Satellite swarms trigger a battle to save the night sky p. 142

Prolonging the benefits of deep brain stimulation pp. 153 & 201

Microbial-derived hormones fuel prostate cancer pp. 154 & 216

Science

$15
8 OCTOBER 2021
science.org

AAAS

BATTERY GUIDES

Chelating ligands efficiently shuttle magnesium ions

pp. 156 & 172

EMPOWER WITH EVIDENCE

AAAS | ANNUAL MEETING

Join us for the hybrid 2022 AAAS Annual Meeting, held in Philadelphia, and online, February 17-20.

AAAS requires that all participants in its in-person meetings be fully vaccinated.

Be there for lectures from these renowned speakers during the hybrid #AAASmtg!

Deepakraj M. Divan
Director, Center for Distributed Energy, Georgia Tech

Katalin Karikó
Senior Vice President, BioNTech RNA Pharmaceuticals
Adjunct Professor, University of Pennsylvania School of Medicine

Susan Amara
President, AAAS
Chair, 2022 Annual Meeting

John O. Dabiri
Centennial Professor of Aeronautics and Mechanical Engineering, California Institute of Technology

Andrea M. Ghez
Professor of Physics and Astronomy, University of California, Los Angeles

Katie Hinde
Associate Professor, School of Human Evolution and Social Change, Arizona State University

Nadya Mason
Rosalyn S. Yalow Professor of Physics, University of Illinois at Urbana-Champaign

Kate Starbird
Associate Professor, University of Washington, Human Centered Design & Engineering

Robin Nelson
Associate Professor, School of Human Evolution and Social Change, Arizona State University

Kyle Whyte
George Willis Pack Professor, School for Environment and Sustainability, University of Michigan

Audra J. Wolfe
Writer and Editor

aaas.org/meetings | #AAASmtg

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

BII | Prize for Science | 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

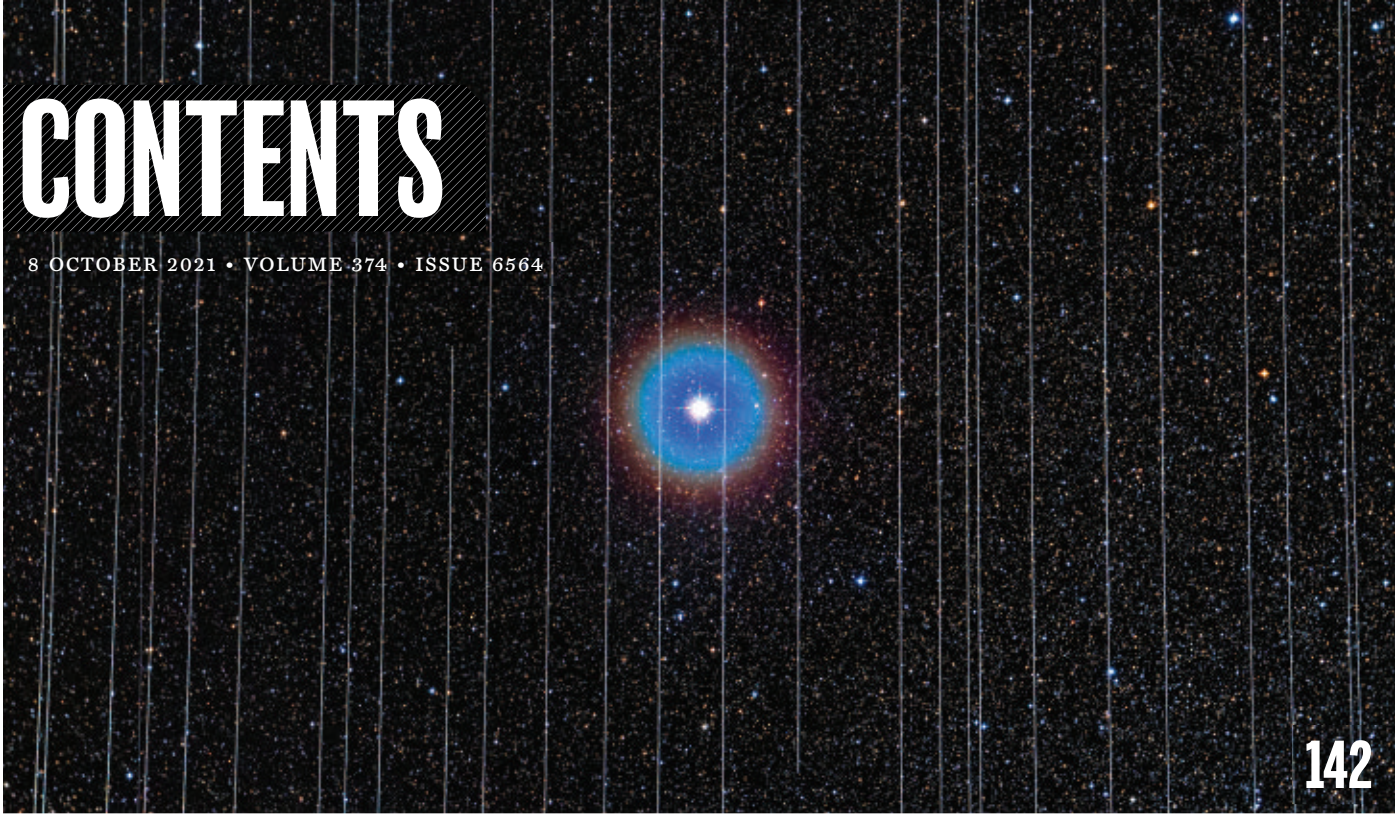
Presented by BII & Science

BII BioInnovation Institute Science

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

CONTENTS

8 OCTOBER 2021 • VOLUME 374 • ISSUE 6564



The bright tracks of satellites in low-Earth orbit mar this 2.5-minute exposure of the double star Albireo taken in December 2019.

PHOTO: RAFAEL SCHMALL

NEWS

IN BRIEF

132 News at a glance

IN DEPTH

135 Pandemic enters transition phase—but to what?
High vaccination rates in some countries may soon make COVID-19 an "endemic" disease
By K. Kupferschmidt

136 Asteroid tour could reveal Solar System's chaotic origin
NASA's Lucy mission will survey Jupiter's Trojans, thought to have been swept up by migrating gas giants *By P. Voosen*

138 Stepping down, NIH director says it's time for 'new vision'
Navigating rough political waters and a pandemic, Francis Collins greatly expanded biomedical research agency
By J. Kaiser

139 Megastudy yields brain cell inventory
Census of motor cortex aims to standardize definitions of cell types *By K. Servick*

140 Plan aims to slash toll from bacterial meningitis
New five-in-one vaccine could help end epidemics in Africa and cut deaths worldwide by 70%, WHO says *By L. Roberts*

141 Vikings in paradise: Did the Norse settle the Azores?
Seafarers may have landed on lush archipelago more than 1000 years ago, sediment cores suggest *By M. Price*

FEATURES

142 The fault in our stars
Satellite swarms are threatening the night sky. Is low-Earth orbit the next great crucible of environmental conflict? *By J. Sokol*
PODCAST

INSIGHTS

PERSPECTIVES

148 Rare metal, precious adhesion
Mussels use microfluidic assembly and vanadium to create an adhesive that anchors them in their natural environment
By J. J. Wilker
REPORT p. 206

150 Pathways to tough yet soft materials
Chemical and physical strategies allow polymer chains in hydrogels to extend or slip
By N. Bosnjak and M. N. Silberstein
REPORTS pp. 193 & 212

152 Filling in the GAPs in understanding RAS
A newly identified regulator increases the efficacy of a new class of targeted anti-RAS drugs *By A. D. Cox and C. J. Der*
REPORT p. 197

153 Revisiting brain stimulation in Parkinson's disease
Electrical burst stimulation rescues movement effectively in parkinsonian mice *By C. A. Haas*
REPORT P. 201

154 Gut bacteria enable prostate cancer growth
Testosterone-synthetizing gut bacteria drive resistance to therapy
By J. A. McCulloch and G. Trinchieri
REPORT P. 216

156 Chelated electrolytes for divalent metal ions
Chelating electrolytes reorganize ion solvation and enable reversible magnesium batteries *By P. Zuo and G. Yin*
RESEARCH ARTICLE P. 172

157 Edmond Fischer (1920–2021)
A biochemical virtuoso
By J. D. Scott et al.

POLICY FORUM

158 Intergenerational inequities in exposure to climate extremes
Young generations are severely threatened by climate change *By W. Thiery* et al.

BOOKS *ET AL.*

161 The infectious diseases that shaped us
An ambitious history traces humanity's entanglements with communicable illness *By G. J. Morgan*

Gordon Research Conferences

Advancing the Frontiers of Science and a Commitment to Diversity

Did you know that GRC offers two diversity fellowship programs for eligible applicants?

The Carl Storm Underrepresented Minority Fellowship (CSURM)

This program was established by the GRC Board of Trustees to honor Director Emeritus Carl Storm's many years of service to GRC and to support his commitment to increasing diversity at the conferences. This fellowship program provides an opportunity for URM graduate students, postdocs, faculty and scientists to receive funding assistance to attend their first GRC.

The Carl Storm International Diversity Fellowship

This program was established by the GRC Board of Trustees to honor Director Emeritus Carl Storm's many years of service to GRC and to support his commitment to increasing diversity at the conferences. This fellowship program allows conference chairs to nