middle of the same granule. See fig. S1A for additional examples of P granules captured in vivo. (**B**) In vitro time-lapse series showing PGL-3 droplets trace-labeled with Dylight 488 wetting the surface of a glass slide with 3 μM PGL-3, with 80 ng/μl *nos-2* RNA, and with or without 0.5 μM MEG-3. Average contact angles at 64 s are indicated. Scale bars are 1 μm, and each scale bar applies to all images in the same row. (**C**) Contact angles measured as in (B). Each dot represents a droplet, and red lines represent the mean. (**D**) In vivo time-lapse series showing single molecules (green) of PGL-3::Halo and MEG-3::Halo in P granules (magenta). Scale bars are 1 μm, and each scale bar applies to all images in the respective set. (**E**) Graph depicting the apparent diffusion coefficients of PGL-3::Halo and MEG-3::Halo molecules in P granules. Each dot represents one trajectory, and the red line represents the mean. (**F**) Graph depicting the dwell time of PGL-3::Halo and MEG-3::Halo molecules in P granules. Each dot represents one trajectory, and the red line represents the mean.

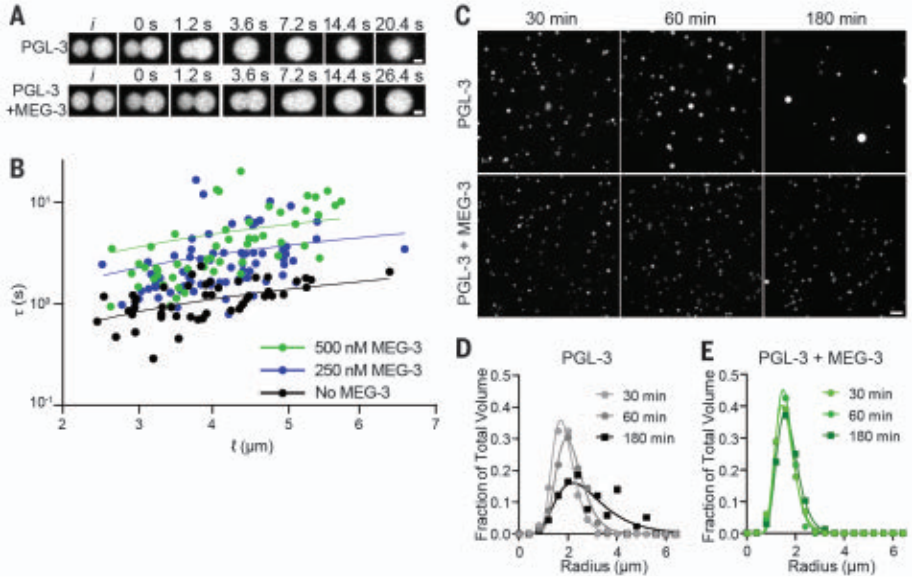


Fig. 2. MEG-3 reduces the surface tension of PGL-3 condensates and prevents coarsening. (**A**) Photomicrographs of PGL-3 droplets (3 μM PGL-3 and 80 ng/μl *nos-2* RNA) coalescing with or without 0.5 μM MEG-3. (**B**) Relaxation time (τ) of fusing PGL-3 droplets (as above) is plotted versus length scale (ℓ) with varying concentrations of MEG-3 as indicated. Each dot represents a single fusion event. The linear slope represents the inverse capillary velocity (η/γ). (**C**) Photomicrographs of a PGL-3 emulsion (max projections) at the indicated time points after assembly. Three micromolar PGL-3 and 80 ng/μl *nos-2* RNA were incubated in condensation buffer in the presence or absence of 0.5 μM MEG-3. Scale bar is 5 μm and applies to all images in the set. (**D** and **E**) Histograms plotting the size distribution of PGL condensates assembled as in (C). Each data point indicates the fraction of total PGL-3 condensate volume represented by condensates binned by radius from 80 images [as in (C)] collected in four replicates. Lines were fit to a log normal distribution.

are "aging Maxwell fluids" whose viscosity increases over days, eventually adopting glass-like properties [see supplementary text and (*2*, *32*)]. To focus on the time scales experienced by PGL-3 condensates during the oocyte-to-embryo transition, we examined PGL-3 condensates within the first 3 hours of assembly for all in vitro studies. To probe internal dynamics, we measured the movement of fluorescent microspheres embedded in PGL-3 condensates with and without MEG-3 (fig. S2, A and B, and movie S3). Fitting the mean squared displacement (MSD) of microspheres to the equation $MSD = 4Dt^{\alpha}$, we calculated a diffusion coefficient (D) and anomalous diffusion exponent (α). We observed an $\alpha \approx 1$ for PGL-3, consistent with newly formed PGL-3 condensates behaving as viscous liquids (fig. S2C). Using the Stokes-Einstein relation, we calculated a viscosity of $\eta = 5.4$ Pa·s in the range estimated previously for PGL-3 condensates

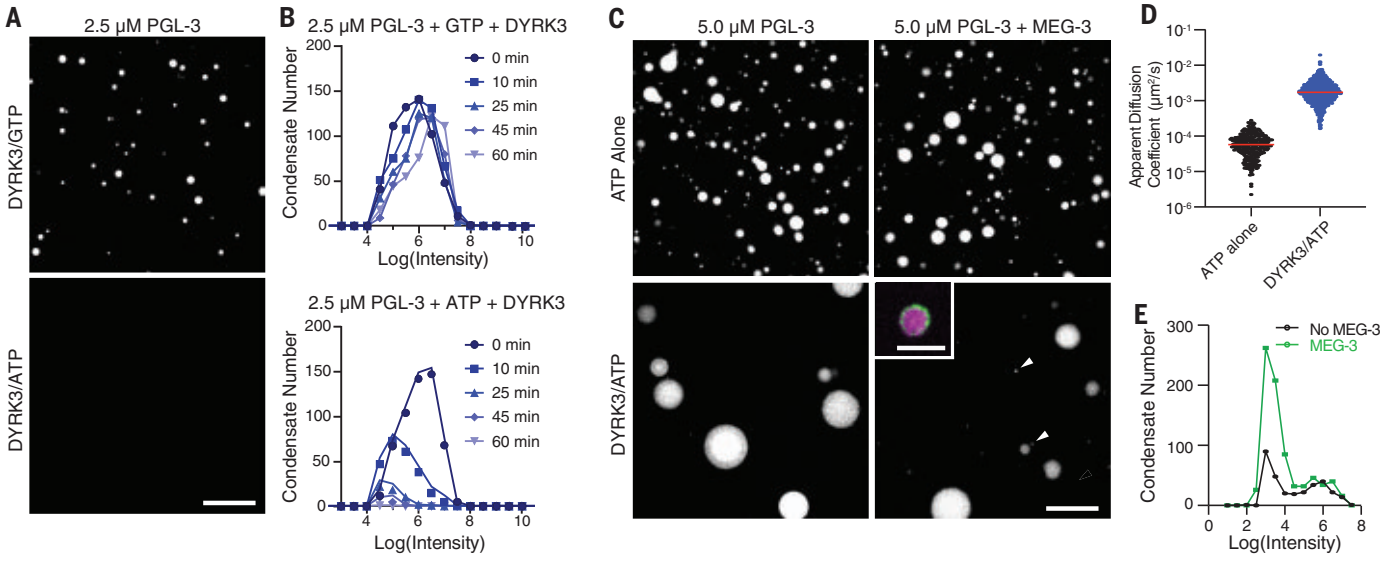


Fig. 3. MEG-3 stabilizes PGL-3 condensates against kinase accelerated coarsening. (**A**) Photomicrographs of a 2.5 μM PGL-3^{488} emulsion after a 60-min treatment with 100 nM DYRK3 kinase in the presence of GTP (top) or ATP (bottom). Scale bar is 50 μm and applies to both images. (**B**) Histograms of PGL condensates assembled as in (A) with GTP and DYRK3 or ATP and DYRK3 at indicated time points. Circles indicate the number of PGL-3 condensates binned by the log(intensity) of each condensate. Colors indicate the time after addition of DYRK3. (**C**) Photomicrographs of PGL-3 condensates (max projections) assembled with and without 70 nM MEG-3 and captured 60 min after addition of 100 μM ATP with and without 100 nM DYRK3. Arrows point to small PGL-3 and MEG-3 co-condensates. The inset is a high-resolution image with PGL-3 in magenta and MEG-3 in green (see fig. S4, F and G, for additional examples). Scale bars are 50 μm (applies to all images in the set) and 5 μm (inset). (**D**) Diffusion coefficients of 200-nm microspheres in PGL-3 condensates (5 μM PGL-3 and 100 μM ATP) with and without 100 nM DYRK3. Each dot represents a single microsphere trajectory. The red line represents the mean. (**E**) Histograms of PGL-3 condensates assembled with and without 70 nM MEG-3 and incubated for 60 min in 100 μM ATP with 100 nM DYRK3. Circles indicate the number of PGL-3 condensates binned by the log (intensity) of each condensate captured from 17 images. See fig. S4, D and E, for additional MEG-3 concentrations and an additional time point.

in vivo and in vitro (fig. S2D) (*2*, *21*). Addition of MEG-3 did not have a notable effect on PGL-3 viscosity ($\eta = 5.6$ Pa·s) (fig. S2, B to D, and movie S3), as expected for a surface agent that, on its own, does not affect internal dynamics.

To measure PGL-3 surface tension, we examined the coalescence behavior of PGL-3 droplets (Fig. 2A). Relaxation time of coalescing condensates can be expressed by $\tau \approx \ell(\eta/\gamma)$, where ℓ is the geometric mean of the diameters of the droplets at the onset of fusion, η is the viscosity of the droplets, γ is the surface tension, and η/γ is the inverse capillary velocity. PGL-3 condensates coalesce with a linear relationship over a range of condensate sizes, yielding an inverse capillary velocity of 0.25 s/μm and surface tension of 19.4 μN/m (Fig. 2B and movie S4), comparable to previous estimates for P granules in vivo (*21*). Addition of MEG-3 slowed coalescence of PGL-3 droplets, yielding an inverse capillary velocity of 1.2 s/μm and a surface tension of 4.7 μN/m (500 nM MEG-3; Fig. 2B and movie S5).

In addition to modulating surface tension, Pickering agents can also inhibit coalescence by steric hindrance and slow rearrangement at the condensate interface (*33*). We observed examples of MEG-3–coated PGL-3 droplets that did not relax to a sphere (movie S6) or remained in contact without fusing during imaging (movie S7). Higher concentrations of MEG-3 (1 μm) increased surface coverage and caused PGL-3 droplets to flocculate without fusing (fig. S2E). We conclude that, as expected for a Pickering agent, MEG-3 clusters lower the surface tension of PGL-3 condensates and also form a physical barrier to coalescence.

MEG-3 prevents coarsening of the PGL-3 emulsion without eliminating surface exchange

To determine whether MEG-3 stabilizes the PGL-3 emulsion against coarsening, we examined the evolution of a newly assembled PGL-3 emulsion over time. Over the course of 180 min, the PGL-3 emulsion coarsened substantially: Droplets increased in size on average and decreased in number without a change in the total volume of PGL-3 in droplets (Fig. 2, C and D, and fig. S2, F to H). Addition of MEG-3 reduced coarsening, stabilizing droplet size and number over the 180 min of the experiment (Fig. 2, C and E, and fig. S2, F to H). MEG-3 concentrations constrained the size of PGL-3 droplets in a dose-dependent manner (fig. S2, I to M).

MEG-3 does not affect the internal dynamics of PGL-3 condensates as measured by FRAP (*29*). To examine whether MEG-3 prevents exchange of PGL-3 molecules at the interface, we used PGL-3 preparations trace-labeled with different fluorophores and examined mixing kinetics. Unlike MEG-3 clusters, which do not mix after assembly, PGL-3 condensates continue to exchange after assembly (fig. S3, A to C). Addition of soluble PGL-3 to a preformed emulsion leads to the formation of new condensates that, over the course of an hour, completely mix with the old condensates (fig. S3, D to F). Addition of MEG-3 prevented coarsening but did not affect the rate of new and old PGL-3 mixing (fig. S3, D to F). We conclude that MEG-3 clusters stabilize the PGL-3 emulsion by lowering surface tension without completely blocking surface exchange, as described for other Pickering agents (*34*).

MEG-3 stabilizes PGL-3 condensates against DYRK3-accelerated coarsening

The relative high viscosity of the PGL-3 emulsion is consistent with the apparent stability of P granules during most of germline development and suggests that active processes must operate to dissolve P granules during oocyte maturation and polarization. MBK-2 kinase is

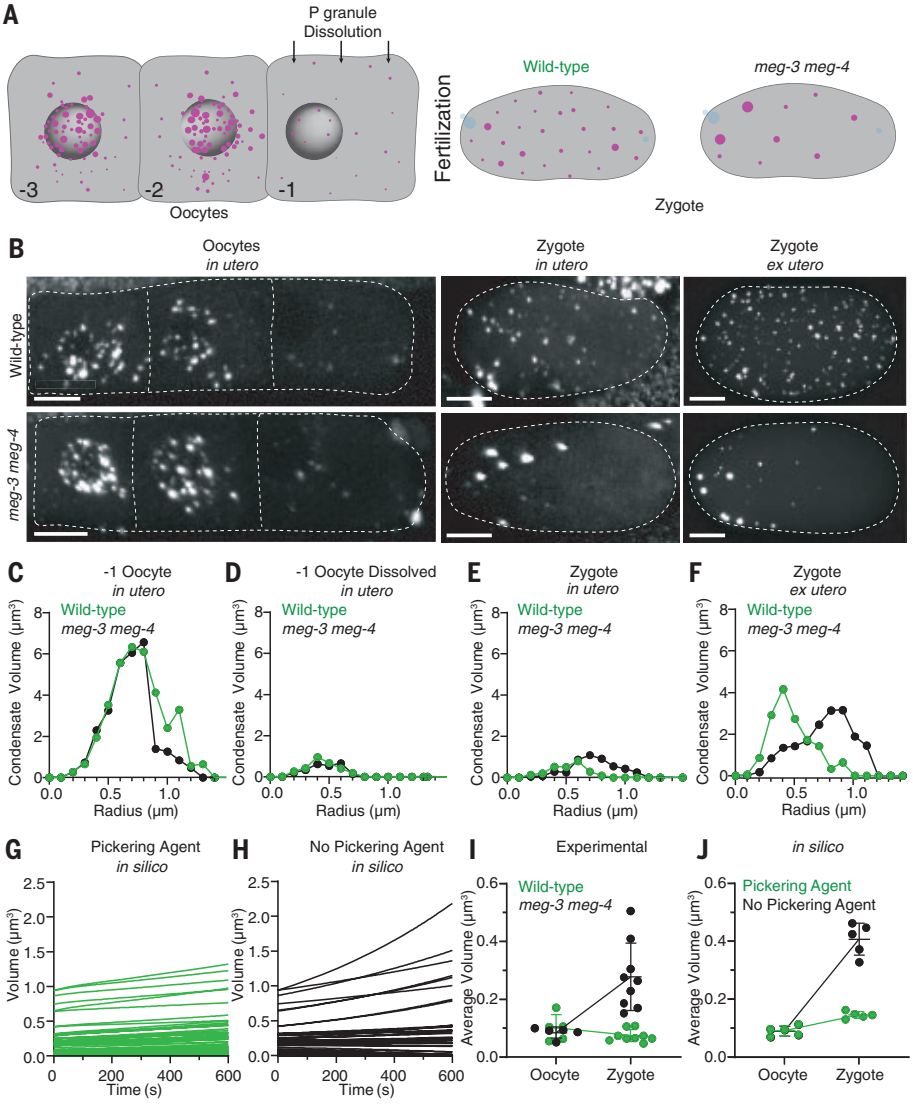


Fig. 4. MEG-3 and MEG-4 stabilize P granules against coarsening during the oocyte-to-zygote transition. (**A**) Schematics depicting the dissolution of P granules (magenta) during the transition from oocyte to fertilized zygote. The numbers indicate the relative position of each oocyte in the germline, and blue represents DNA. (**B**) Photomicrographs of wild-type and *meg-3 meg-4* oocytes and zygotes expressing PGL-3::mCherry (white). Photomicrographs were captured in live adult hermaphrodites (in utero) or after dissection out of the uterus (ex utero). Representative photomicrographs are max projections corresponding to ~20% of oocyte volume and ~80% of zygote volume. The anterior (left) bias for PGL condensates in *meg-3 meg-4* zygotes correlates with anterior displacement of the oocyte nucleus (and associated P granules) that occurs immediately before fertilization. The white dashed lines indicate the boundary of each oocyte or zygote. Scale bars are 10 μm. (**C** to **F**) Histograms of PGL condensate volumes measured from images captured as in (B) representing 100% of oocyte and zygote volumes. Circles indicate the volume of individual PGL-3 in condensates binned by condensate radius in wild-type (green) and *meg-3 meg-4* (black) oocytes and zygotes. Volumes are higher in (F) than in (E) owing to higher detection sensitivity ex utero. (**G** and **H**) Graphs showing the evolution of individual PGL condensates in a 10-min period starting after dissolution in wild-type and *meg-3 meg-4* oocytes and zygotes under simulated conditions. Each line represents the evolution of a single condensate over time. (**I** and **J**) Graphs showing the average volume of individual PGL condensates in oocytes and zygotes under experimental and simulated conditions. For (I), each dot corresponds to an oocyte [same dataset as shown in (D)] or zygote [same data set as shown in (F)] of the indicated genotypes. For (J), each dot corresponds to one simulation. Simulations were run in the presence or absence of the Pickering agent. Horizontal and vertical lines represent the mean SD.

required for P granule dissolution during polarization, and its mammalian homolog DYRK3 has been implicated in the dissolution of condensates in mammalian cells (*28*, *35*, *36*). We found that recombinant DYRK3 phosphorylates PGL-3 efficiently in vitro (fig. S4A). Addition of DYRK3 and adenosine triphosphate (ATP) to preassembled PGL-3 condensates (2.5 μM) led to their complete dissolution over 60 min (Fig. 3, A and B). The effect did not depend on the hydrotrope properties of ATP (*37*) because dissolution was not observed in the presence of guanosine triphosphate (GTP) (Fig. 3, A and B). Sedimentation experiments confirmed that phosphorylation by DYRK3 increases the fraction of PGL-3 in the soluble pool (fig. S4B). At a higher concentration of PGL-3 (5 μM), condensates could be maintained in the presence of DYRK3 (Fig. 3C). We used these supersaturated conditions to conduct microrheology experiments on PGL-3 droplets treated with DYRK3 (Fig. 3D and movie S8). We found that DYRK3 decreases the viscosity of PGL-3 droplets (Fig. 3D). Consistent with accelerated internal dynamics, DYRK3-treated PGL-3 condensates also coarsened rapidly (Fig. 3, C and E). Addition of MEG-3 did not interfere with PGL-3 phosphorylation but led to a ~3-fold increase in the frequency of small PGL-3 droplets (Fig. 3, C and E, and fig. S4D). The small droplets were covered with MEG-3 and remained stable for 120 min, even in the presence of much larger condensates (fig. S4, E to G). These observations indicate that MEG-3 stabilizes the PGL-3 emulsion against coarsening, even under conditions where PGL-3 dynamics have been accelerated by phosphorylation.

MEG-3 and MEG-4 stabilize P granules against coarsening during the oocyte-to-zygote transition

To examine the impact of MEG-3 on P granule dynamics in vivo, we used quantitative live-cell imaging to measure the number and size of PGL-3 condensates in wild type and *meg-3 meg-4* mutants (where *meg-3* and *meg-4* are deleted). We began by imaging eggs in utero as they progress through oocyte maturation and fertilization (Fig. 4, A and B). We found that the volume of PGL-3 in condensates decreased 10-fold during oocyte maturation and remained low in newly fertilized zygotes as they completed the meiotic divisions (Fig. 4, C to F, and fig. S5A). Total PGL-3 levels did not change during this period (fig. S5B), consistent with a transient increase in PGL-3 solubility. We observed the same

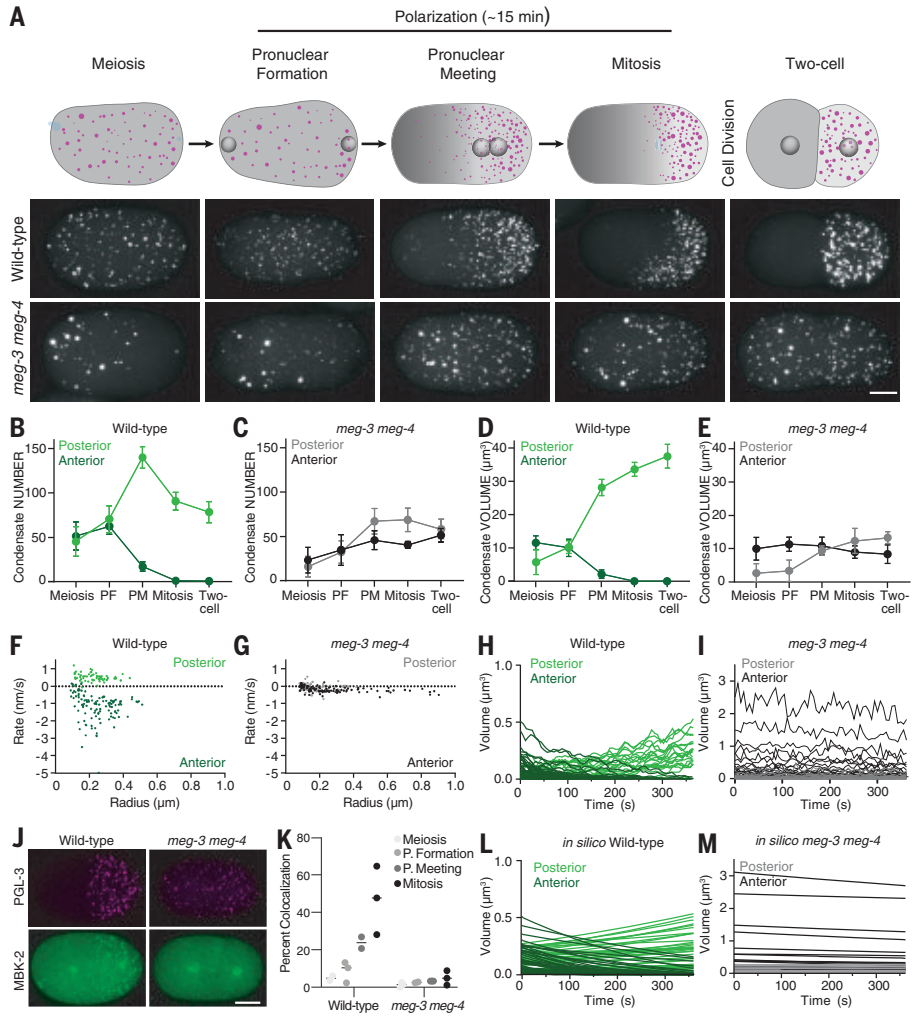


Fig. 5. MEG-3 and MEG-4 drive asymmetric growth of the P granule emulsion during polarization. (**A**) Top row: Cartoons depicting PGL-3 condensates (magenta) and MEX-5 (gray) at different stages during the transition from unpolarized to polarized zygote. Blue represents DNA. Bottom two rows: Photomicrographs of wild-type and *meg-3 meg-4* zygotes (ex utero) expressing PGL-3::mCherry (white) and matching the stages shown in the cartoons above. Scale bar is 10 µm and applies to all images. (**B** to **E**) Graphs showing the total number [(B) and (C)] and total volume [(D) and (E)] of PGL-3::mCherry condensates in the posterior (light color) or anterior (dark color) half of wild-type and *meg-3 meg-4* zygotes calculated from the photomicrographs shown in (A). Circles represent the average from five zygotes, and error bars represent the SD. PF, pronuclear formation; PM, pronuclear meeting. (**F** and **G**) Graphs showing the rate of change in radius of PGL-3::mCherry condensates in wild-type and *meg-3 meg-4* zygotes calculated from traces shown in (H) and (I). (**H** and **I**) Graphs showing the evolution of individual PGL-3::mCherry condensates in anterior and posterior regions during polarization in wild-type and *meg-3 meg-4* zygotes. Traces begin at pronuclear formation and end just before pronuclear meeting. (**J**) Photomicrographs of fixed zygotes (mitosis) of indicated genotypes showing the distribution of MBK-2::OLLAS (green) and PGL-3::mCherry (magenta). Note that, in addition to P granules, MBK-2 localizes to centrosomes. Scale bar is 10 µm and applies to all images in the set. (**K**) Graph showing the percentage of PGL-3:mCherry condensates colocalized with MBK-2::OLLAS puncta in wild-type or *meg-3 meg-4* zygotes at the indicated developmental stages. Each circle represents one zygote (>50 puncta), and each line represents the mean. (**L** and **M**) Graphs showing the evolution of individual anterior (dark color) and posterior (light color) PGL-3 condensates under conditions simulating "wild-type" (starting condensate sizes as in wild-type zygotes, high conversion rates, and Pickering agent) and "*meg-3 meg-4*" (starting condensate sizes as in *meg-3 meg-4* zygotes, low conversion rates, and no Pickering agent). Compare with experimental data in (H) and (I).

decrease in PGL-3 condensate volume in wild-type and *meg-3 meg-4* oocytes, indicating that the increase in PGL-3 solubility during the oocyte-to-zygote transition is not dependent on *meg-3* and *meg-4* (Fig. 4, C to F, and fig. S5, A and B). The size distribution of PGL-3 condensates after dissolution, however, was different in the two genotypes. The PGL-3 emulsion coarsened rapidly in *meg-3 meg-4* zygotes with fewer larger condensates dominating the emulsion (Fig. 4, B and E). By contrast, wild-type zygotes maintained many small PGL-3 condensates, consistent with MEG-3 stabilizing the PGL-3 emulsion against coarsening (Fig. 4, B and E). Zygotes in late meiosis can survive outside of the uterus, allowing for the acquisition of high-resolution images ex utero. These images confirmed that wild-type zygotes contain dozens of <1-µm condensates not observed in *meg-3 meg-4* zygotes (Fig. 4, B and F).

These observations suggest that MEG-3 functions as a Pickering agent for the PGL-3 emulsion. To examine the physical plausibility of this hypothesis, we modeled in silico the kinetics of an idealized PGL-3 emulsion in the presence or absence of a Pickering agent that lowers surface tension. To account for the intrinsically slow dynamics of PGL-3 condensates, we modeled PGL-3 dynamics under a "conversion-limited" scheme, where the soluble-to-condensate conversion rate of PGL-3 molecules is much slower than their diffusion-limited adsorption-desorption rate and is therefore rate limiting for condensate growth and degrowth (see supplementary text). To model dissolution of PGL-3 condensates during oocyte maturation, we assigned a relatively high conversion rate to PGL-3 condensates and a high critical concentration for PGL-3 phase separation because most PGL-3 condensates dissolve during this period. To model MEG-3 as a Pickering agent, we reduced the Gibbs-Thomson length (proportional to surface tension) of MEG-3–coated PGL-3 condensates by 100-fold compared with PGL-3–only condensates. Parameter sweeps revealed that reductions in the Gibbs-Thomson length as low as twofold, within the range observed in vitro (Fig. 2B), were sufficient to show the same quantitative behavior (see supplementary text). We used the model to run simulations tracking the dynamics of condensates with starting sizes matching the distribution of PGL-3 condensates in oocytes after dissolution (Fig. 4, G and H). Coarsening was notably different in the presence or absence of the Pickering agent, with the stabilization of smaller condensates requiring the Pickering agent. The simulations reproduced the increase in average condensate volume observed in *meg-3 meg-4* zygotes in comparison to wild-type zygotes (Fig. 4, I and J). These results support the hypothesis that MEG-3 stabilizes

the PGL-3 emulsion against an increase in PGL-3 dynamics in newly fertilized zygotes by lowering the surface tension of PGL-3 condensates.

MEG-3 and MEG-4 drive asymmetric growth of the P granule emulsion during polarization

Next, we examined PGL-3 dynamics as zygotes transition from unpolarized to polarized. During this period, zygotes exit meiosis, assemble pronuclei that migrate to the center of the zygote, and fuse and initiate the first mitotic division, and MEX-5 redistributes in an anterior-rich gradient (Fig. 5A). During pronuclear migration, we observed new PGL-3 condensates appearing throughout the cytoplasm in both wild-type and *meg-3 meg-4* zygotes (Fig. 5, B and C, and fig. S6A). The total volume of PGL-3 in condensates also increased (fig. S6B), without a change in total PGL-3 (fig. S6C), consistent with a return to low PGL-3 solubility in both genotypes. During this period, total condensate volume in the anterior half of the zygote decreased steadily, eventually reaching zero, whereas total condensate volume in the posterior increased (Fig. 5D). Notably, in *meg-3 meg-4* zygotes, we also observed a decrease and increase in total condensate volume in the anterior and posterior, respectively, but the amplitude of the change was greatly reduced (Fig. 5E). By mitosis, in wild-type zygotes, all PGL-3 condensates were restricted to the posterior cytoplasm; by contrast, in *meg-3 meg-4* zygotes, PGL-3 condensates remained stable throughout the cytoplasm through the first cell division (Fig. 5, A to E), suggesting that in *meg-3 meg-4* mutants, the PGL-3 emulsion does not respond efficiently to the MEX-5–driven solubility gradient.

To examine condensate dynamics directly, we tracked individual PGL-3 condensates in live zygotes from pronuclear formation to pronuclear meeting (Fig. 5, F to I, and movie S9). These observations confirmed that in *meg-3 meg-4* zygotes, condensates experience very slow growth and dissolution rates during polarization, with a slight bias for decay in the anterior [anterior rate ($k^{A,avg}$) = −0.19 nm/s; posterior rate ($k^{P,avg}$) = 0.06 nm/s; Fig. 5, G and I]. Growth and decay rates were more than fivefold faster in wild-type zygotes, with a clear bias for dissolution in the anterior and condensation in the posterior ($k^{A,avg}$ = −1.18 nm/s; $k^{P,avg}$ = 0.50 nm/s; Fig. 5, F and H). These findings suggest that unlike in oocytes where PGL-3 dissolution occurs independently of *meg-3* and *meg-4*, during polarization, *meg-3* and *meg-4* are required to accelerate PGL-3 condensate dynamics.

In addition to *meg-3* and *meg-4*, P granule dissolution during polarization requires the DYRK family kinase MBK-2 (*38*, *39*). Genetic epistasis experiments have shown that dissolution activity of MBK-2 is dependent on *meg-3* and *meg-4* (*28*) (see supplementary text). Consistent with these observations, using an epitope-tagged allele of endogenous MBK-2, we found that MBK-2 is recruited to PGL-3 condensates during polarization and that this recruitment is diminished in *meg-3 meg-4* mutant embryos (Fig. 5, J and K). MEG-3 and MBK-2 levels varied more than fivefold among PGL-3 condensates (fig. S6, D to G). Condensate growth rates during polarization were also heterogeneous in a manner that did not correlate with initial condensate size (Fig. 5, F and G). Together, these observations suggest that heterogeneous recruitment of MEG-3 and MBK-2 variably accelerates PGL-3 condensate dynamics during polarization.

Based on these findings, we built on our theoretical model of the PGL-3 emulsion, adding three new features specific to polarization: (i) higher variable conversion rates for wild-type versus *meg-3 meg-4* PGL-3 condensates to reflect heterogeneous fluidization by MEG proteins and MBK-2, (ii) higher solubility for PGL-3 in anterior cytoplasm to reflect the influence of the MEX-5 gradient, and (iii) lower Gibbs-Thomson lengths for posterior condensates to reflect sustained coverage of posterior condensates by MEG proteins ("asymmetric Pickering agent"). Parameters were benchmarked to allow for complete dissolution of anterior granules in ~6 min, as observed in vivo. We used the model to run simulations tracking the growth and decay of condensates matching the size distribution of PGL-3 condensates in vivo before polarization. The simulations faithfully recapitulated the coarsening-free dissolution of anterior condensates and growth of posterior condensates that were observed in wild-type zygotes (Fig. 5L and movie S10). Simulations mimicking conditions in *meg-3 meg-4* mutants reproduced the slow dynamics

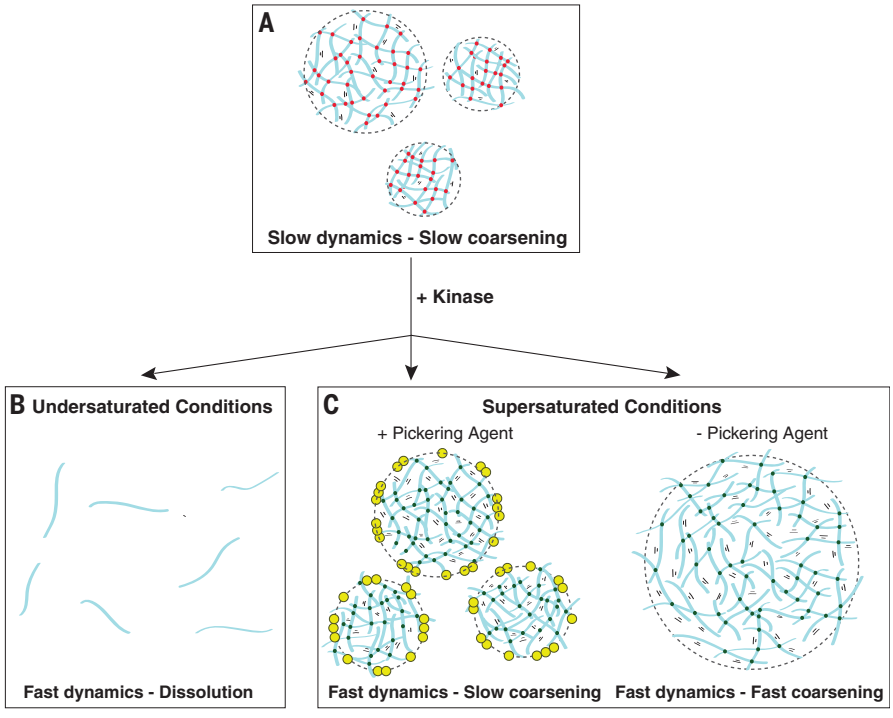


Fig. 6. Pickering agents stabilize dynamic emulsions against kinase-accelerated coarsening. Schematics showing an idealized emulsion of a self-interacting polymer (blue). (**A**) In untreated condensates, polymer-polymer binding and unbinding events are slow (red dots), allowing the emulsion to persist with minimal coarsening over short time scales. (**B** and **C**) Phosphorylation by kinase increases solubility and accelerates internal dynamics [green dots in (C)]. In undersaturated conditions (B), kinase-accelerated dynamics cause the condensates to dissolve, as is observed in the anterior of wild-type polarized zygotes. In saturated conditions (C), kinase-accelerated dynamics allow the condensates to rapidly grow by allowing new molecules from the dilute phase to enter the condensates. In the presence of the Pickering agent (yellow), the condensates are stabilized against coarsening, as observed in the posterior of wild-type polarized zygotes. In the absence of the Pickering agent, the condensates coarsen, as observed in *meg-3 meg-4* mutants during the oocyte-to-embryo transition. During polarization, MEG-3 and MEG-4 function both as Pickering agents and recruiters of the kinase MBK-2. In *meg-3 meg-4* zygotes undergoing polarization, PGL condensates are maintained throughout the cytoplasm with minimal coarsening, because MBK-2 kinase is not recruited to the condensates and PGL dynamics remain slow, as in (A).

observed in those mutants (Fig. 5M and movie S10). The simulations also reproduced the rapid increase in PGL-3 condensate volume (fig. S6H) that is observed during mitosis in wild-type but not *meg-3 meg-4* zygotes (figs. S5A and S6B). The rapid rise is a consequence of the faster approach to equilibrium in the supersaturated environment of the posterior cytoplasm by PGL-3 condensates with accelerated dynamics driven by MEG proteins and MBK-2.

To test the importance of each feature in the model, we reran the wild-type simulations, changing one model feature at a time. Simulations where condensates were assigned uniformly low conversion rates yielded decay rates that were too slow to clear anterior granules under the in vivo time constraints (fig. S6I and movie S10). Simulations with low PGL-3 solubility across the cytoplasm led to the coarsening of anterior condensates (fig. S6J and movie S10). Simulations lacking the Pickering agent led to the coarsening of posterior condensates (fig. S6K and movie S10). Together with the in vitro and in vivo findings, the theory supports a model where MEG proteins facilitate P granule polarization in response to the MEX-5–induced saturation gradient by accelerating PGL-3 conversion dynamics (through recruitment of MBK-2) and by functioning as Pickering agents to lower surface tension on posterior condensates, thus preventing coarsening during periods of fast PGL dynamics (Fig. 6; see supplementary text for a full discussion of theoretical considerations).

Discussion

Since their description by Ramsden and Pickering in the early 1900s (*19*, *20*), Pickering agents have been used widely to stabilize emulsions in the pharmaceutical, energy, and food industries (*40*–*42*). Unlike surfactants (amphiphilic molecules that insert at interfaces), Pickering agents are nanoscale solid particulates that adsorb to interfaces upon partial wetting by both phases. Adsorption is energetically favored and balances the drive to reduce interfacial area, stabilizing the emulsion against coarsening (*43*). Many types of solid particulates have been shown to function as Pickering agents, from silica to denatured proteins (*44*). The first described intrinsically disordered protein, casein, functions as a natural Pickering agent in homogenized milk (*45*). We characterized intrinsically disordered protein MEG-3 as an intracellular Pickering agent, and we speculate that other self-assembling biopolymers will exhibit similar properties. In somatic cells, PGL droplets are covered by EPG-2 clusters that may function like MEG-3 to regulate the size and dynamics of PGL droplets in preparation for autophagy (*46*). Artificial protein-RNA assemblies that adsorb to the surface of stress granules have been reported to influence their size and coalescence (*47*). mRNAs that accumulate on the surface of protein condensates could also serve as stabilizing agents (*48*, *49*). Given the rich diversity of biopolymers in cells, it is tempting to speculate that biopolymers acting as Pickering agents will prove a general organizing principle for biomolecular condensates. By interacting with enzymes like the DYRK family kinase MBK-2, biological Pickering agents also regulate interfacial exchange to control the flux of molecules in and out of condensates in response to environmental changes. In this respect, surface-adsorbed biopolymers fulfill a boundary function for biomolecular condensates, reminiscent of the role of lipid bilayers and associated machineries (e.g., channels) in membrane-bound organelles. Indeed, it has been speculated that lipid membranes arose from liposomes that functioned as Pickering stabilizers for aqueous emulsions (*50*).

REFERENCES AND NOTES

1. A. A. Hyman, C. A. Weber, F. Jülicher, *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
2. L. Jawerth *et al*., *Science* **370**, 1317–1323 (2020).
3. S. Boeynaems *et al*., *Trends Cell Biol.* **28**, 420–435 (2018).
4. E. S. Freeman Rosenzweig *et al*., *Cell* **171**, 148–162.e19 (2017).
5. M. Feric, C. P. Brangwynne, *Nat. Cell Biol.* **15**, 1253–1259 (2013).
6. J. Berry, S. C. Weber, N. Vaidya, M. Haataja, C. P. Brangwynne, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5237–E5245 (2015).
7. S. Elbaum-Garfinkle *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7189–7194 (2015).
8. S. Ranganathan, E. I. Shakhnovich, *eLife* **9**, e56159 (2020).
9. J. Berry, C. P. Brangwynne, M. Haataja, *Rep. Prog. Phys.* **81**, 046601 (2018).
10. D. T. McSwiggen, M. Mir, X. Darzacq, R. Tjian, *Genes Dev.* **33**, 1619–1634 (2019).
11. J. R. Wheeler, T. Matheny, S. Jain, R. Abrisch, R. Parker, *eLife* **5**, e18413 (2016).
12. J. E. Lee, P. I. Cathey, H. Wu, R. Parker, G. K. Voeltz, *Science* **367**, eaay7108 (2020).
13. D. S. W. Lee, N. S. Wingreen, C. P. Brangwynne, *Nat. Phys.* **17**, 531–538 (2021).
14. C. F. Lee, C. P. Brangwynne, J. Gharakhani, A. A. Hyman, F. Jülicher, *Phys. Rev. Lett.* **111**, 088101 (2013).
15. C. A. Weber, C. F. Lee, F. Jülicher, *New J. Phys.* **19**, 053021 (2017).
16. D. Zwicker, A. A. Hyman, F. Jülicher, *Phys. Rev. E* **92**, 012317 (2015).
17. P. C. Bressloff, *Phys. Rev. E* **101**, 042804 (2020).
18. J. D. Wurtz, C. F. Lee, *Phys. Rev. Lett.* **120**, 078102 (2018).
19. W. Ramsden, F. Gotch, *Proc. R. Soc. London* **72**, 156–164 (1904).
20. S. U. Pickering, *J. Chem. Soc. Trans.* **91**, 2001–2021 (1907).
21. C. P. Brangwynne *et al*., *Science* **324**, 1729–1732 (2009).
22. S. Saha *et al*., *Cell* **166**, 1572–1584.e16 (2016).
23. M. Hanazawa, M. Yonetani, A. Sugimoto, *J. Cell Biol.* **192**, 929–937 (2011).
24. U. Sheth, J. Pitt, S. Dennis, J. R. Priess, *Development* **137**, 1305–1314 (2010).
25. C. M. Schubert, R. Lin, C. J. de Vries, R. H. Plasterk, J. R. Priess, *Mol. Cell* **5**, 671–682 (2000).
26. J.-X. Chen *et al*., *Mol. Cell. Proteomics* **15**, 1642–1657 (2016).
27. J. Smith *et al*., *eLife* **5**, e21337 (2016).
28. J. T. Wang *et al*., *eLife* **3**, e04591 (2014).
29. A. Putnam, M. Cassani, J. Smith, G. Seydoux, *Nat. Struct. Mol. Biol.* **26**, 220–226 (2019).
30. Y. Wu *et al*., *Mol. Biol. Cell* **30**, 333–345 (2019).
31. T. Ngai, S. Bon, Eds., *Particle-Stabilized Emulsions and Colloids* (Royal Society of Chemistry, 2014); https://pubs.rsc.org/en/content/ebook/978-1-84973-881-1.
32. L. M. Jawerth *et al*., *Phys. Rev. Lett.* **121**, 258101 (2018).
33. A. B. Pawar, M. Caggioni, R. Ergun, R. W. Hartel, P. T. Spicer, *Soft Matter* **7**, 7710–7716 (2011).
34. C. D. Crowe, C. D. Keating, *Interface Focus* **8**, 20180032 (2018).
35. K. M. Pang *et al*., *Dev. Biol.* **265**, 127–139 (2004).
36. A. K. Rai, J.-X. Chen, M. Selbach, L. Pelkmans, *Nature* **559**, 211–216 (2018).
37. A. Patel *et al*., *Science* **356**, 753–756 (2017).
38. S. Quintin, P. E. Mains, A. Zinke, A. A. Hyman, *EMBO Rep.* **4**, 1175–1181 (2003).
39. J. Pellettieri, V. Reinke, S. K. Kim, G. Seydoux, *Dev. Cell* **5**, 451–462 (2003).
40. C. C. Berton-Carabin, K. Schroën, *Annu. Rev. Food Sci. Technol.* **6**, 263–297 (2015).
41. F. Goodarzi, S. Zendehboudi, *Can. J. Chem. Eng.* **97**, 281–309 (2019).
42. C. Albert *et al*., *J. Control. Release* **309**, 302–332 (2019).
43. D. Gonzalez Ortiz, C. Pochat-Bohatier, J. Cambedouzou, M. Bechelany, P. Miele, *Engineering* **6**, 468–482 (2020).
44. Y. Yang *et al*., *Front. Pharmacol.* **8**, 287 (2017).
45. E. M. Redwan, B. Xue, H. A. Almehdar, V. N. Uversky, *Curr. Protein Pept. Sci.* **16**, 228–242 (2015).
46. G. Zhang, Z. Wang, Z. Du, H. Zhang, *Cell* **174**, 1492–1506.e22 (2018).
47. M. Garcia-Jove Navarro *et al*., *Nat. Commun.* **10**, 3230 (2019).
48. B. Hampoelz *et al*., *Cell* **179**, 671–686.e17 (2019).
49. D. Tauber *et al*., *Cell* **180**, 411–426.e16 (2020).
50. D. C. Dewey, C. A. Strulson, D. N. Cacace, P. C. Bevilacqua, C. D. Keating, *Nat. Commun.* **5**, 4670 (2014).
51. C. F. Lee, chiufanlee/Pickering-stabilization, Version v1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5083065.

ACKNOWLEDGMENTS

We thank the Johns Hopkins Integrated Imaging Center (S10OD023548) for microscopy support. We thank the Lavis lab for HaloTag ligands JF_{549} and JF_{646}, the Griffin lab for the MEG-3::Halo strain, the Waugh lab for TEV protease (pRK793, Addgene), T. Hyman, the Baltimore Worm Club, and the Seydoux lab for many helpful discussions. **Funding:** This work was supported by the National Institutes of Health (grant numbers R37HD037047 to G.S. and F32GM134630 to A.P.). G.S. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** A.W.F., A.P., and G.S. designed the research. A.W.F. and A.P. performed all experiments and collected and analyzed data. C.F.L. performed the theoretical analysis. A.W.F., A.P., C.F.L., and G.S. prepared the manuscript with contributions from all authors. **Competing interests:** G.S. serves on the scientific advisory board of Dewpoint Therapeutics, Inc. A.W.F., A.P., and G.S. are inventors on provisional application #63/094,987 filed on 10/22/2020 held by Johns Hopkins University that covers the use of intrinsically disordered proteins as Pickering agents. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Code for simulations is deposited and accessible at Zenodo (*51*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abg7071
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*52*–*69*)
MDAR Reproducibility Checklist
Movies S1 to S10

22 January 2021; accepted 21 July 2021
10.1126/science.abg7071

RESPIRATORY ENZYMES

The three-spin intermediate at the O–O cleavage and proton-pumping junction in heme–Cu oxidases

Anex Jose[1], Andrew W. Schaefer[1], Antonio C. Roveda Jr.[1], Wesley J. Transue[1], Sylvia K. Choi[2], Ziqiao Ding[2], Robert B. Gennis[2], Edward I. Solomon[1,3]*

Understanding the mechanistic coupling of molecular oxygen reduction and proton pumping for adenosine triphosphate synthesis during cellular respiration is the primary goal of research on heme-copper oxidases—the terminal complex in the membrane-bound electron transport chain. Cleavage of the oxygen-oxygen bond by the heme-copper oxidases forms the key intermediate P_M, which initiates proton pumping. This intermediate is now experimentally defined by variable-temperature, variable-field magnetic circular dichroism spectroscopy on a previously unobserved excited state feature associated with its heme iron(IV)-oxo center. These data provide evidence that the iron(IV)-oxo in P_M is magnetically coupled to both a copper(II) and a cross-linked tyrosyl radical in the active site. These results provide new insight into the oxygen-oxygen bond cleavage and proton-pumping mechanisms of heme-copper oxidases.

Aerobic organisms use O_2 to drive the respiratory electron transport chain that governs production of adenosine triphosphate (ATP), the universal energy currency in biology. The four-electron reduction of O_2 to H_2O underlying this biological energy transduction process is catalyzed by the superfamily of metalloenzymes called heme-copper oxidases (HCOs), most notably cytochrome *c* oxidases (C*c*Os) and ubiquinol oxidases (UbOs). The free energy gained in this exergonic reaction is used by the HCOs to pump protons across the inner mitochondrial or bacterial membrane to create a proton gradient used by ATP synthase, also bound to the membrane, for ATP synthesis (*1*, *2*). Oxygen reduction in HCOs takes place at an active site that consists of a heme center ligated by an axial histidine (His) residue and a copper center (Cu_B) ligated by three His residues, one of which is posttranslationally modified to form a covalent cross-link with a nearby tyrosine (Tyr) residue (fig. S1A). HCOs also have a bis-His ligated low spin (LS) heme and a binuclear Cu_A center (only in C*c*Os) for electron transfer (ET) to the active site.

The current consensus on the mechanism of O_2 reduction by HCOs (fig. S1B) (*1*–*7*) is that when O_2 binds to the mixed valent enzyme (MV), the O–O bond is rapidly cleaved to form intermediate P_M. Because MV comprises a fully reduced heme–Cu active site and fully oxidized other redox centers, this P_M formation indicates that all four requisite electrons for O–O bond cleavage are supplied by the active site. Only the source of two of the electrons has been clearly identified as the heme iron ($Fe^{II}\rightarrow Fe^{IV}$); whereas, one electron each is commonly thought to be donated by Cu_B ($Cu^{I}\rightarrow Cu^{II}$) and the cross-linked Tyr ($Tyr\rightarrow Tyr^{\bullet}$; tyrosyl radical) (*8*–*11*).

Babcock and co-workers have used radioactive iodine to probe the possible formation of the $Tyr^{\bullet}$ in P_M through its labeling (*9*). However, only ~2×10^{-3} mole fraction of the P_M formed was found to be labeled. Later studies have also suggested a $Tyr^{\bullet}$ in P_M (*12*–*15*), and electron paramagnetic resonance (EPR) studies on P_M generated by the reaction of the oxidized enzyme (O) with H_2O_2 have shown ~2 to 5% of $Tyr^{\bullet}$ (not necessarily the cross-linked Tyr) but still lacked the Cu(II) signal (*16*, *17*). Thus, evidence for the presence of Cu(II) and $Tyr^{\bullet}$ during the oxygen reduction reaction in HCOs remains elusive. Studies of intermediates in the catalytic cycle of HCOs are thus far limited to the heme side of the heme–Cu active site using resonance Raman (rR) and ultraviolet-visible (UV-vis) absorption spectroscopies. The intense features of the two hemes present in these enzymes have precluded the study of the oxidation and protonation states of the Cu_B and Tyr in the active site.

P_M is the first intermediate in the catalytic cycle where the O–O bond is fully cleaved, and HCOs pump the first proton across the membrane upon its one-electron reduction (fig. S1B). Thus, an understanding of the geometric and electronic structural contributions of the heme–Cu active site to O–O bond cleavage and how this is coupled to proton pumping are important issues remaining to be explored in HCO literature that can be addressed through the experimental definition of P_M. This is accomplished in this study by variable-temperature, variable-field magnetic circular dichroism (VTVH-MCD) spectroscopy on a previously unobserved excited-state feature associated with the heme Fe^{IV}=O in P_M to study its ground state.

[1]Department of Chemistry, Stanford University, Stanford, CA 94305, USA. [2]Department of Biochemistry, University of Illinois, Urbana, IL 61801, USA. [3]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, USA.
*Corresponding author. Email: solomone@stanford.edu

Results and analysis

Generating P_M in ubiquinol oxidase bo_3

We chose to study P_M in *Escherichia coli* cytochrome bo_3 UbO because it lacks strong absorption features of the Cu_A center in the near-infrared (NIR) energy region important for this study. We performed the photoreaction of the CO-bound MV with O_2 in D_2O at pH 10 (CAPS buffer), thereby generating P_M (supplementary text and figs. S2 and S3), which is stable for several minutes (*10*, *18*–*22*).

MCD spectrum of P_M

We characterized the electronic and magnetic properties of P_M and its decayed form F and O by MCD spectroscopy. At 24 K and 6 T (Fig. 1A), these spectra are dominated by the intense features from $\pi\rightarrow\pi^*$ intraligand and charge transfer transitions of hemes: The LS heme-*b* shows characteristic features at energies <11,000 cm^{-1}, and both hemes (heme-*b* and heme-*o*) have overlapping features at energies >15,000 cm^{-1}. Between these regions, a relatively weak derivative-shaped feature is present in P_M, centered at ~13,900 cm^{-1} (Fig. 1B, red), that is absent in the next intermediate F (Fig. 1B, orange) and in O (Fig. 1B, blue). We therefore conclude that this is a clear characteristic feature of P_M. In C*c*O, this relatively weak feature at ~13,900 cm^{-1} would be obscured by overlap with an intense MCD feature from Cu_A (the $\Psi\rightarrow\Psi^*$ transition) (*23*). A similar MCD feature (~13,500 cm^{-1}) is observed for the alkaline form of myoglobin compound-II (Mb-II) (from horse heart), characteristic of its isolated Fe^{IV}=O heme (*24*). On the basis of the similar natures of these MCD features (band shapes and energies) and the similarities in their heme Fe^{IV}=O centers (supplementary text and fig. S4), we use Mb-II as an isolated Fe^{IV}=O heme reference (i.e., lacking the nearby Cu_B center) for comparison with this center in P_M.

VTVH-MCD of P_M

The VTVH interrogation of this derivative-shaped MCD feature of P_M and Mb-II gives the sets of isofield curves shown in Fig. 2, A and B, respectively. These show similar isofield behaviors at higher temperatures (>15 K; $1/T < 0.07$ K^{-1}) but major differences in their low-temperature behavior (<15 K). Whereas the isofield MCD intensity plots of Mb-II decrease to a temperature-independent value at low temperatures, the isofield MCD plots of P_M decrease below 20 K to a nadir around 8 K before rising again as the temperature is further decreased to 1.7 K. This rise in intensity at low temperature indicates that P_M has a paramagnetic ground state. This

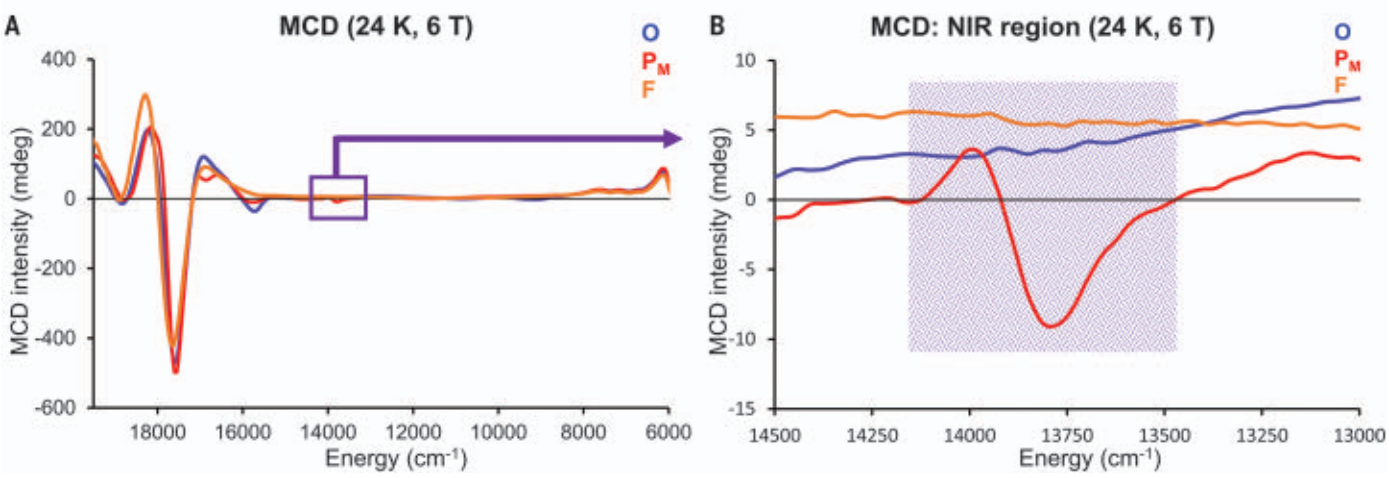


Fig. 1. MCD spectrum of O, P_M, and F. (**A**) MCD spectrum of O, pH 10 (blue); P_M (red); and F (orange) at 24 K and 6 T [the enlarged region shown in (B) is marked with a square]. (**B**) MCD spectrum in the NIR region of O (blue), P_M (red), and F (orange) at 24 K and 6 T (the derivative-shaped MCD feature in P_M is highlighted). mdeg, millidegrees.

can only be true if the Fe^{IV}=O [S = 1, axial zero-field splitting (ZFS), D ~ +20 cm^{-1}] heme in P_M is magnetically coupled to additional paramagnetic center(s).

Fit to the VTVH-MCD data of P_M: Coupling to additional paramagnetic centers

The analysis and fit of the VTVH-MCD data of Mb-II, including orientation averaging and out-of-state magnetic field effects, are presented in the supplementary text and figs. S5 and S6. Because the VTVH-MCD data of P_M reveal the presence of additional paramagnetic center(s) interacting with the Fe^{IV}=O, the influence of isotropic exchange coupling $\left(-2\vec{J}\hat{S}_1 \cdot \hat{S}_2\right)$ of this zero-field split S = 1 center with additional spin center(s) was assessed.

First, the exchange coupling of the Fe^{IV}=O with a single S = ½ center was evaluated. Floating the Fe^{IV}=O ZFS (D) and exchange interaction (J_{12}) gave D = +19.7 cm^{-1} and J_{12} = −4.6 cm^{-1}. This best fit greatly deviates from the data (Fig. 2C) at low-temperature–low-field conditions (T < 2.5 K, H = 2 to 4 T), where the fit intensity continues to increase as the data rapidly saturate, and the data at low-temperature–high-field conditions (T < 10 K, H = 7 T), where the fit greatly underestimates the observed intensity. Thus, the VTVH-MCD data of P_M require the presence of a third spin center (S = ½), where all spin centers have the superexchange pathways required for exchange coupling. We therefore evaluated three-spin models, where the Fe^{IV}=O center is exchange-coupled with an S = ½ center (J_{12}), which is in turn exchange-coupled with the second S = ½ center (J_{23}), and a model with the Fe^{IV}=O center exchange-coupled to both the S = ½ centers (J_{12} and J_{13}). Both models greatly improved the fit to the data, demonstrating the requirement of a three-spin system in P_M (fig. S7). The best fit was obtained when all spin centers were allowed to have exchange interactions with each other. The final fit to the VTVH-MCD data and the spin Hamiltonian parameters of P_M are given in Fig. 2A (contributions to the MCD intensity of P_M relative to Mb-II are compared in fig. S8).

On the basis of the geometric and electronic structural model of P_M obtained through proton-initiated O–O bond cleavage of the generally assumed peroxo-bridged Fe^{III}/Cu^{II} precursor (I_P in fig. S1B), we assign the two additional S = ½ spin centers in P_M revealed by the VTVH-MCD data as the Cu(II) and the cross-linked $Tyr^{\bullet}$ (*7*). One of these S = ½ centers has a larger exchange coupling with the Fe^{IV}=O than the other, which requires better orbital overlap with the Fe^{IV}=O; we therefore assign this to the Cu(II) center. The third spin has superexchange pathways with both the Fe^{IV}=O and the Cu(II) centers, and these exchange interactions triangulate its location to the cross-linked $Tyr^{\bullet}$. On the basis of these data, we obtain the magnetic coupling scheme of P_M (Fig. 2D). The superexchange pathways that enable these exchange interactions are evaluated in the final section of the analysis using the density functional theory (DFT) model of P_M derived from the proton-initiated O–O bond cleavage in (*7*).

Spin Hamiltonian analysis

The best three-spin fit of the VTVH-MCD data provides the energy levels in Fig. 2E, where the magnetic field is along the z axis, which is the dominant contributor to MCD intensity of an xy-polarized transition (*25*). With $D \gg$ all $|J|$ values, the magnetic sublevels form a lower-energy group of four (Fig. 2E, red and black) and a higher-energy group of eight (Fig. 2E, orange and blue). VTVH-MCD intensity is dictated by the spin expectation value and Boltzmann population of each sublevel, so the lowest four-level grouping dominates the observed low-temperature (T < 10 K) MCD behavior. These levels originate from coupling the nonmagnetic M_S = 0 Fe^{IV}=O sublevel with the two S = ½ spin centers, providing direct insight into their exchange couplings. The paramagnetic $M_{S,tot}$ = −1 (S_{tot} = 1, black) is the ground sublevel when H > 2 T, giving rise to MCD intensity at the lowest temperature, 1.7 K in Fig. 2A. As the temperature is increased from 1.7 K, population density is redistributed into two $M_{S,tot}$ = 0 sublevels (from S_{tot} = 1 and 0, black and red) and an $M_{S,tot}$ = +1 sublevel (S_{tot} = 1, black), reducing the MCD intensity and producing the intensity dip at ~8 K in Fig. 2A. Upon further increase in temperature, the higher energy $M_{S,tot}$ = −1 and −2 sublevels (from the Fe^{IV}=O S = 1, M_S = ±1 coupled to the two S = ½, orange and blue) are populated, and the MCD intensity increases, leading to a peak at ~20 K. At still higher temperatures, the MCD intensity decreases according to the Curie behavior.

The two-spin system provided a much poorer best fit to the VTVH-MCD data of P_M in Fig. 2C because it would have an S_{tot} = ½ lowest energy spin state (fig. S9). The Zeeman splitting of the $M_{S,tot}$ = ±½ is half of that of the $M_{S,tot}$ = ±1, and thus the faster rate of saturation of the MCD intensity at lower temperatures observed in the isofield plots is not feasible in a two-spin system (2 to 4 T; Fig. 2C). Therefore, it is the low-temperature, low-magnetic field region of the VTVH-MCD data in P_M that conclusively establishes a magnetically coupled three-spin system of the Fe^{IV}=O, the Cu(II), and the cross-linked $Tyr^{\bullet}$ centers.

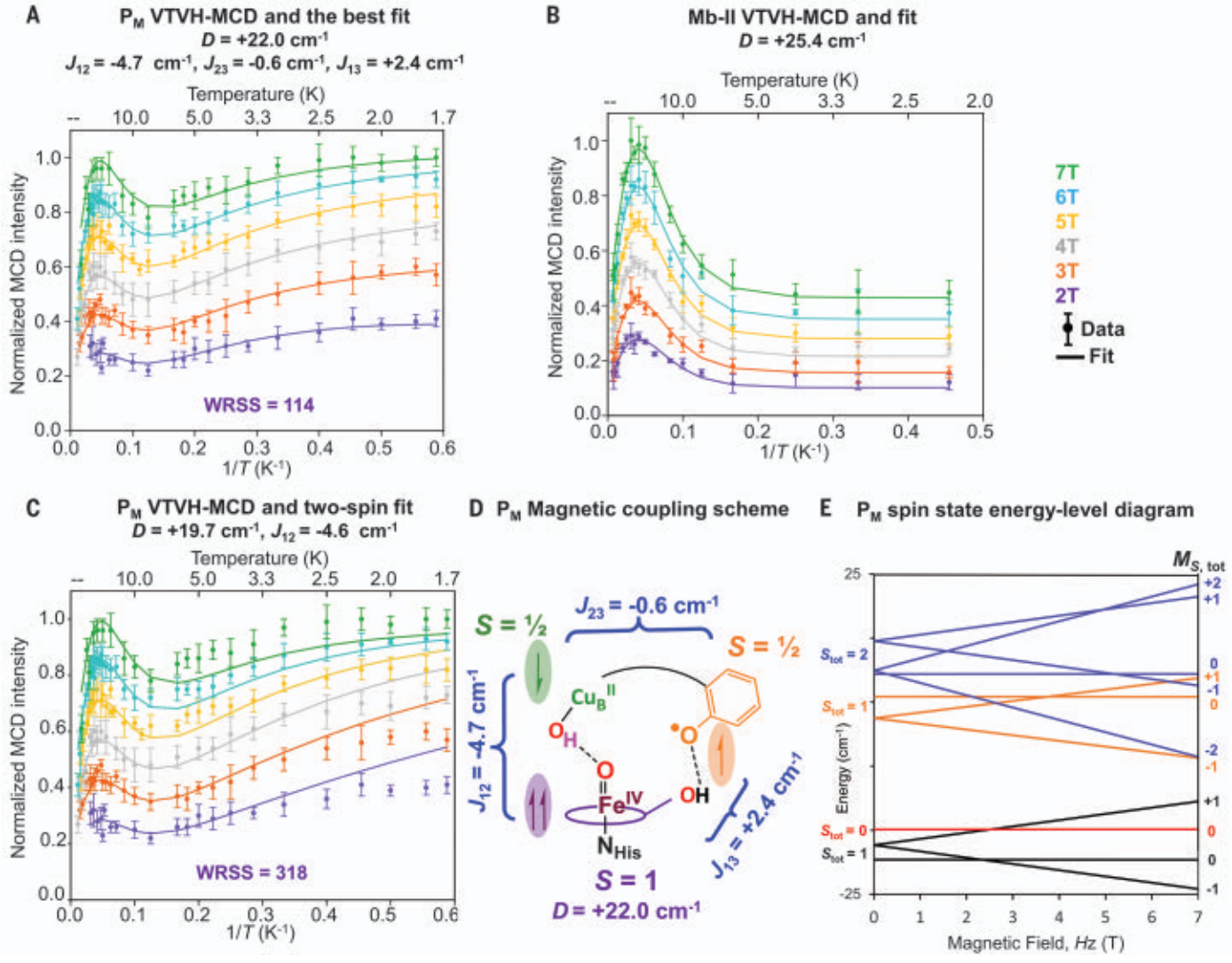


Fig. 2. VTVH-MCD data and analysis of P_M and Mb-II. (**A** and **B**) VTVH-MCD isofield data at 2 to 7 T and the best fit for P_M (three-spin, S_1 = 1 and S_2 = S_3 = ½) (A) and Mb-II (S = 1) (B). (**C**) VTVH-MCD data of P_M fit using two-spin model (S_1 = 1 and S_2 = ½). (**D**) The best fit [three-spin model, solid lines in (A)] derived magnetic coupling scheme for P_M. (**E**) Spin state energy-level diagram of P_M [magnetic field along the z-axis (H_z); S_{tot} and corresponding $M_{S,tot}$ levels are labeled and color coded]. VTVH-MCD intensities are obtained by measuring peak-to-trough intensity of the derivative-shaped MCD features of P_M and Mb-II (fig. S4A). The data are normalized to the highest intensity (P_M: 1.7 K, 7 T; Mb-II: 24 K, 7 T) and are means ± standard deviations. All fits with xy-polarization. The 1-T data have been omitted because of the poor signal-to-noise quality. T, Tesla; T, temperature; WRSS, weighted residual squared-sum. See supplementary text for details of spin Hamiltonian parameters used.

Geometric and electronic structure of P_M

The geometric and electronic structural model of P_M that was obtained in (*7*) had a singlet ground state arising from weak antiferromagnetic (AF) coupling between the Fe^{IV}=O and Cu^{II}–OH and weak ferromagnetic coupling between the Cu^{II}–OH and $Tyr^{\bullet}$. We obtain a lower-energy structure [ΔH = −1.8 kcal/mol with respect to the structure in (*7*)] upon reoptimizing it after including a small structural perturbation based on features present in the crystal structures of HCOs (Fig. 3A; calculated spin densities and J values are given at the bottom). The N(His)–C(Tyr) cross-link has rotated by ~9° and tilted by ~4° with respect to the structure in (*7*). The calculated J values of this structure agree with the results from the VTVH-MCD data and further validate that the second and third spin centers are the Cu(II) and the cross-linked $Tyr^{\bullet}$. To understand the nature of the magnetic interactions and their associated superexchange pathways, we examined the molecular orbitals (MOs) responsible for the exchange interactions (schematic representation of the MOs are shown in Fig. 3, B to E; the corresponding calculated MO contours are given in fig. S10).

The AF coupling between Fe^{IV}=O and Cu^{II}–OH arises from the superexchange pathway mediated by the hydrogen bonding interaction between the oxo(Fe^{IV}) and hydroxo(Cu^{II}) ligands (Fig. 3A, heavy atom distance, 2.85 Å; and Fig. 3, B and C), whereas the AF coupling between Cu(II) and $Tyr^{\bullet}$ is facilitated by the superexchange pathway through the His σ-orbital, which results in the overlap between the Cu(II) and $Tyr^{\bullet}$ magnetic orbitals (Fig. 3D). The distortion of the N(His)–C(Tyr) cross-link described above is responsible for this σ/π overlap. The ferromagnetic coupling between Fe(IV) and $Tyr^{\bullet}$ is promoted by the hydrogen bonding interaction between the hydroxyl of the hydroxyethylfarnesyl of the heme and the $Tyr^{\bullet}$ (Fig. 3A, heavy atom distance, 2.82 Å; and Fig. 3E).

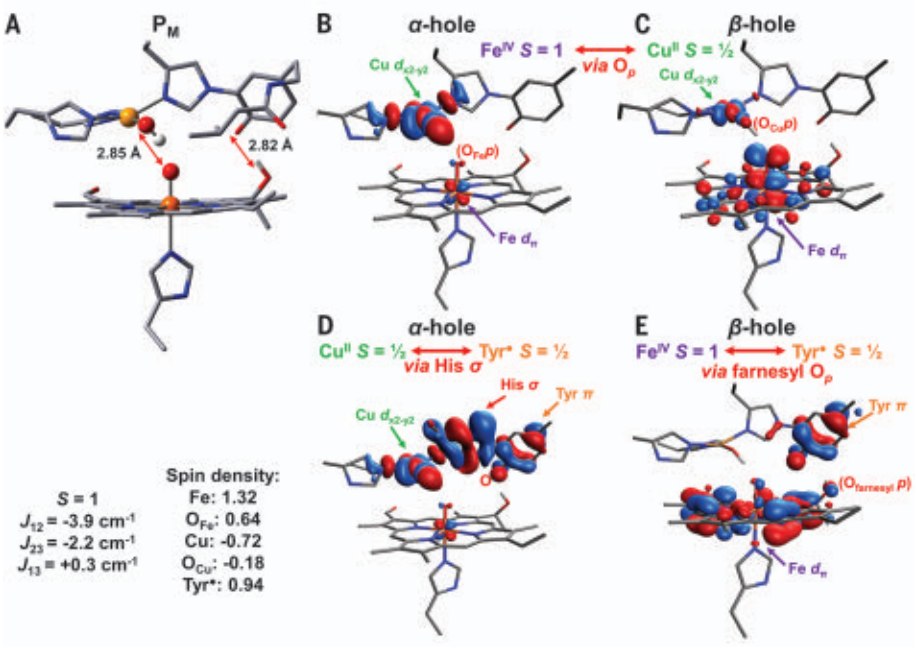


Fig. 3. DFT-optimized structure and MOs involved in the superexchange pathways of P_M. (**A**) DFT-optimized structure of P_M in agreement with the VTVH-MCD data. Hydrogen bonding interactions between Fe^{IV}=O and Cu^{II}–OH and the hydroxyl of hydroxyethylfarnesyl and Tyr• are shown with red arrows; heavy atom (O–O) distances are given. Calculated exchange couplings between magnetic centers and their spin densities are given at the bottom. H_2O hydrogen bonded to Tyr• is removed for clarity. (**B** to **E**) Representative schematics of MOs involved in the superexchange pathway between Fe (IV), Cu(II), and Tyr•: α-hole on Cu^{II} d_{x2-y2} ($+O_{Cu}$ p) mixed into the Fe^{IV} d_π ($+O_{Fe}$ p) (B), β-hole on Fe^{IV} d_π ($+O_{Fe}$ p) mixed into the Cu^{II} d_{x2-y2} ($+O_{Cu}$ p) (C), α-hole on Cu^{II} d_{x2-y2} mixed into the Tyr π-orbital through the His σ-orbital (D), and β-hole on Tyr• π-orbital mixed into Fe^{IV} d_π [orthogonal to the Fe^{IV} d_π in (C)] through the $O_{farnesyl}$ p-orbital (E). The valine residue in (A) is removed for clarity in (B) to (E).

Discussion

The VTVH-MCD on the newly observed excited state feature associated with the heme Fe^{IV}=O in P_M reveals a three-spin exchange-coupled system involving Fe^{IV}=O (S = 1), Cu(II) (S = ½), and Tyr• (S = ½) in the active site. Definition of the geometric and electronic structure of P_M gives further insight into the O–O bond cleavage mechanism in HCOs. Our past DFT results showed that the Tyr donates the fourth electron for the O–O bond cleavage through the His–Tyr cross-link superexchange pathway, which agrees with the VTVH-MCD data showing the exchange interaction between Cu(II) and Tyr•. It has also been concluded that the Tyr supplies the proton required to cleave the O–O bond in I_p (*7*). Because a valine (Val) residue near the active site prevents the direct proton transfer (PT) from the Tyr to the O bound to Cu (O_{Cu}) (*7*), the proton must first transfer to the O_{Fe} (through a water molecule in the active site) and then rotate to transfer to the O_{Cu}. This leads to the O–O bond cleavage resulting in the P_M structure in Fig. 3A, which agrees with the VTVH-MCD results from this study (Fig. 2A).

The ground state exchange coupling constant, J, is related to the electronic coupling matrix element, H_{DA}, for ET as both involve the same superexchange pathway (*26*, *27*). By experimentally determining the ground state exchange interactions between the spin centers in P_M and by defining the superexchange pathways involved using DFT, our results elucidate an ET pathway from the heme Fe^{IV}=O to the cross-linked Tyr• in HCOs. Proceeding along the catalytic cycle, when the Fe^{IV}=O center in P_M receives an electron from the LS heme, the hydrogen bond–enabled superexchange pathway facilitates ET from the Fe^{IV}=O to the Cu^{II}–OH and then to the Tyr• through the pathway enabled by the σ/π overlap of the His and the Tyr. From calculations, the addition of an electron to the structure of P_M in Fig. 3A results in the reduction of Tyr• to tyrosinate (Tyr^-) (fig. S11A).

Intensive efforts by researchers over several decades have contributed to a molecular-level understanding of proton pumping in HCOs (*1*, *2*). This ET from the LS heme to P_M initiates proton pumping (*28*). Coupled to this ET is the arrival of a pump proton ($_pH^+$) to a proton-loading site (PLS) located close to the active site. The PLS has been generally proposed to be a propionate of the active site heme or nearby residues (*1*, *2*, *29*); however, because the calculations show that reduction of P_M involves Tyr•→Tyr^-, this Tyr^- will have higher proton affinity than the propionate and other nearby residues, which indicates that the tyrosinate can be the initial residue that accepts $_pH^+$ to compensate its additional negative charge. A plausible PT pathway from the glutamic acid (Glu) residue to the Tyr^- is shown in fig. S12, A and B [this Glu is at the end of the D proton channel and is thought to be the branching point that supplies chemical protons ($_cH^+$) to the active site and $_pH^+$ to the PLS (*1*, *2*)]. Although the DFT calculations of the one-electron reduced form of P_M show that protonation of the Cu^{II}–OH is thermodynamically favored (ΔG ~9 kcal/mol) relative to the protonation of the Tyr^- (fig. S11, B and C) (*30*, *31*), because of the higher barrier of this PT to the Cu^{II}–OH (*32*) (forming the thermodynamic product), initial PT from the Glu to the Tyr^- would be kinetically favored [PT pathway from the Glu to the Cu^{II}–OH proposed in (*2*) is shown in fig. S12, C and D]. Once the Tyr^- is protonated, this proton likely will not transfer to the Cu^{II}–OH. This is because the direct PT is hindered by the Val near the active site, and the pathway that enabled this PT for O–O bond cleavage in I_P is not available in P_M (as the cleaved O–O bond results in an Fe^{IV}=O with a low proton affinity) (*7*). The subsequent $_cH^+$ transfer from the Glu to the Cu^{II}–OH repels $_pH^+$ from the Tyr to outside the membrane, forming intermediate F. The structure of F with a cross-linked Tyr^- is consistent with Fourier transform infrared studies on F where the Tyr was found to be deprotonated at pH 6.5 to 9.0 (*33*). This model accounts for consumption of two protons from the N (negative) side of the membrane and a proton that is ejected to the P (positive) side during P_M→F (*34*). Thus, Tyr can be directly involved in proton pumping either by serving as an initial residue that accepts $_pH^+$ and transfers it to PLS or by serving as the PLS itself. In this context, it is also notable that the only proton channel in the type-B family of HCOs ends near this Tyr.

The combination of our experimentally calibrated DFT study of O–O cleavage in (*7*) and the VTVH-MCD results presented here for the resultant intermediate P_M also provide initial descriptions of the geometric and electronic structures of the one-electron reduced form of P_M, its protonated forms, and its thermodynamic product F. Further experiments are underway to spectroscopically define these key intermediates for insight into how HCOs couple the O–O bond cleavage to the function of proton pumping for ATP synthesis.

REFERENCES AND NOTES

1. M. Wikström, K. Krab, V. Sharma, *Chem. Rev.* **118**, 2469–2490 (2018).
2. S. Yoshikawa, A. Shimada, *Chem. Rev.* **115**, 1936–1989 (2015).
3. S. M. Adam *et al.*, *Chem. Rev.* **118**, 10840–11022 (2018).
4. A. W. Schaefer, M. T. Kieber-Emmons, S. M. Adam, K. D. Karlin, E. I. Solomon, *J. Am. Chem. Soc.* **139**, 7958–7973 (2017).
5. W.-G. Han Du, L. Noodleman, *Inorg. Chem.* **54**, 7272–7290 (2015).
6. F. Poiana *et al.*, *Sci. Adv.* **3**, e1700279 (2017).
7. A. W. Schaefer, A. C. Roveda Jr., A. Jose, E. I. Solomon, *J. Am. Chem. Soc.* **141**, 10068–10081 (2019).
8. M. Fabian, G. Palmer, *Biochemistry* **34**, 13802–13810 (1995).
9. D. A. Proshlyakov *et al.*, *Science* **290**, 1588–1591 (2000).
10. D. A. Proshlyakov, M. A. Pressler, G. T. Babcock, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8020–8025 (1998).
11. V. R. I. Kaila, M. P. Johansson, D. Sundholm, L. Laakkonen, M. Wiström, *Biochim. Biophys. Acta* **1787**, 221–233 (2009).
12. T. Uchida, T. Mogi, T. Kitagawa, *Biochemistry* **39**, 6669–6678 (2000).
13. M. Iwaki, A. Puustinen, M. Wikström, P. R. Rich, *Biochemistry* **45**, 10873–10885 (2006).
14. V. S. Oganesyan *et al.*, *J. Biol. Inorg. Chem.* **15**, 1255–1264 (2010).
15. D. A. Proshlyakov, *Biochim. Biophys. Acta* **1655**, 282–289 (2004).
16. M. A. Yu *et al.*, *J. Am. Chem. Soc.* **134**, 4753–4761 (2012).
17. K. Budiman *et al.*, *Biochemistry* **43**, 11709–11716 (2004).
18. I. Szundi, J. Ray, A. Pawate, R. B. Gennis, O. Einarsdóttir, *Biochemistry* **46**, 12568–12578 (2007).
19. M. Karpefors, P. Ädelroth, A. Namslauer, Y. Zhen, P. Brzezinski, *Biochemistry* **39**, 14664–14669 (2000).
20. B. Chance, C. Saronio, J. S. Leigh Jr., *J. Biol. Chem.* **250**, 9226–9237 (1975).
21. K. Oda, T. Ogura, E. H. Appelman, S. Yoshikawa, *FEBS Lett.* **570**, 161–165 (2004).
22. E. Pinakoulaki *et al.*, *J. Biol. Chem.* **288**, 20261–20266 (2013).
23. D. R. Gamelin *et al.*, *J. Am. Chem. Soc.* **120**, 5246–5263 (1998).
24. N. Foote, P. M. A. Gadsby, C. Greenwood, A. J. Thomson, *Biochem. J.* **261**, 515–522 (1989).
25. F. Neese, E. I. Solomon, *Inorg. Chem.* **38**, 1847–1865 (1999).
26. T. C. Brunold, D. R. Gamelin, E. I. Solomon, *J. Am. Chem. Soc.* **122**, 8511–8523 (2000).
27. R. G. Hadt, S. I. Gorelsky, E. I. Solomon, *J. Am. Chem. Soc.* **136**, 15034–15045 (2014).
28. I. Belevich, M. I. Verkhovsky, M. Wikström, *Nature* **440**, 829–832 (2006).
29. J. Lu, M. R. Gunner, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12414–12419 (2014).
30. W.-G. Han Du, A. W. Götz, L. Noodleman, *Inorg. Chem.* **58**, 13933–13944 (2019).
31. M. R. A. Blomberg, *Biochemistry* **55**, 489–500 (2016).
32. R. Liang, J. M. J. Swanson, Y. Peng, M. Wikström, G. A. Voth, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7420–7425 (2016).
33. E. A. Gorbikova, M. Wikström, M. I. Verkhovsky, *J. Biol. Chem.* **283**, 34907–34912 (2008).
34. K. Faxén, G. Gilderson, P. Ädelroth, P. Brzezinski, *Nature* **437**, 286–289 (2005).
35. A. Jose, A. W. Schaefer, A. C. Roveda Jr., W. J. Transue, S. K. Choi, Z. Ding, R. B. Gennis, E. I. Solomon, Spectroscopic data and DFT coordinates for "The three-spin intermediate at the O–O cleavage and proton pumping junction in heme–Cu oxidases," Zenodo (2021); https://doi.org/10.5281/zenodo.4961011.
36. W. J. Transue, wtransue/mcdps: Initial particle swarm MCD fitting routine, version 0.4, Zenodo (2021); https://doi.org/10.5281/zenodo.4969304.

ACKNOWLEDGMENTS

We thank the Stanford Research Computing Center for providing computational resources. **Funding:** This study received funding from the National Institutes of Health grant R01DK031450 (to E.I.S.); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Finance Code 001 (to A.C.R.); and the Ruth L. Kirschstein National Research Service Award F32GM131602 (to W.J.T.). **Author contributions:** E.I.S., A.W.S., A.C.R., and A.J. designed the experiments. S.K.C. and Z.D. prepared and purified the protein. A.W.S., A.C.R., and A.J. performed the experiments. W.J.T. wrote the VTVH-MCD fitting program. A.J. and A.W.S. performed the DFT calculations. A.J., A.W.S., A.C.R., and E.I.S. analyzed the data. A.J. and E.I.S. wrote the manuscript with assistance from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All spectroscopic data in the main text and supplementary materials, computer code for fitting the VTVH-MCD data, and atomic coordinates of DFT structures are freely available through Zenodo (*35*, *36*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abh3209
Materials and Methods
Supplementary Text
Figs. S1 to S12
References (*37*–*51*)
MDAR Reproducibility Checklist

1 March 2021; accepted 30 July 2021
10.1126/science.abh3209

REPORTS

MATERIALS SCIENCE

Centrifugation and index matching yield a strong and transparent bioinspired nacreous composite

Ali Amini[1,2], Adele Khavari[1]†, Francois Barthelat[2,3], Allen J. Ehrlicher[1,2,4,5,6,7]*

Glasses have numerous applications because of their exceptional transparency and stiffness; however, poor fracture, impact resistance, and mechanical reliability limit the range of their applications. Recent bioinspired glasses have shown superior mechanical performance, but they still suffer from reduced optical quality. Here, we present a nacreous glass composite that offers a combination of strength, toughness, and transparency. Micrometer-sized glass tablets and poly(methyl methacrylate) (PMMA) were mixed and structured by centrifugation, creating dense PMMA-glass layers. A transparent composite was created by tuning the refractive index of PMMA to that of glass and using chemical functionalization to create continuous interfaces. The fabrication method is robust and scalable, and the composite may prove to be a glass alternative in diverse applications.

Engineering glasses are brittle materials with low fracture toughness (the material's capability to resist the growth of a crack or flaw) and low strength (the material's resistance against failure), which limits the range of their applications (*1*). Thermal or chemical tempering is a common strategy used to increase the strength of glasses (*2*); however, this does not markedly improve fracture toughness (*3*) and can lead to catastrophic, "explosive" types of failures. Laminating creates a polymeric glass sandwich-like composite structure (*4*), the greatest advantage of which is safety: Upon fracture, the polymeric layers prevent small pieces of fractured glass from shattering in catastrophic failure. However, there are only modest mechanical improvements observed in laminated glasses (*5*–*7*).

To improve glass toughness and strength, researchers have explored implementing design principles observed in biology. Nacre, the tough material that makes up the inner layer of mollusk shells, is a classic example of a tough structural biomaterial. Nacre is 3000 times as tough as its individual constituents (*8*); it breaks at 1% of strain, which is substantially higher than the individual ceramic building blocks; and its elastic modulus is ~1000 times as large as that of the connective proteins alone (*9*).

Many techniques of varying complexity have been proposed to fabricate synthetic materials mimicking nacre's structure (*10*, *11*). Some of these have focused on making transparent composites (*12*–*14*), resulting in thin films with enhanced mechanical and optical properties. To extend the applications beyond thin films, a scalable nacreous composite has been developed by infiltrating poly(methyl methacrylate) (PMMA) into a glass tablet scaffold while matching refractive indices of the two phases (*15*). Despite superior fracture resistance properties compared with glass, this composite is less transparent.

Others have used top-down methods, including laser-engraving interlocking jigsaw-shaped three-dimensional arrays in bulk glass

[1]Department of Bioengineering, McGill University, Montreal, Quebec H3A 0C3, Canada. [2]Department of Mechanical Engineering, McGill University, Montreal, Quebec H3A 0C3, Canada. [3]Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO 80309, USA. [4]Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec H3A 0C7, Canada. [5]Department of Biomedical Engineering, McGill University, Montreal, Quebec H3A 2B4, Canada. [6]Centre for Structural Biology, McGill University, Montreal, Quebec H3G 0B1, Canada. [7]Goodman Cancer Research Centre, McGill University, Montreal, Quebec H3A 1A3, Canada.
***Corresponding author. Email: allen.ehrlicher@mcgill.ca**
†Present address: Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada.

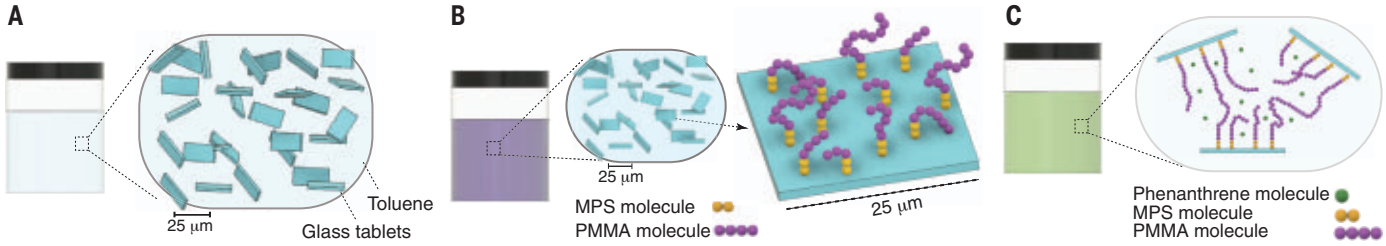


Fig. 1. Centrifuge-based fabrication method of nacreous glass composite. (**A**) Cleaned glass tablets were dispersed in toluene. (**B**) Glass tablets were surface treated with γ-MPS and then with a solution of MMA in toluene to promote polymerization from the glass surface. (**C**) Surface-treated glass tablets were involved in free radical polymerization of PMMA at 50°C.

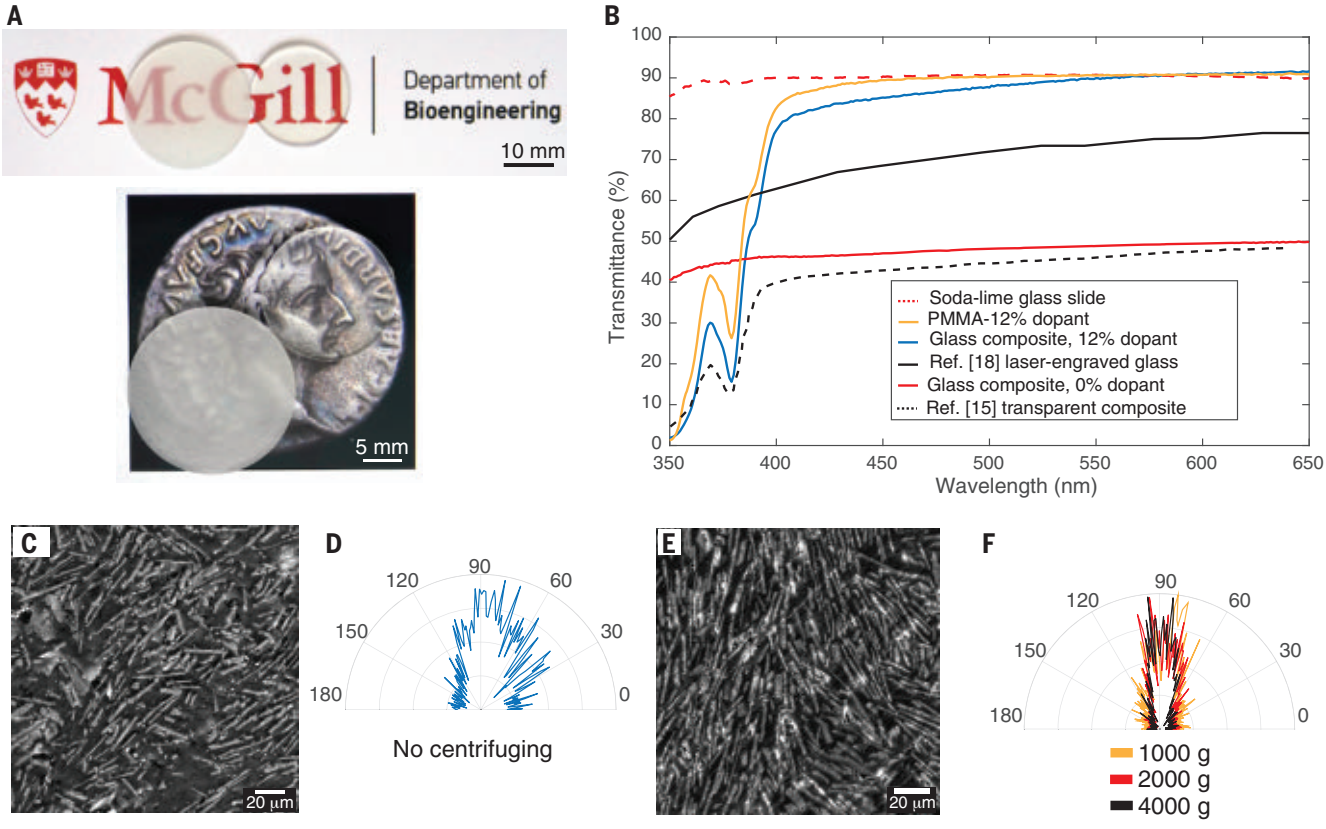


Fig. 2. Adding dopant yields a transparent composite, and centrifugation imposes order and compactness on the structure. (**A**) One-millimeter-thick glass composites with 0% dopant (samples on the left) and 12% dopant (samples on the right). The "Tribute Penny" coin is attributed under CC BY-SA 3.0. (**B**) Transmittance for the soda-lime glass, doped PMMA (12%), 12% glass composite, laser-engraved laminated glass (*18*), and bioinspired transparent composite (*15*). Transmittance values were measured at room temperature (20°C). (**C**) Section SEM image of a noncentrifuged composite (~24% glass volume fraction). We observed a noticeable number of areas with no tablets and also many tablets with random orientations in the material. (**D**) The distribution of tablet orientations also confirms this observation. (**E**) Section SEM image of centrifuged composite (2000*g*, ~43% glass volume fraction). Tablets are more oriented in one direction, and areas with no tablets are rarely observed. (**F**) Polar distribution of orientation in the tablets for different centrifuging speeds.

(*16*) and glass lamination processes of thin glasses with laser-engraved cross-plied (*17*) and tablet-like architectures (*18*). These approaches have resulted in increased composite fracture toughness and impact resistance, but reduced stiffness and strength. Stiffness and strength can generally be improved by decreasing the size of the patterns; however, this reduces transparency and scalability (*18*). This highlights a general trade-off challenge that bioinspired glasses have suffered from between mechanics, transparency, and fabrication scalability.

The structure and composition of our transparent nacre-like material result from stringent requirements and careful material selection and preparation. For high mechanical performance—i.e., high toughness, strength, and stiffness—hard and stiff inclusions with high aspect ratios must be bonded by a more deformable and weaker matrix (*9*, *19*–*22*). For high optical performance, the refraction index of the two phases must be identical. Finally, a strong bonding between the hard and soft phase is required to ensure a high

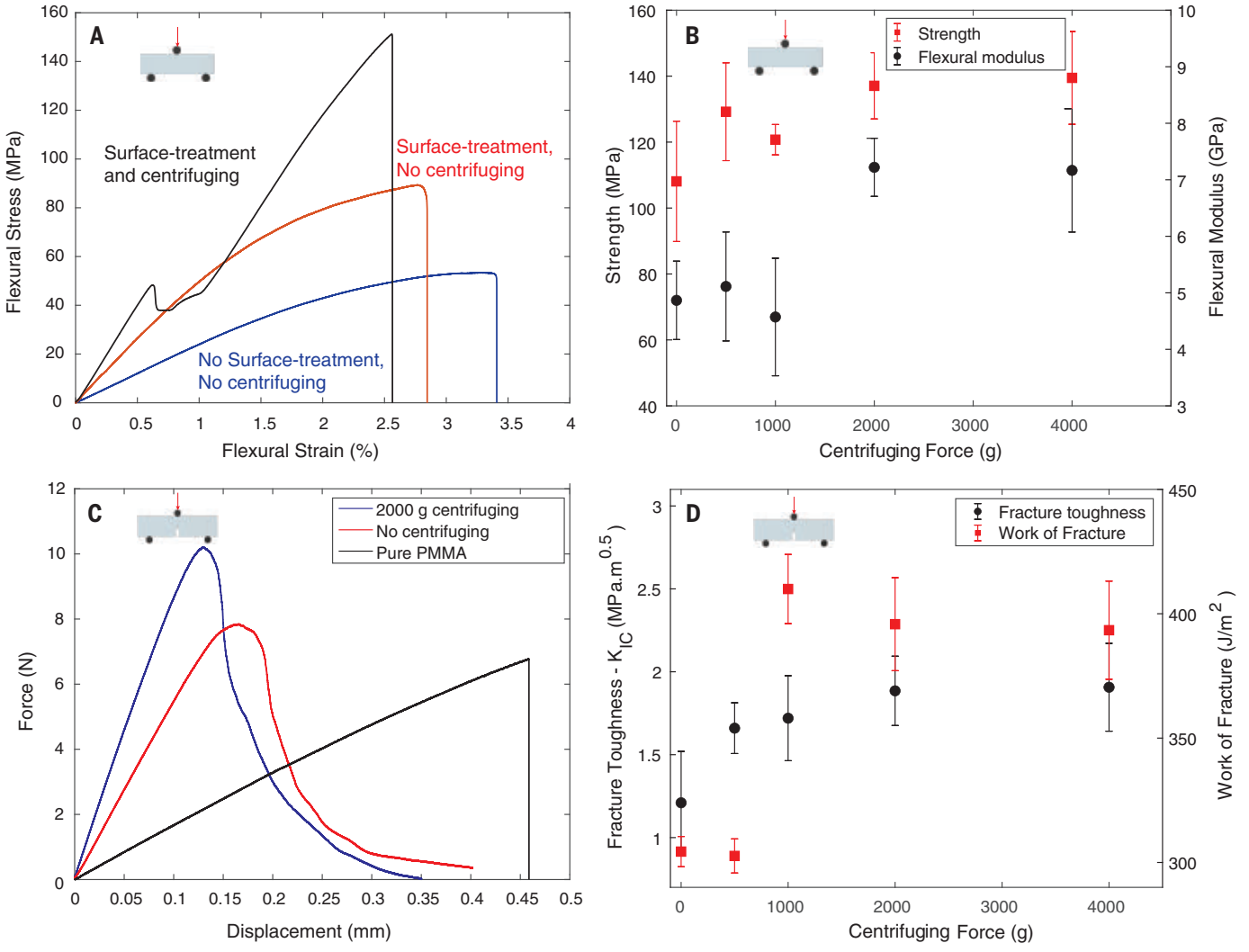


Fig. 3. Surface functionalization and centrifuging process improve the mechanical properties of the glass composite. (**A**) Flexural stress–flexural strain curves for the composites with and without surface-functionalized glass and centrifugation. (**B**) Strength and flexural modulus increase with centrifugation. (**C**) SENB test load-displacement curves for pure PMMA, noncentrifuged, and centrifuged composites, illustrating an increased fracture strength with composite formulation and subsequent centrifugation. (**D**) K_{IC} and WOF values increase with increasing the centrifuging speed up to 2000*g*.

level of interface strength and to prevent light scattering that would reduce optical quality. To fulfill these mechano-optical requirements, we chose micrometer-sized glass tablets (64 to 70% SiO_2; detailed chemical composition is described in the supplementary materials) for the hard inclusions and PMMA for the soft matrix. Glass tablets were utilized as the hard component because of their high diameter-to-thickness aspect ratio, transparency, high stiffness, and well-characterized surface chemistry for surface functionalization. PMMA, an amorphous polymer that polymerizes through a free radical bulk polymerization process (*23*), was selected as the soft phase because of its deformability (*24*) and excellent optical properties (*25*). The optical refraction indices of glass (n_{glass} = 1.52) and PMMA (n_{PMMA} = 1.49) do not exactly match, but they can be made to do so by adding an organic dopant, phenanthrene, to PMMA (*26*).

To achieve a strong and defect-free interface between glass and PMMA, we functionalized the glass tablet surface with a silane [(3-trimethoxysilyl)propyl methacrylate, or γ-MPS] (Fig. 1 and fig. S1). The prepared PMMA and glass tablets were mixed and then centrifuged to induce an aligned brick-and-mortar architecture and high volume fraction of glass inclusions. As a final stage, PMMA polymerization was achieved by baking at 50°C (for 12 hours), 70°C (for 4 hours), and 100°C (for 2 hours).

We measured the refractive index of PMMA-dopant samples as a function of dopant weight percentage (fig. S2A) and then estimated the composition that leads to the highest level of transparency (fig. S2D). Although the composite with no index-matching dopant was very hazy, and the sample was not transparent because of light scattering, our index-matched glass composite demonstrated high levels of transparency (Fig. 2A). The optical transmittance of our glass composite compares well with both soda-lime monolithic glass and PMMA doped with 12% phenanthrene, particularly in the sensitive spectrum of human vision (400 to 700 nm) (Fig. 2B), and its average transmittance is only 16% less than the soda-lime glass (fig. S2E). It also has 24% higher transmittance than similar bioinspired laminated composites (*18*) and has almost 100% higher transmittance than a nacre mimetic composite (*15*). (fig. S2, E and F). Although the method in (*15*) mimics other aspects of nacre by creating mineral bridges through a heat-treatment process, this process also creates many potential light-scattering sites, leading to the poor optical performance reported

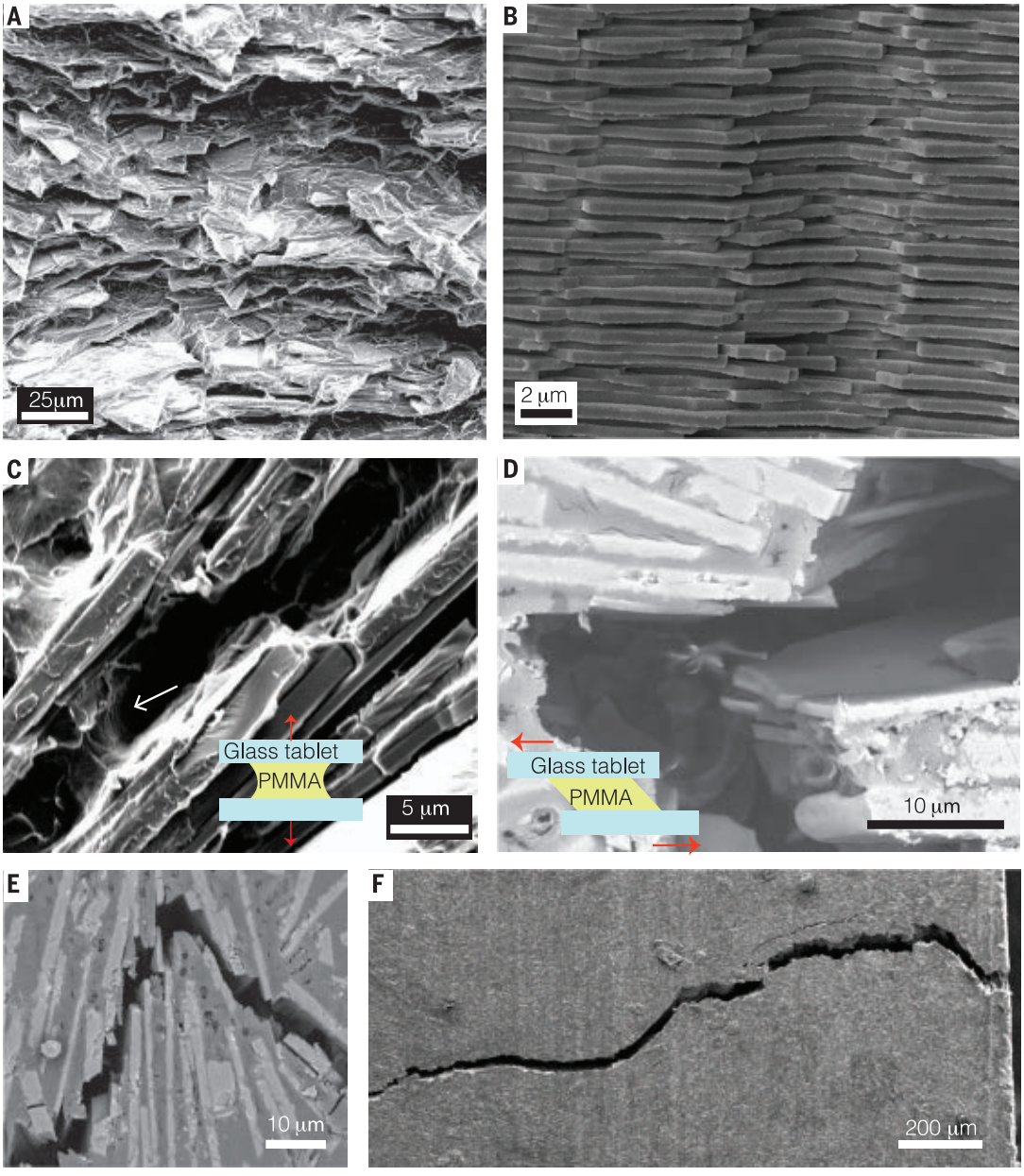


Fig. 4. Transparent nacreous glass mimics key aspects of nacre's microstructure and toughening mechanisms. (**A**) SEM micrograph of the sample cross section in a fractured surface in our glass composite. Staggered glass platelets with polymer layers in between and numerous pulled-out glass tablets are a result of fracture. (**B**) Fracture surface of natural nacre. (**C**) Whereas tablets are experiencing delamination as a result of lateral deformation, the polymer layer acts as a bridge between tablets and contributes to the composite toughness (white arrow). (**D**) Relative displacement of tablets as a result of shear loading. (**E**) Tablet pull-out and deflection of crack growth path in microscale. (**F**) Macroscopic crack deflection in the material as a result of microscopic toughening mechanisms.

for that material. The pale yellow color of our composite is a result of the phenanthrene's molecular structure, which blocks a major part of the blue light. This is mainly because of the conjugation—alternating arrangement of single and double bonds between carbon atoms (*27*)—where the energy spacing between π and π* orbitals in a conjugated system matches the energy range associated with the visible light spectrum. As the conjugated systems become bigger, the maximum absorbance peak shifts to the larger wavelengths. To overcome this effect, one solution would be to use polycyclic aromatic hydrocarbons with smaller conjugated systems. We successfully tested biphenyl, as an example of a smaller conjugated system (fig. S3), and it does not block blue light; therefore, the resulting composite does not have a yellow tint.

In addition to high optical performance, our material needs to be stiff and possess high levels of tensile strength and fracture toughness, which requires a high concentration of well-aligned glass inclusions. As glass and PMMA have different densities, we used centrifugation to increase the fraction of glass in our composite, leading to a high volume fraction of the stiff (glass) phase and consequently a thin interstitial connective (PMMA) phase. The volume fraction of glass inclusions increased from ~24% for noncentrifuged composites to ~43% for the samples centrifuged with 2000*g* (fig. S4). The thickness of the polymer layer between the tablets also decreased from ~35 μm for the simply mixed sample to ~17 μm for the sample centrifuged with 2000*g*. Centrifugation homogenizes the distribution of tablets, preventing the formation of tablet-free regions of PMMA (Fig. 2, C and E). Centrifugation also aligns the glass tablets (Fig. 2, D and F). By comparing the polar orientation distribution graphs of the noncentrifuged and centrifuged composites, we examined the role of the centrifuging process in inducing order in the structure of the composite. Centrifugation up to 2000*g* increased tablet order and alignment; however, further increasing the centrifugation

speed did not seem to considerably improve tablet alignment. Because no drastic change in volume fraction (or mechanical testing data, as in Fig. 3) was observed with further increasing the centrifuging speed, we determined 2000*g* to be an optimal centrifugation force.

The mechanical performance of the composite was evaluated using three-point bending tests. The glass composite displays two distinct linear and nonlinear regimes in flexural response (Fig. 3A). We attribute the nonlinear regime to the relatively large shear deformation of PMMA, which enables the relative displacement of the glass tablet following the shear lag model (*28*). PMMA is considered to be brittle compared with other polymers; however, it exhibits ductile behavior when compared with brittle materials such as glass and ceramics (fig. S5). Besides, PMMA is much more deformable in shear than in tension (*29*). All samples displayed evidence of inelastic deformation, with strains at failure in the 2.5-to-3.5% range. Flexural modulus and strength, however, varied significantly across the different designs we explored. The glass composite surface functionalized with γ-MPS was 1.9 times as stiff as the glass composite without any surface treatment (Fig. 3A and table S1). The composite's flexural strength increased about twofold by functionalizing the glass tablets' surfaces with γ-MPS. This increase in strength, however, only produced a glass composite slightly stronger than pure PMMA (table S1). The flexural strength increased to ~140 MPa by including the centrifuge process as a part of the fabrication process; this aligned the glass tablets into layers of parallel planes and also yielded a denser overall structure. The beneficial strengthening effects of centrifugation appeared to plateau at 2000*g*, with no substantial increase in flexural strength for higher forces (Fig. 3B). The flexural modulus increased from 4.7 GPa for the noncentrifuged sample to ~7.2 GPa for the sample centrifuged with 2000*g* force, with no significant increase of the modulus with higher centrifugation speeds. The effect of surface functionalization and centrifugation appeared to be insignificant on the rupture strain, and most of the samples possessed a rupture strain of ~3% (fig. S6 and table S1).

In addition to modulus, strength, and deformability, we measured fracture toughness using a single-edge notch bending (SENB) configuration. Figure 3C shows typical force-deflection results for these tests, showing a brittle response for pure PMMA (indicating fast and catastrophic crack propagation with little toughening) but a more controlled and "graceful" failure with a more-gradual decrease in force in the postpeak region for the glass composite (indicating stable crack propagation associate with toughening mechanisms). The centrifuged glass composite also showed a higher peak force. We computed the fracture toughness K_{IC} using the maximum force in linear regime (*30*, *31*) and the work of fracture [(WOF) energy required to break a material into two parts as a result of fracture] from the area under the force-deflection curve. By centrifuging (2000*g*), we increased the glass composite's fracture toughness (K_{IC}) and energy absorption (WOF) by ~55 and 30%, respectively (Fig. 3D). To demonstrate the behavior of the material in terms of crack propagation, we also computed the *J* integral values and derived the resistance curves (K_J, fracture toughness against crack extension length). A rising pattern in toughness values was observed as the crack propagated in the material (fig. S7C). The maximum K_J value before crack instability (K_{JC}, based on the criteria in eqs. S10- to S12) for the 2000*g* composite was ~4.51 MPa $m^{1/2}$ (~448.5 J/m^2). This was about twice the estimated K_{IC} value, which suggests that extrinsic toughening mechanisms are activated after crack growth starts.

The results in Fig. 3 demonstrate that centrifuging improves the mechanical and fracture properties. This improvement is achieved by inducing order in the structure and creating a staggered structure of glass tablets and PMMA polymer similar to the structure of nacre (Fig. 4, A and B). Such order promotes some important extrinsic toughening mechanisms, leading to the excellent performance of the composite under fracture.

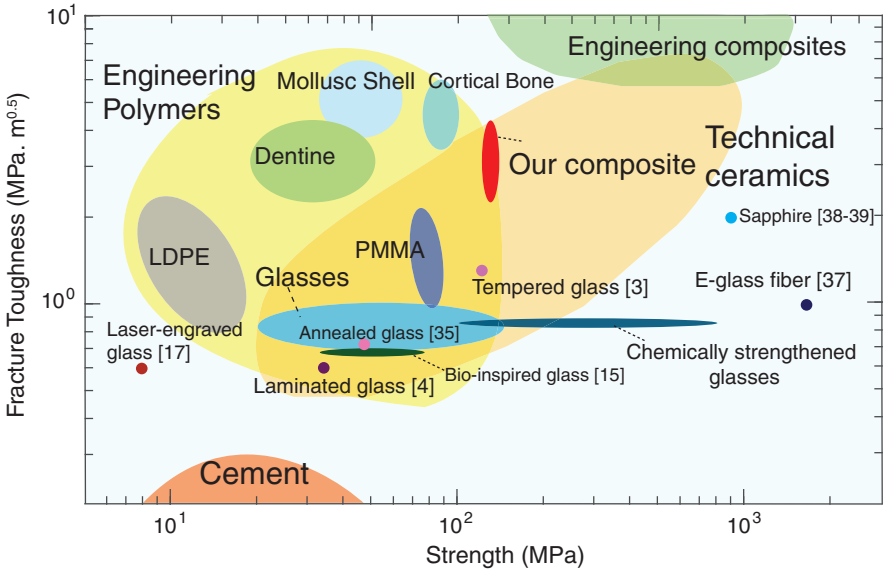


Fig. 5. Ashby plot of fracture toughness versus strength for numerous synthetic and natural materials. Our transparent nacreous composite (centrifuged at 2000*g*) outperforms annealed soda-lime (*35*), tempered (*3*), laminated (*4*), bioinspired transparent composite (*15*), and laser-engraved laminated (*17*) glasses in both strength and fracture toughness. E-glass fiber (*37*) and sapphire (*38*–*39*), as the best examples of glass and transparent ceramic materials respectively, outperform our composite in strength; however, unlike our material, they experience a brittle and catastrophic fracture. The data point representing our material's fracture toughness is made up of values starting from K_{IC} values and ending at the maximum K_{JC} values.

In the absence of mineral bridges and tablet interlocking, a key toughening mechanism in our material is polymer bridging by the formation of polymer ligaments between tablets, activated when the delamination of tablets occurs (Fig. 4C and fig. S7). This mechanism occurred either by complete debonding (Fig. 4C and fig. S8, A and B) or by the formation of microscopic cavities in the soft phase (fig. S8, C and D). Also reported to occur in natural nacre (*32*), tablets experience delamination as a result of lateral displacements if (i) the interface material is deformable and (ii) the bonding between polymer and glass is strong (*33*). This highlights the role of glass surface treatment and consequent strong bonds between the soft and hard phases. The relative displacement of the individual tablets from local tensile stress and interfacial shear stresses (Fig. 4D) is another important micromechanism that not only absorbs energy through plastic deformations in the soft phase, but also leads to tablet pull-out—a mechanism that contributes to high levels of toughness in biological nacre (*21*) (Fig. 4, A and E, and fig. S9, A and B). Scanning electron microscopy (SEM) images suggest that crack deflection and delamination frequently occur, with crack deflection appearing to be a central toughening mechanism in our composite. Tablet pull-out, however, appears to occur less frequently. Previous models have shown that the WOF as a result of pull-out is directly proportional to the volume fraction of

the hard phase (*32*), and therefore we attribute the less-frequent occurrence of tablet pull-out in our material to the relatively bigger polymer layer thickness and consequently lower volume fraction compared with natural nacre. Mechanisms such as crack deflection, however, give rise to large deformations, and high levels of energy absorption manifest as high fracture toughness in the material. Because of the activation of the toughening mechanisms—and consequently the propagation of microcracks through the soft phase with many path diversions—many deflections in the crack propagation path are observed in the micro- and macroscale (Fig. 4, E and F). Figure 5 shows that our composite possesses a strength similar to that of thermally tempered glass (*3*) but has a higher fracture toughness. Although our composite has a strength similar to the lower-range values of chemically strengthened glasses, it outperforms chemically tempered glasses in several key areas: (i) It possesses higher toughness values. (ii) Unlike chemically tempered glasses, it is possible to cut and machine our glass composite using conventional machining techniques and tools (fig. S10). (iii) Our material has a higher degree of damage tolerance, as the rising crack resistance curve (fig. S7C) demonstrates that any propagating crack will be met with increasingly strong resistance from the material, so these cracks will not compromise the overall strength of the system.

Our glass composite also outperforms previously reported bioinspired glass composites (*15*) in both strength and fracture toughness (fig. S7A). The reason for this difference is likely in the glass tablet aspect ratios. For a fixed matrix shear strength, there is a range of aspect ratios for tablets that causes the material to experience tablet pull-out (*19*). In this sense, too-short and too-long glass tablets cause vertical interface and tablet failure to be the prevailing modes of failure, respectively. In other words, keeping the interface shear strength constant while increasing the tablet aspect ratio would lead to a decrease in the WOF of the composite. The aspect ratio of the tablets in our material is about 25, or about one-tenth the size of the ones in (*15*), and closer to the ideal range for strength and toughness (eq. S14) (*34*), which explains the differences in strength and fracture toughness values of the two materials. Considering the WOF as a nonlinear measure of fracture resistance, our composite outperforms annealed (*35*) and laminated glasses (*4*) as well as pure PMMA (fig. S8B). The laser-engraved laminated glass structure (*17*) possesses a very high WOF; however, this has only been achieved with a corresponding compromise in reduced material strength. Our material displays a brick-and-mortar microstructure (Fig. 2C and Fig. 4A) and produces high levels of fracture toughness because of its extrinsic toughening mechanisms, even though the volume fraction of the hard phase is lower than that observed in natural nacre (~95 vol % in natural nacre versus 46 vol % in our material). The lower content of the hard phase probably reduces tablet pull-out at large scales, resulting in our material being in the middle of the natural nacre's toughness range [J = 300 to 1700 J/m^2 for red abalone nacre (*36*)].

Hard-phase alignment has long been recognized as a key strategy, often pursued from a serialized bottom-up approach. Centrifugation is a rapid and scalable approach useful for fabricating any composite geometry and dimensions and may be further enhanced by increasing the density differential between hard and soft phases. This is a fundamental advantage over the serialized layer-by-layer approaches, which sacrifice production for precision. Additionally, this moves composite fabrication out of specialized nanofabrication facilities and into the realm of industrially approachable processes. The dependency of mechanical properties on centrifugation force illustrates the importance of both aligning the hard phase tablets and minimizing the compliant PMMA phase in the overall composite structure, similar to the minimization of protein (~<5%) found in natural nacreous composites. This centrifugation-imposed order on the structure also effectively enables the activation of toughening mechanisms, such as crack deflection, tablet delamination, and tablet pull-out, which together contribute to high levels of fracture toughness in our material. These strategies enable our glass composite to mechanically outperform annealed, thermally tempered, and laminated glasses in fracture toughness and flexural strength. By making the soft and hard phases have the same refractive indices, one can create any number of varied materials in structured composites that have few to no optical defects.

REFERENCES AND NOTES

1. L. Wondraczek *et al*., *Adv. Mater.* **23**, 4578–4586 (2011).
2. F. M. Ernsberger, *Techniques of Strengthening Glasses*, vol. 5 (Academic Press, Inc., 1980).
3. F. Petit, A. C. Sartieaux, M. Gonon, F. Cambier, *Acta Mater.* **55**, 2765–2774 (2007).
4. X. Huang, G. Liu, Q. Liu, S. J. Bennison, *Struct. Eng. Mech.* **52**, 603–612 (2014).
5. J. E. Minor, P. L. Reznik, *J. Struct. Eng.* **116**, 1030–1039 (1990).
6. M. M. El-Shami, S. Norville, Y. E. Ibrahim, *Alex. Eng. J.* **51**, 61–67 (2012).
7. H. S. Norville, K. W. King, J. L. Swofford, *J. Eng. Mech.* **124**, 46–53 (1998).
8. H. D. Espinosa, J. E. Rim, F. Barthelat, M. J. Buehler, *Prog. Mater. Sci.* **54**, 1059–1100 (2009).
9. F. Barthelat, H. Tang, P. D. Zavattieri, C. M. Li, H. D. Espinosa, *J. Mech. Phys. Solids* **55**, 306–337 (2007).
10. N. Almqvist *et al*., *Mater. Sci. Eng. C* **7**, 37–43 (1999).
11. H. Le Ferrand, F. Bouville, T. P. Niebel, A. R. Studart, *Nat. Mater.* **14**, 1172–1179 (2015).
12. P. Podsiadlo *et al*., *J. Phys. Chem. B* **112**, 14359–14363 (2008).
13. T. Ebina, F. Mizukami, *Adv. Mater.* **19**, 2450–2453 (2007).
14. Y. Liu, S.-H. Yu, L. Bergström, *Adv. Funct. Mater.* **28**, 1703277 (2018).
15. T. Magrini *et al*., *Nat. Commun.* **10**, 2794 (2019).
16. M. Mirkhalaf, A. K. Dastjerdi, F. Barthelat, *Nat. Commun.* **5**, 3166 (2014).
17. Z. Yin, A. Dastjerdi, F. Barthelat, *Acta Biomater.* **75**, 439–450 (2018).
18. Z. Yin, F. Hannard, F. Barthelat, *Science* **364**, 1260–1263 (2019).
19. M. R. Begley *et al*., *J. Mech. Phys. Solids* **60**, 1545–1560 (2012).
20. P. Fratzl, O. Kolednik, F. D. Fischer, M. N. Dean, *Chem. Soc. Rev.* **45**, 252–267 (2016).
21. F. Barthelat, R. Rabiei, *J. Mech. Phys. Solids* **59**, 829–840 (2011).
22. F. Barthelat, *J. Mech. Phys. Solids* **73**, 22–37 (2014).
23. T. S. Balke, thesis, McMaster University (1972).
24. T. P. Niebel, F. Bouville, D. Kokkinis, A. R. Studart, *J. Mech. Phys. Solids* **96**, 133–146 (2016).
25. T. Hanemann, J. Boehm, C. Müller, E. Ritzhaupt-Kleissl, *Proc. SPIE* **6992**, 69920D (2008).
26. J. Böhm, J. Hausselt, P. Henzi, K. Litfin, T. Hanemann, *Adv. Eng. Mater.* **6**, 52–57 (2004).
27. L. M. Harwood, T. D. W. Claridge, *Introduction to Organic Spectroscopy* (Oxford Univ. Press, 1997).
28. X.-L. Gao, K. Li, *Int. J. Solids Struct.* **42**, 1649–1667 (2005).
29. J. Zhang, T. Jin, Z. Wang, L. Zhao, *Results Phys.* **6**, 265–269 (2016).
30. B. Lawn, *Fracture of Brittle Solids* (Cambridge Univ. Press, 1993).
31. ASTM Standard E1820-13, Standard Test Method for Measurement of Fracture Toughness (ASTM International, 2013); https://doi.org/10.1520/E1820-13.
32. A. P. Jackson, J. F. V. Vincent, R. M. Turner, *Proc. R. Soc. Lond. B.* **234**, 415–440 (1988).
33. F. Barthelat, Z. Yin, M. J. Buehler, *Nat. Rev. Mater.* **1**, 16007 (2016).
34. H. Gao, B. Ji, I. L. Jäger, E. Arzt, P. Fratzl, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5597–5600 (2003).
35. S. M. Wiederhorn, *J. Am. Ceram. Soc.* **52**, 99–105 (1969).
36. F. Barthelat, H. D. Espinosa, *Exp. Mech.* **47**, 311–324 (2007).
37. M. Herráez, A. Fernández, C. S. Lopes, C. González, *Phil. Trans. R. Soc. A.* **374**, 20150274 (2016).
38. G. E. Scott Jr., *Angle Orthod.* **58**, 5–8 (1988).
39. Y. Mitamura, Y. Wang, *J. Biomed. Mater. Res.* **28**, 813–817 (1994).

ACKNOWLEDGMENTS

The authors thank C. Molter for helpful discussions and a critical review of the manuscript, P. Tirgar for helping with sample preparation, and P. Hubert for providing us with access to his laboratory facility. **Funding:** The authors acknowledge support from NSERC RGPIN/05843-2014 and EQPEQ/472339-2015, a FRQNT team grant, the Canadian Foundation for Innovation Projects nos. 32749 and 33122, and the Canada Research Chairs Program. **Author contributions:** Conceptualization: A.J.E. and A.A.; Methodology - development: A.A., A.K., F.B., and A.J.E.; Methodology - application: A.A.; Investigation: A.A. and A.K.; Formal analysis: A.A. and F.B.; Software: A.A.; Visualization: A.A.; Writing - original draft: A.A.; Writing - review and editing: A.J.E., A.K., A.A., and F.B.; Funding acquisition: A.J.E. and F.B.; Resources: A.J.E. and F.B.; Supervision: A.J.E. **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

https://science.org/doi/10.1126/science.abf0277
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S3
References (*40*–*44*)

30 September 2020; accepted 10 August 2021
10.1126/science.abf0277

SUPERCONDUCTIVITY

Anomalously strong near-neighbor attraction in doped 1D cuprate chains

Zhuoyu Chen[1,2,3]†, Yao Wang[4]†, Slavko N. Rebec[1,2,3], Tao Jia[1,3,5], Makoto Hashimoto[6], Donghui Lu[6], Brian Moritz[1], Robert G. Moore[1,7], Thomas P. Devereaux[1,3,8]*, Zhi-Xun Shen[1,2,3,5]*

In the cuprates, one-dimensional (1D) chain compounds provide a distinctive opportunity to understand the microscopic physics, owing to the availability of reliable theories. However, progress has been limited by the challenge of controllably doping these materials. We report the synthesis and spectroscopic analysis of the 1D cuprate $Ba_{2-x}Sr_xCuO_{3+\delta}$ over a wide range of hole doping. Our angle-resolved photoemission experiments reveal the doping evolution of the holon and spinon branches. We identify a prominent folding branch whose intensity fails to match predictions of the simple Hubbard model. An additional strong near-neighbor attraction, which may arise from coupling to phonons, quantitatively explains experiments for all accessible doping levels. Considering structural and quantum chemistry similarities among cuprates, this attraction may play a similarly important role in high-temperature cuprate superconductors.

Strong correlations in electronic materials such as high-temperature cuprate superconductors lead to complex properties that have been a central challenge in condensed matter physics (*1*). So far, there remains no exact solution of many-body Hamiltonians in two or more dimensions, making a quantitative comparison between theory and experiment a challenge. A promising route to improved understanding may be through the analysis of one-dimensional (1D) systems: The reduced dimension enables exact solutions of microscopic Hamiltonians like the Hubbard model and low-energy field theory without bias. Such possibilities have been recognized in theory, but progress has been hindered by the lack of suitable experimental material systems.

In a 1D system, strong correlations manifest as so-called spin-charge separation (*2*–*7*). As schematized in Fig. 1A, a photoinduced hole fractionalizes into a "spinon" and a "holon," carrying spin and charge, respectively, giving rise to spectral branches with different propagation velocities. For half-filled d^9 systems, such spin-charge separation along with a Mott gap (Fig. 1B) has been confirmed by angle-resolved photoemission spectroscopy (ARPES) (*4*–*6*). However, the simplicity of the spectrum in the undoped case and the lack of information on interactions between doped carriers make it difficult to establish a connection to underlying microscopic models. In particular, it remains undetermined whether fundamental Hamiltonians, like the Hubbard or *t*-*J* models, contain all of the essential ingredients for general cuprate properties, including *d*-wave superconductivity in 2D. To address this question, a doping-dependent study of a 1D cuprate chain compound would be beneficial. However, controllably doped 1D cuprate systems remain elusive after more than two decades of effort (*4*–*6*).

We achieved a wide range of hole doping in a corner-sharing cuprate chain by synthesizing $Ba_{2-x}Sr_xCuO_{3+\delta}$ (BSCO, whose lattice structure is shown in Fig. 1C). Thin films were grown with a recently developed ozone-reactive molecular beam epitaxy (MBE) system connected in situ to ARPES beamline 5-2 at the Stanford Synchrotron Radiation Lightsource (SSRL). Atomic layer-by-layer growth in purified ozone was monitored in real-time via in situ reflective high energy electron diffraction (RHEED; fig. S1, B and C) (*8*). A mixture of isovalent Sr and Ba facilitates the tuning of in-plane lattice constants to mitigate mismatches with $SrTiO_3$ substrates (*9*). Orthorhombic BSCO on cubic $SrTiO_3$ forms twinned domains with different chain orientations (fig. S1, D and E). In different domains, chain directions are perpendicular to each other and parallel to the substrate

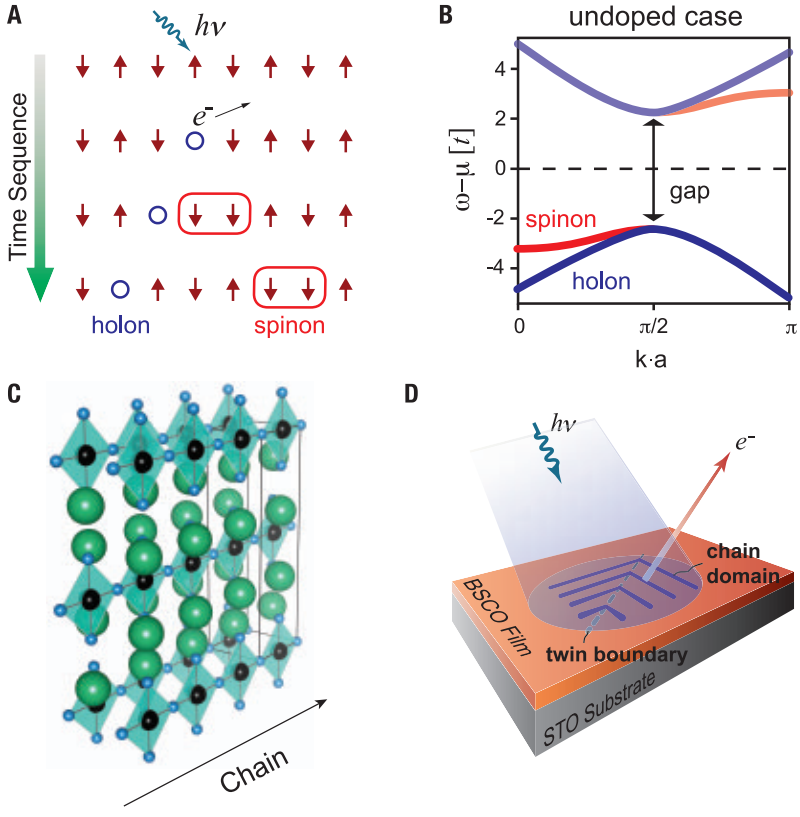


Fig. 1. Strongly correlated antiferromagnetic (AFM) cuprate chain. (**A**) Schematic of an example photoemission process in a 1D AFM chain. Red arrows, blue circles, and red rounded rectangles correspond to spin-full electrons, holons, and spinons, respectively. The green arrow represents the arrow of time. (**B**) Schematic diagram of the spectral dispersion relation for an undoped chain. Blue and red solid lines correspond to the holon and spinon branches, respectively. (**C**) The lattice structure of $Ba_{2-x}Sr_xCuO_{3+\delta}$. Green, blue, and black balls correspond to Ba/Sr, O, and Cu atoms, respectively. (**D**) Schematic of sample setup for photoemission. The BSCO film on the $SrTiO_3$ (STO) substrate is twinned with chain domains smaller than the beam spot.

[1]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [2]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [3]Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA 94305, USA. [4]Department of Physics and Astronomy, Clemson University, Clemson, SC 29631, USA. [5]Department of Physics, Stanford University, Stanford, CA 94305, USA. [6]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [7]Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [8]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: tpd@stanford.edu (T.P.D.); zxshen@stanford.edu (Z.X.S.)
†These authors contributed equally to this work.

surface (Fig. 1D). All ARPES data shown below were collected for temperatures lower than 20 K with 65-eV photons, from one single 2.5 unit-cell sample (equivalent to 5 $Ba_{1-x/2}Sr_{x/2}O$-CuO-$Ba_{1-x/2}Sr_{x/2}O$ layers) with nominal $x = 0.16$ measured by quartz crystal microbalance. Doping by interstitial oxygen was controlled via ozone and vacuum annealing series (*10*).

Figure 2, A1 to A6, present the ARPES Fermi surface maps with controlled hole doping from 9% up to 40%. The two perpendicular Fermi surfaces indicate simultaneous photoelectron detection from both sets of domains, as mentioned above. Within each set, the lack of transverse dispersion in the spectral distribution demonstrates the highly 1D nature of the chains. Figure 2, B1 to B6, exhibit the energy-momentum spectra of cuts along the Brillouin zone edge. These cuts are located at a position where the spectral weight from one set of domains is minimized, enabling analysis of the spectra from chains with a single orientation. Note that features at higher binding energies beyond 1.2 eV are compromised by nonbonding oxygen band intensities (*11*). Figure 2, C1 to C6, show the corresponding second derivatives to highlight the quasiparticle features.

Comparison to cluster perturbation theory (CPT) simulations for the single-band Hubbard model (*12*, *13*), displayed in Fig. 2, D1 to D6, allows for the identification of the following spectral features: (i) The highest-intensity features correspond to main holon branches (blue dashed lines with "h" in B1 and C1) crossing the Fermi level E_F at the Fermi momentum k_F; (ii) spinon branches (red dashed lines with "s" in B1 and C1), which flatten near $k_{||} = 0$ and merge into the main holon branches for higher momenta. These features agree with previous observations in undoped, parent materials (*4*–*6*). The persistence of spin-charge separation up to 40% doping agrees with predictions from the single-band Hubbard model, showing that correlation effects remain very strong in 1D, even for higher dopings (*14*). By fitting the

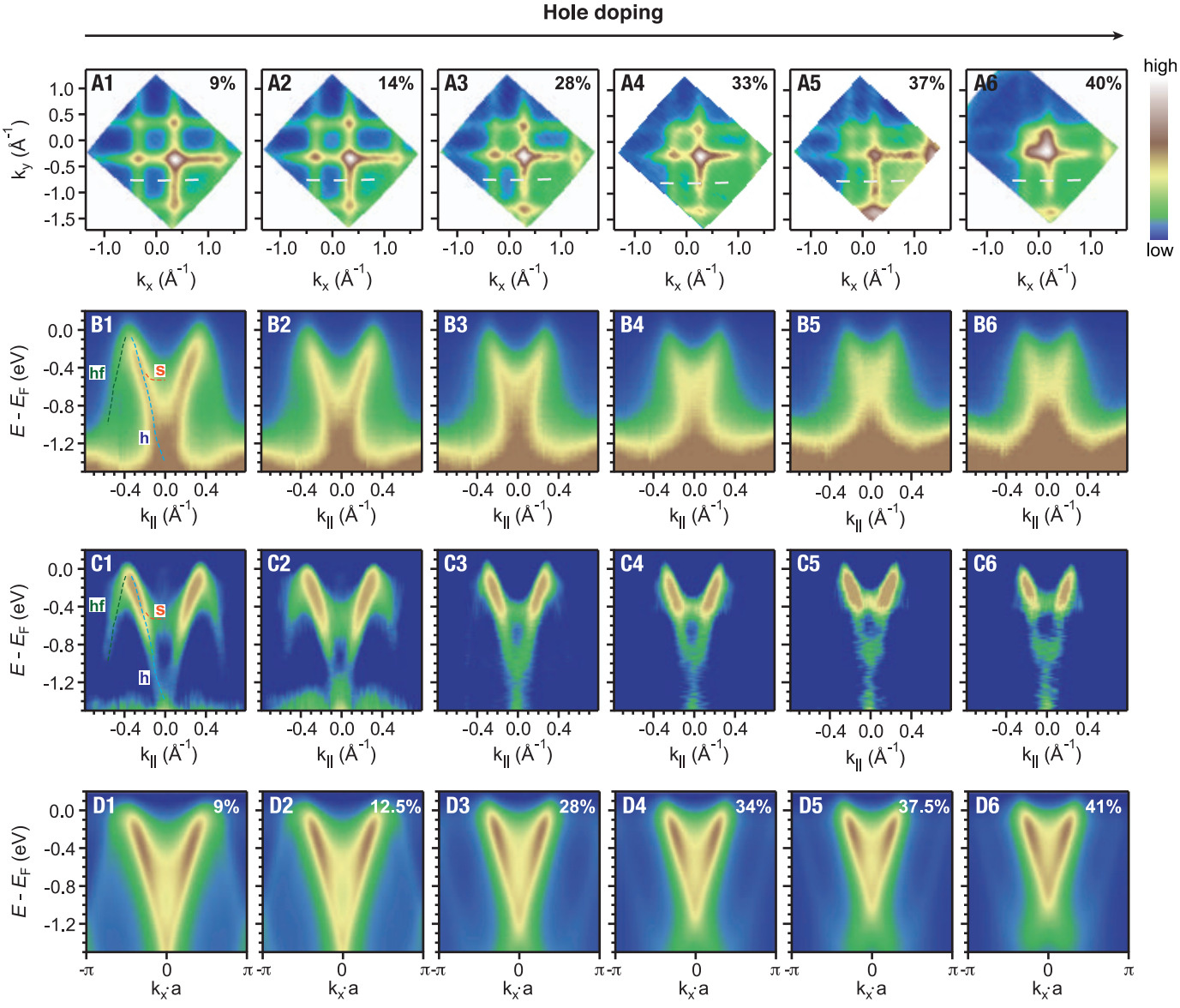


Fig. 2. Hole doping series of ARPES spectra. (**A1** to **A6**) Fermi surface maps with different hole doping levels. Black squares show the first Brillouin zone of the STO substrate. (**B1** to **B6**) Measured ARPES spectra taken along the cuts shown with white dashed lines in (A1) to (A6). (**C1** to **C6**) Second derivatives of energy distribution curves and momentum distribution curves are summed with sign reversed for comparison. Dashed blue, red, and green lines in (B) and (C) indicate the location of the holon ("h"), spinon ("s"), and holon folding ("hf") features. (**D1** to **D6**) Spectral function simulation results based on 16-site cluster perturbation theory (CPT) with varied corresponding hole doping levels.

doping-dependent spinon and holon binding energies at $k_{||} = 0$ (figs. S2 and S3), we obtained Hubbard model parameters $U = 8t$ and $t = 0.6$ eV, with an estimated $J = 4t^2/U = 0.3$ eV, consistent with earlier results (*6*, *7*).

In contrast to the observations in undoped compounds (*4*–*6*), an additional feature at $|k| > |k_F|$, denoted by "hf" in Fig. 2, B1 and C1, fades rapidly with doping, as shown more directly in the momentum distribution curves (MDCs) for 9%, 14%, and 33% doping (see subdominant peaks in Fig. 3, A, B, and C, respectively). As illustrated in Fig. 3D and in comparison with theory (*14*–*16*), the holon dispersion inherited from the undoped cuprate splits into two branches at k_F for doped systems: one folding directly at E_F and k_F represented in green [hereafter the holon folding (hf) branch], and the other crossing E_F and bending back at a larger momentum $2\pi - 3k_F$ (hereafter the $3k_F$ branch) as an extension of the main holon branch. Note that the term "folding" here is simply a description of the spectral feature rather than suggesting existence of translational symmetry breaking. The Lorentzian fitting of our MDCs yields the peak positions plotted in Fig. 3E and confirms the observed prominent "hf" spectral feature as being the holon folding branch (green). The presence of the hf branch agrees with predictions using the single-band Hubbard (*14*–*16*) or *t*-*J* models (*17*, *18*), previously attributed to a holon-holon interaction mediated by the spin superexchange and intrinsic in these models. This interaction occurs in the lowest-order t/U expansion as an attractive form with the magnitude $\sim J/4 \sim 0.1t$ (*19*).

Despite a consistency with the basic dispersion, we find that the simple Hubbard model is fundamentally deficient in accurately addressing additional spectral features. As evident in the MDCs shown in Fig. 4A, the Hubbard model prediction (second curve from top) shows that the relative intensity of the hf feature is rather weak compared to the $3k_F$ branch, whereas the experimental observation is the opposite. This implies that realistic holon-holon attractive interaction is not adequately described by the spin superexchange. To reproduce the experimental spectra, we find that a sizable attractive near-neighbor Coulomb interaction V must be introduced to the Hubbard Hamiltonian (*19*). Figure 4A displays the MDCs extracted from simulations with different V values from positive (repulsive) to negative (attractive): The near-neighbor attractive Coulomb interaction enhances the spectral weight of the hf branch and suppresses the $3k_F$ branches. We conclude that a strong holon-holon attraction $V \approx -1.0t$ matches well with the experimental spectra (bottom curve) and explains why the $3k_F$ branch is not found in experiment. Note that in principle such attraction in 1D primarily affects the charge degree of freedom; thus, the chemical potential is pinned by the spinon dispersions, unlike 2D cases where a quasiparticle gap is usually expected.

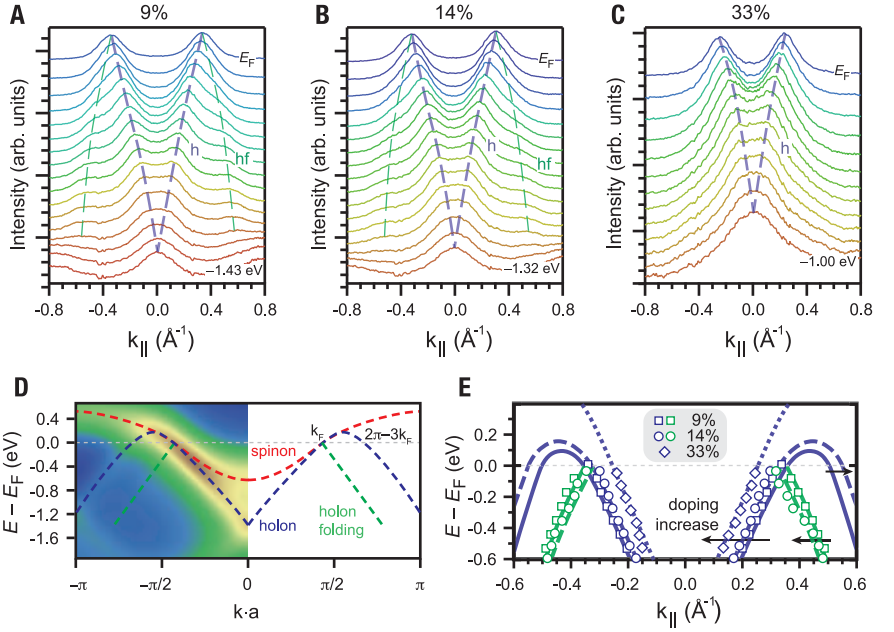


Fig. 3. Identification of the holon folding. (**A** to **C**) MDCs for varied energies at 9, 14, and 33% hole doping. Purple and green dashed lines are guides to the eye for the holon (h) and the holon folding (hf) branches. (**D**) CPT simulation based on single-band Hubbard model without Fermi-Dirac suppression above E_F for the 12.5% doping case, shown in the color-scale plot. Dashed lines with different colors track the traces of different branches as revealed by the CPT simulation. Note that the holon folding feature (faint white) is relatively weak and barely visible in the color-scale plot for the CPT simulation of pure Hubbard model without additional terms. (**E**) Peak positions extracted from MDCs for different dopings (squares: 9%; circles: 14%; diamonds: 33%) overlaid on theoretical lines of single-band Hubbard prediction, where solid, dashed, and dotted lines correspond to increasing doping. Blue and green colors correspond to the major holon (h) and the holon folding (hf) branches, respectively. Black arrows indicate the direction of the shift for different features with increasing dopings.

As mentioned above, the hf spectral feature fades rapidly with increasing doping, as summarized in Fig. 4B, perhaps implying a doping-dependent holon-holon attraction. Comparing both the spectral intensity and the doping level at which it vanishes (figs. S4 and S5), we find the best agreement with a (doping-independent) V between $-1.2t$ and $-0.8t$. To quantify this strong doping dependence, we extract the relative hf peak intensities with respect to the main holon-peak intensities based on Lorentzian fittings for experimental MDCs (circles in Fig. 4C); for comparison, we perform the same fitting on the simulated MDCs with $V = -1.0t$ (squares in Fig. 4C). The doping dependence of the hf peak intensity is captured for a fixed V, whose influence is renormalized by doping, such that this intensity gradually decreases and eventually disappears at ~30% hole doping. That is, the doping dependence is intrinsic in the calculation and does not require parameter tuning. The Coulomb interaction between electrons should always be repulsive, so the only probable origin of this effective attraction, beyond spin superexchange already accounted for in the Hubbard model, would be coupling to some bosonic excitations. Although multiple bosonic modes may play a role in cuprates, here, we postulate that the attractive V is mediated by phonons, considering the evidence of electron-phonon coupling in a variety of cuprates (*20*–*29*).

We emphasize that this additional near-neighbor attraction is an order of magnitude stronger than the inherent attraction in the Hubbard model mediated by spin superexchange (given by $-J/4$ in the limit of $t/U \ll 1$), when considering all interactions on the same footing. This is demonstrated in Fig. 4D, in which a baseline of zero net attraction is found by using the simulated folding peak relative intensity as a phenomenological parameter and back-extrapolating it as a function of V. In the case of the Hubbard model ($V = 0$), the effective attraction is $\sim 0.1t$ above the baseline (consistent with the $J/4$ estimation), whereas the experimental case is at least $\sim 1t$ above the baseline, an order of magnitude larger. This difference implies that the single-band Hubbard model misses a sizable, attractive interaction between neighboring holes, which likely originates from electron-phonon coupling (*19*, *30*, *31*).

Considering the structural similarities among cuprates, our Hamiltonian with near-neighbor attraction should be applicable to CuO_2 planes

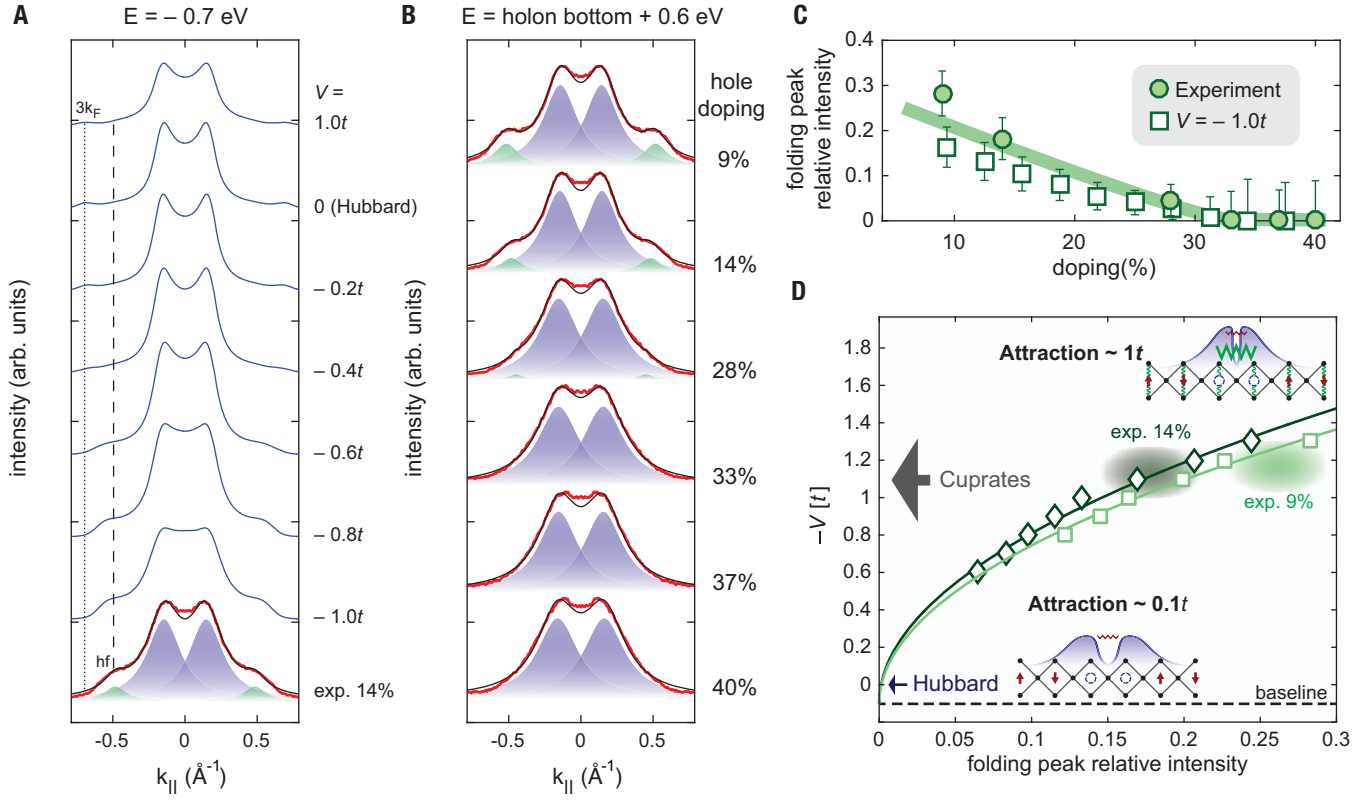


Fig. 4. Strong holon folding indicating attractive interaction beyond Hubbard model. (**A**) CPT simulation incorporating a near-neighbor interaction term V (blue curves), compared with experimental MDC at −0.7 eV below E_F for 14% doping (red curve). Black curve is a fit utilizing two Lorentzian peaks shown in blue (major) and green (hf). Vertical dashed and dotted lines mark the locations of the hf and the $3k_F$ branch peaks, respectively. Curves are shifted and scaled vertically for comparison. (**B**) Doping-dependent experimental MDCs at energies approximately 0.6 eV above the holon bottom, such that the major branch peaks have similar momenta for comparison. (**C**) Folding peak intensity relative to the major peak intensity as a function of doping (circles: experiment; squares: simulations for $V = -1.0t$). The green line is a guide to the eye. (**D**) $-V$ as a function of CPT simulated folding peak relative intensity f as a phenomenological measure of the net near-neighbor attraction (squares: 9%; diamonds: 12.5%). The solid green curves are a fit to the data using $f = a(V - V_0)^2$, where a is a doping-dependent prefactor and V_0 represents the compensation against the effective attraction in the Hubbard model, ~ (0.10 ± 0.05)t given by the fitting. The gray shaded areas correspond to experimental points (and error bars) of 9 and 14% dopings. Insets: Schematic of holon attractions for the pure Hubbard model (lower) and the mediation by bosons (upper). In addition to the symbols introduced in Fig. 1A, green springs, blue shades, and red and green zigzags represent electron-phonon coupling, holon distributions, AFM-, and boson-mediated attractions, respectively.

and favors neighboring electron pairs. Given that the Hubbard model with strong on-site repulsive U and strong near-neighbor attractive V explains the essential physics in 1D—charge, spin, charge-charge attraction and its doping dependence—it is plausible that such an augmented Hubbard model provides a holistic picture for all cuprates, including robust d-wave superconductivity that has been elusive in all 2D Hubbard-like models.

REFERENCES AND NOTES

1. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
2. E. H. Lieb, F. Y. Wu, *Phys. Rev. Lett.* **20**, 1445–1448 (1968).
3. J. Voit, *Rep. Prog. Phys.* **58**, 977–1116 (1995).
4. C. Kim *et al.*, *Phys. Rev. Lett.* **77**, 4054–4057 (1996).
5. H. Fujisawa *et al.*, *Phys. Rev. B* **59**, 7358–7361 (1999).
6. B. J. Kim *et al.*, *Nat. Phys.* **2**, 397–401 (2006).
7. J. Schlappa *et al.*, *Nature* **485**, 82–85 (2012).
8. Z. Liu, T. Hanada, R. Sekine, M. Kawai, H. Koinuma, *Appl. Phys. Lett.* **65**, 1717–1719 (1994).
9. D. R. Lines, M. T. Weller, D. B. Currie, D. M. Ogbome, *Mater. Res. Bull.* **26**, 323–331 (1991).
10. H. Sato, M. Naito, H. Yamamoto, *Physica C* **280**, 178–186 (1997).
11. Y. F. Kung *et al.*, *Phys. Rev. B* **93**, 155166 (2016).
12. D. Sénéchal, D. Perez, M. Pioro-Ladrière, *Phys. Rev. Lett.* **84**, 522–525 (2000).
13. D. Sénéchal, D. Perez, D. Plouffe, *Phys. Rev. B* **66**, 075129 (2002).
14. M. Kohno, *Phys. Rev. Lett.* **105**, 106402 (2010).
15. H. Benthien, F. Gebhard, E. Jeckelmann, *Phys. Rev. Lett.* **92**, 256401 (2004).
16. A. Nocera, F. H. L. Essler, A. E. Feiguin, *Phys. Rev. B* **97**, 045146 (2018).
17. K. Penc, K. Hallberg, F. Mila, H. Shiba, *Phys. Rev. Lett.* **77**, 1390–1393 (1996).
18. K. Penc, K. Hallberg, F. Mila, H. Shiba, *Phys. Rev. B* **55**, 15475–15488 (1997).
19. See supplementary materials.
20. A. Lanzara *et al.*, *Nature* **412**, 510–514 (2001).
21. K. M. Shen *et al.*, *Phys. Rev. Lett.* **93**, 267002 (2004).
22. T. Cuk *et al.*, *Phys. Rev. Lett.* **93**, 117003 (2004).
23. A. S. Mishchenko, N. Nagaosa, *Phys. Rev. Lett.* **93**, 036402 (2004).
24. H. Matsueda, T. Tohyama, S. Maekawa, *Phys. Rev. B* **74**, 241103 (2006).
25. C. Slezak, A. Macridin, G. A. Sawatzky, M. Jarrell, T. A. Maier, *Phys. Rev. B* **73**, 205122 (2006).
26. O. Gunnarsson, O. Rösch, *Phys. Rev. B* **73**, 174521 (2006).
27. Y. He *et al.*, *Science* **362**, 62–65 (2018).
28. W. S. Lee *et al.*, *Phys. Rev. Lett.* **110**, 265502 (2013).
29. J. J. Lee *et al.*, *Phys. Rev. B* **89**, 041104 (2014).
30. H. Suzuura, H. Yasuhara, A. Furusaki, N. Nagaosa, Y. Tokura, *Phys. Rev. Lett.* **76**, 2579–2582 (1996).
31. S. Johnston *et al.*, *Phys. Rev. B* **82**, 064513 (2010).
32. Z. Chen *et al.*, Replication Data for: Anomalously Strong Near-Neighbor Attraction in Doped 1D Cuprate Chains, *Harvard Dataverse* (2021). https://doi.org/10.7910/DVN/5I1WSW.

ACKNOWLEDGMENTS

We thank S. A. Kivelson, D. H. Lee, and D. Orgad for stimulating discussions. ARPES experiments were performed at Beamline 5-2, Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory. **Funding:** This work was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract DE-AC02-76SF00515. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility operated under Contract no. DE-AC02-05CH11231. R.G.M. acknowledges support by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory,

managed by UT-Battelle, LLC, for the US Department of Energy. Y.W. acknowledges support from NSF award DMR-2038011.
Author contributions: Z.C. conceived the experiment, grew the films, performed ARPES measurements, and analyzed data. Y.W. performed theoretical and numerical calculations. S.N.R. assisted in film growth and ARPES measurements. T.J. assisted in film growth. M.H. and D.L. developed the ARPES setup and assisted in ARPES measurements. B.M. assisted in calculations and data interpretations. R.G.M. developed the MBE setup and assisted in film growth. Z.C., Y.W., T.P.D., and Z.X.S. interpreted the data and wrote the manuscript with input from all authors. T.P.D. and Z.X.S. supervised the project on theory and experiment aspects, respectively. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data presented in this study are available at Harvard Dataverse (*32*).

SUPPLEMENTARY MATERIALS
https://science.org/doi/10.1126/science.abf5174
Materials and Methods
Supplementary Text
Figs. S1 to S7

31 October 2020; accepted 10 August 2021
10.1126/science.abf5174

PHYSICS

Pendellösung interferometry probes the neutron charge radius, lattice dynamics, and fifth forces

Benjamin Heacock[1,2,3]*, Takuhiro Fujiie[4,5], Robert W. Haun[6,7], Albert Henins[1], Katsuya Hirota[4,8], Takuya Hosobata[5], Michael G. Huber[1], Masaaki Kitaguchi[4,9], Dmitry A. Pushin[10,11], Hirohiko Shimizu[4], Masahiro Takeda[5], Robert Valdillez[2,3], Yutaka Yamagata[5], Albert R. Young[2,3]

Structure factors describe how incident radiation is scattered from materials such as silicon and germanium and characterize the physical interaction between the material and scattered particles. We used neutron Pendellösung interferometry to make precision measurements of the (220) and (400) neutron-silicon structure factors and achieved a factor-of-four improvement in the (111) structure factor uncertainty. These data provide measurements of the silicon Debye-Waller factor at room temperature and the mean square neutron charge radius $\langle r_n^2 \rangle = -0.1101 \pm 0.0089$ square femtometers. Combined with existing measurements of the Debye-Waller factor and charge radius, the measured structure factors also improve constraints on the strength of a Yukawa modification to gravity by an order of magnitude over the 20 picometer–to–10 nanometer length scale range.

Neutrons, electrons, and x-rays all exhibit Pendellösung interference upon Bragg diffraction in the Laue geometry, where the Bragg planes are perpendicular to the entrance and exit faces of the crystal (Fig. 1A). The phenomenon—characterized by oscillations of diffracted intensity as a function of probe wavelength, crystal thickness, and lattice potential—was predicted by Ewald (*1*) and first observed in electron diffraction from MgO (*2*); it was later demonstrated by using x-rays (*3*) and neutrons (*4*, *5*). X-ray Pendellösung studies in particular (*3*, *6*–*9*) have proved an invaluable tool for measuring the electron density in silicon (*10*–*13*) and provide data that can be compared with ab initio lattice dynamical models (*14*). Neutron Pendellösung measurements yield information complementary to that obtained with other probes; however, achieving high precision has proven challenging.

A neutron wave propagating along the symmetry planes of a Bragg-diffracting crystal forms standing waves with integer periodicity to the Bragg planes. The neutron's kinematic momentum then depends on the incident neutron energy and the overlap of the standing wave with the crystalline potential. The detuning of the incident neutron wave into energetically degenerate "fast" and "slow" modes creates a spatial beating period along the Bragg planes between the diffracted and transmitted intensities that may be resolved as Pendellösung oscillations. The absolute phase shift of the Pendellösung interference fringes is determined by the material's structure factor, which is specific to the radiation species and Miller indices (hkl) of the Bragg planes. For the neutron case, structure factors can be described in terms of the thermally averaged single-atom coherent elastic scattering amplitudes $b(Q) = 2me^{-W}\tilde{V}(Q)/\hbar^2$, where $\hbar$ is the reduced Planck constant, m is the neutron mass, and $\tilde{V}(Q)$ is the Fourier transform of the neutron-atomic potential. The Debye-Waller factor (DWF; e^{-W}) accounts for the thermal motion of the atoms in the lattice, where $W = BQ^2/(16\pi^2)$ and B is determined by the mean-square thermal atomic displacement $\langle u^2 \rangle = B/(8\pi^2)$. The momentum transfer magnitude is $Q_{hkl} = (2\pi/a)\sqrt{h^2 + l^2 + k^2}$, where a is the lattice constant for a material with a cubic unit cell.

The main contribution to $b(Q)$ is the coherent nuclear scattering length, which is Q-independent. The leading-order Q-dependent contribution to $b(Q)$ (Fig. 1F) is the DWF, exhibiting a relative impact of a few percent. A determination of B can provide a benchmark for lattice dynamical models. The next-to-leading-order contribution to the Q-dependence of $b(Q)$ arises from the neutron's spherically symmetric charge distribution (which is a result of its three-quark composite structure) interacting with the very large interatomic electric fields of ~10^8 V cm^{-1} (*15*). This interaction is described by the mean square neutron charge radius $\langle r_n^2 \rangle$, where $-\langle r_n^2 \rangle/6$ is the slope of the neutron electric form factor $G_E^n(Q^2)$ with respect to Q^2 at zero momentum transfer. Determinations of $\langle r_n^2 \rangle$ along with electron scattering data can be used to study the neutron's internal charge distribution (*16*–*18*) or as a parameter in chiral effective field theory (EFT) studies of light nuclei (*19*). As proposed by (*20*), a Pendellösung inteferometry determination of $\langle r_n^2 \rangle$ can weigh in on the slight tension between individual neutron scattering experiments making up the Particle Data Group (PDG)–recommended value of $\langle r_n^2 \rangle_{\mathrm{PDG}} = -0.1161(22)$ fm^2 (*21*).

Aside from the expected B and $\langle r_n^2 \rangle$ contributions, $b(Q)$ is sensitive to interactions beyond the standard model of particle physics, often referred to as BSM interactions. BSM interactions on the atomic length scale would result in a distinctive Q-dependence (*22*). Yet-to-be-detected "fifth" forces arise from several BSM theories seeking to explain mysteries of modern physics, including the incompatibility of general relativity and quantum mechanics (*23*–*25*), dark matter and/or dark energy (*26*–*30*), and the smallness of neutrino masses (*31*). The large variety of BSM theories has motivated a multidisciplinary experimental effort to constrain the yet-undiscovered physics over the entirety of physical observation scales (35 orders of magnitude in length scale) from collider experiments at the shortest length scales and highest energies to astrophysical observations at the largest length scales and lowest energies (*32*).

We tabulated the measured $b(Q_{hkl})$ for silicon and used these data to (i) make a high-precision determination of B for silicon at 295.5 K, (ii) measure the neutron charge radius, and (iii)

[1]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. [2]Department of Physics, North Carolina State University, Raleigh, NC 27695, USA. [3]Triangle Universities Nuclear Laboratory, Durham, NC 27708, USA. [4]Department of Physics, Nagoya University, Nagoya 464-8602, Japan. [5]RIKEN Center for Advanced Photonics, RIKEN, Hirosawa 2-1, Wako, Saitama 351-0198, Japan. [6]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA. [7]Department of Physics and Engineering Physics, Tulane University, New Orleans, LA 70118, USA. [8]High Energy Accelerator Research Organization, Tsukuba 305-0801, Japan. [9]Kobayashi-Maskawa Institute, Nagoya University, Nagoya 464-8602, Japan. [10]Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L3G1, Canada. [11]Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada, N2L3G1
*Corresponding author. Email: benjamin.heacock@nist.gov

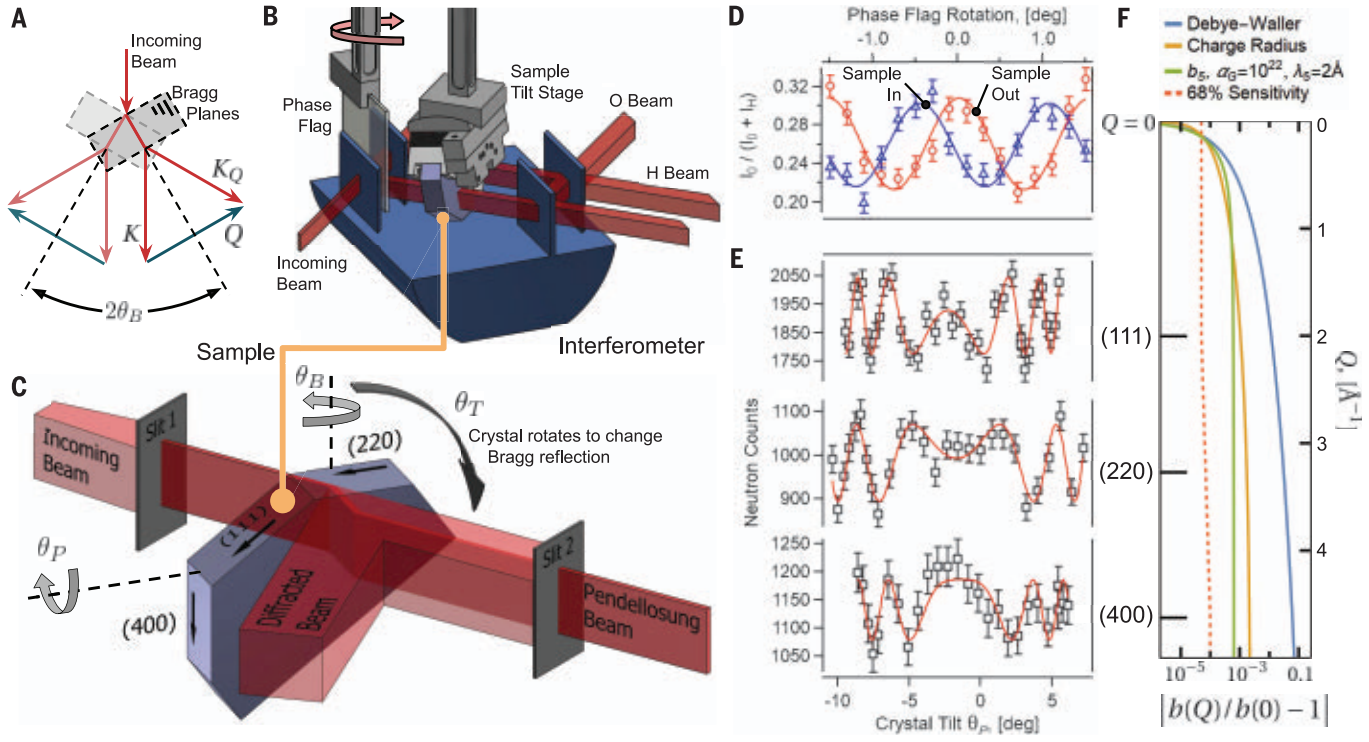


Fig. 1. Pendellösung interference in silicon. (**A**) Bragg diffraction Laue geometry can be measured when the crystal is rotated by $\pm\theta_B$ relative to the incoming monochromatic beam. Here, Q is the momentum transfer, and K and K_Q are the forward-diffracted and diffracted wave vectors, respectively. (**B**) The quantity $b(0)D$ is measured with an interferometer. (**C**) $b(Q)D$ is measured by using Pendellösung oscillations in the forward-diffracted beam. The (400) arrow points in the (004) direction; we indicate Miller indices with the largest value first to emphasize that $b(Q)$ depends on the magnitude of Q. (**D** and **E**) Typical interferograms with lines of best fit for the interferometer and pendellösung setups, respectively. Error bars are from Poisson counting statistics. (**F**) The theoretical relative contributions of each Q-dependent term in $b(Q)$. The Debye-Waller term is negative, whereas the other two terms are positive. The data are sensitive (68% level) to features larger than the dashed line.

place constraints on the strength of a BSM Yukawa modification of gravity over the 20 pm–to–10 nm length scale range.

Pendellösung interference may be resolved in either the diffracted or forward-diffracted beams (Fig. 1A, K_Q and K, respectively) by rotating a crystal slab precisely about the axis perpendicular to the diffraction plane [wavelength method (*5*)] or perpendicular to the Bragg planes [crystal thickness method (*33*), used here], creating an interferogram (Fig. 1E). The interferogram is then fit to a functional form (eq. S1) to determine the phase of the oscillations, which is proportional to $b(Q)$, the neutron wavelength λ, and the thickness of the crystal slab D. We also made forward-scattering ($Q = 0$) measurements for each sample orientation using a perfect-crystal neutron interferometer, a device in which a monolithic crystal connects Bragg-diffracting "blades" that protrude from a common base (Fig. 1B) (*34*). The phase shift between the two interferometer paths is modulated by rotating a fused silica phase flag, creating an interferogram (Fig. 1D). The phase flag interferograms are likewise fit to a functional form (eq. S4), and the phase is extracted (*34*). The forward-scattering phase shift from the sample is then the difference in the fitted phase between the sample-in and sample-out of the beam (Fig. 1D) and is proportional to $b(0)$, λ, and D.

The extraction of $b(Q)D$ from the measured phase shift was enabled by an in situ determination of λ in which Pendellösung interferograms were obtained with the crystal rotated by both positive and negative Bragg angles $\pm\theta_B$ (Fig. 1A). The average phase shift is then quadratic in the slight difference in wavelength between the two interferograms and small enough to be neglected. The average wavelength was computed by using $\lambda = 4\pi\sin\theta_B/Q_{hkl}$, the silicon lattice constant, and $2\theta_B$ as given by an angular encoder embedded in the crystal's rotational positioning stage. A similar approach for measuring wavelength was used for the forward scattering measurements, with the interferometer and accompanying optics rotated by $\pm\theta_B$ of the interferometer.

Prior measurements of $b(Q_{111})$ were subject to strain fields in the crystal slab, observed as a one-directional phase shift with distinct wavelength dependence (*35*). However, the reported strain gradients were much larger than what is observed in neutron interferometers generally (*36*, *37*) and specifically in a neutron interferometer that used a very similar machining and post-fabrication process (*38*) to the Pendellösung sample used here. Reducing strain gradients invariably comes at the cost of increased variation in D, but a strain-relieving acid etch now typical for neutron interferometers was not performed in the prior measurement of $b(Q_{111})$ (*35*). To separate $b(Q)$ from D, we normalized the Pendellösung phase shift with the forward scattering phase shift measured over the same relevant crystal volume for each Bragg reflection. To this end, flats were cut on the Pendellösung sample so that it could fit between the interferometer optical components, with the neutron beam still illuminating the required sample area. The ratio of the two measurements forms $b(Q)/b(0)$, naturally isolating the Q-dependence of the scattering amplitude while also eliminating the need to measure D by other means. This normalization relaxes the requisite crystal flatness, enabling measurements of $b(Q)/b(0)$ using an acid-etched, strain-free crystal.

Both Pendellösung and forward scattering measurements were performed at the neutron interferometer and optics facility auxiliary beamline (NIOFa) at the National Institute of Standards and Technology (NIST) Center for Neutron Research (NCNR). The NIOFa

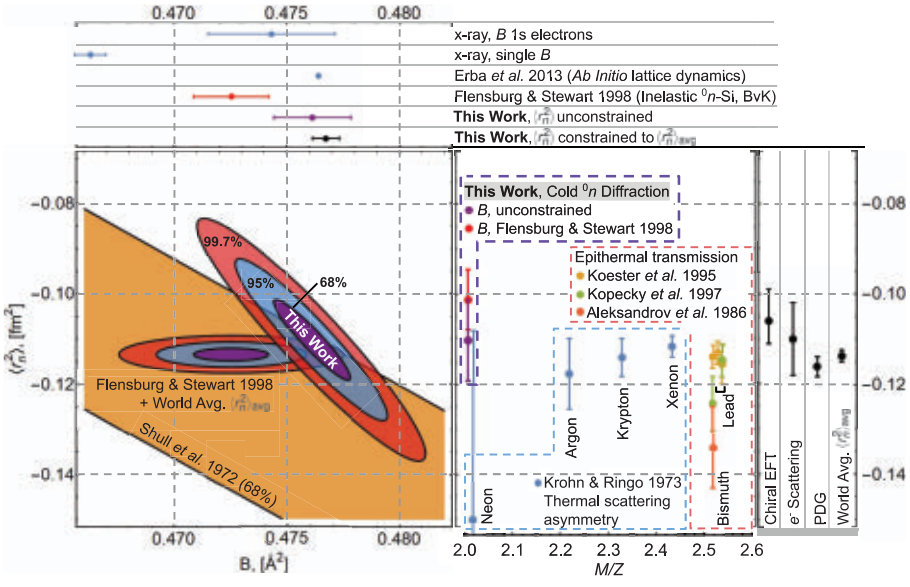


Fig. 2. Confidence regions over $\langle r_n^2 \rangle$ and *B*. Data have been rescaled to 295.5 K and compared with previous work. Our data are consistent with the previous (111) silicon measurement Shull *et al*. 1972 (*35*) but have four times the precision. The Erba *et al*. 2013 (*14*) determination of *B* is a theoretically computed value that does not have an estimated uncertainty. Other previous work includes that of Flensburg and Stewart 1998 (*42*), Krohn and Ringo 1973 (*45*), Koester *et al*. 1995 (*58*), Kopecky *et al*. 1997 (*43*), Aleksandrov *et al*. 1986 (*59*), as well as chiral EFT (*19*), e^- scattering (*18*), PDG (*21*), and world average (World Avg.) (*41*). A detailed discussion is provided in (*41*).

Table 1. Measured structure factors and uncertainties (68% confidence intervals). All results are scaled to 295.5 K.

hkl	*b*(*Q*)/*b*(0) − 1	Uncertainty statistical	Uncertainty systematic	Uncertainty total
(111)	−0.011075	3.5×10^{-5}	3.8×10^{-5}	5.1×10^{-5}
(220)	−0.030203	5.9×10^{-5}	3.6×10^{-5}	6.9×10^{-5}
(400)	−0.060596	8.0×10^{-5}	4.9×10^{-5}	9.4×10^{-5}

provides both 2.2 and 4.4 Å monochromatic ($\Delta\lambda/\lambda \sim 0.5\%$) neutrons (*39*). The (220) and (400) Pendellösung interferograms were measured by using 2.2 Å, and the (111) Pendellösung interferograms and interferometer forward scattering measurements used 4.4 Å. The resulting ratios *b*(*Q*)/*b*(0) for the (111), (220), and (400) Bragg reflections are reported in Table 1.

The theoretical shape versus *Q* of the contributions to *b*(*Q*)/*b*(0) from the DWF and $\langle r_n^2 \rangle$ are shown in Fig. 1F. To extract these two parameters, the three measured values of *b*(*Q*)/*b*(0) were fit to

$$b(Q)/b(0) = e^{-W}\left[1 - Z\frac{b_{ne}}{b(0)}\right] + f_e(hkl)\frac{b_{ne}}{b(0)} \quad (1)$$

by minimizing the χ^2 sum of weighted residuals, where *Z* is the crystal's atomic number and *B* and b_{ne} are treated as fit parameters. The room-temperature x-ray silicon form factors $f_e(hkl)$ in units of elementary charge are given by x-ray scattering measurements (*13*). The neutron-electron scattering length b_{ne} is related to the charge radius by $b_{ne} = \langle r_n^2 \rangle m_n/(3m_e a_0) = \langle r_n^2 \rangle/(86.34\ \text{fm})$, with the electron mass m_e and Bohr radius a_0. The fit function's sensitivity to the forward scattering length *b*(0) (*40*) and $f_e(hkl)$ (*13*) is small compared with experimental uncertainties.

If *B* and $\langle r_n^2 \rangle$ are treated as free parameters, then our results are $B = 0.4761(17)$ Å^2 and $\langle r_n^2 \rangle = -0.1101(89)$ fm^2 with a correlation coefficient of −0.94; the χ^2-surface of *B* and $\langle r_n^2 \rangle$ versus prior determinations of both parameters is shown in Fig. 2. Detailed comparisons are made in (*41*), but of particular interest is the lattice dynamical result from fitting neutron inelastic scattering data to a Born–von Kármán (BvK) model $B_{\text{BvK}} = 0.4725(17)$ Å^2 when scaled to 295.5 K (*42*) and the world average $\langle r_n^2 \rangle_{\text{avg}} = -0.1137(13)$ fm^2 from prior neutron scattering experiments, the computation of which is described in (*41*). The confidence region formed by B_{BvK} and $\langle r_n^2 \rangle_{\text{avg}}$ shows slight tension with our results.

In regard to $\langle r_n^2 \rangle$, most prior determinations come from epithermal neutron total transmission through lead or bismuth, with potentially correlated systematic uncertainties from solid-state and nuclear resonance effects (*43*). The only other type of experiment contributing to $\langle r_n^2 \rangle_{\text{PDG}}$ and $\langle r_n^2 \rangle_{\text{avg}}$ are scattering asymmetry measurements in noble gasses (*44*, *45*). Pendellösung interferometry constitutes the only determination of $\langle r_n^2 \rangle$ that uses cold neutrons and contains entirely different systematic uncertainties compared with these previous methods (*20*, *46*). Furthermore, a recent determination of $\langle r_n^2 \rangle_{\text{EFT}} = -0.106^{+0.007}_{-0.005}$ fm^2 from chiral EFT combined with measurements of the hydrogen-deuterium isotope shift (*47*) is consistent with our results and shifted in the same direction relative to $\langle r_n^2 \rangle_{\text{PDG}}$, with $\langle r_n^2 \rangle_{\text{EFT}}$ showing a 1.8σ difference with $\langle r_n^2 \rangle_{\text{PDG}}$ and a less-concerning 1.4σ difference with $\langle r_n^2 \rangle_{\text{avg}}$.

It may be possible to decrease uncertainties in the extracted $\langle r_n^2 \rangle$ by constraining *B*. However, the tendency for x-ray structure factor determinations of *B* to increasingly deviate from theory for higher-order structure factors (*14*, *48*), as well as the disagreement between *B* as determined with x-rays versus neutrons, suggests a breakdown in the rigid atom approximation (*41*). On the other hand, neutron inelastic scattering data, as well as fully anharmonic lattice dynamical models, have shown that harmonic approximations used by the BvK and other models are insufficient for silicon's strong anharmonicity (*49*–*51*). Furthermore, the BvK model was unable to predict the measured frequencies on the phonon dispersion curves for relative uncertainties less than 0.5% (*42*), suggesting an inadequacy of the BvK model at this level of precision, in which case the quoted relative uncertainty for B_{BvK} of 0.34% may be too small. Whether semi-empirically fitting the interatomic force constants from other lattice dynamical models would produce better fits to neutron inelastic scattering data, and whether the resulting determination of *B* in turn agrees with our results, could provide a means for evaluating different thermodynamic lattice models. Because the reduced χ^2 for the BvK model fit was $\chi^2_{\text{red}} = 33.5$ (193 − 21 degrees of freedom) before the expansion of any experimental uncertainties (*42*), a particularly successful lattice model may be able to reduce uncertainties in *B* by up to a factor of $\sqrt{\chi^2_{\text{red}}} = 5.8$ relative to B_{BvK}. If the resulting 0.06% measurement of *B* were applied as a constraint to our evaluation of $\langle r_n^2 \rangle$, then the uncertainty thereof would be reduced to $\lesssim 0.004$ fm^2.

A future test of lattice dynamical models by using Pendellösung interferometry may be achieved from the sensitivity of odd structure

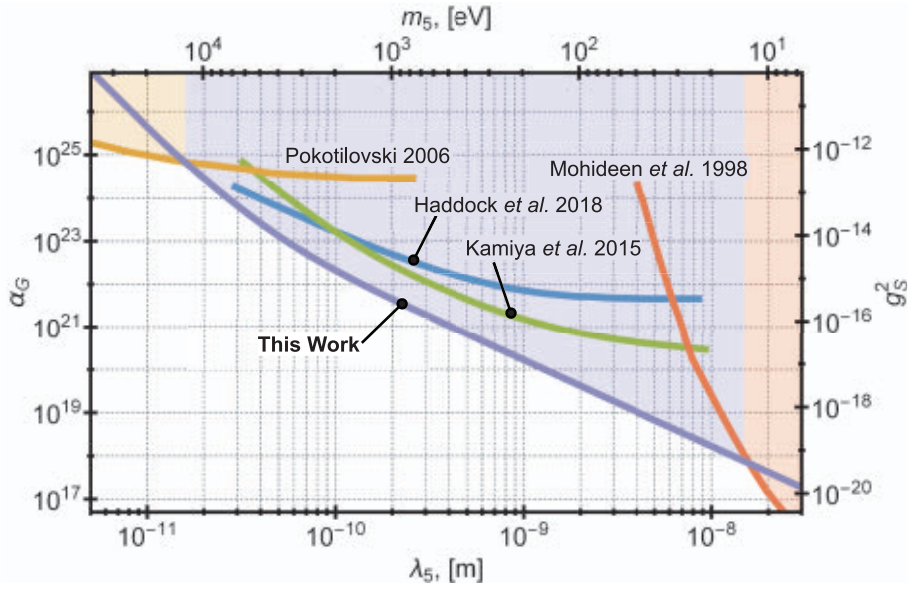


Fig. 3. Limits (95% confidence) on the strength of a Yukawa modification to gravity a_G compared with previous experiments as a function of the force's range λ_5. The same limits constrain the coupling to nucleon number g_S^2 of a yet-undiscovered scalar with mass m_5. The shaded region is excluded. Other limits shown are those of Pokotilovski 2006 (*56*), Mohideen *et al*. 1998 (*60*), Haddock *et al*. 2018 (*55*), and Kamiya *et al*. 2015 (*54*).

factors to three-phonon terms by means of the anharmonic DWF (aDWF). Of the reflections in Table 1, only the (111) reflection is affected by the aDWF and only at the 3×10^{-5} level (*41*). However, the contribution grows like $h \times k \times l$ (*42*), and a measurement of the (333) structure factor with a relative precision similar to Table 1 is capable of providing a determination of the average cubic displacement along the silicon bond direction $\langle u_{111}^3 \rangle$ at the 5% level. The near-room-temperature (288 K) $\langle u_{111}^3 \rangle$ has only been measured with 10% relative uncertainty through the "forbidden" (222) structure factor, which itself deviated from the expected T^2 temperature dependence extrapolated from higher temperatures (>650 K) by a factor of two (*52*), suggesting that quantum zero-point anharmonic vibrations may contribute meaningfully to $\langle u_{111}^3 \rangle$ at room temperature. Pendellösung interferometry is distinctive in its capability to be extended to cryogenic temperatures to measure $\langle u_{111}^3 \rangle$ in the regime at which the anharmonic zero-point vibrations are the most relevant. This is of particular interest for silicon, given how the anharmonic force constants determine silicon's electrical properties and anomalous thermal expansion at low temperatures (*51*), providing motivation to measure higher-order odd $b(Q_{hkl})$ at a variety of temperatures.

A BSM "fifth" fundamental force or other BSM physics adds an additional term

$$\delta\left[\frac{b(Q)}{b(0)}\right] = e^{-W}\frac{b_5(Q) - b_5(0)}{b(0)} \qquad (2)$$

to Eq. 1, given a BSM contribution to the neutron-atomic potential $b_5(Q) = 2m\tilde{V}_5(Q)/\hbar^2$. The method for using neutron structure factors to constrain BSM physics is general: A χ^2 sum of weighted residuals for the measured structure factors—as well as B, the aDWF, and $\langle r_n^2 \rangle$—is minimized with respect to BSM model parameters, given a functional form of $b_5(Q)$. The weights and values assigned to the B, $\langle r_n^2 \rangle$, and the aDWF residuals are to be based on an estimation of how the BSM physics would bias the experiments from which the residuals are derived. The tabulated structure factors can thus continue to constrain future BSM theories as they are developed, with improved constraints as the number of measured structure factors increases.

A common parameterization for constraining BSM physics is a Yukawa modification to gravity (*24*), leading to $\delta[b(Q)/b(0)]e^W = \alpha_G(3.9 \times 10^{-27}\ \text{Å}^{-2}\ \lambda_5^2)Q^2\lambda_5^2/(1 + Q^2\lambda_5^2)$ for the case of silicon in Eq. 2 (*22*); this expression depends on an interaction strength relative to gravity α_G and length scale λ_5. The coupling to nucleon number g_S of an undiscovered massive scalar with particle mass $m_5 = \hbar/(c\lambda_5)$, where c is the speed of light, proposed by a number of BSM theories is constrained by the same limits (*24*, *25*, *53*). The constraints on the strength of an undiscovered Yukawa potential over the relevant range of λ_5 from this work compared with previous experiments is shown in Fig. 3. Our results constitute an order-of-magnitude improvement over nearly three decades in λ_5. Previous limits constrain $\langle r_n^2 \rangle$ according to the PDG value and uncertainty (*54*, *55*), except for a higher-energy experiment that uses an expanded uncertainty (*56*), in which a systematic shift in the measured $\langle r_n^2 \rangle$ from an unaccounted-for Yukawa modification to gravity could be relevant for $\lambda_5 \lesssim 10$ pm. In keeping with other previously published constraints, we likewise set the $\langle r_n^2 \rangle$ residual according to the PDG value and uncertainty when computing the limits in Fig. 3. In addition to measuring $b(Q)$ for higher-order reflections in silicon, measuring multiple $b(Q)$ from germanium, for which precision measurements only exist for the (111) reflection (*57*), is expected to improve constraints on α_G by a factor of five, regardless of whether $\langle r_n^2 \rangle$ is constrained because of the differing M/Z for germanium and silicon. Germanium measurements would also reduce the uncertainty in our measured $\langle r_n^2 \rangle$ to ~0.005 fm^2 with B from both atomic species unconstrained, in which case Pendellösung interferometry could weigh in more strongly on the $\langle r_n^2 \rangle$ landscape. If determinations of B for silicon and germanium can be achieved at the 0.1% level by using neutron inelastic scattering and lattice dynamical models, then an achievable uncertainty for $\langle r_n^2 \rangle$ from Pendellösung interferometry of 0.002 fm^2 would be comparable with the uncertainty of $\langle r_n^2 \rangle_{\text{PDG}}$. Ultimately, whereas x-ray structure factors in silicon as large as (12120) and (880) have been measured at room and liquid-nitrogen temperatures, respectively (*12*, *13*), an expanding set of tabulated neutron structure factors for silicon, germanium, and other crystals at multiple temperaures would provide increasingly sensitive measures of material-specific thermal displacement parameters and the neutron charge radius while continuing to improve BSM constraints.

REFERENCES AND NOTES

1. P. P. Ewald, *Ann. Phys.* **354**, 1–38 (1916).
2. R. D. Heidenreich, *Phys. Rev.* **62**, 291–292 (1942).
3. N. Kato, A. Lang, *Acta Crystallogr.* **12**, 787–794 (1959).
4. D. Sippel, K. Kleinstück, G. Schulze, *Phys. Lett.* **14**, 174–175 (1965).
5. C. Shull, *Phys. Rev. Lett.* **21**, 1585–1589 (1968).
6. H. Hattori, Hideo Kuriyama, T. Katagawa, N. Kato, *J. Phys. Soc. Jpn.* **20**, 988–996 (1965).
7. S. Tanemura, N. Kato, *Acta Crystallogr. A* **28**, 69–80 (1972).
8. M. Wada, N. Kato, *Acta Crystallogr. A* **33**, 161–168 (1977).
9. M. Deutsch, M. Hart, *Acta Crystallogr. A* **41**, 48–55 (1985).
10. B. Willis, *Acta Crystallogr. A* **25**, 277–300 (1969).
11. P. Aldred, M. Hart, *Proc. R. Soc. London Ser. A* **332**, 223–238 (1973).
12. M. Spackman, *Acta Crystallogr. A* **42**, 271–281 (1986).
13. S. Cummings, M. Hart, *Aust. J. Phys.* **41**, 423 (1988).
14. A. Erba, M. Ferrabone, R. Orlando, R. Dovesi, *J. Comput. Chem.* **34**, 346–354 (2013).
15. V. Fedorov *et al*., *Phys. Lett. B* **694**, 22–25 (2010).
16. T. R. Gentile, C. Crawford, *Phys. Rev. C Nucl. Phys.* **83**, 055203 (2011).
17. H. Atac, M. Constantinou, Z.-E. Meziani, M. Paolone, N. Sparveris, *Eur. Phys. J. A* **57**, 65 (2021).
18. H. Atac, M. Constantinou, Z.-E. Meziani, M. Paolone, N. Sparveris, *Nat. Commun.* **12**, 1759 (2021).
19. E. Epelbaum, H. Krebs, P. Reinert, *Front. Phys.* **8**, 98 (2020).
20. J.-M. Sparenberg, H. Leeb, *Phys. Rev. C Nucl. Phys.* **66**, 055210 (2002).

21. M. Tanabashi *et al.*, *Phys. Rev. D* **98**, 030001 (2018).
22. G. L. Greene, V. Gudkov, *Phys. Rev. C Nucl. Phys.* **75**, 015501 (2007).
23. N. Arkani-Hamed, S. Dimopoulos, G. Dvali, *Phys. Rev. D Part. Fields* **59**, 086004 (1999).
24. E. G. Adelberger, B. R. Heckel, A. E. Nelson, *Annu. Rev. Nucl. Part. Sci.* **53**, 77–121 (2003).
25. V. M. Mostepanenko, G. L. Klimchitskaya, *Universe* **6**, 147 (2020).
26. A. Zee, *Phys. Lett. B* **594**, 8–12 (2004).
27. P. Fayet, *Phys. Rev. D Part. Fields Gravit. Cosmol.* **75**, 115017 (2007).
28. S. Fichet, *Phys. Rev. Lett.* **120**, 131801 (2018).
29. P. Brax, S. Fichet, G. Pignol, *Phys. Rev. D* **97**, 115034 (2018).
30. P. Brax, S. Fichet, P. Tanedo, *Phys. Lett. B* **798**, 135012 (2019).
31. N. Arkani-Hamed, S. Dimopoulos, *Phys. Rev. D Part. Fields* **65**, 052003 (2002).
32. J. Murata, S. Tanaka, *Class. Quantum Gravity* **32**, 033001 (2015).
33. V. Somenkov, S. S. Shilstein, N. Belova, K. Utemisov, *Solid State Commun.* **25**, 593–595 (1978).
34. H. Rauch, S. A. Werner, *Neutron Interferometry: Lessons in Experimental Quantum Mechanics, Wave-particle Duality, and Entanglement*, vol. 12 (Oxford Univ. Press, 2015).
35. C. Shull, J. Oberteuffer, *Phys. Rev. Lett.* **29**, 871–874 (1972).
36. B. Heacock *et al.*, *Phys. Rev. A* **95**, 013840 (2017).
37. B. Heacock *et al.*, *Rev. Sci. Instrum.* **89**, 023502 (2018).
38. B. Heacock *et al.*, *Acta Crystallogr. A Found. Adv.* **75**, 833–841 (2019).
39. C. Shahi *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **813**, 111–122 (2016).
40. A. Ioffe *et al.*, *Phys. Rev. A* **58**, 1475–1479 (1998).
41. Materials and methods are available as supplementary materials.
42. C. Flensburg, R. F. Stewart, *Phys. Rev. B Condens. Matter Mater. Phys.* **60**, 284–291 (1999).
43. S. Kopecky *et al.*, *Phys. Rev. C Nucl. Phys.* **56**, 2229–2237 (1997).
44. V. Krohn, G. Ringo, *Phys. Rev.* **148**, 1303–1311 (1966).
45. V. Krohn, G. Ringo, *Phys. Rev. D Part. Fields* **8**, 1305–1307 (1973).
46. J.-M. Sparenberg, H. Leeb, *J. Electron Spectrosc. Relat. Phenom.* **129**, 315–317 (2003).
47. A. A. Filin *et al.*, *Phys. Rev. Lett.* **124**, 082501 (2020).
48. M. Deutsch, M. Hart, P. Sommer-Larsen, *Phys. Rev. B Condens. Matter* **40**, 11666–11669 (1989).
49. O. Hellman, I. A. Abrikosov, *Phys. Rev. B Condens. Matter Mater. Phys.* **88**, 144301 (2013).
50. D. Kim *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 014307 (2015).
51. D. S. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1992–1997 (2018).
52. J. Roberto, B. Batterman, D. Keating, *Phys. Rev. B* **9**, 2590–2599 (1974).
53. V. Nesvizhevsky, G. Pignol, K. Protasov, *Phys. Rev. D Part. Fields Gravit. Cosmol.* **77**, 034020 (2008).
54. Y. Kamiya, K. Itagaki, M. Tani, G. N. Kim, S. Komamiya, *Phys. Rev. Lett.* **114**, 161101 (2015).
55. C. C. Haddock *et al.*, *Phys. Rev. D* **97**, 062002 (2018).
56. Y. N. Pokotilovski, *Phys. Lett. B* **719**, 341–345 (2013).
57. C. Shull, W. Shaw, *Z. Naturforsch. A* **28**, 657–661 (1973).
58. L. Koester *et al.*, *Phys. Rev. C Nucl. Phys.* **51**, 3363–3371 (1995).
59. Y. Aleksandrov, M. Vrana, G. Manrique, T. Machekhina, L. Sedlakova, *Sov. J. Nucl. Phys.* **44**, 900 (1986).
60. U. Mohideen, A. Roy, *Phys. Rev. Lett.* **81**, 4549–4552 (1998).
61. B. Heacock, M. Huber, Open data for "Pendellösung interferometry probes the neutron charge radius,lattice dynamics, and fifth forces". Zenodo (2021). doi:10.5281/zenodo.5080965.

ACKNOWLEDGMENTS

We thank the NIST Center for Neutron Research for providing the neutron facilities used in this work. **Funding:** This work was supported by The National Institute of Standards and Technology, the US Department of Energy (DOE) under grants 89243019SSC000025 and DE-FG02-97ER41042, the National Science Foundation (NSF) under grant PHY-1307426, the Natural Sciences and Engineering Council of Canada (NSERC) Discovery program, and the Canada First Research Excellence Fund (CFREF). **Author contributions:** B.H., M.G.H., H.S., and A.R.Y. conceived the project. B.H., T.F., R.W.H., M.G.H., D.A.P., and R.V. collected neutron data under the supervision of M.G.H.; A.H. performed the sample crystallographic alignment and initial fabrication. T.F., T.H., M.T., and Y.Y. machined and characterized the sample surfaces under the supervision of Y.Y.; B.H., M.G.H., and D.A.P. chemically etched the sample. M.G.H., K.H., M.K., H.S., Y.Y., and A.R.Y. were responsible for overall supervision. B.H. performed the main analysis and wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The datasets and analysis code are available through Zenodo (*61*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abc2794
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 and S2
References (*62*–*67*)

8 June 2020; resubmitted 3 March 2021
Accepted 6 August 2021
10.1126/science.abc2794

PHASE-CHANGE MEMORY

Ultralow–switching current density multilevel phase-change memory on a flexible substrate

Asir Intisar Khan[1]†, Alwin Daus[1]†, Raisul Islam[1], Kathryn M. Neilson[1], Hye Ryoung Lee[1,2], H.-S. Philip Wong[1], Eric Pop[1,2,3]*

Phase-change memory (PCM) is a promising candidate for data storage in flexible electronics, but its high switching current and power are often drawbacks. In this study, we demonstrate a switching current density of ~0.1 mega-ampere per square centimeter in flexible superlattice PCM, a value that is one to two orders of magnitude lower than in conventional PCM on flexible or silicon substrates. This reduced switching current density is enabled by heat confinement in the superlattice material, assisted by current confinement in a pore-type device and the thermally insulating flexible substrate. Our devices also show multilevel operation with low resistance drift. The low switching current and good resistance on/off ratio are retained before, during, and after repeated bending and cycling. These results pave the way to low-power memory for flexible electronics and also provide key insights for PCM optimization on conventional silicon substrates.

Numerous emerging flexible electronics for the internet of things (IoT) (*1*, *2*) benefit from low-power embedded memory (*3*–*5*); examples include electronic skin (*6*), paper-like deformable displays (*7*), and smart IoT sensors for food and drug monitoring (*2*). Many of these applications could revolutionize personalized health care and global supply chains, particularly if they can rely on flexible substrates, which are easier to integrate with various form factors and surfaces. However, integrating conventional silicon-based memory into flexible electronics is difficult owing to the limited bendability of thinned silicon, complicated transfer processes, and potential thermal challenges on substrates with low thermal conductivity (*4*, *8*, *9*). As an alternative, memory devices could be fabricated directly on the flexible substrates, as long as this can be done at sufficiently low temperature (typically below 250°C). For example, flash and ferroelectric memory have been directly demonstrated on flexible substrates, but at the expense of higher operating voltage and power consumption than needed for their rigid counterparts (*4*, *10*).

Two other memory candidates that can be processed directly on flexible substrates are resistive random-access memory (RRAM) and phase-change memory (PCM) (*11*–*13*). Although many demonstrations exist with RRAM, its filamentary operation causes variability issues (*4*, *12*). By contrast, PCM offers lower variability than RRAM while providing faster speed, larger memory window, and longer write endurance than flash (*14*). These attributes and its potential for multilevel memory make PCM a candidate of interest for data storage in flexible electronics (*15*, *16*). The same properties are also important for IoT devices, which must often process data locally because of limited communication bandwidth with cloud servers (*17*).

However, a fundamental challenge of PCM has been its relatively high switching current and power, posing an obstacle for widespread adoption on both rigid (silicon) and flexible substrates (*12*, *13*, *18*). PCM on flexible substrates has been relatively unexplored, and previous studies (*13*, *18*) operated with high current owing to the use of conventional phase-change materials (e.g., $Ge_2Sb_2Te_5$) and lithography limitations (e.g., limited ability to focus) on uneven flexible substrates. Although

[1]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. [2]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305 USA. [3]Precourt Institute for Energy, Stanford University, Stanford, CA 94305, USA.
***Corresponding author. Email: epop@stanford.edu**
†These authors contributed equally to this work.

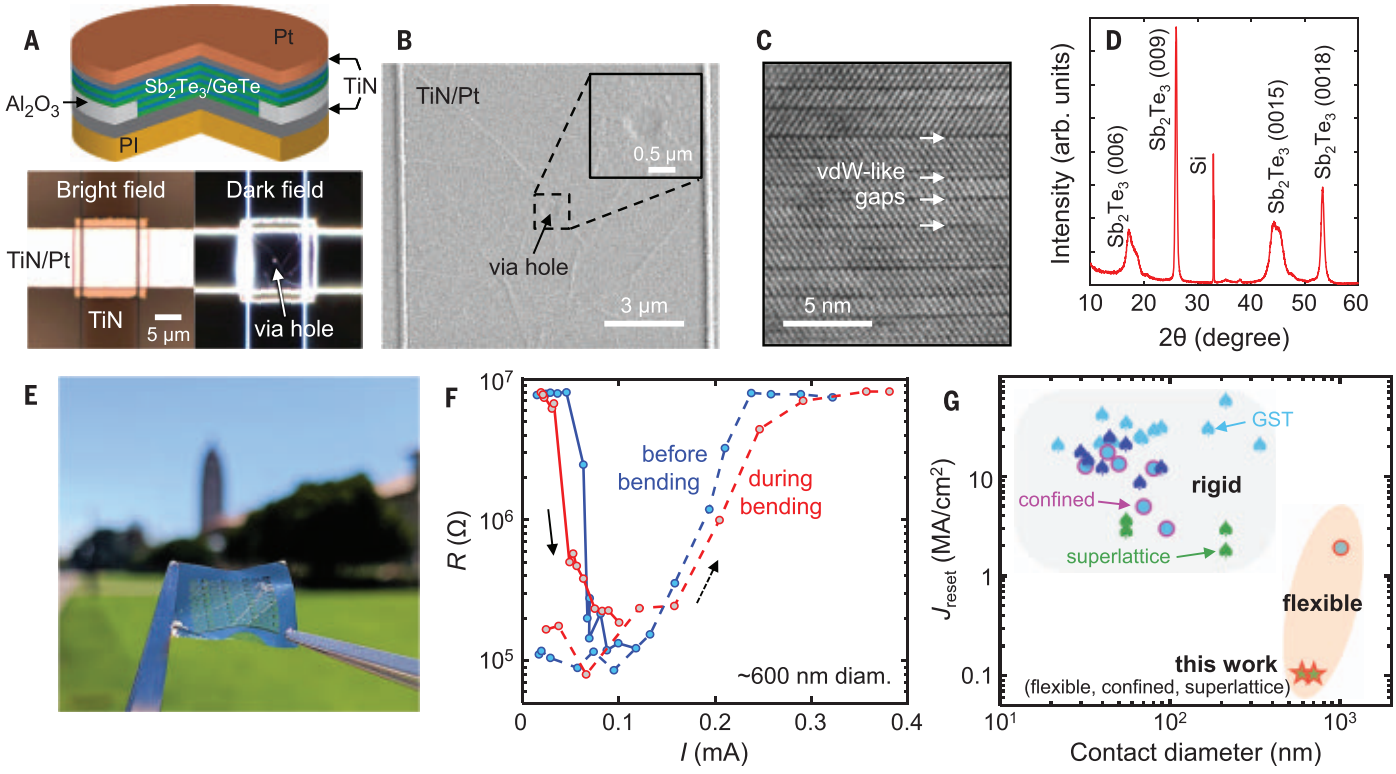


Fig. 1. Flexible superlattice PCM. (**A**) (Top) Schematic device cross section. (Bottom) Top-view optical images of fabricated devices. (**B**) Top-view scanning electron microscopy image of a device with a ~600-nm via hole (i.e., a hole in the Al_2O_3 insulating layer, defining the active device region). The inset shows a zoomed-in view. (**C**) High-resolution STEM cross section of the superlattice stack deposited on a Si substrate, using the same conditions as for the flexible PCM devices. STEM shows 5-atom Sb_2Te_3 layers, as well as 7- and 11-atom layers with Ge atoms diffused into $Ge_xSb_yTe_z$ (*25*), all separated by clear vdW-like gaps (see also fig. S7). (**D**) XRD of the same superlattice stack as in (C), demonstrating the polycrystallinity of the as-deposited films (see also fig. S2). (**E**) Photograph of one of our flexible substrates with memory devices. (**F**) Read resistance versus current for our flexible PCM (with a ~600-nm via hole) before and during bending to a radius of 4 mm, demonstrating low-current switching in both the flat and bent conditions. Solid arrow, from high-resistance state (HRS) to low-resistance state (LRS); dashed arrow, from LRS to HRS. Ω, ohms. (**G**) Benchmarking of reset current density (J_{reset}) reveals that J_{reset} is about two orders of magnitude lower than that of conventional PCM on rigid silicon substrates (*20*, *21*, *29*, *30*) and one order of magnitude lower than that of other flexible PCM (*13*). Some trends can be discerned—e.g., J_{reset} is reduced from mushroom-type cells (spades) to confined cells (circles) and is also reduced from conventional $Ge_2Sb_2Te_5$ (GST; light blue) or doped GST (dark blue) to superlattice PCM (green). Flexible PCM also has lower J_{reset} than its rigid counterparts, and our combination of flexible substrate with superlattice PCM in a confined cell yields the lowest $J_{reset} \approx 0.1$ MA/cm^2 (red stars).

the reset current of PCM is known to decrease with the cell size (*14*), such scaling is more difficult to achieve on flexible substrates; thus, reducing the reset current density becomes of greater importance. Reducing current density also benefits PCM on rigid substrates (*19*), where the area used by memory selector devices (which supply the current in high-density memory arrays) is valuable and must be minimized (*14*, *20*).

Here, we achieved ultralow reset current density and multilevel operation, which are difficult to realize for PCM on flexible substrates, with a combined approach that includes (i) use of a superlattice phase-change material, which limits heat loss into the metal electrodes; (ii) taking advantage of the very low thermal conductivity of the flexible substrate, which blocks heat loss from the bottom electrode (BE); and (iii) Joule heat confinement in a pore-like geometry. Our PCM devices were directly fabricated on a flexible polyimide (PI) substrate (Fig. 1A), without any layer transfers and a maximum process temperature of 200°C. The superlattice phase-change material (*21*) was deposited as 12 periods of alternating Sb_2Te_3 (4 nm) and GeTe (1 nm) layers. The ~600-nm-diameter pore-like device is surrounded by ~35-nm Al_2O_3 (Fig. 1B) on a TiN bottom contact [see (*22*) for fabrication details, fig. S1, and table S1].

We used high-resolution scanning transmission electron microscopy (STEM) to image the superlattice cross section on a silicon substrate, with the same deposition conditions as for the flexible PCM (Fig. 1C). [STEM sample preparation is difficult on the weak, electrically and thermally insulating polymer, which can warp during focused ion beam milling (*23*, *24*).] STEM imaging displays Sb_2Te_3 and $Ge_xSb_yTe_z$ layers, revealing some interfacial reconfiguration known to occur during deposition kinetics (*25*) but preserving clear and parallel van der Waals–like (vdW-like) gaps (*26*). X-ray diffraction (XRD) data in Fig. 1D show sharp polycrystalline peaks from the same superlattice film used for the STEM cross section. (Additional XRD data are shown in fig. S2 for a similar superlattice on TiN/PI/Si.) The pronounced (00*l*) out-of-plane XRD peaks correspond to highly oriented layers parallel to the substrate. A small broadening and shifting of some Sb_2Te_3 peaks and the presence of $Ge_xSb_yTe_z$ peaks (fig. S2) are consistent with partial diffusion of Ge atoms into Sb_2Te_3 (*27*, *28*).

The resulting PCM devices on the ~5-μm PI substrate are flexible (Fig. 1E), and we measured their resistance (*R*) versus current (*I*) before and during bending (Fig. 1F). For set and reset programming, respectively, we used 1-, 20-, and 500-ns and 1-, 60-, and 1-ns rise, width, and fall pulses (see fig. S3 for the setup

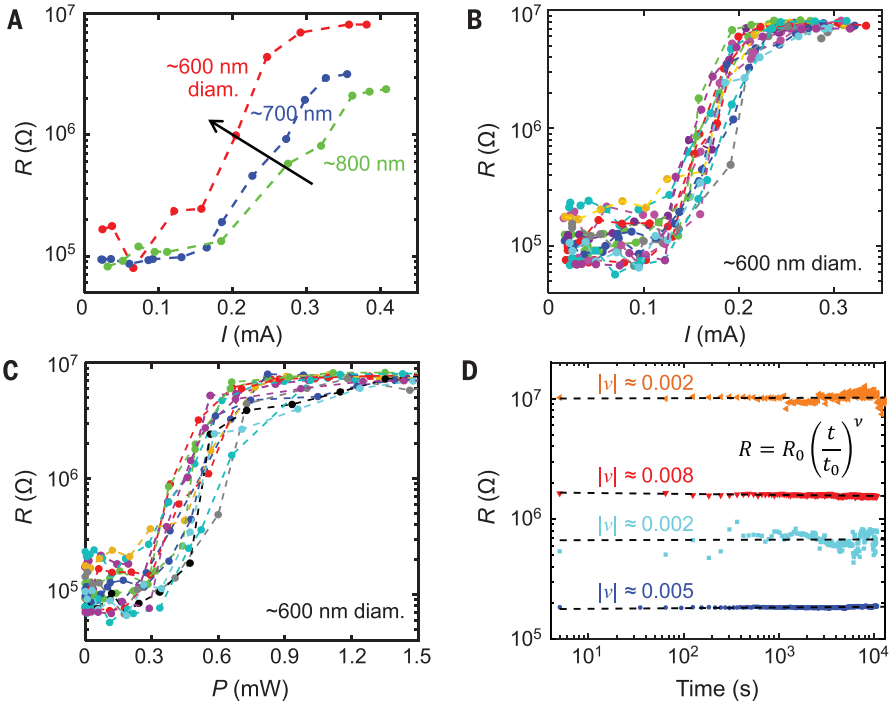


Fig. 2. Scaling, power, and multilevel measurements. (**A**) Read resistance versus current for our devices, showing the scaling of I_{reset} with cell area. (**B**) Resistance versus current for 15 cycles, showing low cycle-to-cycle variability of these devices (see also fig. S6B). (**C**) Resistance versus power (P) for 15 cycles, showing a switching power (P_{reset}) of ~0.8 mW. (**D**) Four stable-resistance states in the flexible PCM device with 600-nm diameter, showing multilevel capability with ultralow resistance drift over 10^4 seconds [see supplementary text section I (*22*)]. The resistance of the highest state (~10 megohms) is slightly higher than in the other panels because the multilevel states were reached with single-shot pulses from the lowest-resistance state, whereas the data in (A) to (C) were obtained with gradually increasing pulse magnitudes. R_0 is the resistance at an arbitrary time t_0, and t is the time after the last switching event.

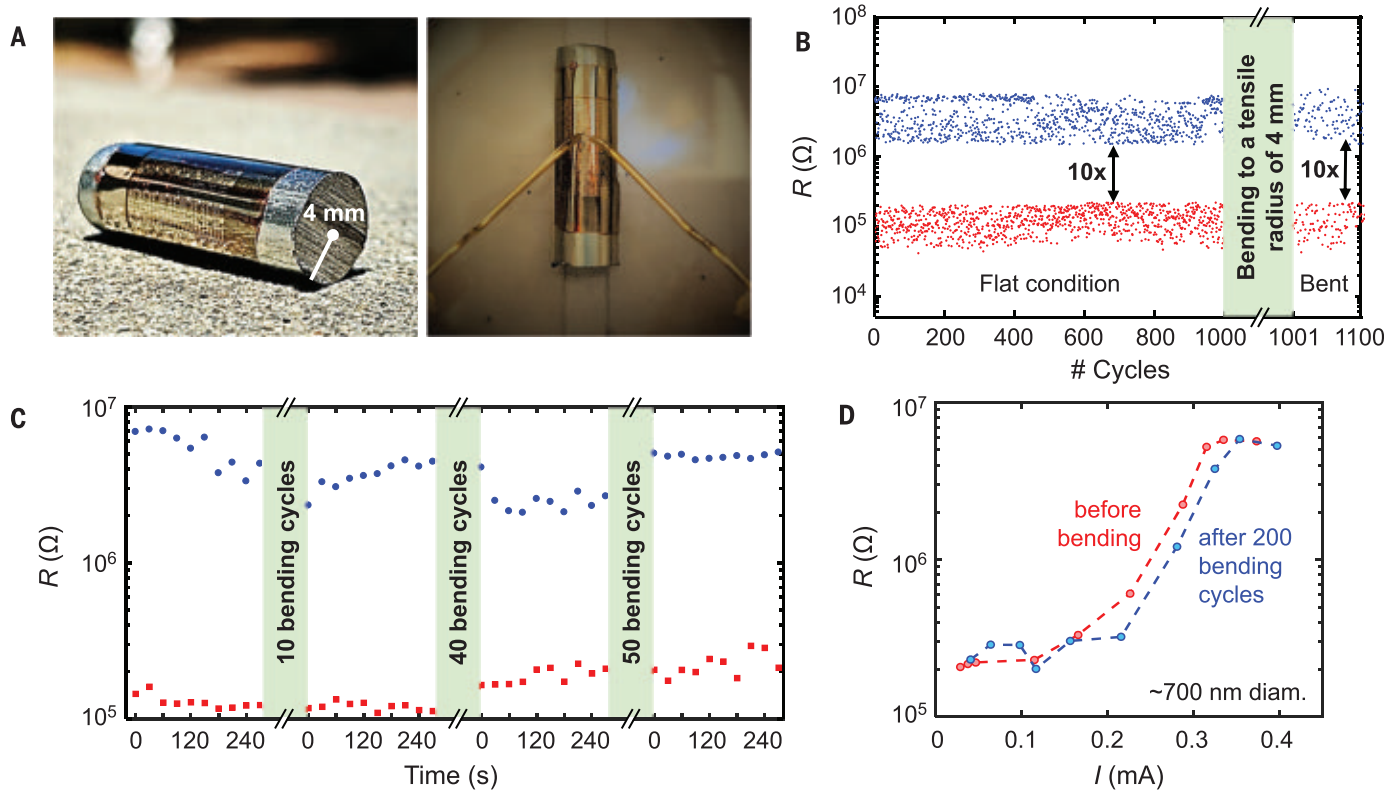


Fig. 3. Characterization before and after bending. (**A**) Measurement setup for our flexible PCM devices under a bending radius of 4 mm. (**B**) Resistance versus number of switching cycles before (i.e., flat condition) and during bending, showing that a 10-fold resistance window (10×) is maintained. (**C**) A resistance on/off ratio >10 is maintained over time upon cyclic bending. (**D**) Resistance versus current before bending and after 200 bending cycles, showing that low I_{reset} and the resistance on/off ratio are maintained after bending cycles.

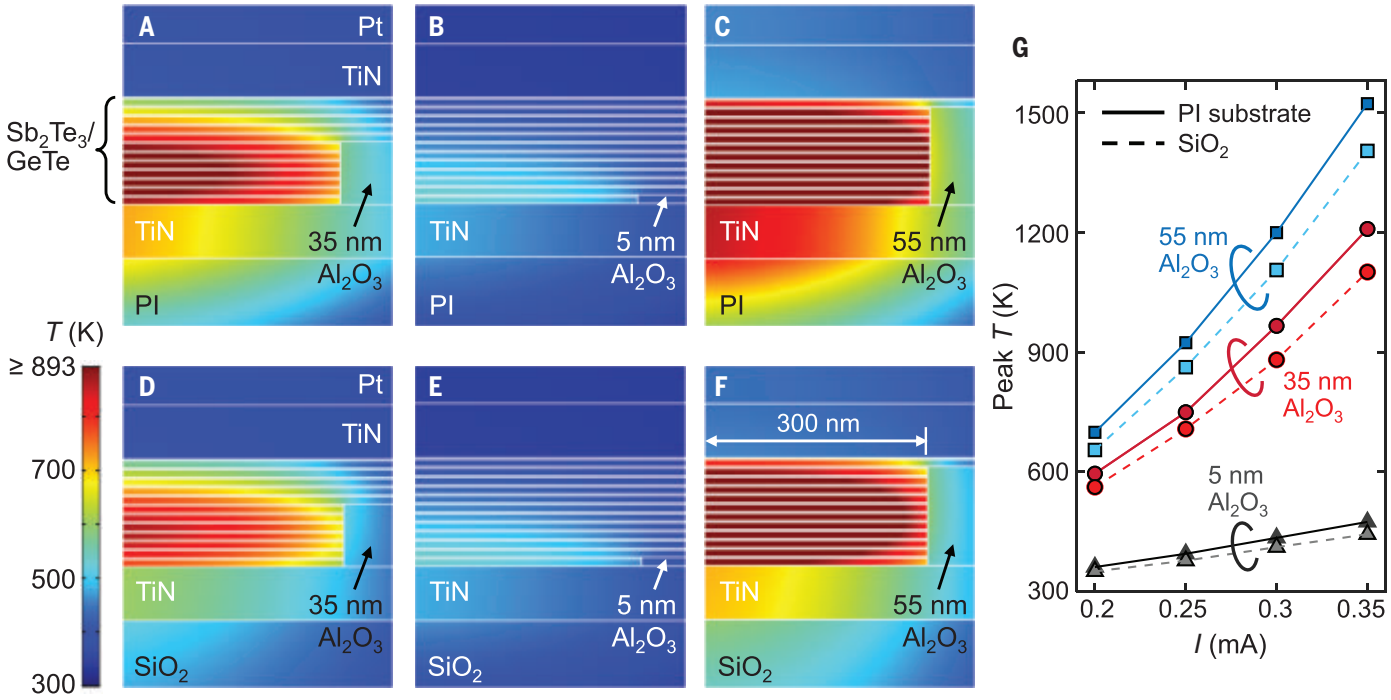


Fig. 4. Electrothermal simulations. Temperature distributions after a reset current pulse (0.3 mA, 60 ns) in superlattice PCM devices on flexible PI with (**A**) 35-nm Al_2O_3 insulation (see fig. S9 for an extended temperature profile into the PI substrate), (**B**) 5-nm Al_2O_3 insulation, and (**C**) 55-nm Al_2O_3 insulation, demonstrating the effect of thermal confinement from the superlattice layers and lateral Joule heat confinement from the Al_2O_3 insulation. Temperature distributions of similar devices on a rigid SiO_2 substrate with (**D**) 35-nm Al_2O_3 insulation, (**E**) 5-nm Al_2O_3 insulation, and (**F**) 55-nm Al_2O_3 insulation, showing less thermal confinement on a rigid substrate with higher thermal conductivity. The left edge is the axis of symmetry in (A) to (F), device diameters are 600 nm, and the vertical and horizontal scales are unequal. (**G**) Peak temperature versus current at the end of a 60-ns pulse for the devices whose temperature distributions are shown in (A) to (F).

and fig. S4 for an example output pulse). We read the resistance of all devices with a 50-mV direct current (dc) bias (*22*). Our ~600-nm diameter flexible PCM devices switch at ~0.2 to 0.25 mA of reset current (I_{reset}), with a resistance ratio up to ~100, maintaining the switching characteristics during both flat and bent substrate conditions. The estimated reset current density (J_{reset}) of ~0.1 MA/cm^2 is about two orders of magnitude lower than that of conventional PCM on rigid silicon substrates (*29*, *30*) and more than an order of magnitude lower than that of existing flexible PCM (*13*) (Fig. 1G).

We found that the reset current scales with the pore cell area (from ~600- to ~800-nm diameters; Fig. 2A), indicating the scalability of our technology, further supported by additional benchmarking (fig. S5). Our devices also demonstrate reasonably low cycle-to-cycle variation (Fig. 2B). We calculated resistance versus power (Fig. 2C), which we estimated from R versus I (Fig. 2B) and R versus voltage (V) (fig. S6A). We further noted low cycle-to-cycle variation for >450 cycles (fig. S6B). The peak reset power is ~0.8 mW; this could be further reduced with smaller cell diameters but is limited here by the optical lithography. The reset energy of ~48 pJ (*22*) is nearly an order of magnitude lower than that of other devices, which used a filamentary bottom contact (*18*), owing to our shorter reset pulses of ~60 ns. Our reset power is two orders of magnitude lower than that of earlier nonfilamentary flexible PCM (*13*).

We also explored the multilevel capability and resistance drift of our flexible superlattice PCM, demonstrating four stable-resistance states (Fig. 2D). The same figure also reveals a low resistance drift coefficient (ν ~ 0.002 to 0.008), which is more than an order of magnitude lower than that of conventional PCM (*31*) with clearly distinguishable resistance states [see supplementary text section I and table S2 (*22*)]. We applied several separate reset pulses of the same shape (1, 60, and 1 ns) and increasing amplitude (2.1, 2.6, 3.1, and 3.5 V) to achieve the intermediate-resistance states. Figure 2D also reflects retention of the low- and high-resistance states beyond 10^4 s. The low resistance drift is attributed to the presence of vdW-like gaps in the superlattice material even after electrical cycling (fig. S7), which can limit structural relaxation and long-range atomic diffusion. This behavior is a notable advantage of superlattice-type PCM, regardless of the substrate choice (*32*).

We measured our flexible PCM under a tensile bending radius of 4 mm (Fig. 3A). The devices maintained a resistance on/off ratio of >10 for 10^3 cycles while flat, then for 10^2 cycles while bent (Fig. 3B; fig. S8 shows additional endurance measurements taken over 10^4 cycles). R-versus-I measurements in flat and bent conditions (Fig. 1F) indicate that low I_{reset} is preserved upon bending, with a resistance on/off ratio of nearly 100. Further, we measured the bending cyclability of our PCM devices (Fig. 3C). We read their dc resistance after several bending cycles and measured the resistance states for a consecutive 300 s, with 30-s intervals between each resistance readout. These results show that both high- and low-resistance states are maintained with low drift before bending, as well as after 100 bending cycles. Notably, low switching current is preserved even after 200 bending cycles (Fig. 3D). The robustness of our devices during bending can be attributed to the PI thickness of 5 μm, which leads to a small strain of ~0.06% even at a 4-mm bending radius (*33*). Encapsulation with another few-micrometer-thick polymer layer would place our devices along the neutral mechanical plane, further reducing their strain (*34*). A comparison with other flexible memory technologies—in terms of retention, endurance,

and choice of bending radius—is provided in table S3.

Finally, we sought to understand what leads to the ultralow current density in our flexible superlattice PCM devices. Earlier work had suggested that Ge atom movement may be responsible for switching in similar superlattice PCM on rigid substrates (*21*). However, we have recently found that the superlattice has low cross-plane thermal conductivity and very high electrical anisotropy (*35*) owing to its numerous parallel vdW-like interfaces. Thus, we exploited these properties to conceive our design, which includes confinement of the superlattice PCM cell in a pore-like geometry on a substrate with ultralow thermal conductivity (the flexible PI). This leads to the pronounced reduction of switching current density, which points to a strong thermal component of the switching mechanism.

To gain deeper insight into the role of heat confinement in the pore device, we performed electrothermal simulations [Fig. 4 and supplementary text section II (*22*)]. These reveal simultaneous effects of thermal confinement due to the numerous interfaces, higher electrical resistivity across the superlattice, lateral Joule heat confinement in the pore geometry, and the thermally resistive PI substrate (Fig. 4 and fig. S9). By contrast, the temperature profile of a pore-type $Ge_2Sb_2Te_5$ device on PI (fig. S10) shows less thermal confinement and a much lower peak temperature (~368 K) than in the superlattice PCM device (~966 K). This simulation further confirms the efficacy of electrothermal confinement provided by the superlattice material with numerous vdW-like interfaces (*36*).

We also simulated devices with different Al_2O_3 side-wall thicknesses and found that heat is more concentrated in the active PCM cell region with increasing Al_2O_3 thickness. Thus, the peak temperature required for phase change can be achieved with a lower current pulse (Fig. 4, A to G). We attribute this behavior to improved electrical heating confinement (further visualized in fig. S11, where the peak vertical current density appears across a substantially larger portion of the PCM superlattice for 35-nm versus 5-nm Al_2O_3). For the same input current, 5 nm of Al_2O_3 thickness provides negligible confinement (Fig. 4B), resulting in a substantially lower (by a factor of ~2.2) peak temperature than in our nominal devices (Fig. 4A), whereas 55-nm Al_2O_3 provides better confinement (Fig. 4C).

The flexible PI substrate is very effective for achieving thermal isolation and lower reset current, in contrast to conventional rigid substrates that often include a combination of oxides (such as SiO_2) and silicon. The ultralow thermal conductivity of the flexible PI substrate ($k_{PI} \approx 0.12$ W m^{-1} K^{-1}) (*37*) is an order of magnitude lower than that of amorphous oxides or nitrides (*38*) and two orders of magnitude lower than that of silicon. As a result, it prevents heat loss from the TiN BE. For our nominal device structure with 35-nm Al_2O_3, the peak temperature in the BE is ~20% higher with the PI substrate than with a SiO_2 substrate (Fig. 4, A and D). This indicates that in our laterally confined pore-type PCM cell, the thermal conductivity of the substrate (e.g., PI versus SiO_2/Si) plays an important role, especially for heat loss from the BE, lowering the reset current density in flexible PCM compared with its rigid counterparts (Fig. 1G). Our conclusions are supported by the peak temperature–versus–current plot (Fig. 4G) that shows the importance of confinement in our devices for obtaining low reset current density. These findings should also enable future thermal engineering of PCM devices on silicon by introducing heat-blocking layers or even air gaps below the BE.

In summary, we developed a flexible, superlattice memory with much lower switching current density (~0.1 MA/cm^2) than in all other PCM types, as well as very low reset energy and power relative to other flexible PCMs. These characteristics were enabled by heat confinement in a superlattice material within a pore-type device on an ultralow–thermal conductivity substrate. The flexible PCM devices display multilevel capability with low resistance drift, indicating promise for emerging in-memory computing applications. These results usher in data storage in flexible IoT electronics and also provide key insights for thermal engineering of conventional PCM on commercial, rigid silicon substrates.

REFERENCES AND NOTES

1. C. Wang *et al*., *Nat. Mater.* **12**, 899–904 (2013).
2. G. A. Salvatore *et al*., *Adv. Funct. Mater.* **27**, 1702390 (2017).
3. T. Sekitani *et al*., *Science* **326**, 1516–1519 (2009).
4. M. T. Ghoneim, M. M. Hussain, *Electronics* **4**, 424–479 (2015).
5. A. A. Bessonov *et al*., *Nat. Mater.* **14**, 199–204 (2015).
6. K. Xu, Y. Lu, K. Takei, *Adv. Mater. Technol.* **4**, 1800628 (2019).
7. S.-I. Park *et al*., *Science* **325**, 977–981 (2009).
8. A. M. Hussain, M. M. Hussain, *Adv. Mater.* **28**, 4219–4249 (2016).
9. S. Gupta, W. T. Navaraj, L. Lorenzelli, R. Dahiya, *npj Flex. Electron.* **2**, 8 (2018).
10. S.-J. Kim, J.-M. Song, J.-S. Lee, *J. Mater. Chem.* **21**, 14516–14522 (2011).
11. A. Daus, S. Han, S. Knobelspies, G. Cantarella, G. Tröster, *Materials* **11**, 1672 (2018).
12. S.-T. Han, Y. Zhou, V. A. L. Roy, *Adv. Mater.* **25**, 5425–5449 (2013).
13. B. H. Mun *et al*., *ACS Nano* **9**, 4120–4128 (2015).
14. S. Raoux, F. Xiong, M. Wuttig, E. Pop, *MRS Bull.* **39**, 703–710 (2014).
15. D. Ielmini, H.-S. P. Wong, *Nat. Electron.* **1**, 333–343 (2018).
16. C. Li *et al*., *Nat. Electron.* **1**, 52–59 (2018).
17. F. Samie *et al*., in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)* (IEEE, 2016), pp. 7–12.
18. D. H. Kim *et al*., *Adv. Funct. Mater.* **29**, 1806338 (2019).
19. A. I. Khan *et al*., *IEEE Electron Device Lett.* **41**, 1657–1660 (2020).
20. H.-S. P. Wong *et al*., *Proc. IEEE* **98**, 2201–2227 (2010).
21. R. E. Simpson *et al*., *Nat. Nanotechnol.* **6**, 501–505 (2011).
22. Materials and methods are available as supplementary materials.
23. A. Daus *et al*., *Adv. Electron. Mater.* **3**, 1700309 (2017).
24. F. Rivera, R. Davis, R. Vanfleet, *Microsc. Microanal.* **19**, 1080–1091 (2013).
25. J. Momand *et al*., *Nanoscale* **9**, 8774–8780 (2017).
26. Y. Cheng *et al*., *Adv. Mater.* **31**, 1904316 (2019).
27. T. Ohyanagi *et al*., *Jpn. J. Appl. Phys.* **52**, 05FF01 (2013).
28. J. Feng *et al*., *ACS Appl. Mater. Interfaces* **12**, 33397–33407 (2020).
29. J. Shen *et al*., *ACS Appl. Mater. Interfaces* **11**, 5336–5343 (2019).
30. H.-S. P. Wong *et al*., "Stanford Memory Trends," https://nano.stanford.edu/stanford-memory-trends [accessed 2 April 2021].
31. M. Suri *et al*., in *2013 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)* (IEEE/ACM, 2013), pp. 140–145.
32. K. Ding *et al*., *Science* **366**, 210–215 (2019).
33. A. Daus *et al*., *Nat. Electron.* **4**, 495–501 (2021).
34. Y. J. Park *et al*., *ACS Nano* **13**, 3023–3030 (2019).
35. H. Kwon *et al*., *Nano Lett.* **21**, 5984–5990 (2021).
36. M. Boniardi *et al*., *Phys. Status Solidi RRL* **13**, 1800634 (2019).
37. Parameters to make Polyimide Film (PI-2600) with Thickness more than 100 μm, www.researchgate.net/post/Parameters_to_make_Polyimide_Film_PI-2600_with_Thickness_more_than_100_um [accessed 2 April 2021].
38. S.-M. Lee, D. G. Cahill, *J. Appl. Phys.* **81**, 2590–2595 (1997).

ACKNOWLEDGMENTS

The authors are grateful to K. Yamamoto and S. Fujii from Kioxia Corporation, Japan, for obtaining the high-resolution STEM images and K. Ishimaru of Kioxia for management support. A.I.K. thanks J. McVittie for lab support, C. Neumann for the electrical measurement setup, and R. Paul for assistance with Fig. 1A. **Funding:** This work was performed at the Stanford Nanofabrication Facility (SNF) and Stanford Nano Shared Facilities (SNSF) and is supported in part by member companies of the Stanford Non-volatile Memory Technology Research Initiative (NMTRI). A.I.K. and K.M.N. acknowledge support from the Stanford Graduate Fellowship and A.D. from the Swiss NSF Early Postdoc Mobility fellowship (grant P2EZP2_181619) and the Beijing Institute of Collaborative Innovation (BICI). **Author contributions:** A.I.K., A.D., and E.P. conceived the idea and designed the experiments. A.I.K. and A.D. fabricated the devices, carried out the measurements, and analyzed the data. R.I. and A.I.K. performed the electrothermal simulations. K.M.N. performed XRD and H.R.L. performed scanning electron microscopy. A.I.K., A.D., and E.P. wrote the manuscript, with input from H.-S.P.W., and all authors discussed the results and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in this paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abj1261
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S3
References (*39*–*67*)

22 April 2021; accepted 6 August 2021
10.1126/science.abj1261

ATMOSPHERIC DYNAMICS

Hydraulic jump dynamics above supercell thunderstorms

Morgan E O'Neill[1]*, Leigh Orf[2,3], Gerald M. Heymsfield[4], Kelton Halbert[2]

The strongest supercell thunderstorms typically feature an above-anvil cirrus plume (AACP), which is a plume of ice and water vapor in the lower stratosphere that occurs downwind of the ambient stratospheric flow in the lee of overshooting deep convection. AACP-origin hydration of the stratosphere has a poorly constrained role in ozone destruction and surface warming. In this study, we use large eddy simulations corroborated by radar observations to understand the physics of AACP generation. We show that the overshooting top of a simulated supercell can act as a topographic obstacle and drive a hydraulic jump downstream at the tropopause, similar to a windstorm moving down the slope of a mountain but without solid topography. Once the jump is established, water vapor injection deep into the stratosphere may exceed 7 tonnes per second.

Above-anvil cirrus plumes (AACPs) are a regular feature of strong mid-latitude thunderstorms, particularly over the Great Plains (*1*–*4*). AACPs exhibit a high correspondence with the most severe conditions on the ground, including major tornado and hail events (*4*–*6*). An AACP consists of a thin cirrus plume, often visible by satellite, which shoots several kilometers above the much larger anvil cloud directly into the lower stratosphere where it spreads downwind. AACPs can also aid in the predictability of severe hail and tornadoes, as AACPs are visible above the cloud tops an average of 31 min before severe conditions are reported (*6*).

AACPs may also play a major role in water vapor injection into the lower stratosphere (*2*, *5*, *7*), which is hydrated by deep convection (*8*, *9*). Increased stratospheric water vapor increases stratospheric radiative cooling directly and through ozone destruction (*10*–*12*) but acts to warm the troposphere and surface (*13*, *14*). Model-based estimates of the water injection rate into the lower stratosphere above a deep thunderstorm range from 1 to a few tonnes of water per second on average (*2*, *15*–*18*), suggesting deep convective origin of total stratospheric water vapor, as high as 18% from tropical storms alone (*18*). Thunderstorm frequency may increase owing to a rise in atmospheric specific humidity in a warming climate (*19*), which suggests the potential for positive feedback between more mid-latitude thunderstorms and increased lower stratospheric water vapor in response to climate change (*20*).

AACP dynamics and associated features, particularly the physical mechanism responsible for irreversibly injecting tropospheric water into the lower stratosphere, are an issue of debate, though there is broad agreement regarding the role of breaking gravity waves excited by deep convection (*1*, *2*, *4*, *5*, *21*). It has long been recognized that deep convection and overshooting tops (OTs) can effectively act as topography that can block or redirect oncoming upper tropospheric or stratospheric air (*1*, *21*). This can lead to the breaking of waves downstream, which lofts water into the lower stratosphere (*2*). However, there are a half dozen concurrently observed features for which there is still no consistent physical model, and existing hypotheses appear to contradict each other. We resolve these conflicts by showing that the mechanism responsible is likely a hydraulic jump, forced by and interacting with the dynamic OT. Hydraulic jumps occur when fluid flows too fast for gravity waves to travel upstream, wherein the fluid becomes unstable and suddenly readjusts to a turbulent, slower flow.

Homeyer *et al.* (*5*) were the first to identify the mean environmental conditions that determine whether an AACP forms. They identified the importance of upper-level wind magnitude and, through idealized numerical simulations, showed that stronger constant oncoming winds in the upper troposphere and lower stratosphere induce a strong AACP, whereas winds of half that magnitude do not.

In this study, we conducted two large eddy simulation (LES) experiments using modifications to the Homeyer *et al.* (*5*) soundings (fig. S1): one with strong upper winds, simU_{str}, and one with weak upper winds, simU_{wk} (*22*). These idealized wind profiles are based on reanalysis results (*5*). The Bryan Cloud Model 1 (*23*) is run at 50-m cubic resolution in a storm-following domain of size 243.2 km by 240.8 km by 30 km. We find that most of the competing AACP hypotheses can be simultaneously correct, each explaining a different aspect of a downslope windstorm and hydraulic jump at the tropopause. Regarding the simU_{wk} storm, we provide an observation of a similar storm observed in May 2011 that likely exhibits an example of flow separation above the OT, which precludes a hydraulic jump. Finally, we quantify how much water is irreversibly injected into the lower stratosphere by each simulated storm.

Both simulations rapidly develop a mature supercell thunderstorm with a cold OT that reaches a height of ~15 km, generating gravity waves in its environment (Fig. 1 and figs. S2 and S3). The storm-relative upper winds are 33.7 m s^{-1} for simU_{str} and 14.5 m s^{-1} for simU_{wk}. Storm simU_{str} eventually develops a steady AACP, and simU_{wk} does not.

The upper flow behavior of simU_{str} and simU_{wk} differs noticeably by minute 35, approximately when the OT in each storm becomes established well above the 12-km tropopause (fig. S3). simU_{wk} exhibits flow separation of stratospheric air away from the storm anvil shortly after cresting the OT, similar to flow separation in the lee of real topography. This separation prevents the development of a strong standing gravity wave, and the free-slip–like conditions instead permit weak transient overturning circulations (rotors) (*24*). By contrast, simU_{str} exhibits a deepening standing gravity wave in the lee of the OT.

Before the OT altitude becomes relatively steady, the simU_{str} storm develops a turbulent filament of breaking gravity waves aloft, in a narrow region 2 to 3 km above and ~5 km downstream of the simU_{str} OT (*25*). It is narrowly rooted in the cloud top but spreads out forward and laterally such that it lies above nearly laminar flow (Fig. 1A and fig. S4). By minute 40, this turbulent region is identified by turbulent kinetic energy dissipation values >5 m^2 s^{-2}, indicating substantial mixing. Concurrently, the simU_{str} storm develops a broad, rapid jet exceeding speeds of 75 m s^{-1} at the tropopause. By minute 45, simU_{str} exhibits the onset of a hydraulic jump and a type II rotor (Fig. 1C) (*26*). Meanwhile, the flow separation in the lee of the simU_{wk} storm's OT allows the development and formation of a milder type I rotor between the tropospheric jet below and smooth stratospheric flow above (Fig. 1 and fig. S3).

Lagrangian parcel analysis of simU_{str} reveals that the fastest wind region at the tropopause (speed >110 m s^{-1}) is sourced primarily from the core of the overshooting updraft but also from the lower stratosphere (Fig. 2), a mixing hypothesized by Heymsfield *et al.* (*21*). Most upstream stratospheric air in simU_{str} below the peak altitude of the OT diverges around it. However, some stratospheric air above the OT's central axis plunges downward from 17- to 12.5-km altitude in the lee

[1]Department of Earth System Science, Stanford University, Stanford, CA 94305, USA. [2]Space Science and Engineering Center, University of Wisconsin, Madison, WI, USA. [3]Cooperative Institute for Meteorological Satellite Studies, Madison, WI, USA. [4]NASA Goddard Space Flight Center, Greenbelt, MD, USA.
*Corresponding author. Email: morgan.e.oneill@gmail.com

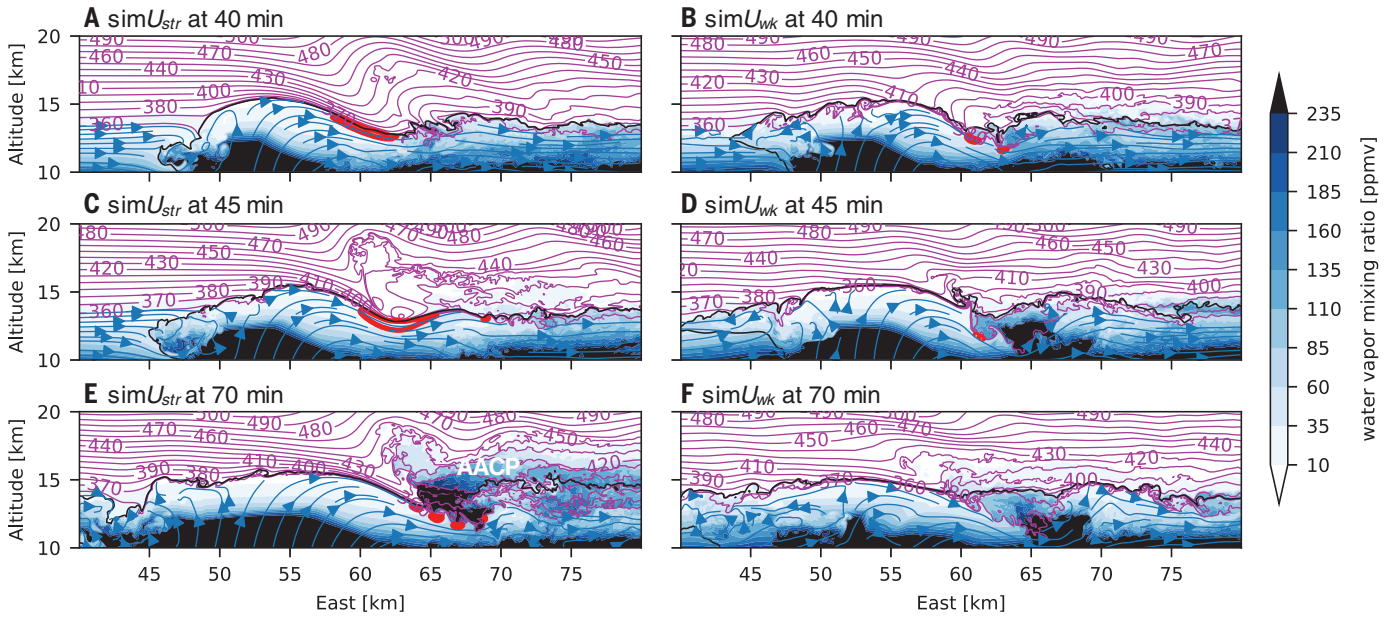


Fig. 1. Storm evolution. Cross-section of simU_{str} (**A**, **C**, and **E**) and simU_{wk} (**B**, **D**, and **F**) at *y* locations indicated in fig. S2. Magenta lines, potential temperature above 360 K; blue lines, streamlines below 360 K; red contour, region where total wind speed >75 m s^{-1}; blue shading, water vapor mixing ratio in parts per million by volume (ppmv); black line, cloud and frozen water mass mixing ratio of 5 × 10^{-2} g kg^{-1}. The general location of the AACP is labeled in (E).

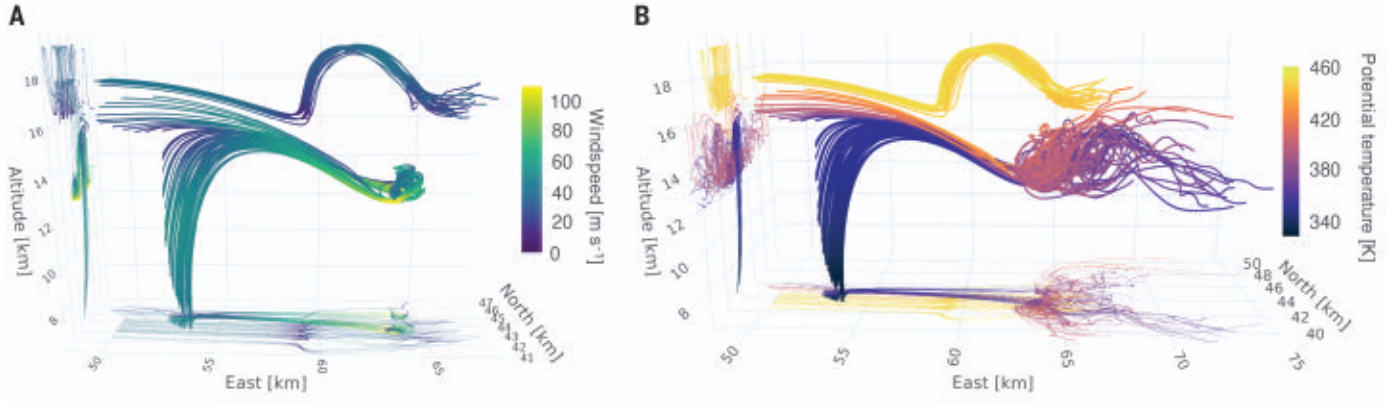


Fig. 2. Lagrangian parcels that meet one of the following conditions: Total wind speed exceeds 110 m s^{-1} within the jet before entering the hydraulic jump ("jump parcels"), or maximum height exceeds 19.25 km ("lid parcels"). (**A**) Parcels are shaded by wind speed, lid parcels are integrated from 40 to 50 min, and jump parcels are integrated from 40 to 46 min to visualize the breaking wave. (**B**) Same parcels are shaded by potential temperature; jump parcels are integrated from 40 to 51.7 min to visualize mixing. For text and interactive 3D versions, see supplementary materials.

of the OT. Negatively buoyant, descending tropospheric air from the OT converges in the same jet, and a tropospheric parcel reaches a storm-relative peak speed of 119 m s^{-1}.

The simU_{str} OT is exceedingly cold, with the coldest parcels nearing 190 K. Frozen water in the moist tropospheric updraft precipitates out (graupel) or descends in the subsiding jet region (ice), and tropospheric air is diabatically warmed as a result of mixing (Fig. 2B). The potential temperature of tropospheric parcels exceeding 90 m s^{-1} increases in the range of 0 to 100 K in total, with most of the increase occurring rapidly once the now-dehydrated air enters the turbulent jump (fig. S5). Jet-bound stratospheric air plummets down the leeside of the OT, where it is adiabatically warmed by more than 25 K within 15 km downstream of the OT (fig. S6), consistent with (*27*) though over half the distance. It is also substantially hydrated through mixing with tropospheric air.

Using airborne radar, we observed a feature that bears similarity to breaking waves in the OT lee of both simulations (though primarily simU_{wk}), lending credence to the realism of our simulations. The NASA ER-2 aircraft flew over an intense, nontornadic hailstorm in Oklahoma on 24 May 2011 (Fig. 3, A, C, and E, and fig. S7) during the Mid-latitude Continental Convective Clouds Experiment (MC3E) (*28*, *29*). The aircraft was instrumented with a dual-frequency nadir-pointing Doppler radar. One of the storms, which had an extremely strong updraft estimated at 50 m s^{-1} and an OT at 16-km altitude (tropopause height ~14 km), also had an apparent Kelvin-Helmholtz (K-H) wave on the lee side of the OT. Rising and sinking motions from the Doppler velocities (Fig. 3E) are consistent with the presence of a K-H wave.

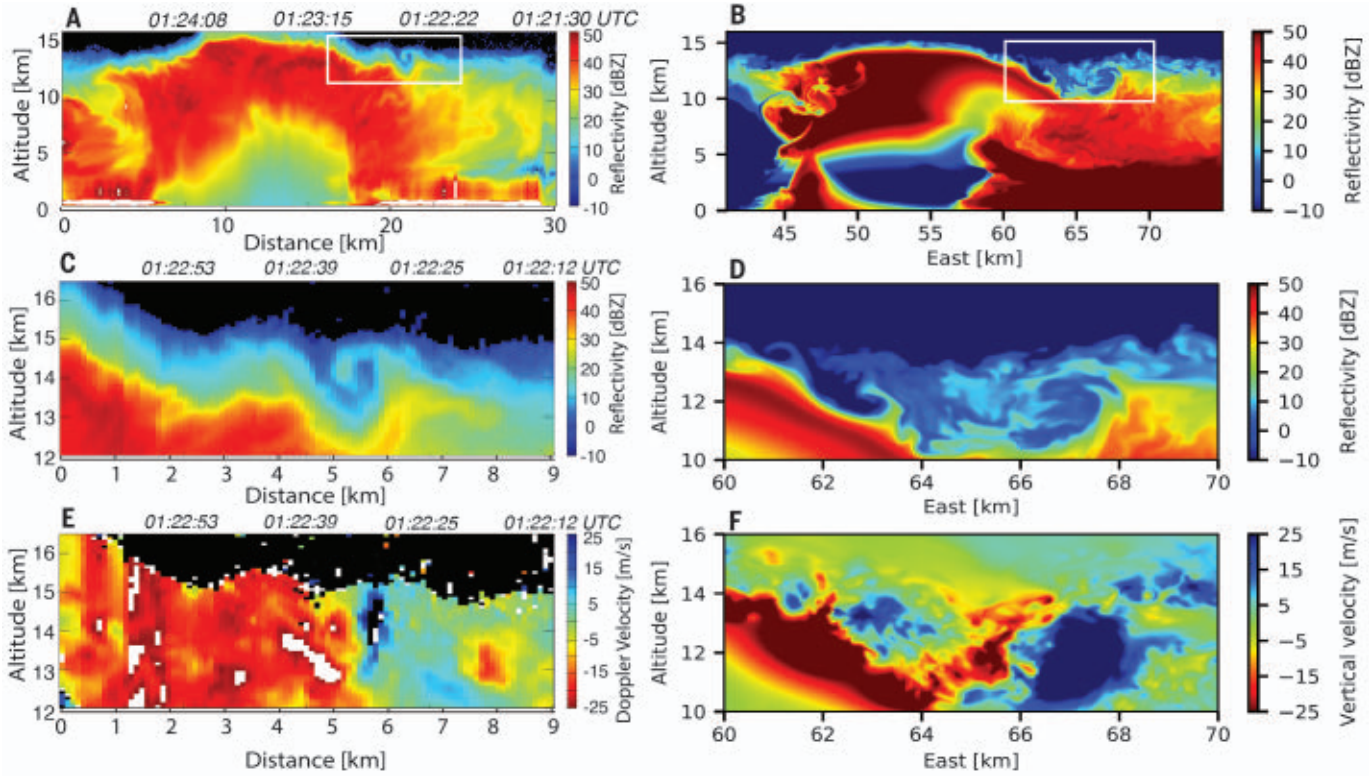


Fig. 3. Comparison of K-H wave at storm top observed by ER-2 HIWRAP Radar during storm transect on 24 May 2011 at 0125 UTC (Universal Time Coordinated) and simulation simU_{wk}. In both cases the stratospheric wind blows from left to right. (**A**) Doppler Ku-band radar reflectivity of the entire transect. (**B**) Simulated reflectivity of simU_{wk}, minute 55 in *xz* cross section at y = 34.925 km. (**C**) Doppler Ku-band radar reflectivity of K-H wave; (**D**) same as (C) but for simU_{wk} candidate K-H feature. (**E**) Doppler estimated vertical velocity (saturated at ±25 m s^{-1}) of K-H wave in (C). (**F**) Simulated vertical velocity of K-H wave in (D) (color bar saturated at ±25 m s^{-1}).

Similar transient K-H features are found in simU_{wk}, and we show a candidate in Fig. 3, B, D, and F. The storm-relative environmental wind speed at 14 km is ~14 m s^{-1} (fig. S8), similar to simU_{wk}'s wind speed of 14.5 m s^{-1}, though the observed storm also had wind shear in that region. The corresponding flow field in simU_{wk} demonstrates that in the lee of the OT, a layer of stratospheric air is stagnant above a region of faster tropospheric air (fig. S9). This permits K-H wave breaking of the descending tropospheric air from the OT to occur in the absence of a hydraulic jump, as well as formation of a weak type I rotor just above it.

The impact of the AACP on local stratospheric hydration is substantial. Simulated water injection into the stratospheric middleworld (between 360 and 380 K potential temperature surfaces) and overworld (above 380 K) is measured by a time series of the domain-integrated perturbation water vapor and frozen water for both storms (Fig. 4, A and B). In the unsaturated stratosphere, sublimation of ice is an irreversible process, so the water vapor time series integrates the contribution of both direct injection and sublimation of injected ice. Overall, simU_{wk} is a broader storm and a better hydrator of the stratosphere. By the end of the simulation, simU_{wk} has injected slightly more water vapor than simU_{str} into the middleworld and ~50% more ice. However, the bulk of this water vapor reaches only the thin lower layer of the stratospheric middleworld.

Though both storms induce gravity wave breaking at the tropopause early on, there is a substantial divergence of overworld water injection at $t \approx 45$ min in simU_{str}, coincident with the occurrence of the peak wind in the jet region and the jump onset (Fig. 4). Upon reaching the jump onset, the stratospheric overworld in simU_{str} experiences an increase in water vapor mass at a rate of ~7.6 tonnes s^{-1}, which is the highest yet quantified in a simulation. By contrast, simU_{wk} exhibits an overworld vapor injection rate of ~1.3 tonnes s^{-1}, which is itself unexpectedly high.

The linear trends of water injection indicate that both storms had much future potential to inject water into the stratosphere. Greatly heightened stratospheric water vapor mixing ratios have been observed downstream of deep mid-latitude convection up to a day later (*9*), suggesting that this short model integration is irreversibly injecting water into the stratosphere.

The severe storm community has long considered OTs to be "topographic" obstacles, inducing lee waves and potentially rotors. This work suggests the presence of hydraulic control, supercritical flow, and a hydraulic jump in the lee of a thunderstorm's OT, at the tropopause. Similar behavior has been observed and modeled for decades by the mountain meteorology community. Theoretical work motivated by severe downslope windstorms in the lee of mountains (*30*–*33*) may be applicable to the prediction of severe thunderstorm behavior.

Internal hydraulic theory has been developed for continuously stratified flow above real topography (*30*, *33*). There are three mechanisms by which a continuous, unbounded atmosphere above an obstacle may initiate a downslope windstorm that terminates in a hydraulic jump (*34*). All three require a vertical length scale that serves the place of an upper boundary from classic hydraulic theory, and this is commonly provided by either vertical shear or a change in the environmental buoyancy frequency with height. Because the ratio of the buoyancy frequency (N) to the mean wind speed (U) is approximately constant with height in the stratosphere of the present simulations, the only capping mechanism available is that sufficient gravity wave breaking occurs aloft of a topographic obstacle and

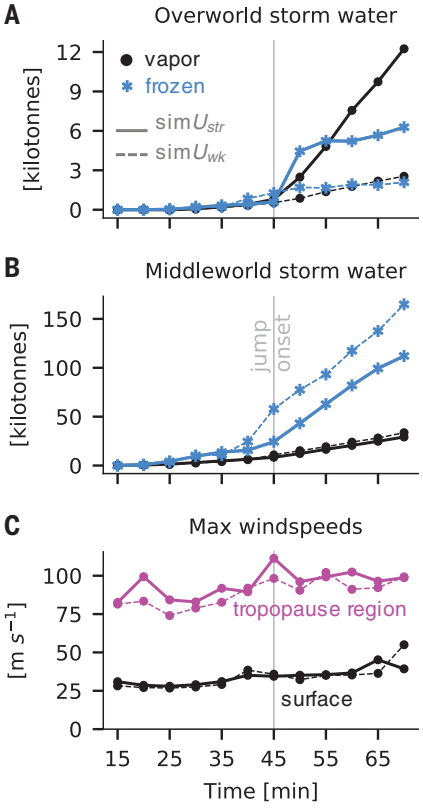


Fig. 4. Temporal evolution of water mass and wind in simU_{str} and simU_{wk} in kilotonnes. Mass unit is 1 tonne = 1000 kg. (**A**) Perturbation water content (black circles for vapor and blue snowflakes for solid phases) in the model domain in the stratospheric overworld above the 380 K potential temperature surface for simU_{str} (solid line) and simU_{wk} (dashed line). (**B**) Similar to (A) except in the stratospheric middleworld 360 to 380 K layer. (**C**) Peak wind speed in the storm vicinity in the tropopause region (magenta, 11- to 14-km altitude) and near the surface (black, 25 m altitude). Jump onset in simU_{str} is indicated by the gray vertical line in each panel.

forms a wave-induced critical layer at that height. This layer can then serve to trap and reflect vertically propagating gravity waves back downward and eventually lead to the development of a hydraulic jump in continuously stratified flow (*35*).

One might measure a nondimensional obstacle height $\hat{h}_m = Nh_m/U$ to understand lee flow behavior as a function of incoming horizontal wind speed U, buoyancy frequency N, and topographic height h_m (*4*). Though the OT reaches 15 km, 3 km above the base state tropopause, stratospheric air that is lifted and advected over the OT rises no more than 200 m ($h_m \approx 200$ m). Most air is deflected around the storm laterally. Though this yields $\hat{h}_m < 1$ for simU_{str}, $\hat{h}_m$ is not the only control that determines wave breaking aloft or downstream, and a hydraulic jump can still occur when $\hat{h}_m < 1$ (*31*).

The primary departure from the downslope windstorm mechanism is the finding that air exiting the OT speeds up into the jet region as it subsides downwind in both simulations, apparently independent of whether the stratosphere converges into it or separates aloft. The jet can exceed speeds of 90 m s^{-1} even in simU_{wk} (Fig. 4C). This feature has no counterpart in real topographic obstacles because it is the topography itself that is flowing, and h_m varies with the evolving OT shape. A jet in the OT lee of both storms seems inevitable as the collapsing OT must accelerate downward in the lee toward its level of neutral buoyancy (*1*).

This work demonstrates a likely internal hydraulic jump in the absence of solid topography. More realistic wind profiles near and above the tropopause can have considerable shear, which can be a more favorable environment for the occurrence of a hydraulic jump (*34*). Though the presence of water species does not seem to have much effect on AACP dynamics, subsequent water injection to the stratosphere greatly increases stratospheric humidity during the duration of the simulation.

The realism of the rapid tropopause jet found in an AACP-producing storm can be validated using the existing high-altitude ER-2 airborne platform. Improvements in the radars aboard the ER-2 include scanning for 3D winds and higher sensitivity. Additionally, a cloud lidar that detects small ice particles often accompanies these radars. A dedicated field campaign that targets AACP-generating convective storms at close range would better constrain storm top physics and quantify their water vapor contribution to the stratosphere; this would be complementary to the ongoing NASA Dynamics and Chemistry of the Summer Stratosphere (DCOTSS) campaign taking place 2021–2022.

REFERENCES AND NOTES

1. T. T. Fujita, *J. Meteorol. Soc. Jpn. Ser. II* **60**, 355–368 (1982).
2. P. K. Wang, *J. Geophys. Res. Atmos.* **108**, 4194 (2003).
3. K. Bedka *et al.*, *J. Appl. Meteorol. Climatol.* **49**, 181–202 (2010).
4. C. R. Homeyer, *J. Atmos. Sci.* **71**, 332–348 (2014).
5. C. R. Homeyer, J. D. McAuliffe, K. M. Bedka, *J. Atmos. Sci.* **74**, 1617–1633 (2017).
6. K. Bedka, E. M. Murillo, C. R. Homeyer, B. Scarino, H. Mersiovsky, *Weather Forecast.* **33**, 1159–1181 (2018).
7. M. E. E. Hassim, T. P. Lane, *Atmos. Chem. Phys.* **10**, 9833–9849 (2010).
8. T. Corti *et al.*, *Geophys. Res. Lett.* **35**, 35 (2008).
9. J. B. Smith *et al.*, *J. Geophys. Res. Atmos.* **122**, 9529–9554 (2017).
10. D. T. Shindell, *Geophys. Res. Lett.* **28**, 1551–1554 (2001).
11. C. A. Smith, J. D. Haigh, R. Toumi, *Geophys. Res. Lett.* **28**, 179–182 (2001).
12. J. G. Anderson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4905–E4913 (2017).
13. P. M, *Geophys. Res. Lett.* **26**, 3309–3312 (1999).
14. S. Solomon *et al.*, *Science* **327**, 1219–1223 (2010).
15. J.-P. Chaboureau *et al.*, *Atmos. Chem. Phys.* **7**, 1731–1740 (2007).
16. C. Chemel, M. R. Russo, J. A. Pyle, R. S. Sokhi, C. Schiller, *Mon. Weather Rev.* **137**, 2493–2514 (2009).
17. X. M. Liu *et al.*, *Atmos. Chem. Phys.* **10**, 8267–8286 (2010).
18. T. Dauhut, J.-P. Chaboureau, J. Escobar, P. Mascart, *Atmos. Sci. Lett.* **16**, 135–140 (2015).
19. R. J. Trapp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19719–19723 (2007).
20. A. E. Dessler, M. R. Schoeberl, T. Wang, S. M. Davis, K. H. Rosenlof, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18087–18091 (2013).
21. G. M. Heymsfield, R. H. Blackmer, S. Schotz, *J. Atmos. Sci.* **40**, 1740–1755 (1983).
22. Materials and methods are available as supplementary materials.
23. G. H. Bryan, J. M. Fritsch, *Mon. Weather Rev.* **130**, 2917–2928 (2002).
24. J. D. Doyle, D. R. Durran, *J. Atmos. Sci.* **59**, 186–201 (2002).
25. T. P. Lane, R. D. Sharman, T. L. Clark, H.-M. Hsu, *J. Atmos. Sci.* **60**, 1297–1321 (2003).
26. R. F. Hertenstein, J. P. Kuettner, Rotor types associated with steep lee topography: Influence of the wind profile. *Tellus Ser. A.* **57**, 117–135 (2005).
27. G. M. Heymsfield, R. H. Blackmer Jr.., *Mon. Weather Rev.* **116**, 2200–2224 (1988).
28. G. M. Heymsfield, L. Tian, L. Li, M. McLinden, J. I. Cervantes, *J. Appl. Meteorol. Climatol.* **52**, 1851–1867 (2013).
29. M. P. Jensen *et al.*, *Bull. Am. Meteorol. Soc.* **97**, 1667–1686 (2016).
30. R. B. Smith, *J. Atmos. Sci.* **42**, 2597–2603 (1985).
31. D. R. Durran, *J. Atmos. Sci.* **43**, 2527–2543 (1986).
32. R. B. Smith, *J. Atmos. Sci.* **44**, 269–297 (1987).
33. D. R. Durran, J. B. Klemp, *J. Atmos. Sci.* **44**, 3402–3412 (1987).
34. D. R. Durran, in *Mountain Waves and Downslope Winds* (American Meteorological Society (1990), pp. 59–81.
35. T. L. Clark, W. R. Peltier, *J. Atmos. Sci.* **34**, 1715–1730 (1977).
36. M. E O'Neill, L. Orf, G. M. Heymsfield, K. Halbert, *Zenodo* (2021); https://doi.org/10.5281/zenodo.4584086.
37. L. Orf, Numerical Simulation of Supercell Thunderstorms (at 50 meter resolution) Associated with Above Anvil Cirrus Plumes, *Dryad* (2021); https://doi.org/10.5061/dryad.xsj3tx9fw.

ACKNOWLEDGMENTS

We are grateful to D. Durran and C. Homeyer for helpful discussions and to the developers of the xarray and plotly python software packages. We also thank three anonymous reviewers for their helpful reviews. **Funding:** L.O. was supported in part by NSF grants AGS-1832327 and OAC-1663954. L.O.'s access to the NSF-sponsored Frontera supercomputer was provided by the Texas Advanced Computing Center via grant ATM20006. G.M.H. was funded by the NASA Precipitation Measurement Mission (PMM) and Ground Validation program. **Author contributions:** M.E.O. and L.O. designed the numerical experiments. L.O. ran, processed, and visualized the simulations, and K.H. computed the Lagrangian parcels. G.M.H. provided the MC3E radar measurements and data analysis. M.E.O. conducted the analyses and wrote the paper, and all coauthors assisted in reviewing and editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** CM1 is publicly available (www2.mmm.ucar.edu/people/bryan/cm1/). Modifications to source code and routines to run model are publicly available (https://github.com/leighorf/cm1r19.8-LOFS and https://github.com/leighorf/LOFS-read). Simulation input files are included in supplementary materials. Plotting codes (*36*) and 203 GB of model output and trajectory data (*37*) are publicly available. HIWRAP radar data are available at the Global Hydrology Resource Center – Distributed Active Archive Center (https://ghrc.nsstc.nasa.gov/home/field-campaigns/mc3e).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abh3857
Materials and Methods
Figs. S1 to S9
Data S1 and S2
References (*38*–*43*)

4 March 2021; accepted 10 August 2021
10.1126/science.abh3857

NEUROSCIENCE

Synaptic mechanism underlying serotonin modulation of transition to cocaine addiction

Yue Li[1], Linda D. Simmler[1], Ruud Van Zessen[1], Jérôme Flakowski[1]†, Jin-Xia Wan[2,3], Fei Deng[2,3], Yu-Long Li[2,3], Katherine M. Nautiyal[4], Vincent Pascoli[1], Christian Lüscher[1,5]*

Compulsive drug use despite adverse consequences defines addiction. While mesolimbic dopamine signaling is sufficient to drive compulsion, psychostimulants such as cocaine also boost extracellular serotonin (5-HT) by inhibiting reuptake. We used SERT Met172 knockin (SertKI) mice carrying a transporter that no longer binds cocaine to abolish 5-HT transients during drug self-administration. SertKI mice showed an enhanced transition to compulsion. Conversely, pharmacologically elevating 5-HT reversed the inherently high rate of compulsion transition with optogenetic dopamine self-stimulation. The bidirectional effect on behavior is explained by presynaptic depression of orbitofrontal cortex–to–dorsal striatum synapses induced by 5-HT via 5-HT$_{1B}$ receptors. Consequently, in projection-specific 5-HT$_{1B}$ receptor knockout mice, the fraction of individuals compulsively self-administering cocaine was elevated.

With chronic consumption, ~20% of cocaine users lose control and are eventually diagnosed as addicted (*1*). An increase in dopamine (DA) levels is typical of all addictive drugs (*2*) and sufficient to trigger forms of synaptic plasticity underlying adaptive behaviors (*3*–*5*). This is exemplified by optogenetic DA neuron self-stimulation (oDASS), which induces neuronal adaptations similar to addictive drugs via selective release of DA from ventral tegmental area (VTA) neurons and yields a bimodal distribution of compulsive and noncompulsive individuals (*6*). Cocaine also inhibits the serotonin [5-hydroxytryptamine (5-HT)] transporter (SERT), causing 5-HT transients in the striatum. While pharmacological reduction of 5-HT in the entire forebrain can favor compulsive cocaine seeking (*7*) and differential efficacy of the 5-HT system may be involved in the vulnerability to drug addiction (*8*, *9*), the relevant circuits and underlying cellular mechanisms remain elusive. To parse the locus of 5-HT modulation, we took advantage of SERT Met172 knockin (SertKI) mice carrying a transporter that does not bind cocaine without altering the basal 5-HT levels (*10*–*12*). Genetically encoded 5-HT sensors (fig. S1) confirmed the absence of cocaine-evoked transients in the dorsal striatum (DS) of SertKIs [15 mg/kg ip (intraperitoneally)] (Fig. 1, A to C). Mice were trained to press a lever that triggered a cocaine intravenous infusion (0.5 mg/kg per infusion) accompanied by a cue light, followed by a progressive ratio (PR) session and four punishment (0.2-mA foot shock) sessions (Fig. 1D and fig. S2A). There was no difference between the SertKI and wild-type (WT) groups during the acquisition period (Fig. 1E). Facing punishment, however, some mice reduced cocaine self-administration (SA), while others continued unabated (Fig. 1F). An unbiased clustering analysis integrating four behavioral parameters over the last two punishment sessions yielded two clusters: renouncers and perseverers (Fig. 1, G and H). Out of 25 SertKI mice, 14 (56%) were classified as perseverers, in stark contrast to the 3 out of 26 (12%) in WT littermate mice (Fig. 1I). Perseverance was correlated neither to baseline cocaine SA (Fig. 1J) nor to the break point (fig. S2B), and the success rate accomplishing increasing break points did not differ between renouncers and perseverers of either genotype (Fig. 1K). Perseverers and renouncers across genotypes perceived pain similarly, as

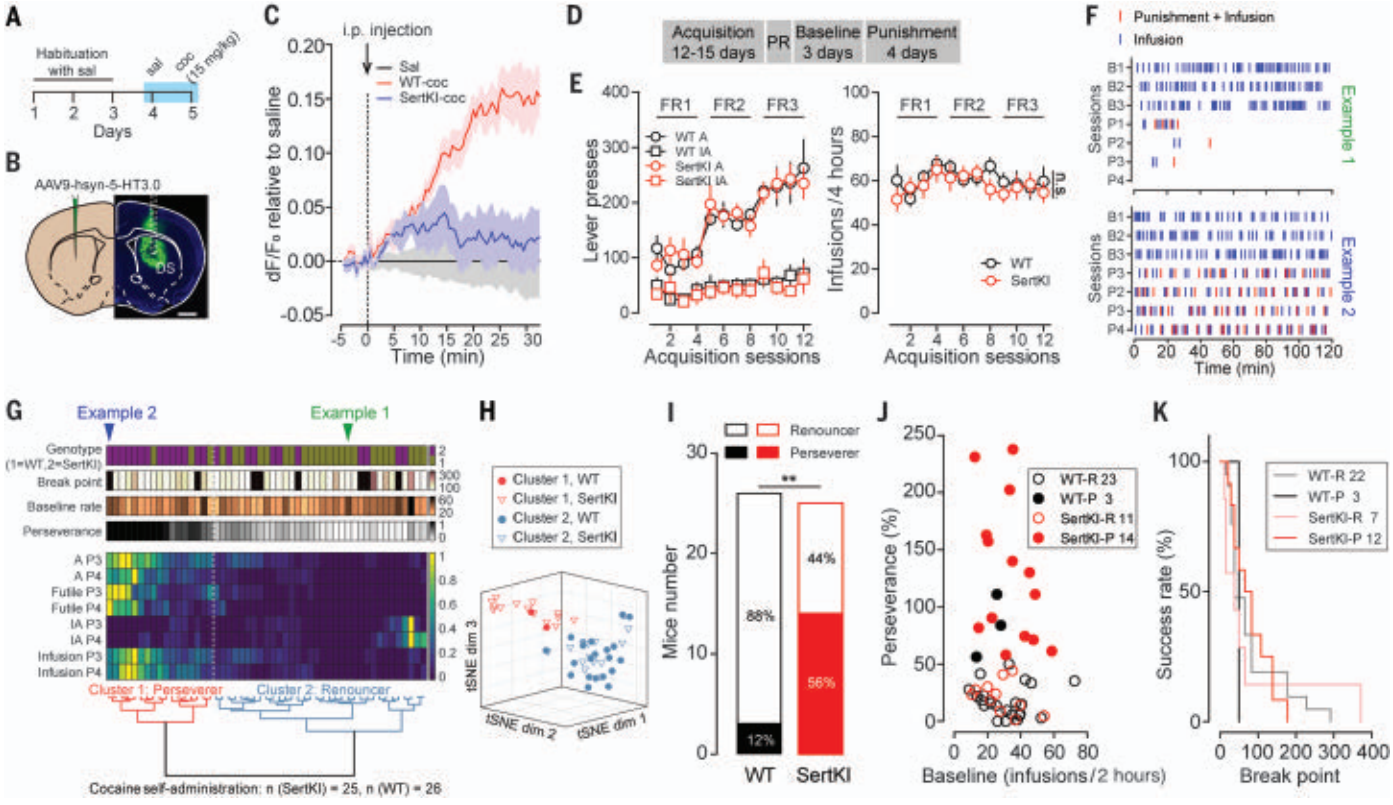


Fig. 1. SertKI animals are more compulsive for cocaine self-administration. (**A**) Schedule of saline and cocaine injections. (**B**) GRAB (G protein–coupled receptor activation–based) 5-HT sensor expression indicated with green fluorescent protein staining in the DS. Scale bar, 1 mm. (**C**) 5-HT transients in the DS induced by saline or cocaine (15 mg/kg) intraperitoneal injection in WT and SertKI mice (n = 3 mice for WT and SertKI group; data from saline-injected WT and SertKI are pooled). (**D**) Timeline of cocaine SA. (**E**) (Left) Number of active (A) and inactive (IA) lever presses of WT (black) and SertKI (red) animals (n = 26 and 25 mice for WT and SertKI groups, respectively) in acquisition sessions. (Right) Cocaine infusions obtained from WT (black) and SertKI (red) mice in acquisition sessions [two-way analysis of variance (ANOVA); $F_{1,588}$ = 1.996, P = 0.1583; n = 26 and 25 mice for WT and SertKI groups, respectively]. n.s., not significant. (**F**) Raster plots for infusions (blue lines) and punishments (red lines) in baseline and punishment sessions of a renouncer (WT mouse, top) and a perseverer (SertKI mouse, bottom). (**G**) Hierarchical clustering based on t-distributed stochastic neighbor embedding (tSNE) projection of different parameters of punishment sessions 3 and 4 (P3 and P4) of cocaine SA. Blue and green arrow heads indicate the examples presented in (F). (**H**) tSNE three-dimensional representation of clusters of perseverers (cluster 1) and renouncers (cluster 2) in cocaine SA. (**I**) Percentage of perseverers and renouncers among WT and SertKI groups (Fisher's exact test; $^{**}P$ = 0.001). (**J**) Perseverance rate as a function of the baseline rate (Pearson correlation coefficient r = −0.08; P = 0.55). (**K**) The success rate of performance as a function of the last progressive ratio value achieved by the mice (logrank test; P = 0.95). PR, progressive ratio; FR, fixed ratio; A, active lever presses; IA, inactive lever presses; R, renouncer; P, perseverer; sal, saline; coc, cocaine; WT, wild type; SertKI, SERT Met172 knockin. Data presented as means ± SEM.

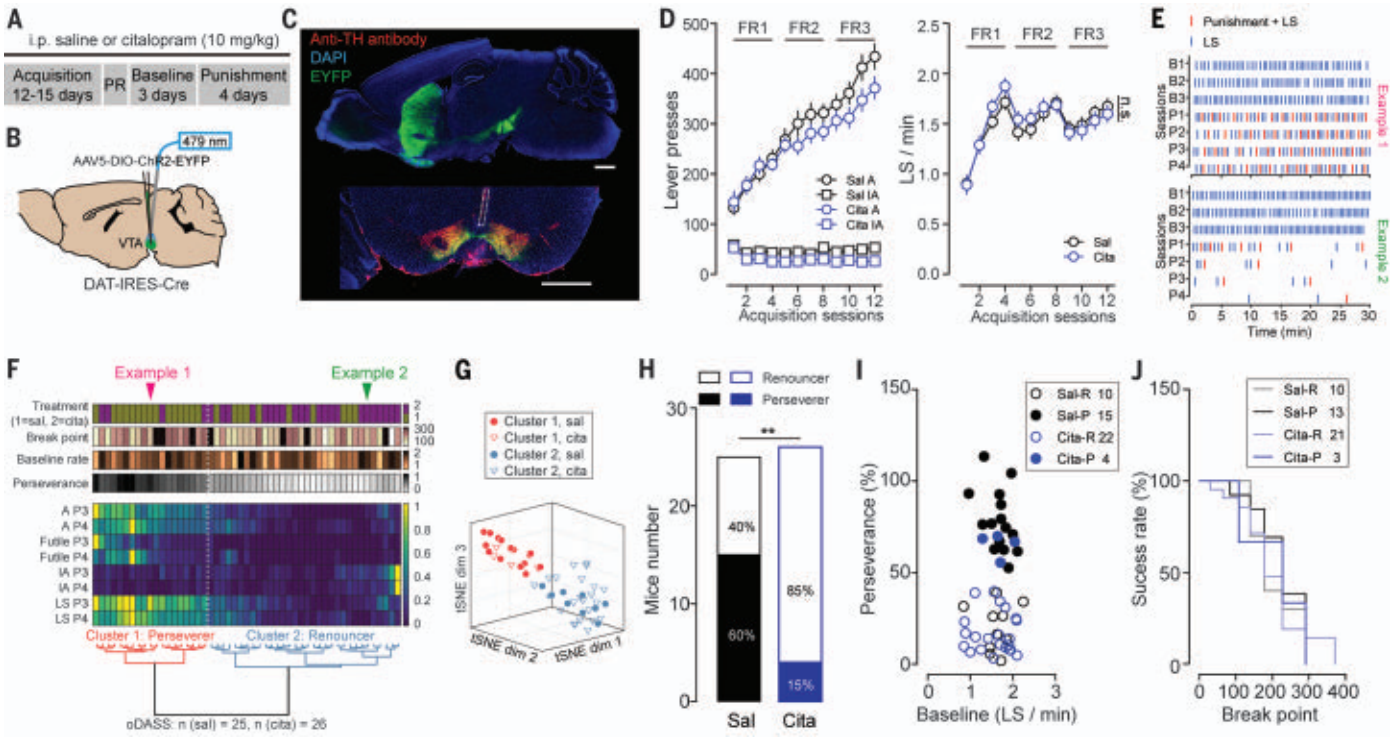


Fig. 2. Citalopram decreases compulsive dopamine self-stimulation. (**A**) Timeline of oDASS and saline or citalopram treatments. (**B**) Schematic of virus injection sites and optic fiber implantation sites. (**C**) ChR2-EYFP expression in the VTA from a sagittal slice (top image) and a coronal slice colabeled with tyrosine hydroxylase (TH; bottom image). Scale bars, 1 mm. (**D**) (Left) Number of active (A) and inactive (IA) lever presses of saline- (black) and citalopram-treated (blue) mice (n = 25 and 26 mice for saline and citalopram groups, respectively) in acquisition sessions. (Right) Laser stimulation per minute obtained from saline- (black) and citalopram-treated (blue) mice in acquisition sessions (two-way ANOVA; $F_{1,588}$ = 0.73, P = 0.39; n = 25 and 26 mice for saline and citalopram groups, respectively). (**E**) Raster plots for laser stimulations (blue lines) and punishments (red lines) in baseline and punishment sessions of a perseverer (top, from saline group) and a renouncer (bottom, from citalopram group). (**F**) Hierarchical clustering based on different parameters of punishment sessions 3 and 4 (P3 and P4) of oDASS. Red and green arrow heads indicate examples presented in (E). (**G**) tSNE three-dimensional representation of clusters of perseverers (cluster 1) and renouncers (cluster 2) in oDASS. (**H**) Percentage of perseverers and renouncers among saline- and citalopram-treated groups (Fisher's exact test; $^{**}P$ = 0.001). (**I**) Perseverance rate as a function of the baseline rate (Pearson r = 0.05; P = 0.71). (**J**) The success rate of performance as a function of the last progressive ratio value achieved by the mice (logrank test; P = 0.95). DAPI, 4′,6-diamidino-2-phenylindole; EYFP, enhanced yellow fluorescent protein; LS, laser stimulation; cita, citalopram. Data presented as means ± SEM.

[1]Department of Basic Neurosciences, Medical Faculty, University of Geneva, CH-1211 Geneva, Switzerland. [2]State Key Laboratory of Membrane Biology, Peking University School of Life Sciences, Beijing 100871, China. [3]IDG McGovern Institute for Brain Research, Peking University, Beijing 100871, China. [4]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA. [5]Clinic of Neurology, Department of Clinical Neurosciences, Geneva University Hospital, CH-1211 Geneva, Switzerland.
***Corresponding author. Email: christian.luscher@unige.ch**
†Deceased.

hot plate latency was not affected by cocaine (fig. S2C).

Next, we did the converse. We gave mice access to oDASS, which leads to compulsion in more than half of individuals (*6*) and pharmacologically elevated 5-HT levels with citalopram (Fig. 2, A to C). Citalopram (10 mg/kg) induced robust 5-HT transients of a magnitude comparable to cocaine (fig. S1D). Active lever presses in mice that expressed ChR2 in VTA DA neurons induced a brief train of laser stimulations (LS; see methods in the supplementary materials). All mice readily mastered oDASS (Fig. 2D) regardless of pharmacological treatment, but major differences emerged when punishment was introduced (Fig. 2E). Clustering analysis, as described above, led to the emergence of perseverers and renouncers in both treatment groups (Fig. 2, F and G). However, only 4 out of 26 (15%) citalopram-treated mice were classified as perseverers, whereas in the saline-treated group, 60% were perseverers—a fraction similar to previous reports (*13*) (Fig. 2H). Again, the perseverance rate was uncorrelated to the baseline oDASS rate (Fig. 2I) or the break point (fig. S2D), and the success rate accomplishing each break point did not differ between renouncers and perseverers across treatment groups (Fig. 2J). For all groups and conditions, pain perception was similar (fig. S2E).

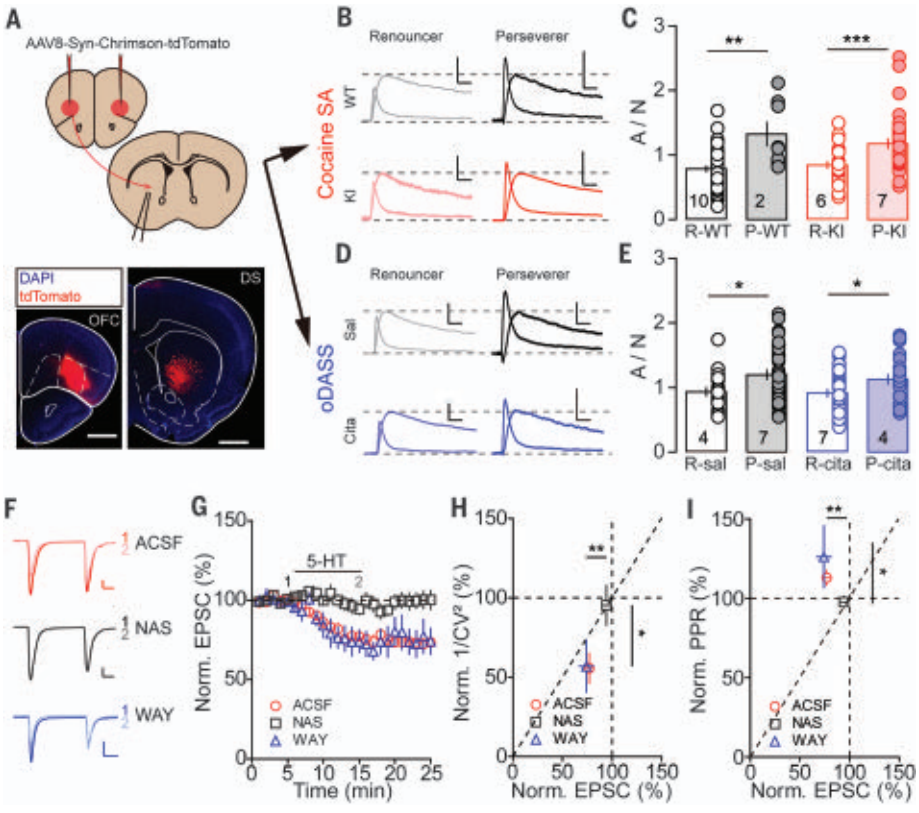


Fig. 3. The OFC-DS pathway is modulated by 5-HT. (**A**) (Top) Schematic of virus injection and recording sites. (Bottom) Chrimson-tdTomato–expressing cell bodies in the OFC (left) and terminals in the DS (right). Scale bars, 1 mm. (**B**) AMPA and NMDA currents at +40 mV of a renouncer (left) and a perseverer (right) from WT (upper) and SertKI (lower) group after cocaine SA. Scale bars, 200 pA, 15 ms. (**C**) Average AMPA/NMDA (A/N) of WT and KI renouncers and perseverers after cocaine SA (Mann Whitney test; $U = 74$, $^{**}P = 0.003$; $n = 53$ and 8 cells from ten and two mice for renouncer and perseverer in the WT group; $U = 252$, $^{***}P = 0.0001$; $n = 32$ and 35 cells from six and seven mice for renouncer and perseverer in SertKI group, respectively). (**D**) AMPA and NMDA currents at +40 mV of a renouncer (left) and a perseverer (right) from saline (upper) and citalopram (lower) treated group after oDASS. Scale bars, 200 pA, 15 ms. (**E**) Average A/N of saline- and citalopram-treated renouncers and perseverers after oDASS (two-tailed t test; $t_{54} = 2.54$, $^{*}P = 0.01$; $n = 21$ and 35 cells from four and seven mice for renouncer and perseverer in the saline-treated group, respectively; $t_{55} = 2.38$, $^{*}P = 0.02$; $n = 34$ and 23 cells from seven and four mice for renouncer and perseverer in the citalopram-treated group, respectively). (**F**) Traces before and after bath application of 5-HT in the presence of artificial cerebrospinal fluid (ACSF; red), NAS181 (gray), and WAY100635 (blue). Scale bars, 200 pA, 10 ms. (**G**) Average traces of EPSC before, during and after bath application of 5-HT in the presence of ACSF (red), NAS181 (gray), and WAY100635 (blue) ($n = 13$ and 14 cells from three mice for the ACSF and NAS181 groups, respectively, and seven cells from four mice for the WAY100635 group). (**H**) Normalized coefficient of variation ($1/CV^2$) versus normalized EPSC (one-way ANOVA; $F_{2,31} = 3.682$, $^{*}P = 0.0367$ for Norm. $1/CV^2$; $F_{2,31} = 7.948$, $^{**}P = 0.0016$ for Norm. EPSC; $n = 13$ and 14 cells from three mice for the ACSF and NAS181 groups, respectively, and seven cells from four mice for the WAY100636 group). (**I**) Normalized pair pulse ratio (PPR) versus normalized EPSC (Kruskal-Wallis test; $^{*}P = 0.041$ for Norm. PPR; $n = 13$ cells from three mice for the ACSF group, 10 and 6 cells from two mice for the NAS181 and WAY100636 groups, respectively). Data presented as means ± SEM.

Given that, for oDASS, the synaptic potentiation of afferents from the orbitofrontal cortex (OFC) to the DS drives perseverance (*13*), we wondered whether this is also the case for cocaine SA. To selectively stimulate the OFC-DS projection, we expressed the red-shifted opsin Chrimson in OFC neurons (Fig. 3A) and evoked excitatory postsynaptic currents (EPSCs) by illuminating the terminals in brain slices of the DS 24 to 48 hours after the last punishment session. The AMPA/NMDA (*N*-methyl-D-aspartate) ratio was higher in perseverers than in renouncers of cocaine SA as well as oDASS, regardless of genotypes and treatment (Fig. 3, B to E), confirming that a potentiated OFC-DS pathway reflects perseverance both in oDASS and cocaine SA. In no condition were the EPSCs rectifying (fig. S3), suggesting a potentiation by an increase in the number of AMPA receptors without a change in subunit composition, akin to the expression mechanism observed in individuals with compulsive oDASS (*13*).

We next examined the effect of 5-HT on synaptic transmission at the OFC-DS pathway in naïve mice. Bath application of 5-HT (4 μM) induced a presynaptic depression of excitatory transmission (Fig. 3, F and G), which could be blocked by 5-HT_{1B} receptor antagonist NAS181 (20 μM) but not by 5-HT_{1A} receptor antagonist WAY100*635* (1 μM) (Fig. 3, F and G). In line with the Gi/Go coupling of presynaptically located 5-HT_{1B} receptors, we observed a decreased coefficient variance ($1/CV^2$) and an increased paired pulse ratio (PPR), suggesting that the presynaptic depression was expressed by a reduction of glutamate release probability (Fig. 3, H and I). For confirmation, we evoked quantal events (qEPSCs) after replacing extracellular calcium (Ca^{2+}) with strontium (Sr^{2+}), thus desynchronizing light-evoked transmitter release (*14*), and found that the qEPSC frequency decreased

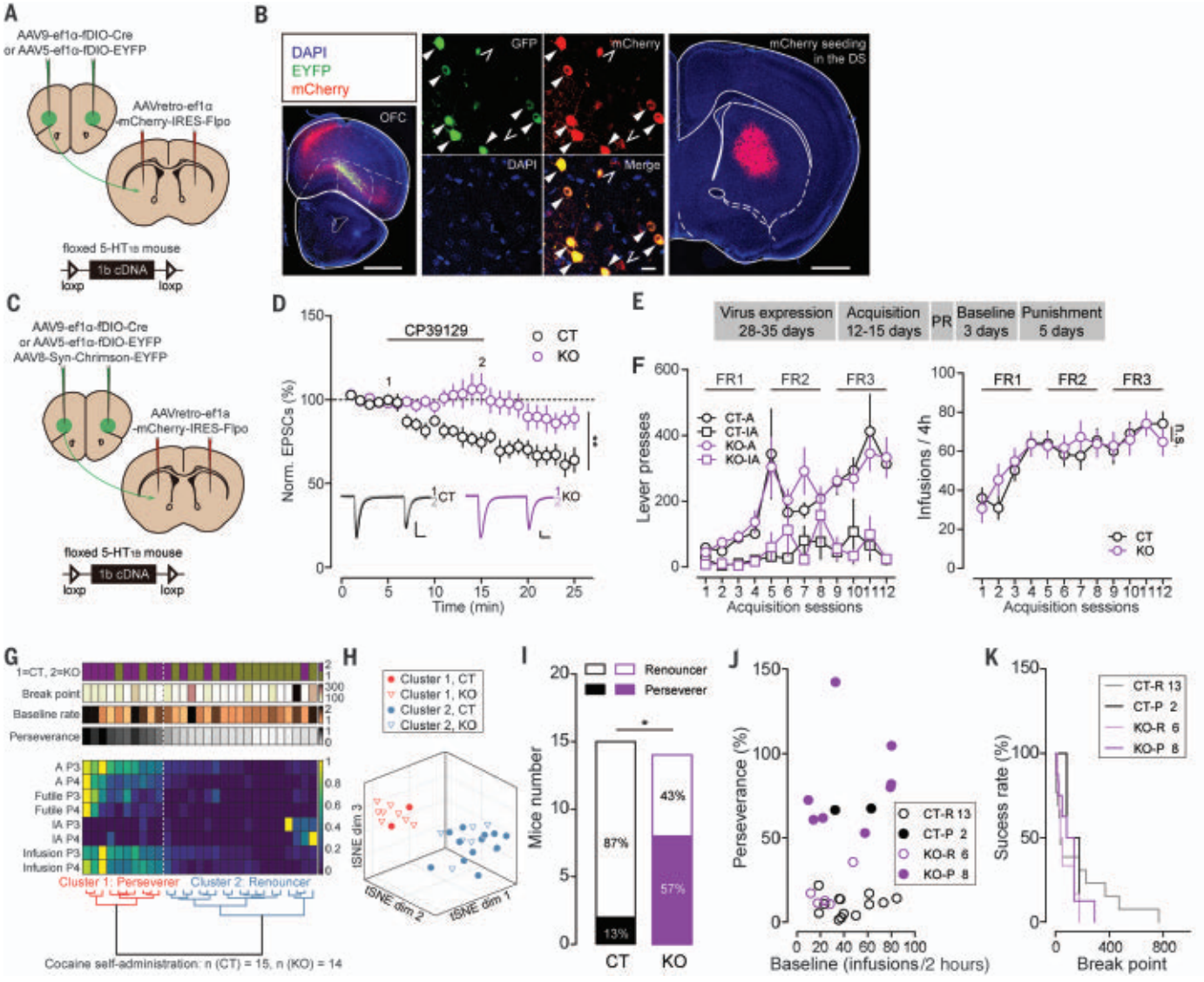


Fig. 4. Knocking out 5-HT$_{1B}$ receptors promotes compulsive cocaine self-administration. (**A**) Schematic of virus injections for cocaine SA. (**B**) (Left and middle) EYFP coexpressing with mCherry in the OFC. (Right) mCherry seeding in the DS. Scale bars, 1 mm (left and right) and 20 μm (middle). Arrows indicate DS projectors in the OFC expressing both EYFP and mCherry, open arrows indicate DS projectors expressing mCherry not infected by AAV-fDIO-EYFP. (**C**) Schematic of virus injections for patch clamp recording. (**D**) Average traces of EPSC before, during, and after bath application of 5-HT$_{1B}$ receptor agonist CP39129 in control (CT, black) and pathway-specific knockout 5-HT$_{1B}$ receptor (KO, violet) groups (compared Norm. EPSC recorded on last 5 min; two-tailed *t* test; t_{25} = 2.86, $^{**}P$ = 0.008; *n* = 15 and 12 cells from three mice for the CT and KO groups, respectively). Scale bars, 200 pA, 10 ms. (**E**) Timeline of virus injection and cocaine SA experiments. (**F**) (Left) Number of active (A) and inactive (IA) lever presses of CT (black) and KO (violet) mice in acquisition sessions. (Right) Cocaine infusions obtained by CT (black) and KO (violet) mice in acquisition sessions (two-way ANOVA; $F_{1,324}$ = 0.198, P = 0.66; *n* = 15 and 14 mice for the CT and KO groups, respectively). (**G**) Hierarchical clustering based on different parameters of punishment sessions 3 and 4 (P3 and P4) of cocaine SA. (**H**) tSNE three-dimensional representation of clusters of perseverers (cluster 1) and renouncers (cluster 2) in cocaine SA. (**I**) Percentage of perseverers and renouncers among CT and KO groups (Fisher's exact test; $^{*}P$ = 0.02). (**J**) Perseverance rate as a function of the baseline rate (Pearson r = 0.20; P = 0.30). (**K**) The success rate of performance as a function of the last progressive ratio value achieved by the mice (logrank test; P = 0.84). Data presented as means ± SEM.

but the amplitude stayed unchanged (fig. S4, A to C). Furthermore, 5-HT induced presynaptic depression in both D1-positive and D1-negative neurons obtained from D1-tdTomato mice (fig. S4D), consistent with previous reports (*15*, *16*).

We next hypothesized that presynaptic depression may reduce the likelihood for long-term potentiation (LTP) at the OFC- DS synapse, which in turn would prevent the transition to compulsion in cocaine SA. This seems plausible since chemogenetic reduction of OFC activity also reduced the fraction of perseverers in oDASS (*6*). Therefore, to establish a causal link between 5-HT–induced presynaptic depression and compulsive cocaine use, we aimed to abolish 5-HT$_{1B}$ receptors selectively in the OFC neurons targeting the DS. We injected

retroAAV-ef1α-mCherry-IRES-Flpo in the DS of 5-HT_{1B} floxed mice (*17*). This led to Flpo expression in OFC neurons targeting the DS. AAV9-ef1α-fDIO-Cre or control virus was then injected in the OFC to express Cre in OFC cells targeting the DS (Fig. 4, A and B), which then led to recombination and abolishment of 5-HT_{1B} receptors. We confirmed the successful receptor knockout functionally by the inability of 5-HT_{1B} agonist CP39129 (2 μM) to induce a presynaptic depression (Fig. 4, C and D).

A month after virus injection, the 5-HT_{1B} knockout mice learned to self-administer cocaine (Fig. 4E). We observed an acquisition period (Fig. 4F) similar to the one described above (Fig. 1E). Once punishment was introduced, we again observed persevering and renouncing mice (Fig. 4, G and H), with a higher fraction of perseverers in the projection-specific 5-HT_{1B} knockout compared to the control group (57% versus 13%) (Fig. 4I). Perseverance was unrelated to baseline performance (Fig. 4J), or break point in the two groups, in both control and knockout mice (Fig. 4K). The percentage of perseverers in the 5-HT_{1B} knockout group was very close to the fraction of perseverers in the SertKI group and in saline-treated oDASS mice.

A synaptic mechanism thus emerges that underlies a modulatory role of 5-HT reducing the likelihood of transition to compulsion and eventually addiction (see fig. S5). In WT mice, cocaine binds to SERT to block 5-HT reuptake. The elevated extracellular 5-HT activates 5-HT_{1B} receptors and causes presynaptic depression of OFC terminals. This reduces the likelihood of inducing postsynaptic potentiation at OFC-DS synapses that ultimately drives compulsion. In SertKI mice, cocaine cannot bind to SERT, and extracellular 5-HT remains unaffected by cocaine infusions (*10*). An OFC-DS transmission not undergoing presynaptic depression may thus enhance the likelihood of LTP induction, and the stochastic process would then show as a higher fraction of compulsive individuals. In 5-HT_{1B} knockout mice, although cocaine still inhibits 5-HT reuptake, the OFC-DS transmission is not depressed, which again may favor LTP induction. This interpretation is in line with the report that genome-wide 5-HT_{1B} receptor knockout mice are more impulsive (*17*). These mice are also more vulnerable to cocaine (*18*), which raises the possibility that individual addiction liability may be determined by 5-HT signaling. Variation in 5-HT synthesis, synaptic release, efficiency of reuptake, and extracellular levels could be additional determinants of overall vulnerability. Here we reveal addiction liability once 5-HT modulation has been eliminated, which is in line with the general idea that 5-HT opposes DA effects to inhibit behavior (*19*). However, this model is challenged by the observation that selective activation of 5-HT neurons allows high motivation to be maintained during complex tasks (*20*, *21*).

While our study focused on cocaine, 5-HT may also counteract the transition to compulsion when other addictive drugs are consumed (*8*, *9*). Amphetamine, despite having a relatively low SERT affinity, increases nonvesicular release of 5-HT, and opioids may indirectly activate 5-HT neurons in the dorsal raphe (*22*–*24*). In fact, the ratio between dopamine transporter and SERT affinity may predict the addiction liability of emerging drugs (*25*). This may also apply to natural rewards such as food and sex, but these have low addiction liabilities, making empirical testing a challenge. Finally, it may also be interesting to explore whether 5-HT modulation levels may not only prevent the transition to compulsion but also facilitate regaining control, as suggested by pharmacological interventions in rodents in a distinct behavioral paradigm. Forebrain 5-HT_{2C} receptors may inhibit compulsive cocaine seeking after compulsion is established (*7*), possibly via modulation of acute effects and early adaptive behaviors (*26*). By contrast, our data show that pathway-specific knockout of 5-HT_{1B} receptors does not affect acquisition or motivation for cocaine and specifically modulates compulsion. Three days of 5-HT_{1B} agonist (CGS12066B; 10 mg/kg, 30 min intraperitoneal injection before each session) treatment after the last punishment session of oDASS left perseverance to oDASS intact (fig. S6), suggesting that 5-HT_{1B} receptor activation prevents transition to compulsion but cannot reverse it once it is established.

The present mechanistic investigation may help to refine approaches to overcome the limitations and diverging findings on the efficacy of 5-HT reuptake blockers in pilot studies with human addicts (*27*–*29*) or to design selective agonists complementing the empirical use of hallucinogens in addiction treatments (*30*). In summary, 5-HT emerges as a modulator of the progression to compulsion via the convergence on key synaptic mechanisms in the framework of the current circuit model for addiction (*5*).

REFERENCES AND NOTES

1. M. Yücel *et al*., *Addiction* **114**, 1095–1109 (2019).
2. C. Lüscher, M. A. Ungless, *PLOS Med.* **3**, e437 (2006).
3. C. Lüscher, R. C. Malenka, *Neuron* **69**, 650–663 (2011).
4. C. Lüscher, *Annu. Rev. Neurosci.* **39**, 257–276 (2016).
5. C. Lüscher, P. H. Janak, *Annu. Rev. Neurosci.* **44**, 173–195 (2021).
6. V. Pascoli, J. Terrier, A. Hiver, C. Lüscher, *Neuron* **88**, 1054–1066 (2015).
7. Y. Pelloux, R. Dilleen, D. Economidou, D. Theobald, B. J. Everitt, *Neuropsychopharmacology* **37**, 2505–2514 (2012).
8. L. G. Kirby, F. D. Zeeb, C. A. Winstanley, *Neuropharmacology* **61**, 421–432 (2011).
9. C. P. Müller, J. R. Homberg, *Behav. Brain Res.* **277**, 146–192 (2015).
10. B. J. Thompson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3785–3790 (2011).
11. L. D. Simmler, R. D. Blakely, *ACS Chem. Neurosci.* **10**, 3053–3060 (2019).
12. L. D. Simmler *et al*., *Br. J. Pharmacol.* **174**, 2716–2738 (2017).
13. V. Pascoli *et al*., *Nature* **564**, 366–371 (2018).
14. S. H. Oliet, R. C. Malenka, R. A. Nicoll, *Science* **271**, 1294–1297 (1996).
15. G. Dölen, A. Darvishzadeh, K. W. Huang, R. C. Malenka, *Nature* **501**, 179–184 (2013).
16. B. N. Mathur, N. A. Capik, V. A. Alvarez, D. M. Lovinger, *J. Neurosci.* **31**, 7402–7411 (2011).
17. K. M. Nautiyal *et al*., *Neuron* **86**, 813–826 (2015).
18. B. A. Rocha *et al*., *Nature* **393**, 175–178 (1998).
19. Y. L. Boureau, P. Dayan, *Neuropsychopharmacology* **36**, 74–97 (2011).
20. Z. Liu *et al*., *Neuron* **81**, 1360–1374 (2014).
21. E. Lottem *et al*., *Nat. Commun.* **9**, 1000 (2018).
22. P. Scholze *et al*., *J. Pharmacol. Exp. Ther.* **293**, 870–878 (2000).
23. L. D. Simmler *et al*., *Br. J. Pharmacol.* **168**, 458–470 (2013).
24. R. Tao, S. B. Auerbach, *J. Pharmacol. Exp. Ther.* **303**, 704–710 (2002).
25. M. Liechti, *Swiss Med. Wkly.* **145**, w14043 (2015).
26. P. J. Fletcher, A. J. Grottick, G. A. Higgins, *Neuropsychopharmacology* **27**, 576–586 (2002).
27. S. Vayalapalli, M. Vaughn, K. Salles-Shahid, J. Byrd-Sellers, K. Drexler, *Am. J. Addict.* **20**, 485–486 (2011).
28. F. G. Moeller *et al*., *Am. J. Drug Alcohol Abuse* **33**, 367–378 (2007).
29. J. Grabowski *et al*., *J. Clin. Psychopharmacol.* **15**, 163–174 (1995).
30. L. P. Cameron *et al*., *Nature* **589**, 474–479 (2021).
31. Y. Li *et al*., Synaptic mechanism underlying serotonin modulation of transition to cocaine addiction, Version 1, Zenodo (2021); https://doi.org/10.5281/zenodo.5091499.

ACKNOWLEDGMENTS

In memory of our dear colleague Jérôme Flakowski, who died on 11 May 2021. We thank R. D. Blakely from the Florida Atlantic University Brain Institute for providing SertKI mice. We thank all lab members for suggestions and comments on the experiments and manuscript. We also thank C. Bellone, B. Everitt, and F. Markopoulos for critical comments on the manuscript. **Funding:** This work was supported by the Swiss National Science Foundation and the European Research Council (C.L.) and R01 MH126178 (K.M.N.). **Author contributions:** Conceptualization: Y.L., V.P., and C.L. Investigation: Y.L., L.D.S., and V.P.; Data analysis: Y.L., R.V.Z., and J.F. Visualization: Y.L. and J.F. Reagents: K.M.N., J.-X.W., F.D., and Y.-L.L. Funding acquisition: C.L. Supervision: C.L. Writing: Y.L., V.P., and C.L. Senior authors: Y.-L.L., K.M.N., V.P., and C.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data, code, and materials are available on Zenodo (*31*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abi9086
Materials and Methods
Figs. S1 to S6
References (*32*, *33*)
MDAR Reproducibility Checklist

9 April 2021; accepted 20 July 2021
10.1126/science.abi9086

2D MATERIALS

Correlated electron-hole state in twisted double-bilayer graphene

Peter Rickhaus[1]*, Folkert K. de Vries[1], Jihang Zhu[2], Elías Portoles[1], Giulia Zheng[1], Michele Masseroni[1], Annika Kurzmann[1], Takashi Taniguchi[3], Kenji Watanabe[3], Allan H. MacDonald[2], Thomas Ihn[1,4], Klaus Ensslin[1,4]

When twisted to angles near 1°, graphene multilayers provide a window on electron correlation physics. Here, we report the discovery of a correlated electron-hole state in double-bilayer graphene twisted to 2.37°. At this angle, the moiré states retain much of their isolated bilayer character, allowing their bilayer projections to be separately controlled by gates. We use this property to generate an energetic overlap between narrow isolated electron and hole bands with good nesting properties. Our measurements reveal the formation of ordered states with reconstructed Fermi surfaces, consistent with a density-wave state. This state can be tuned without introducing chemical dopants, enabling studies of correlated electron-hole states and their interplay with superconductivity.

Fermi surface nesting refers to electron and hole Fermi surfaces that map onto each other under translation by a nesting wave vector $\vec{Q}$. Because nesting implies a small band-energy cost for coherent superposition between electrons and holes, it favors interaction-driven broken-symmetry states. The nesting condition $\epsilon\left(\vec{k}_F + \vec{Q}\right) = -\epsilon\left(\vec{k}_F\right)$, where epsilon is a small energy offset from the Fermi energy and $\vec{k}_F$ is the Fermi wave vector, implies (*1*) that if two closed Fermi surfaces are perfectly nested, then they enclose the same area. Mixing two Fermi surfaces nested by wave vector $\vec{Q}$ leads to density-wave (DW) order with wavelength $2\pi/Q$. In the seminal theoretical work on DW states by Peierls (*2*), nesting occurs between like-spins in half-filled bands, the interactions are lattice mediated, and the DW is accompanied by a lattice distortion. DW states can also be favored by Coulomb interactions between electrons, in which case the lattice distortions (*3*) play only a parasitic role; the DW is then commonly referred to as an excitonic insulator (*4*). This term derives from viewing the order as a condensation of bosonic electron-hole pair states. Evidence for equilibrium excitonic condensation has been reported in 1T-$TiSe_2$ (*5*), in Sb nanoflakes (*3*), and in double quantum wells at high magnetic fields (*6*, *7*). Furthermore, notable progress toward quasi-equilibrium exciton condensation has been achieved in a $MoSe_2$/WSe_2 heterostructure under the application of a large interlayer bias (*8*). Condensation of nonequilibrium excitons and polaritons in optically pumped electron-hole fluids has also been studied extensively (*9*). Correlated electron-hole states continue to attract attention owing to their rich intrinsic physics and their close relationship to superconductivity (*10*).

Twisting van der Waals materials, including graphene bilayers (twisted bilayer graphene) (*11*–*16*) and double Bernal bilayers [twisted double-bilayer graphene (TDBG)] (*17*–*23*), is a proven strategy to engineer moiré bands that favor strongly correlated electronic states. In this study, we seek to realize electron-hole bands that are both nested and relatively narrow. To this end, we have studied the properties of TDBG (*17*–*23*) at intermediate twist angles, where the layer coupling is strong enough to form moiré bands but weak enough to retain the polarizability of decoupled bilayers (*24*–*27*). This polarizability can be quantified by the probability prob(κ_t) to find a band state at the κ_t-point (Brillouin zone corner of the top bilayer) in the top bilayer. If prob(κ_t) is large, which is the case at large twist angles, applying a displacement field D will charge the top bilayer with electrons and the bottom bilayer with holes (Fig. 1A). But correlated electron-hole states are fragile and, if they occur, only indirectly observable (*26*). Thus, the correlation strength needs to be maximized by approaching smaller angles where the narrowed moiré bands have a larger effective mass m^* (Fig. 1B). In figs. S18 and S19, we discuss the dependence of prob(κ_t) and m^*

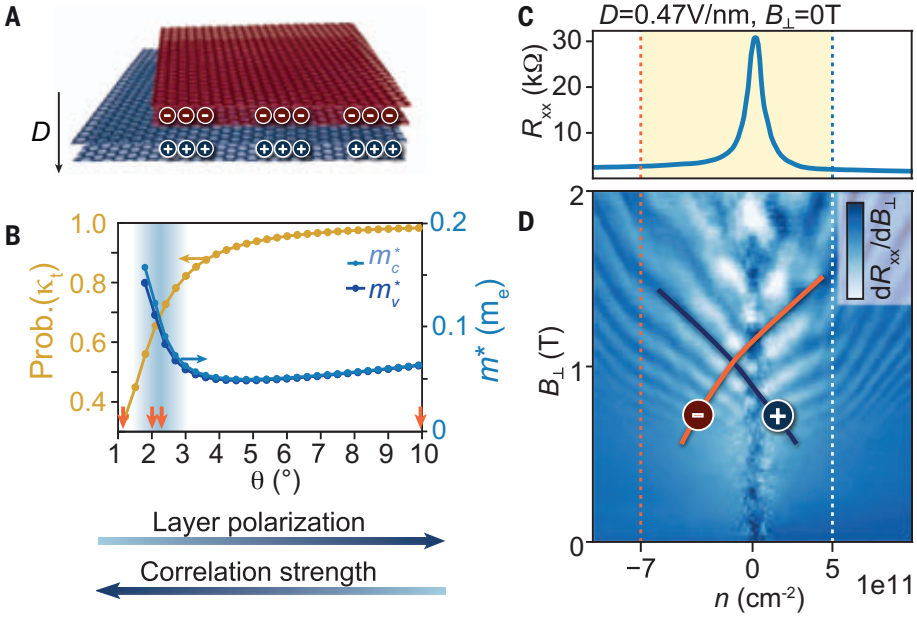


Fig. 1. Designed DW state in a moiré double bilayer. (**A**) Applying an electric field D to TDBG will charge the top and bottom bilayers with electrons and holes, respectively. (**B**) The layer polarization is given by prob(κ_t), and the correlation strength scales with the effective mass m^*. m_c^* (m_v^*) is the effective mass of the conduction (valence) band. Prob(κ_t)(θ) and $m^*(\theta)$ (at layer energy difference Δ_V = 30meV, see below) are obtained from band structure calculations at θ = 2.3°, showing that twists in the intermediate range (shaded region) combine strong polarizations and large masses and are favorable for DW formation. The red arrows mark twist angles at which measurements in fig. S4 are presented. (**C**) A peak in R_{xx} occurs at D = 0.47 V/nm and carrier density n = 0 in a TDBG device with θ = 2.37° and is consistent with a thermally activated gap (fig. S5). The yellow-shaded region represents the density range in which electron and hole bands coexist. (**D**) The bilayer's Landau fan diagram demonstrates that the gap appears inside a regime with both electron and hole bands, as confirmed by the presence of SdH resistance peaks (overlayed by orange and blue lines) that have opposite n versus $B_\perp$ slopes. Measurements are performed at temperature T = 1.2 K.

[1]Solid State Physics Laboratory, ETH Zürich, CH-8093 Zürich, Switzerland. [2]Department of Physics, University of Texas at Austin, Austin, TX 78712, USA. [3]National Institute for Material Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [4]Quantum Center, ETH Zürich, 8093 Zürich, Switzerland.
*Corresponding author. Email: peterri@phys.ethz.ch

on twist angle and external field using the Bistritzer-MacDonald model (*28*), which is extended to the TDBG case.

In this study, we show that correlated electron-hole states are favored to form in the intermediate twist angle regime, roughly between 2° and 3°. Figure 1, C and D, summarizes the main experimental findings at θ = 2.37°. We observe a resistance peak at zero total density $n = 0$ and displacement field $D = 0.47$ V/nm (Fig. 1C), which appears when electrons and holes with approximately the same density coexist. This coexistence is evident in Landau fan measurements (Fig. 1D), which reveal electron and hole minibands that cross in energy.

A side-view schematic of our device is shown in Fig. 2A. We tune the density $n = (C_t V_t + C_b V_b)/e$ and displacement field $D = (C_b V_b - C_t V_t)/2\epsilon_0$ by applying voltages V_t and V_b to top and back gate electrodes. Here, C_t and C_b are the corresponding capacitances per unit area. The resistance peak in Fig. 1C is measured using the contact geometry in Fig. 2B. From the decrease of the resistance peak with increasing temperature (fig. S5), we extract a gap Δ using the Arrhenius law $R_{xx} \propto \exp(\Delta/2k_B T)$, where k_B is the Boltzmann constant. The extracted dependence of Δ on D at total density $n = 0$ is shown in Fig. 2C. The data reveal a gap around $D = 0$ (gray line) that closes with increasing $|D|$. Another gap is opened at large $|D|$ (yellow line).

To identify the conditions for the formation of the gap, we present measurements of Shubnikov–de Haas (SdH) oscillations for different n and D. In Fig. 2, D and E, we plot the numerical derivative of the inverse resistance $\partial/\partial n\ (1/R_{xx})$ (this quantity is chosen for best visibility of the relevant features) measured in a perpendicular magnetic field $B_\perp = 1.5$T. Along the dashed line at $D = 0$ in Fig. 2E, the spacing between SdH oscillations corresponds to a band degeneracy $g = 8$. The degeneracy is lifted by changing D, and SdH lines with two slopes emerge (indicated with arrows) in the dark blue and dark purple regions. In the light blue and light purple regions, the pattern changes, and lines with $g = 4$ that are parallel to the $n = 0$ line are seen. The slope of the SdH lines then changes again in the yellow region, and the oscillations become weaker.

These observations suggest the presence of two subbands in the dark regions. Their energy offset is tuned by D, and thus the subbands are related to the top and bottom bilayers (*27*, *29*). In the light blue and light purple regions, only one subband exists, and SdH oscillations are independent of D. Because the displacement field changes the energetic offset of the subbands, the phase at large D and small n (yellow region), which contains two subbands, corresponds to an electron-hole fluid, in agreement with the slopes in the Landau fan diagram Fig. 1D. Importantly, the observed gap at $n = 0$ and large D occurs in the midst of this two-subband region, thus strongly suggesting that it originates from charge carrier correlations. The region boundaries are well described using an electrostatic model based on parabolic subbands (figs. S8 and S9, where we also discuss the observed asymmetry with respect to n).

The experimental findings are in line with the single-particle band structures shown in Fig. 2, F to H, at θ = 2.3° for different adjacent layer on-site energy differences Δ_V [for conversion of D to Δ_V and the slight difference in θ = 2.37° between theory and experiment, see (*30*)]. At $\Delta_V = 0$ (Fig. 2F), two subbands of the fourfold spin and valley degenerate bilayer graphene emerge near the κ_t and κ_b points, leading to a degeneracy $g = 8$. Changing Δ_V (or D) breaks the layer degeneracy. For finite Δ_V, single-band regions with $g = 4$ emerge (Fig. 2, G and H). Around charge neutrality, bands with opposite carrier type coexist and the bandgap is closed. In contrast to the two gaps that we observe in Fig. 2C, only the gap at $n = 0$ (band energy $E = 0$) and $\Delta_V = 0$ is captured by the single-particle band structure (Fig. 2F). This gap originates from a combination of crystal fields, which leads to electron transfer from the outer to the inner layers (*31*) and localized states in the moiré lattice (*32*). The absence of a gap at $n = 0$ and finite Δ_V in the single-particle band structure agrees with our previous suggestion that the experimental gap in this region emerges from electron-hole correlations.

We now discuss the dependence of the correlated gap on the parameters D and n, which change the sizes of the electron-hole Fermi surfaces. We show the two-terminal conductance G as a function of dc source-drain bias voltage V_{sd} in Fig. 3A and observe coherence peaks at $n = 0$ and large D, suggesting the formation of a coherent ground state. In agreement with thermal activation measurements, the gap size Δ increases with D. At $D = 0.47$ V/nm

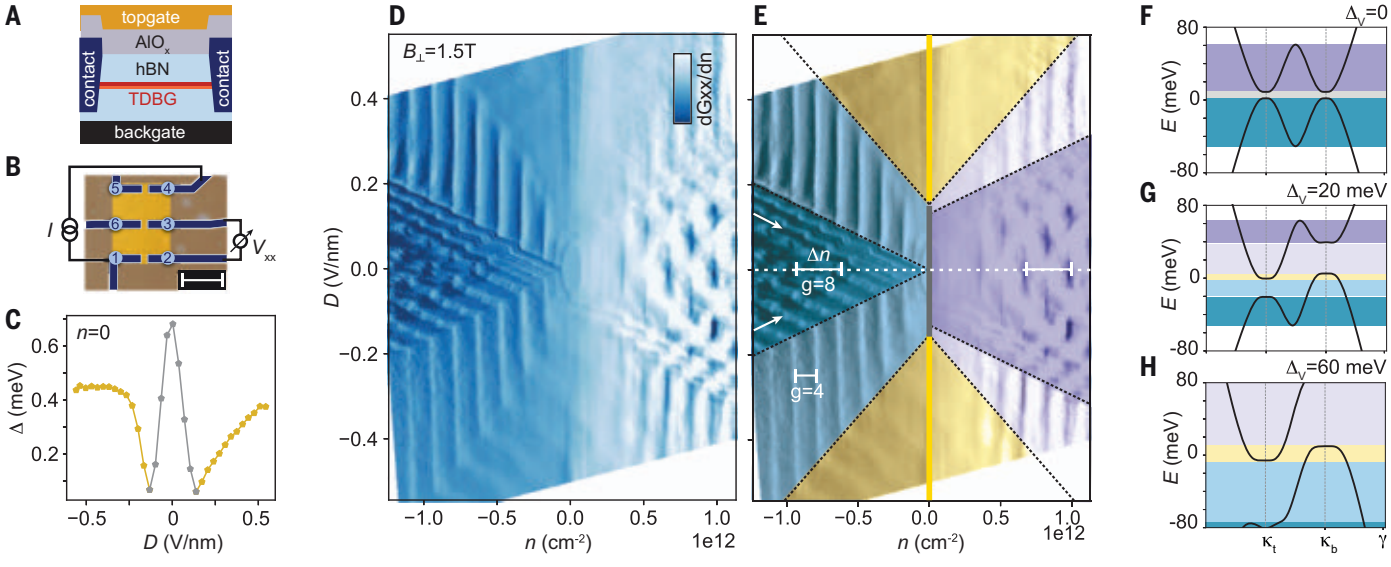


Fig. 2. Establishing electron-hole fluids. (**A** and **B**) Side view (A) and top view (B) of the stack. $R_{xx} \equiv V_{23}/I_{14}$, where V is the measured voltage drop and I is the applied current. Scale bar is 2 μm. hBN, hexagonal boron nitride. (**C**) From thermal activation measurements, we obtain Δ(D). Data highlighted with yellow color are taken in a regime where electron and hole bands coexist, whereas the gray colored data are taken for parameters where a single-particle gap is expected. (**D** and **E**) $\partial/\partial n\ (1/R_{xx})(D, n)$ at $B_\perp = 1.5$T and $T = 1.2$ K exhibits SdH oscillations that distinguish the regions highlighted in (E). In the yellow regions, electrons and holes coexist. The gray dashed lines mark the band edges at the κ_t and κ_b points. (**F** to **H**) Single-particle band structures for three values of $\Delta_V \propto D$.

and $n = 0$, $\Delta \approx 5$ meV (Δ is determined by the inflection points). We note that the thermal activation energy and the bias gap differ by an order of magnitude here. We speculate that this originates from complex thermal breakdown of the correlated gap and estimate that the bias measurement overestimates Δ by ~10% by additional series resistances (*30*). In Fig. 3B, we show the evolution of $\Delta(n)$ at $D = 0.47$ V/nm. Upon increasing n, the gap smears out and vanishes at $n \approx 2.5 \times 10^{11}$ cm^{-2}.

The correlated electron-hole gap vanishes with decreasing $|D|$ and increasing $|n|$. The first effect indicates the importance of effective mass and Fermi surface size, both of which increase with increasing $|D|$, as sketched in Fig. 3C. Conversely, the $|n|$ dependence suggests that an increasing asymmetry between electron and hole Fermi surfaces (blue and red) weakens the gap, as depicted in Fig. 3D. The observation that a correlated state emerges once Fermi surfaces match in size (at $n = 0$), and that the gap falters for small asymmetry, strongly suggests that the nesting of Fermi surfaces (i.e., contours in two dimensions) plays a crucial role for the formation of the correlated state.

We can probe this hypothesis by shifting the Fermi surfaces using a magnetic field. The concept of such a measurement in parallel field is depicted in Fig. 4A. At finite $B_\parallel$, the bands shift by the Zeeman energy. Now, nesting (equal Fermi surface) between opposite magnetic moments is possible at $n = 0$, whereas nesting between same magnetic moments is possible at $n \neq 0$. Indeed, the measured $R_{xx}(n, B_\parallel)$ traces are consistent with this concept (Fig. 4B). At $B_\parallel = 8$T, a shoulder in resistance at finite n is visible, consistent with a partially gapped band structure where half of the carriers form a correlated state while the other magnetic moment-bands remain conducting. The width of the shoulder (in n) increases with the Zeeman energy. How the region of enhanced resistance changes as a function of $|B_\parallel|$ and D can be observed in Fig. 4C. We can model the dependence on Zeeman energy by a basic model, yielding the dashed lines in Fig. 4C (for details, see fig. S10)

The behavior in perpendicular magnetic field $B_\perp$ is more complex. In Fig. 4D, we show $G(V_{sd}, D = 0.47\text{V/nm})$ traces for different $B_\perp$. We see that the gap is closed at a critical field $B_c = 4$T (orange trace) whose $B_c(D)$-dependence is plotted in the inset. In a semiclassical picture, the bands shift with $B_\perp$ owing to the valley-Zeeman effect. The valley g-factor in bilayer graphene $g_v \sim 20$ to 120 (*33*) has opposite signs in conduction and valence bands (see fig. S16 for details). $B_\perp$ therefore increases both electron and hole densities in one valley and decreases both in the other valley, as illustrated in Fig. 4E. If the correlated electron-hole gap is formed by wave function nesting, the gap in one valley (dashed bands–yellow gap) increases with $B_\perp$ and decreases in the other valley (solid bands–red gap). Thus, the completely gapped part of the band structure is decreasing with $B_\perp$, in agreement with the observations. In this picture, the gap opening for $B_\perp > B_c$ (green) can be interpreted as a single-particle gap at K' (green in Fig. 4E) and a correlated gap at K, i.e., as the formation of a valley-polarized correlated state.

The observations in magnetic field are consistent with basic models that assume Fermi surface nesting. We now confirm nesting of electron-hole surfaces by Hartree-Fock (HF) calculations where we incorporate the Coulomb potential and calculate the correlated gap Δ self-consistently (see fig. S13). Note that the moiré potential is not relevant for the emerging correlations because tunneling between top and bottom bilayers is weak at the intermediate twist in the relevant energy range. From HF calculations, we obtain the correlated bands, as shown in fig. S13, A and B. A gap is opened by electron-hole correlation and increases with adjacent layer on-site energy difference Δ_V. Both the gap size and the increase with displacement field agree with experiments, as shown in fig. S13C. We point out that this behavior is a result of decreasing static dielectric constant $\epsilon(\mathbf{q})$ with D (see fig. S20).

We note that Fermi surfaces in the same valley (e.g., κ_t and κ_b) are not perfectly circular even at $n = 0$, owing to interlayer tunneling between the middle two monolayers. However, the Fermi surfaces of opposite valleys (e.g., κ_t and κ'_b) match, suggesting that the correlated state is formed out of charge carriers from opposite valleys at $n = 0$. We show the calculated Fermi surfaces in fig. S14.

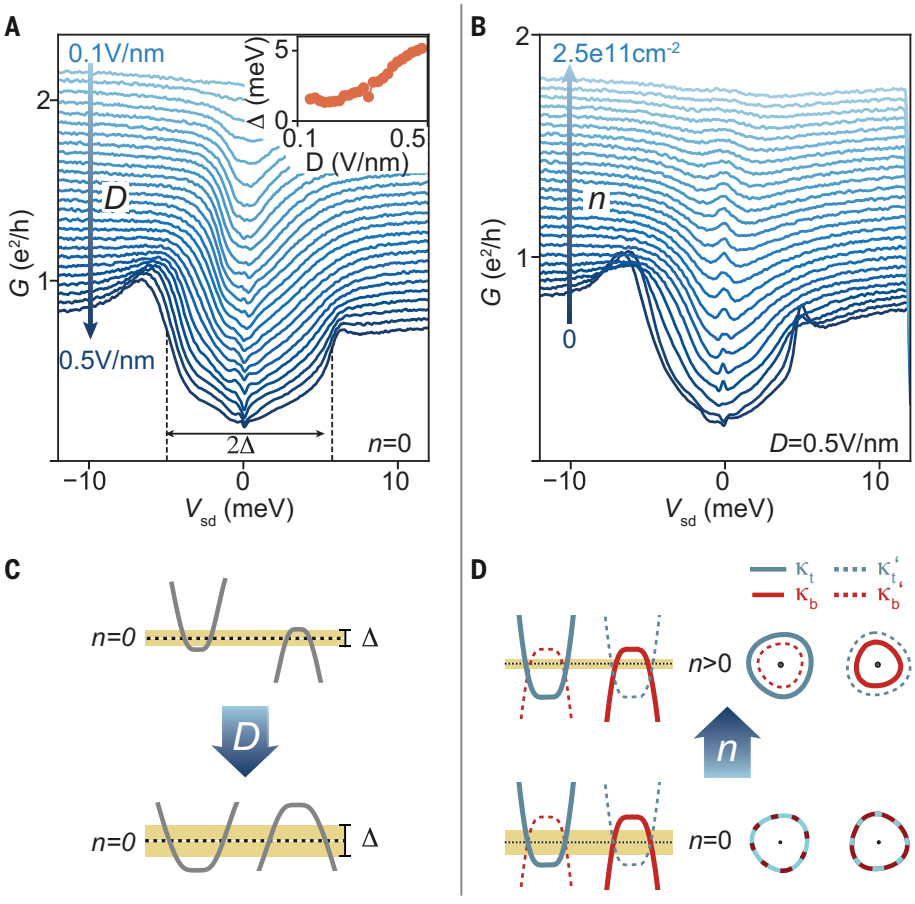


Fig. 3. Nesting. (**A**) Measurements of $G(V_{sd}, D)$ at $T = 0.1$ K reveal the appearance of a gap Δ. Curves are offset from the $D = 0.47$ V/nm trace. The inset shows extracted $\Delta(D)$, revealing that Δ increases with D. (**B**) $G(V_{sd}, n)$ at $D = 0.47$ V/nm showing that the gap vanishes with increasing n, where Fermi surface nesting falters. Curves are offset for clarity. (**C**) The top graphic shows a situation of finite D where electron and hole bands overlap and open a correlated gap Δ. By increasing D (bottom), Δ, the size of Fermi surfaces at $n = 0$, and masses m^* increase. (**D**) Calculated dispersion and Fermi surface around κ_t (blue) and κ_b (red) in the K (solid) and K' (dashed) valley, shown for a zero and finite n (for details, see fig. S14).

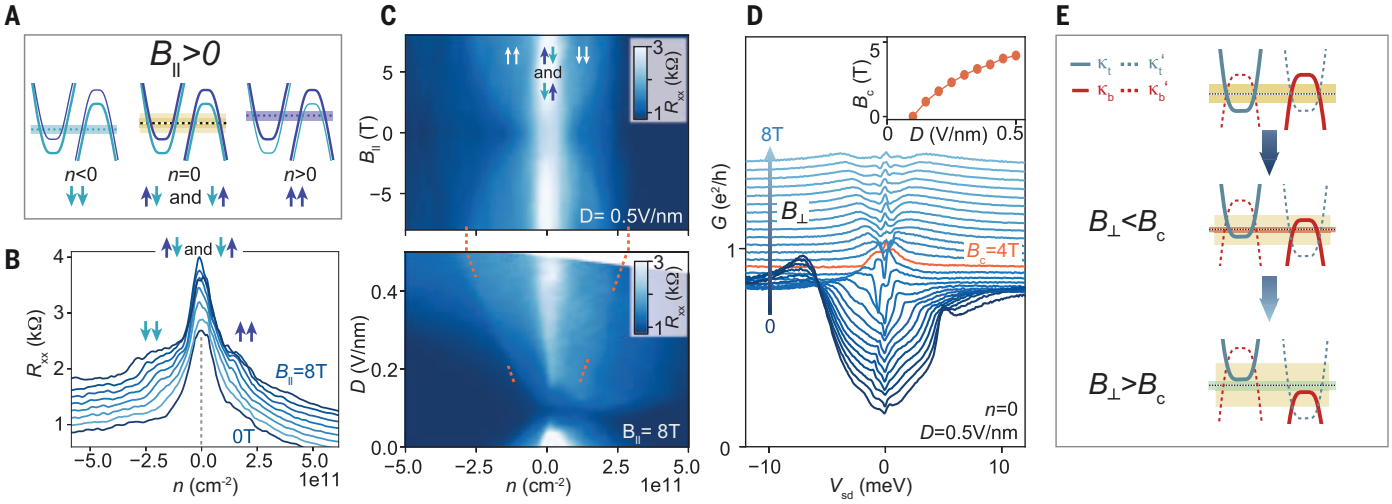


Fig. 4. Spin-polarized correlated gap. (**A**) Sketch of band structures under an applied $B_\parallel$ at different n. For a given $n < 0$ ($n > 0$), the turquoise (blue) bands with same magnetic moment have equal Fermi surface and thus open a gap, which is shaded turquoise (blue). At $n = 0$, bands with opposite magnetic moment (turquoise and blue) can form a gap (yellow-shaded region). The arrows indicate the magnetic moments of the charge carriers. (**B**) The partially gapped regions [turquoise-, yellow-, and blue-shaded regions in (A)] lead to an increase in R_{xx} with $B_\parallel$ at finite n. Curves are offset for clarity. (**C**) $R_{xx}(B_\parallel, n, D = 0.47\text{V/nm})$ and $R_{xx}(D, n, B_\parallel = 8\text{T})$ maps. The orange lines are calculated using $\Delta(D)$ and m^* from the experiment, without fitting parameters. (**D**) $G(V_{sd}, D = 0.5\text{V/nm})$ for increasing $B_\perp$. The gap is closed at $B_\perp = B_c$ (orange line). Curves are offset for clarity. In the inset, we plot the critical field, B_c, as a function of D. (**E**) The microscopic valleys K (solid) and K' (dashed) around κ_t (blue) and κ_b (red) shift in energy with $B_\perp$. For $B_\perp < B_c$, solid lines form a small gap (red-shaded region) and dashed lines a large gap (yellow-shaded region) corresponding to same-valley pairing at $n = 0$. For $B_\perp > B_c$, a single-particle gap (green-shaded region) is formed between the solid lines. For simplicity, Zeeman splitting is not sketched. All measurements in the figure are taken at $T = 0.1$ K.

Our model allows us to adjust the interlayer hopping parameter γ_1 between dimer sites in each bilayer to change m^*. The effective mass m^* generally depends on Δ_V, interlayer tunneling between the middle layers w, tunneling between the dimer sites within a Bernal bilayer γ_1, twist angle θ, and Fermi velocity v_F. If $w \neq 0$, changing w, θ, or Δ_V will change both m^* and layer polarization prob(κ_t). If $w = 0$, changing θ or Δ_V will have no effect on m^*, whereas changing γ_1 or v_F tunes m^* without any effect on layer polarizations. To investigate the effect of m^* solely on the correlated gap, we adjust the interlayer hopping parameter γ_1 (eq. S30) between the dimer sites in each bilayer. We observe an almost linear increase $\Delta(m^*)$ (see fig. S13D), supporting our previous statement that m^* determines the correlation strength (Fig. 1B). Theory does not show a strong dependence of pairing on the momentum difference between Fermi surfaces, only on the effective mass. This is in agreement with the excitonic character of a DW. In summary, the gap opening in HF calculations suggests that the correlated electron-hole state is consistent with a DW.

The correlated electron-hole state can be viewed as an excitonic insulator and is expected to exhibit counterflow superfluidity. Quantum phase transitions between DWs and disordered states can be controlled without chemical doping by varying n, or by varying D at $n = 0$, and such tunability could provide a new window on non–Fermi liquid physics.

REFERENCES AND NOTES

1. D. I. Khomskii, *Basic Aspects of The Quantum Theory of Solids* (Cambridge Univ. Press, 2010).
2. R. Peierls, *Ann. Phys.* **396**, 121–148 (1930).
3. Z. Li *et al.*, *Nano Lett.* **19**, 4960–4964 (2019).
4. D. Jérome, T. M. Rice, W. Kohn, *Phys. Rev.* **158**, 462–475 (1967).
5. A. Kogar *et al.*, *Science* **358**, 1314–1317 (2017).
6. I. B. Spielman, J. P. Eisenstein, L. N. Pfeiffer, K. W. West, *Phys. Rev. Lett.* **84**, 5808–5811 (2000).
7. L. Du *et al.*, *Nat. Commun.* **8**, 1971 (2017).
8. Z. Wang *et al.*, *Nature* **574**, 76–80 (2019).
9. H. Deng, H. Haug, Y. Yamamoto, *Rev. Mod. Phys.* **82**, 1489–1537 (2010).
10. J. Chang *et al.*, *Nat. Phys.* **8**, 871–876 (2012).
11. Y. Cao *et al.*, *Phys. Rev. Lett.* **117**, 116804 (2016).
12. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
13. A. L. Sharpe *et al.*, *Science* **365**, 605–608 (2019).
14. M. Yankowitz *et al.*, *Science* **363**, 1059–1064 (2019).
15. P. Stepanov *et al.*, *Nature* **583**, 375–378 (2020).
16. Y. Saito, J. Ge, K. Watanabe, T. Taniguchi, A. F. Young, *Nat. Phys.* **16**, 926–930 (2020).
17. M. Koshino, *Phys. Rev. B* **99**, 235406 (2019).
18. N. R. Chebrolu, B. L. Chittari, J. Jung, *Phys. Rev. B* **99**, 235417 (2019).
19. Y. W. Choi, H. J. Choi, *Phys. Rev. B* **100**, 201402 (2019).
20. X. Liu *et al.*, *Nature* **583**, 221–225 (2020).
21. C. Shen *et al.*, *Nat. Phys.* **16**, 520–525 (2020).
22. G. W. Burg *et al.*, *Phys. Rev. Lett.* **123**, 197702 (2019).
23. M. He *et al.*, *Nat. Phys.* **17**, 26–30 (2021).
24. A. Luican *et al.*, *Phys. Rev. Lett.* **106**, 126802 (2011).
25. J. D. Sanchez-Yamagishi *et al.*, *Phys. Rev. Lett.* **108**, 076601 (2012).
26. G. W. Burg *et al.*, *Phys. Rev. Lett.* **120**, 177702 (2018).
27. F. K. de Vries *et al.*, *Phys. Rev. Lett.* **125**, 176801 (2020).
28. R. Bistritzer, A. H. MacDonald, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12233–12237 (2011).
29. P. Rickhaus *et al.*, *Sci. Adv.* **6**, eaay8409 (2020).
30. See materials and methods in the supplementary materials.
31. P. Rickhaus *et al.*, *Nano Lett.* **19**, 8821–8828 (2019).
32. F. J. Culchac, R. R. Del Grande, R. B. Capaz, L. Chico, E. S. Morell, *Nanoscale* **12**, 5014–5020 (2020).
33. Y. Lee *et al.*, *Phys. Rev. Lett.* **124**, 126802 (2020).
34. P. Rickhaus, Script repository: Electron-hole correlated state in twisted double bilayer graphene. ETH Library, ETH Zürich (2021); http://doi.org/10.5905/ethz-1007-405.

ACKNOWLEDGMENTS

Funding: We acknowledge financial support from the European Graphene Flagship and the Swiss National Science Foundation through NCCR Quantum Science and the ERC Syngery Grant Quantropy. P.R. acknowledges financial support from the ETH Fellowship program. Growth of hexagonal boron nitride crystals was supported by the Elemental Strategy Initiative conducted by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, and the Core Research for Evolutional Science and Technology (CREST) (JPMJCR15F3), Japan Science and Technology Agency (JST). A.H.M. and J.Z. were supported by the National Science Foundation (NSF) through the Center for Dynamics and Control of Materials, an NSF Materials Research Science and Engineering Center (MRSEC) under cooperative agreement no. DMR-1720595; and by the Welch Foundation under grant TBF1473. **Author contributions:** P.R. fabricated the devices; P.R., F.K.d.V., E.P., and G.Z. performed measurements; and P.R., F.K.d.V., E.P., G.Z., M.M., and A.K. were involved in data analysis. T.T. and K.W. provided hexagonal boron nitride crystals. P.R. wrote the manuscript, and P.R., F.K.d.V., E.P., G.Z., J.Z., A.H.M., T.I., and K.E. were involved in reviewing the manuscript. J.Z. performed simulations. A.H.M., T.I., and K.E. supervised the work. **Competing interests:** The authors have no competing interests. **Data and materials availability:** All data and code are available at ETH Library (*34*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abc3534
Materials and Methods
Supplementary Text
Figs. S1 to S20
References (*35*–*41*)

24 April 2020; accepted 11 August 2021
10.1126/science.abc3534

OPTICS

Coherent perfect absorption at an exceptional point

Changqing Wang[1], William R. Sweeney[2,3], A. Douglas Stone[2,3], Lan Yang[1]*

Recently, exceptional points, a degeneracy of open wave systems, have been observed in photonics, acoustics, and electronics. They have mainly been realized as a degeneracy of resonances; however, a degeneracy associated with the absorption of waves can exhibit distinct and interesting physical features. Here, we demonstrate such an absorbing exceptional point by engineering degeneracies in the absorption spectrum of dissipative optical microcavities. We experimentally distinguished the conditions to realize an absorbing exceptional point versus a resonant exceptional point. Furthermore, when the optical loss was tuned to achieve perfect absorption at an absorbing exceptional point, we observed its signature, an anomalously broadened line shape in the absorption spectrum. The distinct scattering properties of the absorbing exceptional point create opportunities for both fundamental study and applications of non-Hermitian degeneracies.

The field of non-Hermitian physics provides a fertile ground for studying unconventional physical phenomena in open systems (*1*, *2*). Many of the most intriguing phenomena occur at exceptional points (EPs), singularities in the parameter space of a wave system at which two or more discrete eigenvalues and the associated eigenstates of a non-Hermitian system coalesce (*3*). The most familiar experimental examples are EPs associated with the degeneracies of the resonance frequencies of wave systems corresponding to the discrete, purely outgoing solutions of the relevant wave equations. Such resonant EPs, which often are described within the temporal coupled-mode theory (TCMT) (*4*) but exist in exact theories as well (*5*–*7*), lead to interesting physical effects such as enhanced sensitivity as a parameter is swept through an EP (*8*–*11*), nontrivial topological phenomena (*12*–*16*), and excess spontaneous emission noise (*6*, *17*). Moreover, because the onset of laser emission corresponds to a resonance at a real frequency, an EP for two resonances at or very near the real axis leads to unusual lasing behavior, including chiral laser emission (*4*, *18*), laser linewidth broadening (*17*, *19*), single-mode lasing (*20*, *21*), lasing and/or perfect absorption switching (*22*), etc. Except for the case of lasing, the resonant EPs involved are described by a complex frequency, but they are observed by their effects under quasi–steady-state (real frequency) excitation (*23*, *24*). Because the onset of lasing leads to a linear instability that is stabilized by nonlinear saturation, it is difficult to study the purely linear behavior near such a lasing EP experimentally.

A different class of discrete complex frequency wave solutions arise from purely incoming boundary conditions, which are referred to as scattering zeros. It has been shown that the analog of adding gain to achieve lasing was to add absorption to the same structure to make the incoming solution occur at real frequency (*25*). Such a structure is known as a coherent perfect absorber (CPA), the time-reversed version of a laser at threshold, and the input wave front must be tuned to be the incoming version of the corresponding lasing mode to be fully absorbed (*26*, *27*). Unlike lasing and resonance, there is no intrinsic instability or divergence associated with a CPA. Following the analogy to resonant EPs, it was recently proposed (*28*) that the eigenfrequencies of two incoming solutions of a wave equation could become degenerate, leading to an absorbing EP or EP zero (generally complex) or to a CPA EP when this occurs at a real frequency. A number of interesting phenomena distinct from resonant EPs were predicted for such absorbing EPs.

Here, we studied absorbing EPs in a system of coupled optical microcavities associated with perfect absorption, which has no instability at real frequency and presents interesting EP-related behavior not previously observed. This was achieved by distinguishing the roles of absorption loss and nondissipative scattering and coupling loss. Although increasing both types of loss causes resonances to move away from the real frequency axis, scattering loss does this for the zeros and absorption loss does the opposite.

Following (*28*), we consider a model of coupled standing-wave slab cavities with input and output channels on both sides (Fig. 1A). The reflection ($r_{1,2}$) and transmission ($t_{1,2}$) for such a two-channel scattering system lead to a 2×2 *S*-matrix $S = \begin{pmatrix} r_1 & t_2 \\ t_1 & r_2 \end{pmatrix}$. We constructed an equivalent model with a different geometry through coupled whispering gallery mode (WGM) microcavities, $\mu R_{1,2}$, with resonant frequencies $\omega_{1,2}$, respectively (Fig. 1B). This was realized experimentally by coupling two silica microtoroid resonators fabricated on the edges of two silicon chips. The intercavity coupling strength (κ) was tuned by adjusting the gap between the microtoroids, which determines the overlapping evanescent fields of WGMs (fig. S1). The two cavities are coupled to two single-mode fiber taper waveguides with the coupling strengths γ_{c1} and γ_{c2}, respectively. The intrinsic losses $\gamma_{1,2}$ are dominated by the absorption loss from the material and surface impurities, whereas the radiation and scattering loss was negligible in comparison. Transmission here is defined as the ratio of light passing through the coupled resonators (1→4, 3→2) and reflection as remaining in the same waveguide as the input wave (1→2, 3→4). Because of the negligible backscattering of this structure, only the clockwise mode in μR_1 and the counterclockwise mode in μR_2 are excited as long as only ports 1 and 3 are used for input. In the experiment, at most two pairs of resonances and zeros will be probed at a time. In this case, the system can be described by TCMT with an effective Hamiltonian (see supplementary text S1)

$$H_{\text{eff}} = \begin{pmatrix} \omega_1 - i\frac{\gamma_1 + \gamma_{c1}}{2} & \kappa \\ \kappa & \omega_2 - i\frac{\gamma_2 + \gamma_{c2}}{2} \end{pmatrix} \quad (1)$$

The corresponding 2×2 *S*-matrix in the frequency domain is given by

$$S(\omega) = \begin{pmatrix} 1 - i\frac{\gamma_{c1}\Delta_2}{\Delta_1\Delta_2 - \kappa^2} & -i\frac{\sqrt{\gamma_{c1}\gamma_{c2}}\kappa}{\Delta_1\Delta_2 - \kappa^2} \\ -i\frac{\sqrt{\gamma_{c1}\gamma_{c2}}\kappa}{\Delta_1\Delta_2 - \kappa^2} & 1 - i\frac{\gamma_{c2}\Delta_1}{\Delta_1\Delta_2 - \kappa^2} \end{pmatrix} \quad (2)$$

where $\Delta_{1,2} = \omega - \omega_{1,2} + i\frac{\gamma_{1,2}+\gamma_{c1,2}}{2}$. *S* is a symmetric matrix due to the reciprocity of light transport in such structures. To simplify our analysis, we choose $\omega_1 = \omega_2 = \omega_0$ in the following discussion.

Generically, at any complex frequency ω, two complex eigenvalues $\lambda_{1,2}$ can be found for *S* and the corresponding two-by-one eigenvectors $v_{1,2}$ (satisfying $Sv_{1,2} = \lambda_{1,2}v_{1,2}$) define the two eigenchannels, representing specific superpositions of physical channels for which the waveform remains invariant during the scattering processes. Generically, at a resonance of *S*, one eigenvalue of *S* diverges, and at a zero of *S*, one eigenvalue vanishes; in each case, the other eigenvalue is finite. However, at a resonance, each element of the *S*-matrix diverges, which, as noted, makes experimental study complicated. In Fig. 1, C to G, we illustrate the different kinds of EPs that can arise through parameter tuning in this model. In lossless (no absorption) cavities (Fig. 1C), both the poles and zeros are simultaneously degenerate at an EP with the coupling $\kappa = \left|\frac{\gamma_{c1} - \gamma_{c2}}{4}\right|$. Once absorption is added, the pole

[1]Department of Electrical and Systems Engineering, Washington University, St. Louis, MO 63130, USA. [2]Departments of Applied Physics and Physics, Yale University, New Haven, CT 06520, USA. [3]Yale Quantum Institute, Yale University, New Haven, CT 06520, USA.
*Corresponding author. Email: yang@seas.wustl.edu

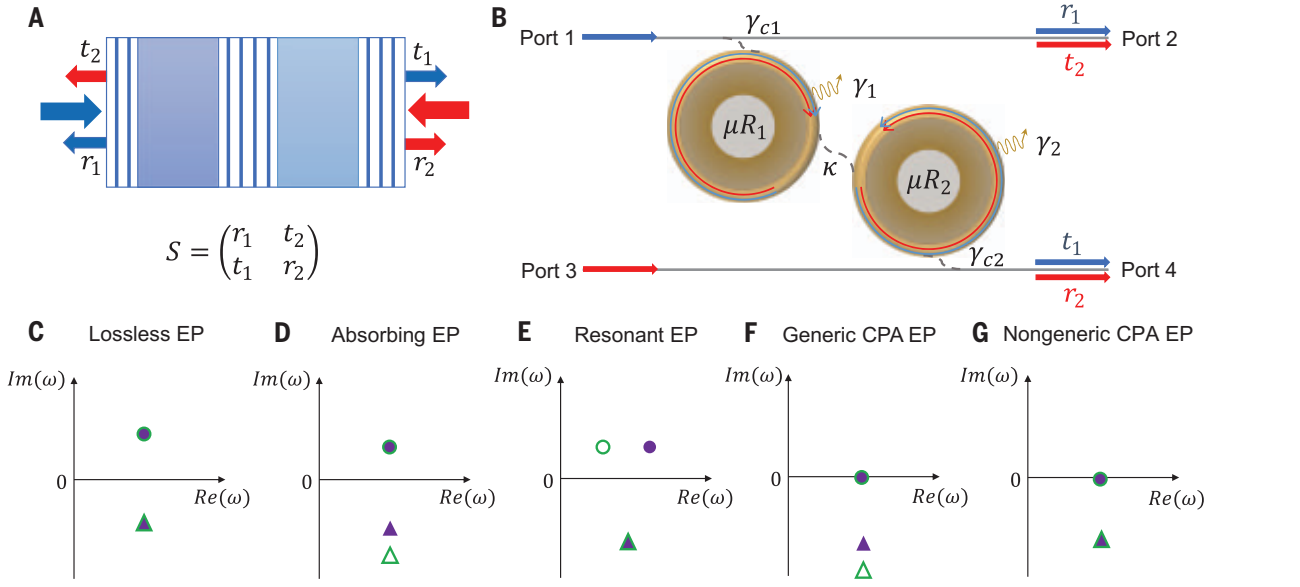


Fig. 1. Scattering properties of the coupled microcavity systems. (**A**) Scattering from two one-dimensional cavities formed by external Bragg mirrors coupled through a central, partially reflecting Bragg mirror, studied in (*28*). The scattered outputs from the incident waves in the left and right channels determine the *S*-matrix. (**B**) Schematic diagram of the analogous model to (A) realized with coupled WGM microcavities (μR_1 and μR_2), with double fiber taper waveguides as the input and output channels. (**C** to **G**) Zeros (circles) and poles (triangles) of the system's *S*-matrix in the complex frequency plane for different types and configurations of EPs. (C) EP in a system without absorption or gain. (D) Absorbing EP. (E) Resonant EP. (F) Generic CPA EP. (G) Nongeneric CPA EP.

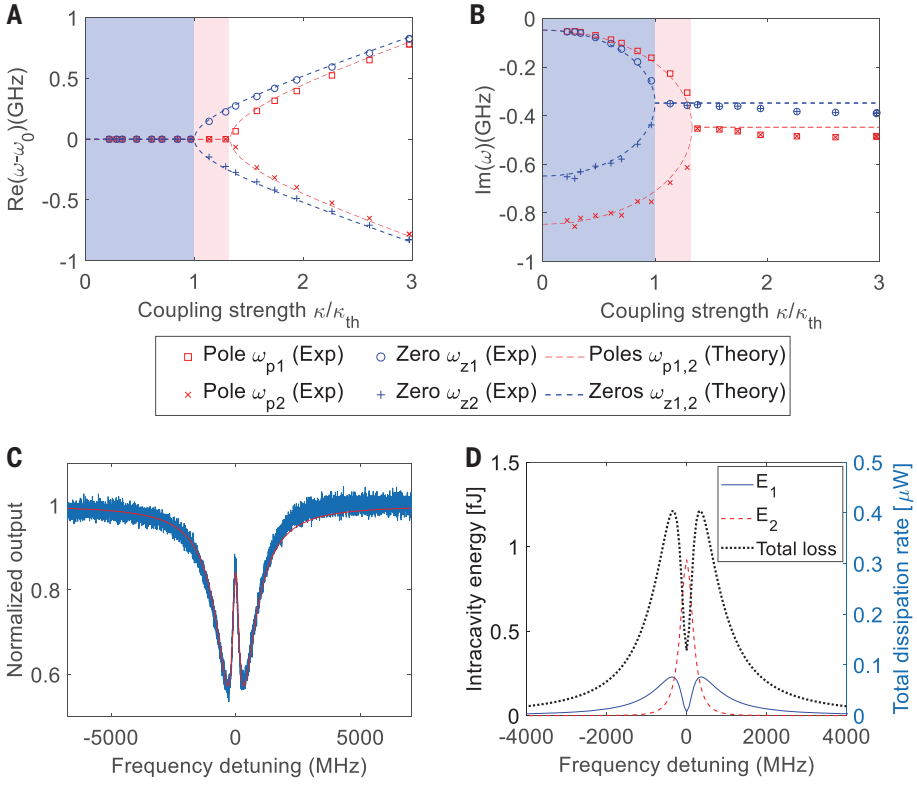


Fig. 2. Characterization of resonant EPs and absorbing EPs. (**A** and **B**) Phase transition diagrams for the real parts (A) and imaginary parts (B) of the poles and zeros as a function of normalized coupling strength κ/κ_{th}. "Exp"/"Theory" refer to experimental/theoretical results. The second waveguide is decoupled from μR_2 (γ_{c2} = 0). The parameters γ_1, γ_2, γ_{c1}, and κ were extracted by curve fitting the data to $|r_1|^2$ (see eq. S41 in supplementary text S6), and then used to derive the complex poles (see eq. S10 in supplementary text S2) and zeros (see eq. S17 in supplementary text S3). (**C**) Normalized output spectrum near an absorbing EP. The blue and red curves are experimental and curve fitting results, respectively. (**D**) Simulation results of the energy distribution in μR_1 (E_1) and μR_2 (E_2) and the total dissipation rate as a function of the frequency detuning at the absorbing EP. Input power: 1 μW.

and zero frequencies are not complex conjugates; one can realize a resonant EP when $\kappa = \left|\frac{(\gamma_{c1}-\gamma_{c2})+(\gamma_1-\gamma_2)}{4}\right|$ and an EP zero when $\kappa = \left|\frac{(\gamma_{c1}-\gamma_{c2})-(\gamma_1-\gamma_2)}{4}\right|$ (supplementary text S2 and S3). This implies that as the coupling varies, the EP zero will occur before the resonant EP if the differential coupling has the same sign as the differential absorption between the two cavities (Fig. 1, D and E) and vice versa if they have opposite signs. The case of special interest is when the two zeros meet on the real axis, corresponding to steady-state perfect absorption at an EP, i.e., a CPA EP (Fig. 1F). This happens, within TCMT, at the critical coupling

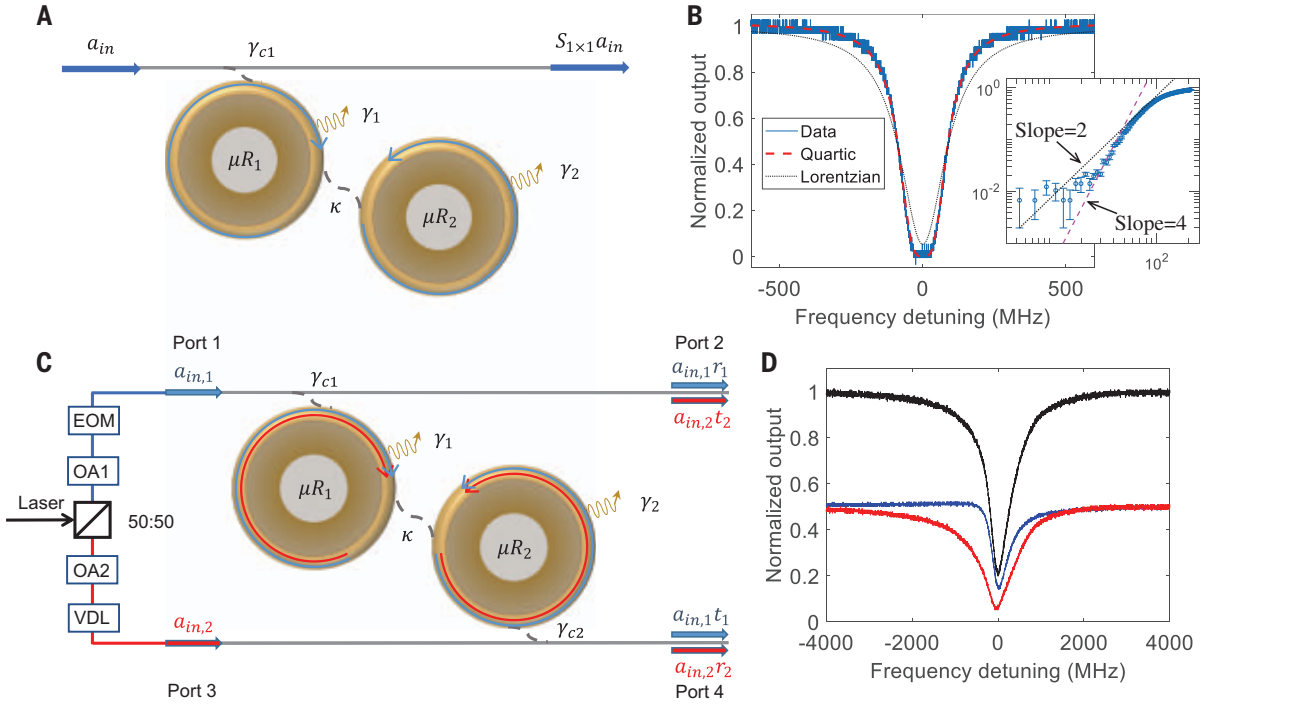


Fig. 3. CPA EPs. (**A**) Schematic diagram of the experimental setup for a one-channel CPA EP. The *S*-matrix is reduced to one element, $S_{1\times1}$. (**B**) Experimentally obtained normalized output spectrum (blue) at the one-channel CPA EP and the curve-fitting results using quartic (red dashed) and Lorentzian (black dotted) functions. The inset displays a double-logarithmic plot of the output versus the frequency detuning. (**C**) Schematic diagram of the experimental setup for a two-channel CPA EP. The magnitudes of the two inputs are controlled by two variable optical attenuators (OA1 and OA2). The relative phase of the two inputs is controlled by an electro-optic phase modulator (EOM). The lengths of the two optical paths are balanced by a variable optical delay line (VDL). (**D**) Experimentally obtained spectra of the output from port 2 (blue curve), the output from port 4 (red curve), and the total output (black curve) normalized to the total input power at the nongeneric CPA EP. To excite the system by the eigenvector of *S*, the amplitudes and relative phase of the two input fields are adjusted by OA1, OA2, and EOM.

condition, when the total intrinsic loss $\gamma_1 + \gamma_2$ is balanced by the total radiative coupling loss $\gamma_{c1} + \gamma_{c2}$, with $\kappa = \left|\frac{\gamma_{c1}-\gamma_1}{2}\right|$ (supplementary text S4). In general, this is referred to as a generic CPA EP, which leads to an anomalous quartic absorption line shape (*28*). However, if the coupling rates are equal ($\gamma_{c1} = \gamma_{c2}$), then both the zeros and poles coalescence at the same value of κ within the TCMT model and we realize simultaneous resonant and absorbing EPs (Fig. 1G and supplementary text S5); this is referred to as a nongeneric CPA EP. The resulting singular behavior of the *S*-matrix leads to a distinct scattering behavior from the generic case to be discussed and demonstrated experimentally below.

Here, we experimentally realize and compare the cases illustrated in Fig. 1, D to G, for the resonant and absorbing EPs in the setup shown in Fig. 1B. Initially, for simplicity, we study the case with the bottom waveguide decoupled from μR_2, which implies, in the *S*-matrix terminology, that only reflection or absorption of the input is possible. The resonant frequencies are aligned by the temperature control of μR_2 through a thermoelectric cooler. The reflection spectrum is measured to derive the complex poles and the zeros. The resonant EP and absorbing EP are distinguished by the phase transitions of the complex poles and zeros when the intercavity coupling strength κ is varied (Fig. 2, A and B). For the parameter scenario here, the zeros meet at a lower coupling, $\kappa = \kappa_{th}$, and then move apart; the coupling is increased further and then the poles meet at $\kappa \approx 1.32\kappa_{th}$. For each case, after the transition, the zeros and poles no longer have the same real part of the frequency but instead share approximately the same imaginary part. This clearly confirms that resonant EPs and absorbing EPs generally occur at different coupling strengths (or equivalently, different differential loss rates). In the alternative scenario, the absorbing EP occurs after the appearance of resonant EP as κ increases (fig. S2).

It is worth noting that the single dip or peak line shape of the output signal spectrum is often regarded as a signature of resonant EPs. Particularly, in a single cavity with coupled clockwise and counterclockwise modes (*8*) or parity-time (PT) symmetric coupled microcavities (*9*), the transmission spectrum evolves from a doublet to a single dip at EPs. This is correct for systems without absorption (or gain); however, for general non-Hermitian systems, EPs may not be accompanied by a singlet in the transmission and reflection spectra. This was confirmed experimentally by the spectra of the scattered signal at the absorbing EP (Fig. 2C) and the resonant EP (fig. S3), which both exhibit an absorbing doublet. This is attributed to the fact that the zeros and poles are not complex conjugate pairs and do not coalesce simultaneously, so that the minimum output is not reached at the frequencies $Re(\omega_{z1,2})$ or $Re(\omega_{p1,2})$ (see supplementary text S7). This observation can also be explained by the simulated field redistribution versus the frequency detuning (δ) at the absorbing EP (Fig. 2D). The total dissipation reaches its maximum at nonzero detuning, where the laser frequency is not equal to $Re(\omega_{z1,2})$ or $Re(\omega_{p1,2})$.

We next study CPA EPs, where the two degenerate zeros occur at a real frequency. With the scheme in Fig. 3A, we realize a one-channel generic CPA EP (equivalent to critical coupling at an EP) by tuning the cavity-waveguide coupling so that $\gamma_{c1} = 119.2814$ MHz $\approx \gamma_1 + \gamma_2$ and adjusting the intercavity gap to achieve an

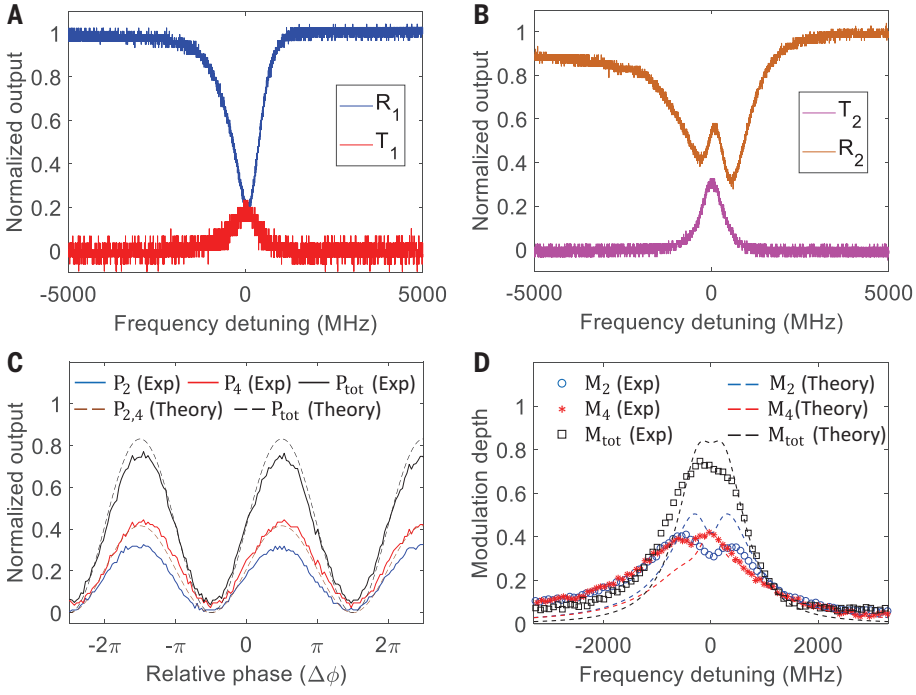


Fig. 4. Scattering properties and phase modulation at the two-channel CPA EP. (**A**) Experimentally measured spectra of $R_1 = |r_1|^2$ and $T_1 = |t_1|^2$ by exciting port 1 only. (**B**) Experimentally measured spectra of $R_2 = |r_2|^2$ and $T_2 = |t_2|^2$ by exciting port 3 only. (**C**) Experimentally obtained output power from port 2 (P_2) and port 4 (P_4) at the zero detuning ($\delta = \omega - \omega_0 = 0$) normalized to the total input power as a function of the relative phase. P_2 and P_4 are equal in TCMT. The total normalized output power (P_{tot}) is obtained by adding P_2 and P_4. (**D**) Modulation depth, the difference between the maximum and the minimum power normalized to the total input power under varying relative phase, of the output from port 2 (M_2) and the output from port 4 (M_4) versus frequency detuning. The modulation depth of the total output power (M_{tot}) is not the sum of M_2 and M_4 because at each frequency detuning, the relative phases that maximize (or minimize) the two outputs can be different.

absorbing EP ($\kappa = \left|\frac{\gamma_{c1} + \gamma_2 - \gamma_1}{4}\right|$). As predicted, the measured spectrum is well fit by a quartic function, which verifies the anomalous broadband absorption line shape for CPA EPs (Fig. 3B and supplementary text S8). By comparison, the Lorentzian curve-fitting result based on the measured γ_{c1} shows a large deviation from the data (Fig. 3B). The quartic relation is further confirmed by a slope of 4 in the logarithmic plot (inset of Fig. 3B).

To study two-channel CPA EPs, we restored the coupling to the second waveguide (Fig. 3C). A nongeneric CPA EP can be more easily found because it occurs simultaneously with a resonant EP (supplementary text S5). Thus, in the experiment, we set $\gamma_{c1} = \gamma_{c2} = (\gamma_1 + \gamma_2)/2$, and then reposition μR_2 in parallel to the waveguide to find the resonant EP. We confirmed experimentally the prediction of a universal form of S at a nongeneric CPA EP, with identical amplitudes in the four elements (Fig. 4, A and B), consistent with the theory (fig. S4) (*28*). $|r_1|^2$ and $|t_1|^2$ are found to be almost equal at $\delta = 0$, whereas $|r_2|^2$ and $|t_2|^2$ approach each other with a small discrepancy because of imperfections such as backscattering effects in the cavity modes. Furthermore, the perfect absorption of the input wave is achieved by coupling the input into the absorbing eigenchannel defined by the eigenstate of S ($[1,-i]^T$). With balanced input power and the optimized relative phase (*29*), strong absorption is observed experimentally (Fig. 3D; for simulation results, see fig. S5); the nonideal perfect absorption is caused by the fluctuation of optical phases and coupling parameters. Moreover, the total absorption line shape for the eigenchannel input is quadratic instead of quartic because the eigenvalues of S are not analytic around the zero detuning (see supplementary text S8).

The degree of absorption in the case of two-channel CPA is sensitive to the relative phase of the inputs and should be maximal when the phase difference corresponds to the relevant eigenchannel of S. To characterize the phase response of our system operating at a CPA EP, we measure the output power from port 2 and port 4 around zero detuning normalized to the total input power (P_2 and P_4) as a function of the relative phase $\Delta\phi$ (figs. S6 and S7) (*29*). The two output signals oscillate in phase with $\Delta\phi$, consistent with the theoretical prediction, and add up to the total output (P_{tot}) with a large oscillation amplitude (Fig. 4C). Furthermore, the modulation depth, defined as the difference between the maximum and minimum output power with varying $\Delta\phi$ normalized to the total input power, is measured versus the detuning. It was observed that the modulation depth of the output from port 2 (M_2) behaved differently from that from port 4 (M_4) (Fig. 4D). This is because the input vectors that maximize and minimize the output power at a nonzero detuning are not always the eigenstate of S, and thus the feature of the nongeneric CPA EP is hidden. Furthermore, the spectrum of the total output modulation depth (P_{tot}) displayed a flattened window with $P_{tot} > 0.8$ in a frequency range above 500 MHz (comparable with γ_c), which indicates a broadband large phase sensitivity at the CPA EP.

Our study suggests the importance of properly choosing the input to reveal the unique scattering properties of EPs. This is critical for non-Hermitian applications such as EP-enhanced sensing, where the amplified spectral splitting can only be observed with a proper probing scheme. Moreover, the absorbing EPs can be potentially found in various non-Hermitian systems including metamaterials and metasurfaces, acoustic systems, electrical circuits, and, more interestingly, open quantum systems. It has been pointed out that realizing PT-symmetric quantum optics is nontrivial because optical gain will unavoidably introduce additional noise that breaks the PT symmetry (*30*). This hinders the exploration of resonant EPs through emission features in quantum systems. However, by engineering the quantum dissipation, the absorbing EPs can be realized without the need for gain and therefore can be potentially exploited to engineer scattering properties of nonclassical light.

REFERENCES AND NOTES

1. L. Feng, R. El-Ganainy, L. Ge, *Nat. Photonics* **11**, 752–762 (2017).
2. R. El-Ganainy *et al.*, *Nat. Phys.* **14**, 11–19 (2018).
3. M.-A. Miri, A. Alù, *Science* **363**, eaar7709 (2019).
4. B. Peng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6845–6850 (2016).
5. Y. D. Chong, L. Ge, A. D. Stone, *Phys. Rev. Lett.* **106**, 093902 (2011).
6. A. Pick *et al.*, *Opt. Express* **25**, 12325–12348 (2017).
7. R. Fleury, D. Sounas, A. Alù, *Nat. Commun.* **6**, 5905 (2015).
8. W. Chen, Ş. Kaya Özdemir, G. Zhao, J. Wiersig, L. Yang, *Nature* **548**, 192–196 (2017).
9. H. Hodaei *et al.*, *Nature* **548**, 187–191 (2017).
10. Y.-H. Lai, Y.-K. Lu, M.-G. Suh, Z. Yuan, K. Vahala, *Nature* **576**, 65–69 (2019).
11. M. P. Hokmabadi, A. Schumer, D. N. Christodoulides, M. Khajavikhan, *Nature* **576**, 70–74 (2019).
12. H. Xu, D. Mason, L. Jiang, J. G. E. Harris, *Nature* **537**, 80–83 (2016).
13. J. Doppler *et al.*, *Nature* **537**, 76–79 (2016).
14. J. W. Yoon *et al.*, *Nature* **562**, 86–90 (2018).
15. W. Tang *et al.*, *Science* **370**, 1077–1080 (2020).
16. H. Zhou *et al.*, *Science* **359**, 1009–1012 (2018).
17. H. Wang, Y. H. Lai, Z. Yuan, M. G. Suh, K. Vahala, *Nat. Commun.* **11**, 1610 (2020).
18. P. Miao *et al.*, *Science* **353**, 464–467 (2016).
19. J. Zhang *et al.*, *Nat. Photonics* **12**, 479–484 (2018).

20. L. Feng, Z. J. Wong, R. M. Ma, Y. Wang, X. Zhang, *Science* **346**, 972–975 (2014).
21. H. Hodaei, M. A. Miri, M. Heinrich, D. N. Christodoulides, M. Khajavikhan, *Science* **346**, 975–978 (2014).
22. Z. J. Wong *et al*., *Nat. Photonics* **10**, 796–801 (2016).
23. C. Shi *et al*., *Nat. Commun.* **7**, 11110 (2016).
24. C. Wang *et al*., *Nat. Phys.* **16**, 334–340 (2020).
25. Y. D. Chong, L. Ge, H. Cao, A. D. Stone, *Phys. Rev. Lett.* **105**, 053901 (2010).
26. D. G. Baranov, A. Krasnok, T. Shegai, A. Alù, Y. Chong, *Nat. Rev. Mater.* **2**, 17064 (2017).
27. W. Wan *et al*., *Science* **331**, 889–892 (2011).
28. W. R. Sweeney, C. W. Hsu, S. Rotter, A. D. Stone, *Phys. Rev. Lett.* **122**, 093901 (2019).
29. Materials and methods are available as supplementary materials.
30. S. Scheel, A. Szameit, *Europhys. Lett.* **122**, 34001 (2018).

ACKNOWLEDGMENTS

Funding: This work was supported by the National Science Foundation (NSF grant no. EFMA1641109). A.D.S. was supported by NSF Condensed Matter and Materials Theory (CMMT) grant DMR-1743235. C.W. acknowledges fellowship support through the McDonnell International Scholars Academy. **Author contributions:** L.Y. and A.D.S. conceived the joint project. C.W. and L.Y. designed the experiments with help from W.R.S. and A.D.S. C.W. performed the experiments and analyzed experimental data with help from W.R.S. C.W. and W.R.S. provided theoretical background and simulations. All authors discussed the results and wrote the manuscript. L.Y. supervised the project.

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

https://science.org/doi/10.1126/science.abj1028
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*31*–*73*)

25 April 2021; accepted 11 August 2021
10.1126/science.abj1028

OLIGONUCLEOTIDES

A P(V) platform for oligonucleotide synthesis

Yazhong Huang[1]†, Kyle W. Knouse[2,3]†, Shenjie Qiu[4]†, Wei Hao[2]‡, Natalia M. Padial[2]‡§, Julien C. Vantourout[2]‡, Bin Zheng[4], Stephen E. Mercer[1], Javier Lopez-Ogalla[2]¶, Rohan Narayan[2], Richard E. Olson[1], Donna G. Blackmond[2], Martin D. Eastgate[4]*, Michael A. Schmidt[4]*, Ivar M. McDonald[1]*, Phil S. Baran[2]*

The promise of gene-based therapies is being realized at an accelerated pace, with more than 155 active clinical trials and multiple U.S. Food and Drug Administration approvals for therapeutic oligonucleotides, by far most of which contain modified phosphate linkages. These unnatural linkages have desirable biological and physical properties but are often accessed with difficulty using phosphoramidite chemistry. We report a flexible and efficient [P(V)]–based platform that can install a wide variety of phosphate linkages at will into oligonucleotides. This approach uses readily accessible reagents and can install not only stereodefined or racemic thiophosphates but any combination of (*S*, *R* or *rac*)–PS with native phosphodiester (PO_2) and phosphorodithioate (PS_2) linkages into DNA and other modified nucleotide polymers. This platform easily accesses this diversity under a standardized coupling protocol with sustainably prepared, stable P(V) reagents.

In traditional small-molecule drug discovery, the features that determine target specificity and DMPK (distribution, metabolism, and pharmacokinetics) are usually inextricably linked (*1*). Conversely, oligonucleotide-based therapeutics have been referred to as "informational drugs" in which the pharmacophore and PK properties can, in theory, be separately optimized because the nucleoside sequence directly determines the former whereas the unifying chemistry (the phosphate linkages used to couple the nucleosides) largely affects the latter (*2*). Advances in organic synthesis have had a profound impact on the ability of modern medicinal chemistry to target and rapidly access increasingly complex small-molecule leads. By contrast, as the range of oligonucleotide sequences and conceivable phosphate linkages has expanded, the fundamental chemistry used to enable their synthesis has largely remained unchanged despite numerous refinements and improvements. The incorporation of varied phosphate linkages has been documented to have a profound impact on both the properties and efficacy of the resulting structures (*3*). The hypothetical chimeric sequence (**1**) illustrated in Fig. 1, adorned with four different phosphorus-based linkages and multiple sugar backbones, tests the limits of the scope of existing methods. The impact of judiciously designing backbone chemical modifications in therapeutic oligonucleotides is only beginning to be realized (*4*–*7*). The opportunity to install broader combinations and variations at will and in any order thus presents an important step in the evolution of oligonucleotide therapeutics and requires the invention of enabling science. The commercialization of oligonucleotides and, more specifically, phosphorothioate antisense oligonucleotides (PS-ASOs) faces major challenges (*8*–*14*). The ability to easily make single-isomer species from stable species and/or fragments will aid control over product quality and has the potential to enable commercialization of these new therapeutics sustainably (i.e., through enhanced regulatory control of these complex molecules). Thus, improved methods of oligonucleotide synthesis amenable to standard methods of automation would have a near-immediate translational impact, enabling both the interrogation of greater linker permutations in drug discovery and better, more sustainable commercialization (see the supplementary materials for process mass intensity analysis) (*3*).

We recently demonstrated a practical approach to address the synthesis of stereopure phosphorothioates (*R*-PS and *S*-PS, mainly in the context of DNA), with the disclosure of the phosphorus sulfur incorporation (dubbed PSI or ψ) reagents. However, this initial study did not address the installation of other linkages such as phosphorodithioates and native phosphate diesters (*15*), other sugar chemistries [such as locked nucleic acid (LNA)], or applicability to a modern automated synthesizer protocol.

With these goals in mind, several challenges were apparent, with substantial lore and reactivity concerns to address when using a P(V)–based approach (*16*–*20*). For example, the rates associated with P(V) reagents have historically been viewed as too sluggish to ever compete with the P(III) manifold (*21*–*23*). Such approaches also had chemoselectivity challenges in that the guanine and thymidine bases would interfere in sequential couplings, reacting with the coupling reagents (*24*). Using P(III)–based reagents to install phosphorodithioate (PS_2) linkages is less than ideal. For instance, Caruthers reported a protocol based on protected thiophosphoramidites that requires discreet oxidation and deprotection steps (*25*) and suffers from the formation of a nearly inseparable phosphorothioate byproduct in ~5 to 10% yield. Despite this serious issue, this method is still commonly used because there are simply no viable alternatives (*26*). Finally, chimeric sequences with multiple modifications could be desired by medicinal

[1]Small Molecule Drug Discovery, Bristol Myers Squibb, Cambridge, MA 02142, USA. [2]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA. [3]Elsie Biotechnologies, San Diego, CA 92121, USA. [4]Chemical Process Development, Bristol Myers Squibb, New Brunswick, NJ 08903, USA.
***Corresponding author. Email: martin.eastgate@bms.com (M.D.E.); michael.schmidt@bms.com (M.A.S.); ivar.mcdonald@bms.com (I.M.M.); pbaran@scripps.edu (P.S.B.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
§Present address: Institute of Molecular Science (ICMol), Universitat de València, Paterna 46980, Spain.
¶Present address: Clarochem Ireland Ltd., Damastown, Mulhuddart, Dublin 15, Republic of Ireland.

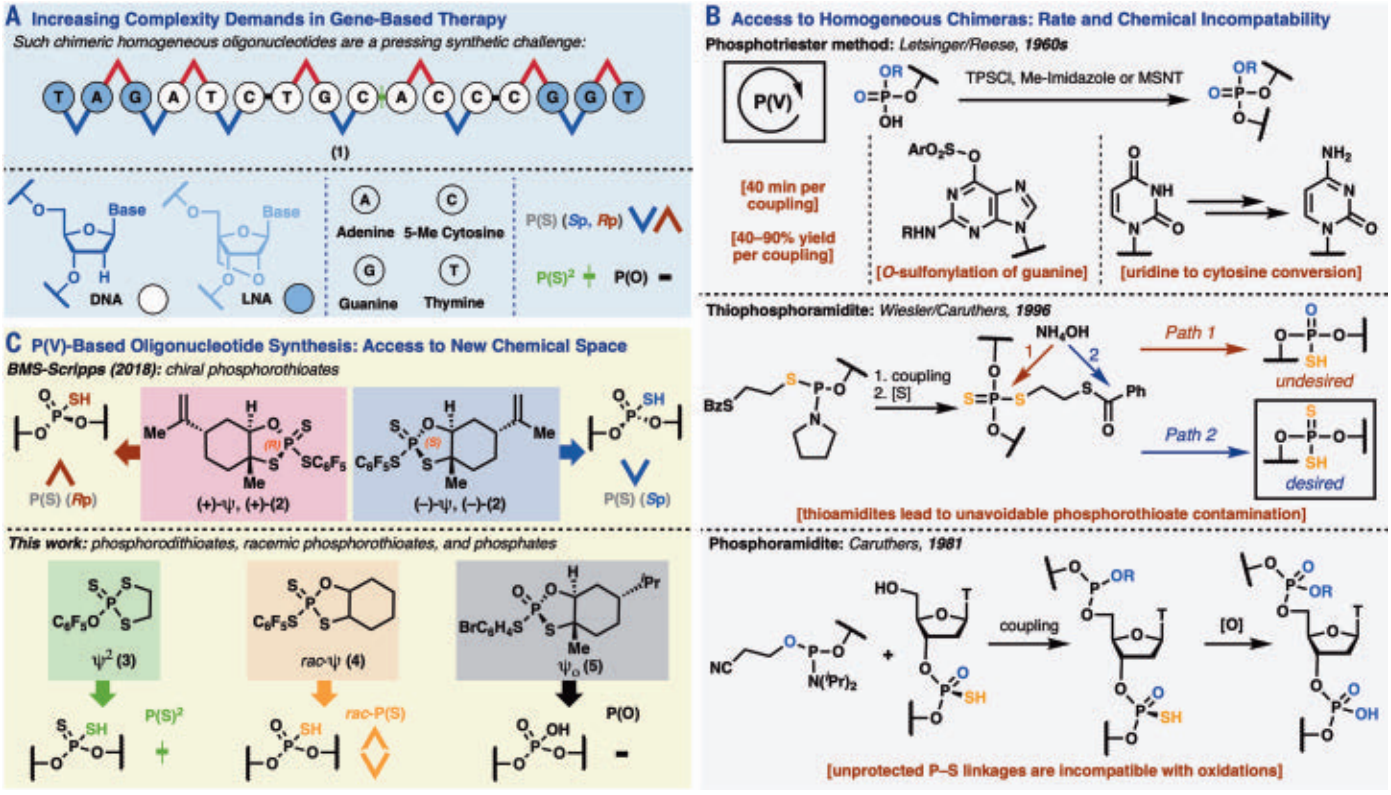


Fig. 1. Introduction and background. (**A**) Inspiration. (**B**) Challenges facing P(V)–based synthesis. (**C**) Recent developments and this work. ^{i}Pr, iso-propyl; Me, methyl; Bz, benzoyl; [O], oxidation; [S], sulfurization.

chemists, but only certain combinations (PS/PO, PS/PS_2, PO/PS_2) are present in the literature. Even a hybrid synthetic approach that merges P(III)–based phosphoramidite chemistry with P(V) reagents is unworkable because it suffers from a lack of chemoselectivity (vide infra) as oxidation of P(III) to P(V) requires a protecting group on exposed PS/PS_2 linkages to avoid desulfurization. The use of organic oxidants was ruled out because of the extreme number of equivalents required and their extended reaction times (*27*).

In this disclosure, three new reagent systems are described, ψ^2 (**3**), *rac*-ψ (**4**), and ψ^O (**5**), which, when combined with the previously reported systems [(+)-ψ, (+)-**2**] and [(−)-ψ, (−)-**2**], provide a unified P(V) approach that entirely departs from the rubric of P(III)–based oligonucleotide synthesis and enables the at will and controlled synthesis of specific chimeric oligonucleotides (Fig. 1C). Along with these new reagents are protocols for their unified application in commercial synthesizers using a single coupling protocol that spans all coupling types. This redox-neutral platform, which is based on the native P(V) oxidation state, challenges past assumptions of sluggish reactivity and enables access to five relevant P linkages across a range of sugar backbones (DNA and LNA) and bases (A, T, G, and mC) in one oligonucleotide construct. Aside from enabling straightforward access to a wide range of chimeric oligonucleotides, the implementation of this new protocol benefits from a reduced reliance on protecting group chemistry (which eliminates the labile cyanoethyl group and therefore acrylonitrile production upon deprotection) (*28*, *29*), bespoke additives (*30*), and redox fluctuations. This new P(V) platform also eliminates one full step in the standard solid-phase oligonucleotide synthesis (SPOS) protocol (namely the phosphorus oxidation).

Figure 2 outlines the synthesis of ψ^2 and ψ^O reagents for the incorporation of phosphorodithioate and native phosphodiester linkages, respectively. The development of these reagents required extensive experimentation, whereas the synthesis of *rac*-ψ was relatively trivial and proceeded by analogy to ψ using cyclohexene oxide (see the supplementary materials). Fully sulfurized versions of diphosphate esters, known as phosphorodithioate linkages, are isopolar and isostructural analogs of phosphates that are completely stable to nucleases (*31*) but maintain the ability to form stable duplexes and elicit desirable mRNA cleavage by RNase H without the complexity of chirality at phosphorus (*32*, *33*). The pioneering work of Stec and colleagues on phospholane heterocycles, which inspired the development of (+)-ψ and (−)-ψ, was an essential precedent for the present work (*19*). In 1995, it was disclosed that dithiaphospholanes could be installed onto nucleosides and coupled to afford dinucleotides incorporating a phosphorodithioate linkage, albeit requiring a separate oxidation step to install sulfur, and a toxic and unstable (explosive) reagent (Fig. 2A) (*18*). Our P(V)–centric study thus built on the lessons of these two precedents with the goals of eliminating the extraneous oxidation step and dangerous reagents. Upon identification of the optimal leaving group (21 evaluated; see the supplementary materials for Hammett correlations of leaving groups) and ring size (two evaluated), inexpensive P_2S_5 could be combined with pentafluorophenol, followed by capping of P(V) intermediate (**6**) with thiirane to generate ψ^2 on large scale (>100 g). In this way, unsafe P(III) chemistry was avoided, and a stable and viable P(V) reagent was developed.

Oligonucleotides with native phosphodiester linkages have poor pharmacokinetics and are rapidly degraded by nucleases, but their use is invaluable in routine molecular biology and diagnostic settings (*34*). In addition, limited PO linkages are incorporated into current antisense oligonucleotides. To the best of our knowledge, there are no known P(V)–based reagents for achieving P(III)–competitive reactivity and chemoselectivity that are stable and amenable to automated synthesis (*34*). Although the design of such a reagent benefited

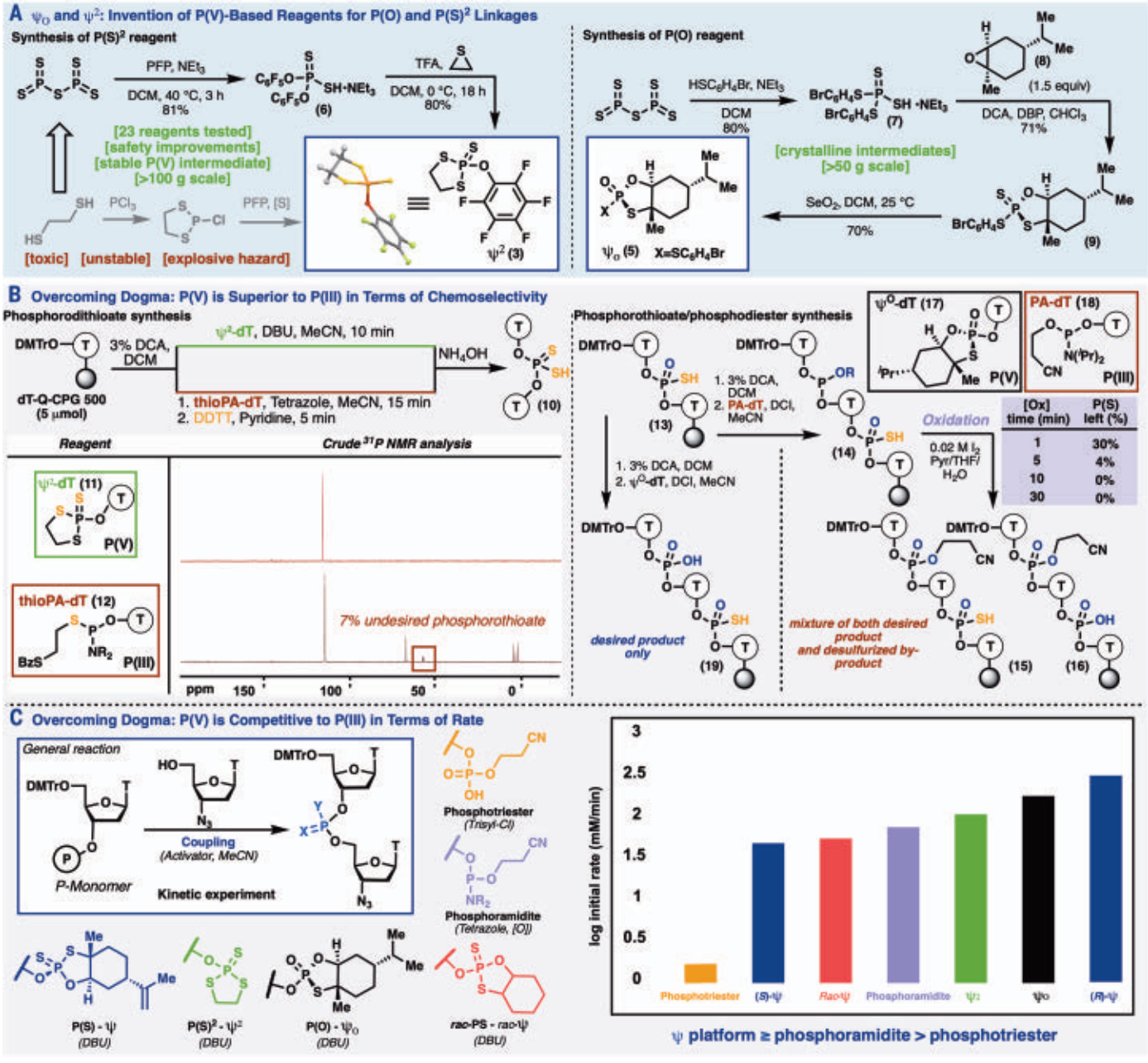


Fig. 2. Development of a fully P(V) platform for oligonucleotide synthesis. (**A**) Reagent development. (**B**) Overcoming chemoselectivity challenges. (**C**) Overcoming historical P(V) rate challenges. PFP, pentafluorophenol; TFA, trifluoroacetic acid; DCA, dichloroacetic acid; DBP, dibutylphosphate; DDTT, 3-[(dimethylamino-methylidene)amino]-3H-1,2,4-dithiazole-3-thione; DCI, 4,5-dicyanoimidazole, DMTr, 4,4'-dimethoxytrityl.

from the lessons of our previous studies, it was repeatedly thwarted by the challenge of identifying both a highly reactive and stable entity upon loading to a monomer. We evaluated nearly 30 different backbones along with three different leaving groups before arriving at ψ^O (see the supplementary materials for a comprehensive summary). The backbone optimization systematically evaluated ring size, substituents, electronic effects, and stereochemistry to probe effects on loading, coupling, and overall stability, with ψ^O emerging as the only viable candidate. The synthesis outlined in Fig. 2A requires three simple, scalable steps (>50 g). In a similar fashion to ψ^2, inexpensive P_2S_5 was reacted with 4-bromothiophenol, yielding P(V) intermediate (**7**), which, when combined with hydrogenated *cis*-limonene oxide (**8**), yields the PS reagent (**9**); desulfurization with SeO_2 yields ψ^O.

With these new reagents in hand, a side-by-side comparison with state-of-the-art P(III) chemistry was conducted (Fig. 2B). For the installation of the PS_2 linkage, the three-step P(III) approach to access simple dimer (**10**) through P(III) adduct (**12**) resulted in ~7% of the PS impurity resulting from desulfurization during the deprotection event. By contrast, using ψ^2, dimer (**10**) was cleanly accessed in two steps through P(V) adduct (**11**) in >99% purity. To field test and contrast the synthesis of mixed PO-PS backbones by the two platforms, PS dimer (**13**) was subjected to a standard phosphoramidite coupling with (**18**) to yield protected trimer (**14**), which, upon oxidation of P(III) to the requisite P(V), resulted in

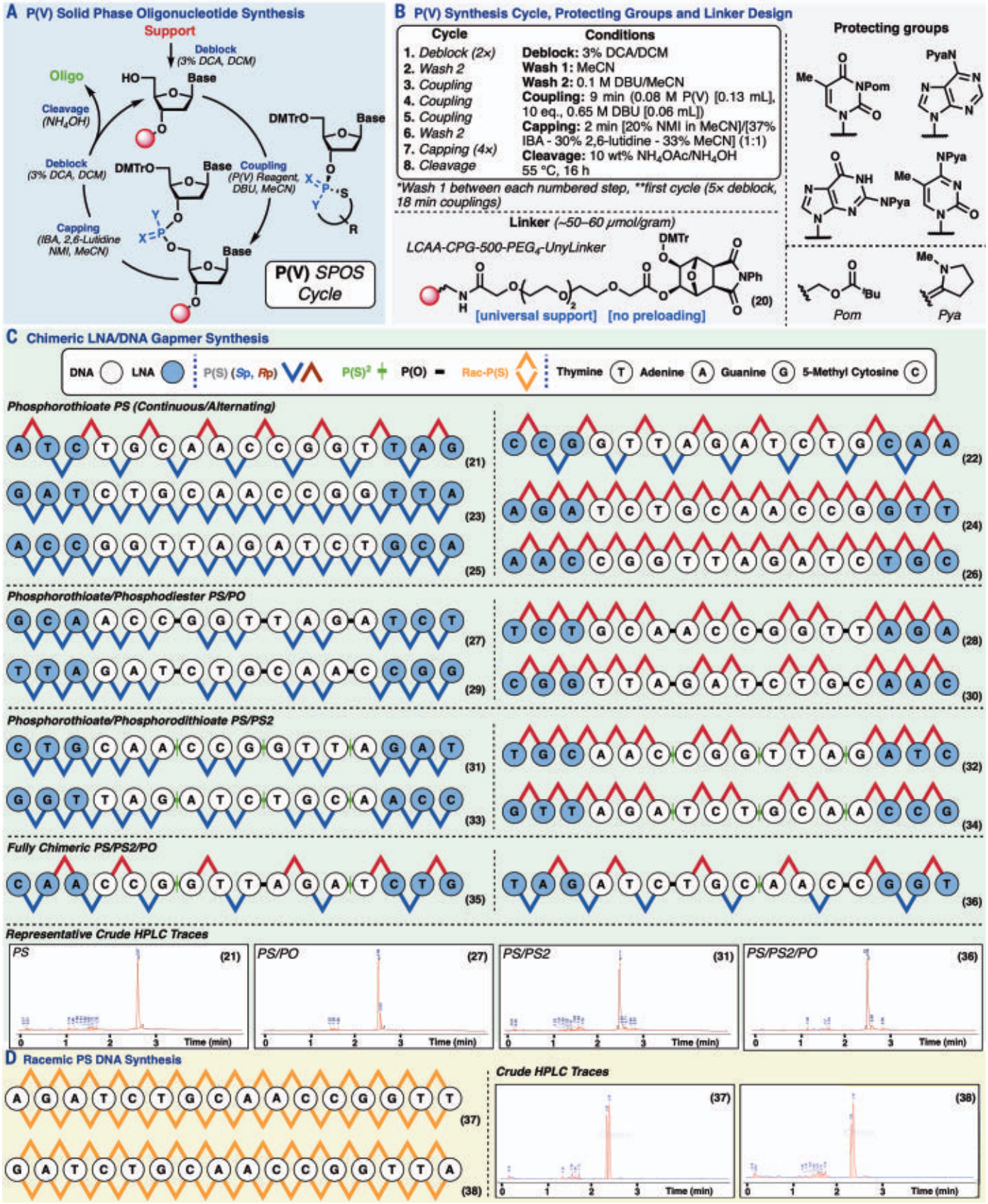


Fig. 3. Automated P(V) oligonucleotide synthesis. (**A**) P(V) solid-phase oligonucleotide synthesis cycle. (**B**) Synthetic cycle conditions, protecting groups, and linker chemistry. (**C**) 3-10-3 LNA/DNA gapmer synthesis. (**D**) Racemic phosphorothioate oligonucleotide synthesis. IBA, isobutyric anyhydride; NMI, *N*-methyl imidazole.

rapid desulfurization [see ratio of (**15**)/(**16**) over time]. Conversely, the redox-neutral P(V) approach using P(V) adduct (**17**) cleanly delivered the unprotected mixed PS-PO trimer (**19**) without any loss of sulfur.

The final challenge that P(V)–based reagents face is the longstanding perception that their diminished coupling rates preclude them from being used in traditional automated oligonucleotide synthesis regimes. With a full suite of reagents in hand based on P(V), we evaluated their coupling performance side by side with canonical P(III) chemistry using kinetics studies (Fig. 2C). In addition, the original P(V)–based coupling using phosphotriester chemistry was included because this was the initial benchmark. Temporal reaction progress monitoring was performed by taking aliquots of the mixture coupling P-loaded-T with AZT and monitoring product formation by high-performance liquid chromatography–mass spectrometry using an internal standard. Full kinetic profiles are included in the supplementary materials, and here we present the main trends by considering relative rates calculated from these profiles. Consistent with the literature, the classic P(V)–based phosphotriester approach was extremely sluggish (Fig. 2C, orange bar). However, the P(V) reagent suite detailed herein performed equally well as the industry standard P(III) protocol, with all reactions reaching full conversion in under two minutes.

Extensive optimization of the SPOS methodology for phosphoramidite-based synthesis has been achieved over the course of the past 30+ years. Although some of these methods could be leveraged in this new context, there were areas in which existing solutions were not compatible with the P(V) synthesis protocol (Fig. 3A). Existing universal supports were not sufficient stable toward 1,8-diazabicyclo [5.4.0]undec-7-ene (DBU), prompting our development of a universal support (**20**) with substantially improved base stability. Guided by Stec's earlier work, Pya-protecting groups were used instead of the standard amide-protecting groups (Fig. 3B) (*18*). Improved results were also obtained when a Pom protection was used for thymidine (*35*). With all P(V) reagents in hand and the chemoselectivity and relative coupling rates established, a systematic interrogation was undertaken to test the efficiency of this redox-neutral P(V) -platform on automated SPOS (Fig. 3A). The cycle commences with the deblocking of the DMT-protecting group of a resin-bound nucleoside to afford a free 5 -alcohol that is primed to react with any P(V)–loaded nucleotide in the subsequent coupling step. This key step was carefully optimized for all reagents through systematic reactivity and hydrolysis studies using ultraviolet and ^{31}P–nuclear magnetic resonance (NMR) analyses to enable a robust double-coupling protocol for each P(V) reagent. The subsequent capping and deblocking steps complete the solid-phase cycle and set the stage for the next coupling (Fig. 3A). The utility of any new reagent system for an oligonucleotide synthesis platform is wedded to its fidelity and robustness in the context of preparing diverse sequences with a single protocol. Thus, a matrix was designed to incorporate all possible nucleobase (A, C, G, and T) and sugar (DNA and LNA) combinations templated onto a 3-10-3 DNA/LNA gapmer scaffold, the current state of the art in RNase H–activating ASOs (Fig. 3C) (*3*). LNA modifications that have a marked effect on binding affinity were coupled in near quantitative yields (*36*). A single protocol was used regardless of the P(V) monomers used to assess the generality of the method versus sequence-specific optimization. First, the general method was field tested to produce homogeneous, chiral PS-ASOs with both alternating (**21**, **22**) and continuous stereochemical patterns (**23**–**26**). This approach is the second industrially viable platform to produce stereopure PS-ASOs and uses redox-neutral, sustainably derived P(V)–based reagents (*37*). With this milestone achieved, the incorporation of PO_2 linkages into these constructs was pursued. These chimeric sequences (**27**–**30**) could be accessed in high purity with no substantial loss of sulfur during synthesis. Next, sequences bearing both PS and PS_2 (**31**–**34**) linkages were prepared. Finally, constructs containing all four possible linkages were cleanly produced (**35**, **36**). Thus, these protocols describe a convenient single platform for probing linkage SAR that could enable systematic tuning of physical and biological properties (*38*, *39*). Given the differences in scale, chemical sequence, method of purification, and quantification, a direct comparison between yields of this P(V) platform and those of other stereopure methods is outside the scope of this report. In its present manifestation, these homogeneous, chimeric oligonucleotides (**21**–**38**) were produced in 12 to 27% isolated yield using a sequence- and linkage-agnostic protocol. When compared directly with the stereorandom constructs (produced using state-of-the-art chemistry) that were produced in 30 to 60% yield and accounting for the ability to prepare distinctive chimeric systems, the difference in yield is a relatively small gap to fill to bring homogeneous, stereopure oligonucleotides on par with their stereorandom counterparts. From a pragmatic perspective, the observed yields even in this first disclosure are more than enough to progress a medicinal chemistry program.

Current methods to produce PS oligonucleotides generate a statistical mixture of isomers depending on the specific conditions used (*40*). Homogeneous, well-defined ASOs are of value, but no true P(V) oligonucleotide synthesis platform would be complete without rapid access to these stereorandom variants (*41*). Thus, a final need was to produce *rac*-ψ (Fig. 1), a derivative of ψ that retains high reactivity and provides mixtures of diastereomers comparable to those obtained with P(III) methodology. *rac*-ψ was loaded onto DNA cores, and the corresponding monomers were cleanly implemented (Fig. 3D) into the aforementioned workflow, thus enabling racemic PS oligos to be produced on this platform (**37**, **38**).

Oligonucleotide therapeutics, so-called informational drugs, target essentially every level of the central dogma of molecular biology. Given the immense chemical space conceivable, the number of nucleic acid modifications that have been investigated to date is quite narrow, with even fewer represented in the clinic. Although P(III)–based phosphoramidites have revolutionized access to vast swaths of this space, new linkages and chemical modifications have pushed the limits of what is currently accessible. Oligonucleotide synthesis originated with P(V)–based reagents because nature only uses this oxidation state to craft its building blocks of life, but this method was soon abandoned because of perceptions of the reduced reactivity and selectivity of these reagents. This work provides a compelling justification for renewed research in this field because it can enable democratized access to a wide range of medicinally promising chimeric sequences. Finally, this new approach offers a realistic opportunity to connect the exploration of new chemical space directly to patients through a platform containing improved environmental sustainability, enhancing our ability to discover, develop, and commercialize this promising class of therapeutics.

REFERENCES AND NOTES

1. J. P. Hughes, S. Rees, S. B. Kalindjian, K. L. Philpott, *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
2. A. Khvorova, J. K. Watts, *Nat. Biotechnol.* **35**, 238–248 (2017).
3. W. B. Wan, P. P. Seth, *J. Med. Chem.* **59**, 9645–9667 (2016).
4. M. Byrne *et al.*, *Transl. Vis. Sci. Technol.* **10**, 23 (2021).
5. Y. Liu *et al.*, *Nat. Commun.* **12**, 847 (2021).
6. K. Bleicher *et al.*, Gapmer oligonucleotides comprising a phosphorodithioate internucleoside linkage and therapeutic use. International Patent WO/2019/122282 (2019).
7. L. Bethge, J. Hauptmann, C. Frauendorf, A. Weingärtner. Nucleic acids for inhibiting expression of a target gene comprising phosphorodithioate linkages. International Patent WO/2019/092280 (2019).
8. T. Tredenick, "Oligonucleotides: Opportunities, pipeline and challenges" (Pharma Manufacturing, 2016); https://www.pharmamanufacturing.com/articles/2016/oligonucleotides-opportunities-pipeline-and-challenges/ .
9. F. Eckstein, V. W. Armstrong, H. Sternbach, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2987–2990 (1976).
10. A. Wilk, W. J. Stec, *Nucleic Acids Res.* **23**, 530–534 (1995).
11. B. I. Andrews *et al.*, *J. Org. Chem.* **86**, 49–61 (2021).
12. M. Shimizu, T. Wada, Asymmetric auxiliary group. U.S. Patent US/10696711 (2020).
13. N. Oka, M. Yamamoto, T. Sato, T. Wada, *J. Am. Chem. Soc.* **130**, 16031–16037 (2008).
14. R. Suchsland, B. Appel, P. Virta, S. Müller, *RSC Advances* **11**, 3892–3896 (2021).

15. K. W. Knouse *et al*., *Science* **361**, 1234–1238 (2018).
16. W. J. Stec, A. Grajkowski, M. Koziolkiewicz, B. Uznanski, *Nucleic Acids Res.* **19**, 5883–5888 (1991).
17. W. J. Stec *et al*., *J. Am. Chem. Soc.* **117**, 12019–12029 (1995).
18. A. Okruszek, A. Sierzcha-la, K. L. Fearon, W. J. Stec, *J. Org. Chem.* **60**, 6998–7005 (1995).
19. W. J. Stec *et al*., *J. Am. Chem. Soc.* **120**, 7156–7167 (1998).
20. P. Guga, W. J. Stec, *Curr. Protoc. Nucleic Acid Chem.* **14**, 4.17.11–14.17.28 (2003).
21. F. Eckstein, I. Rizk, *Angew. Chem. Int. Ed.* **6**, 695–697 (1967).
22. C. B. Reese, R. Saffhill, *Chem. Commun.* 767–768 (1968).
23. R. L. Letsinger, W. B. Lunsford, *J. Am. Chem. Soc.* **98**, 3655–3661 (1976).
24. C. B. Reese, P. A. Skone, *J. Chem. Soc., Perkin Trans. 1* 1263–1271 (1984).
25. W. T. Wiesler, M. H. Caruthers, *J. Org. Chem.* **61**, 4272–4281 (1996).
26. X. Yang, *Curr. Protoc. Nucleic Acid Chem.* **70**, 4.77.71–74.77.13 (2017).
27. E. Radzikowska, J. Baraniak, *Org. Biomol. Chem.* **13**, 269–276 (2015).
28. X. Guo, J. A. Stolee, Y. A. Fillon, L. Zou, *Org. Process Res. Dev.* **25**, 318–326 (2021).
29. D. C. Capaldi *et al*., *Org. Process Res. Dev.* **7**, 832–838 (2003).
30. N. Oka, T. Wada, K. Saigo, *J. Am. Chem. Soc.* **125**, 8307–8317 (2003).
31. G. M. Porritt, C. B. Reese, *Tetrahedron Lett.* **31**, 1319–1322 (1990).
32. M. H. Caruthers *et al*., *Nucleosides Nucleotides* **10**, 47–59 (1991).
33. W. S. Marshall, M. H. Caruthers, *Science* **259**, 1564–1570 (1993).
34. M. H. Caruthers, *J. Biol. Chem.* **288**, 1420–1427 (2013).
35. M. C. de Koning *et al*., *Org. Process Res. Dev.* **10**, 1238–1245 (2006).
36. S. K. Singh, A. A. Koshkin, J. Wengel, P. Nielsen, *Chem. Commun.* (4): 455–456 (1998).
37. N. Iwamoto *et al*., *Nat. Biotechnol.* **35**, 845–851 (2017).
38. S. T. Crooke, X. H. Liang, B. F. Baker, R. M. Crooke, *J. Biol. Chem.* **296**, 100416 (2021).
39. S. T. Crooke, B. F. Baker, R. M. Crooke, X. H. Liang, *Nat. Rev. Drug Discov.* **20**, 427–453 (2021).
40. H. Jahns *et al*., *Nat. Commun.* **6**, 6317 (2015).
41. M. E. Østergaard *et al*., *Nucleic Acids Res.* **48**, 1691–1700 (2020).

ACKNOWLEDGMENTS

We thank D.-H. Huang and L. Pasternack (Scripps Research) for assistance with NMR spectroscopy; J. Chen (Automated Synthesis Facility, Scripps Research) for purification of compounds and acquisition of HRMS data; A. L. Rheingold, C. E. Moore, and M. Gembicky (UCSD) for x-ray analysis; and S. Mukherjee (BMS) for calculating the process mass intensity. **Funding:** This work was supported by Bristol Myers Squibb, NIGMS (GM106210), NIH (grant number GM-118176), Marie Skłodowska-Curie Global Fellowships (749359-EnanSET, N.M.P) within the European Union research and innovation framework program (2014-2020), and the NSF SURF Program (TSRI, R.N). **Author contributions:** M.D.E., M.A.S., P.S.B., K.W.K., I.M.M., and R.E.O. conceived of this research program. K.W.K., W.H., S.Q., Y.H., B.Z., S.E.M., J.C.V., J.L.O., R.N., N.M.P., M.A.S., and I.M.M. performed the experiments. K.W.K. and P.S.B. wrote the manuscript. J.C.V., M.D.E., M.A.S., R.E.O., I.M.M., and D.G.B. assisted in editing the manuscript. **Competing interests:** M.A.S., B.Z., K.W.K., M.D.E., P.S.B., R.E.O., and I.M.M. are listed as inventors on US Patent Application US 2019/0322694, "Novel phosphorous (V)-based reagents, processes for the preparation thereof, and their use in making stereo-defined organophosphorous (V) compounds." K.W.K. and P.S.B. are co-founders and shareholders of Elsie Biotechnologies. The remaining authors declare no competing interests. **Data availability:** The data that support the findings of this study are available in the main text or supplementary materials. Crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre under reference no. CCDC 2093530.

SUPPLEMENTARY MATERIAL

https://science.org/doi/10.1126/science.abi9727
Materials and Methods
Figs. S1 to S7
NMR Spectra
References (*42*–*46*)

12 April 2021; accepted 12 July 2021
10.1126/science.abi9727

Vial Sampling Kits for PFAS Analysis

J.G. Finneran Associates introduces new Convenience Kits for per- and polyfluoroalkyl substances (PFAS) testing. There is increasing evidence that exposure to PFAS can adversely affect human health. Because PFAS are widely used in materials such as polytetrafluoroethylene (PTFE), commonly used in chemical-analysis consumables, laboratories need new labware products to minimize background contamination from these chemicals. J.G. Finneran Associates R.A.M. screw thread vials are purpose-designed for PFAS analyses. Manufactured from chemically resistant polypropylene, they are ideal for "PFAS-free" sampling, testing, storing, and transporting tasks. These lightweight vials are an economical alternative to glass and work in most autosamplers. When you are unable to use PTFE as a vial liner, J.G. Finneran Associates new polyimide/silicone-lined closure is the perfect alternative. This new closure operates up to 130°C, while the liner performs well up to 10 injections. The polyimide is amber translucent, and the white silicone rubber is a high-purity rubber composition resulting in low leachables. The new Convenience Kits for PFAS testing contain 100 R.A.M. screw thread polypropylene vials and 100 of the new closures/septa, packaged in a clear tray that fits a bench drawer for easy access.

J.G. Finneran Associates
For info: +1-856-696-3605
www.jgfinneran.com

Antigen Retrieval System

EZ-Retriever System is a microwave-based antigen retrieval pretreatment system invented by BioGenex. It is used for dewaxing, rehydration, and antigen retrieval of formalin-fixed, paraffin-embedded tissue sections to enhance exposure of antigenic epitopes. This system not only resolves inconsistencies due to different types of antigen retrieval methods by applying different maximum microwave power levels, solution volumes, and heating time, but also facilitates standardization of dewaxing and antigen protocols for high-quality, reproducible staining. Dewaxing, rehydration, and antigen retrieval are performed in one step. The system is high throughput and can handle 96 slides in 35 min.

BioGenex
For info: +1-800-421-4149
www.biogenex.com

Biosimilar Products for Research Use

AMS Biotechnology has launched a new range of research-grade biosimilars that enable analysis of biological processes without the need to purchase expensive therapeutic-grade biologics. Investigating the biological effects of a drug or doing proof-of-concept assay development requires a source for the drug. Typically, it is difficult to obtain original pharmaceutical-grade biotherapeutic drugs for research use rather than for clinical applications. Consequently, biological compounds almost identical to a reference biologic drug (biosimilars) that have been approved by governmental regulatory authorities for medical treatment are vital research tools. To develop an effective biologic drug, it is important to synthesize a reference antibody. A reference antibody can greatly facilitate the downstream development of a new biologic drug and can help identify critical cell lines, reagents, and performance characteristics. Our new range of biosimilar antibodies for research use avoid the need to source an expensive therapeutic product and enable fast, inexpensive assay development.

AMS Biotechnology
For info: +1-617-945-5033
www.amsbio.com/biosimilars

Lab Management Software

To meet new demands and overcome challenges in laboratories worldwide, Eppendorf offers smart lab management software solutions for laboratory management, such as the VisioNize Lab Suite, which enhance productivity for all routine tasks in the laboratory. VisioNize is the answer for effective lab management in the areas of remote monitoring, alarm notifications, device management, and task management. The service suite is continually updated, representing a sustainable investment in the laboratory of the future. The SARS-CoV-2 pandemic has a wide range of effects on laboratories across the globe, which include not only the current high volume of requests for PCR tests for the novel coronavirus, but also the challenges laboratories face in quickly adopting digital technologies to cope with the new norms of social distancing and remote work away from laboratory benches. These past months have proven that digitalization is central to every interaction—in the private realm as well as in business—and have fast-tracked the acceptance of digital solutions.

Eppendorf
For info: +1-800-645-3050
www.eppendorf.com/visionize

Microfluidic Hydrogel Kit

Sphere Fluidics and ClexBio announce the launch of the biocompatible CYTRIX Microfluidic Hydrogel Kit. The kit combines ClexBio's novel CYTRIX Hydrogel with Sphere Fluidics' specially designed Pico-Gen double aqueous biochip, to allow plug-and-play generation of defined, reproducible, and tailorable hydrogel microstructures for 3D cell culture, organoids, single-cell analysis, and many other applications. This technology makes it possible for researchers to study valuable cells over weeks rather than days, in microenvironments that mimic the natural extracellular matrix. The CYTRIX Hydrogel overcomes the challenges of existing microfluidic hydrogel formation techniques, such as clogging, finicky temperature control, and time-critical mixing procedures. It offers cytocompatibility and microfluidic-optimized gelation kinetics, providing researchers with a cell-friendly solution to reduce the effect of harmful temperatures, pH changes, and use of UV light required for existing hydrogel formation techniques. The system enables efficient encapsulation of mammalian cells, bacteria, and other microorganisms in a homogeneous, defined extracellular matrix that supports cell viability.

Sphere Fluidics
For info: +44-(0)-1223-628890
spherefluidics.com/microfluidic-hydrogel

Electronically submit your new product description or product literature information! Go to www.science.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.

The Paris Region Fellowship Program

The second call for 26 post-doctoral positions in a large panel of disciplines for international researchers of all nationalities,

Professional training programme and career development opportunities for recruited researchers,

Call deadline October, 21st, 2021

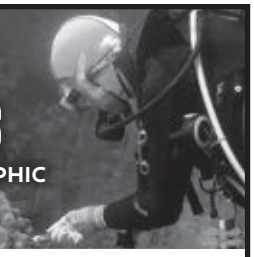
Further information :
parisregion.eu/parisregionfp.html

Application Website :
https://parisregionfp.sciencescall.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie actions Grant Agreement n° 945298-ParisRegionFP.

Région îledeFrance

Fellowships for Postdoctoral Scholars



AT WOODS HOLE OCEANOGRAPHIC INSTITUTION

New or recent doctoral recipients are encouraged to submit **applications prior to October 15**, 2021, to start after January 1 and before December 1, 2022.

Awards will be in the following areas: **Applied Ocean Physics & Engineering; Biology**; **Geology & Geophysics**; **Marine Chemistry & Geochemistry**; **Physical Oceanography**; **interdepartmental research**; **The Ocean Bottom Seismic Instrument Center**; **The Ocean Twilight Zone Project**; and a **joint USGS/WHOI award**.

Awards are aimed at advancing applicants' research careers. Scholarships are 18-months (annual stipend $63,300, plus health allowance and research budget). Recipients will pursue their own research interests, mentored by resident staff. Communication with potential WHOI advisors prior to submitting applications is encouraged. The Postdoctoral Scholar Program is committed to broadening participation in ocean science and engineering.

For further information: ***https://go.whoi.edu/pdscholarship***

WOODS HOLE **OCEANOGRAPHIC** INSTITUTION

USDA ars Agricultural Research Service
www.ars.usda.gov

Supervisory Research Molecular Biologist/ Geneticist
Research Leader (Center Director)
Plant Gene Expression Center
Albany, CA

The U.S. Department of Agriculture, Agricultural Research Service, Plant Gene Expression Center (PGEC) in Albany, California invites applications for a permanent, full time **Supervisory Research Molecular Biologist/ Geneticist** who will serve as the Research Leader (Center Director) for the program. The successful candidate will have expertise in plant molecular biology and/or plant genetics, and will be responsible for developing a research program in predictive agriculture that combines aspects of genomics and genetics, molecular biology, physiology, biochemistry, and/or metagenomics with computational biology to generate new tools for breeding that improve performance of agricultural crops. Areas of research interest include enhancing plant productivity, adaptation to climate change, biotic and/or abiotic stress tolerance, signal transduction, plant metabolism, and plant-microbiome interactions. The PGEC is a world-renowned institute that conducts foundational and translational plant biology research to improve agriculture. The candidate will direct PGEC research as a whole, hire and mentor new scientists, and manage resources. PGEC research is conducted through a close collaborative partnership with the University of California, Berkeley, where PGEC scientists have the opportunity for Adjunct Faculty status in the Department of Plant and Microbial Biology. The salary range is **$132,822.00 to $172,500.00** per year. A comprehensive Federal benefits package is available.

For more details and to apply, please visit **USAJOBS.gov, Announcement Number ARS-D21Y-11193664-AF.** Open period is from **Friday, September 3, 2021 to Friday, October 1, 2021**. If you have questions, please contact Dr. Jennifer Lewis at **jennifer.lewis@usda.gov**. For information about the PGEC visit **https://www.ars.usda.gov/pacific-west-area/albany-ca/pgec/**.

U.S. Citizenship is required.
The USDA/ARS is an Equal Opportunity Employer and Provider.

Science Careers helps you advance your career. Learn how !

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.

Visit ScienceCareers.org today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

WE'RE HIRING

Science Director & Research Professor

About VIB-UGent PSB

The cornerstone of PSB is curiosity-driven plant research on fundamental cellular mechanisms. The unique combination of world-class basic research, efficient translation pipelines toward field application, and an established ecosystem of companies make PSB a leading European Plant Science Center at the forefront of plant science-based solutions for sustainability and climate change.

The challenge

VIB and Ghent University are jointly searching for a dynamic, internationally recognized leader to inspire and lead the research center. The selected candidate will receive a dual appointment: a VIB appointment as Science Director of the center and an appointment as a tenured professor at Ghent University, Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics. It is expected that the candidate will maintain an active interest in leading an independent research lab.

The profile

Due to the retirement of the current Science Director, VIB is seeking an experienced, visionary scientist who is widely recognized in their academic field. This candidate has a strong record of scientific publications and desire to develop a common vision and mission for PSB, with sensitivity toward building a diverse and inclusive working environment. Subsequently, VIB is looking for a Research Professor with didactical skills to teach university students to develop academic competencies.

VIB

Interested candidates can visit jobs.vib.be or scan here to learn more.

Science Careers online @sciencecareers.org

By **Karishma Bisht**

A ghost author speaks up

When I stumbled on the research paper from one of my former labs, related to a project I had contributed to, I was taken by surprise. No one had told me our work was going to be published! A quick glance showed I was acknowledged, which was gratifying. However, when I looked at the figures, my surprise turned to shock. I had generated two of the figures, as well as the underlying data, a couple years earlier, when I was a research fellow in the lab after finishing my master's degree. Based on my contribution, I should have been included as an author. Then I discovered the group had published other papers about our work without even acknowledging me. I had become a ghost author, my contributions used without credit.

I knew this wasn't right. Still, I was hesitant to speak up. What if it went badly?

Luckily, I had my Ph.D. adviser to guide me. Not only is she a great mentor, she also taught a responsible conduct of research (RCR) course—inspired by her postdoc adviser, who ran a similar discussion group—which I had taken early in my Ph.D. journey. Since then, we've discussed authorship practices, peer review, data management, and more. Her course and our conversations have been invaluable for me, especially as someone who came to the United States for my Ph.D. after training elsewhere. They helped me adjust to a new research environment and learn the nuances of conducting research to the highest ethical standards.

"Another co-author had known about my contribution but chose not to credit me."

My adviser suggested I contact the corresponding author, but urged me to choose my words wisely. I knew I risked burning bridges by complaining to my former group, but I took comfort in the fact that I have moved to a different research area and will not be needing any recommendation letters from that lab.

I mustered the courage to email the corresponding author. I had a good relationship with them, and I guessed they had simply forgotten about my work on the project. I started by congratulating them on the publication; noted my contribution, attaching the original data and figures while keeping my tone respectful; and requested they submit a formal correction to add me as a co-author on the paper.

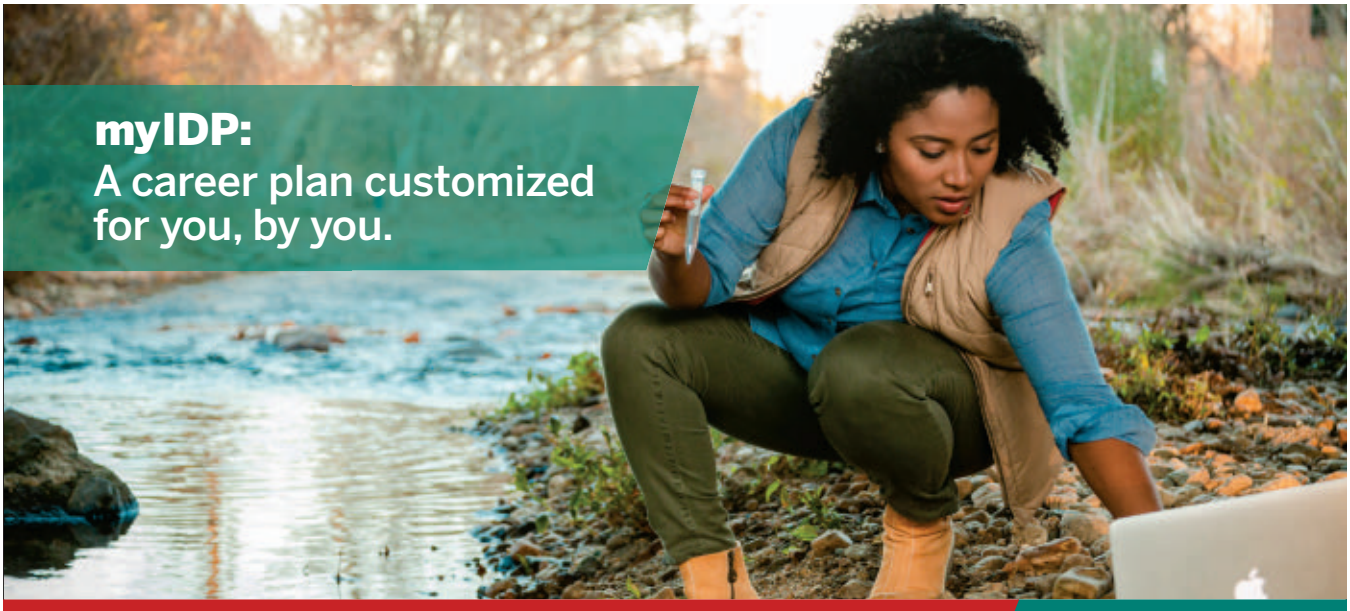
My former colleague was quick to respond, and I was relieved that they shared my concerns. They assured me they would work on fixing the issue. But after 1 month of silence, I followed up—only to find out that another co-author had known about my contribution but chose not to credit me. The corresponding author assured me that things would be taken care of, but 9 months have now passed with no progress.

Still, I am proud that I spoke up, and I've resolved not to become a ghost author again. Staying in touch with my former lab and standing up for myself prior to publication might have made a difference. I realize that won't always work: A few months before the paper that made me a ghost was published, I had emailed one of the authors to ask about the status of the work, only to receive a brush off. And sometimes there's no need to take the initiative: When my work from a different former lab was ready to be published, the group leader and co-author did a great job keeping me in the loop throughout the manuscript submission and revision process. But I've learned that I can't expect that, and it is on me—and early-career researchers more broadly—to be proactive and persistent in reaching out to discuss authorship and credit on future publications.

When I recently graduated from my Ph.D. program, I initiated a thorough discussion with my adviser about my contributions to potential manuscripts from the lab to avoid future conflicts. Luckily, my adviser's RCR focus and our strong relationship helped the conversation go smoothly, and I don't have any concerns about being recognized in the future for my work in her lab. Still, I'll make sure to keep in touch. ■

Karishma Bisht is a postdoctoral research associate at Princeton University. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER



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.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology

UMASS University of Massachusetts Medical School

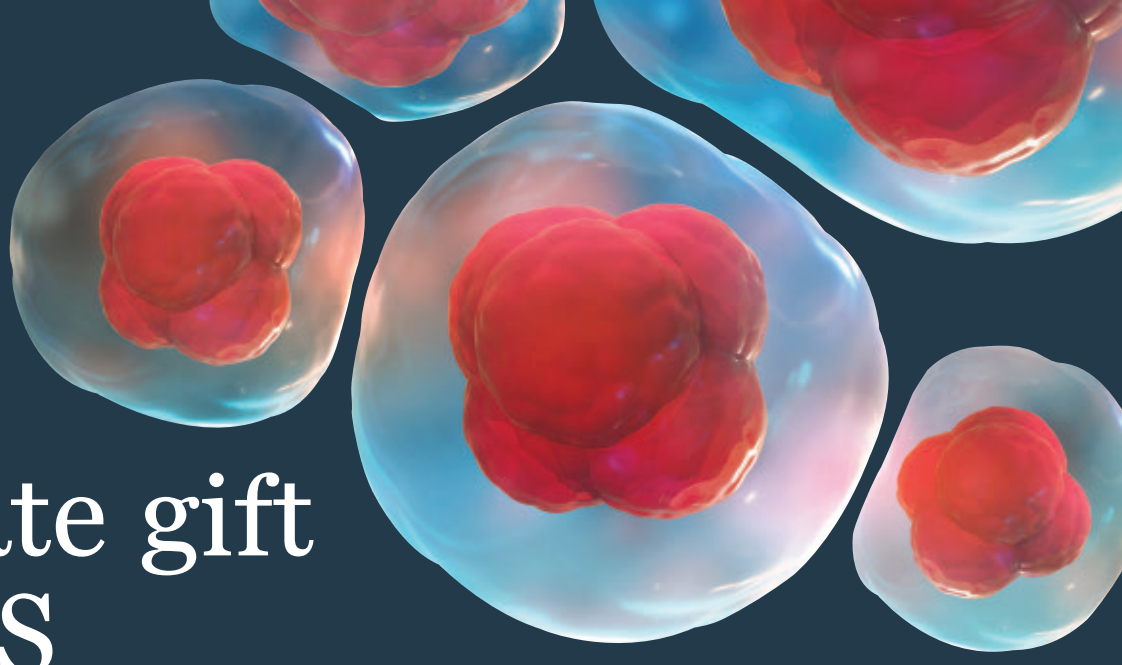
UCSF University of California San Francisco

UIC

BURROUGHS WELLCOME FUND

1848 SOCIETY

AAAS



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

1848 SOCIETY

AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA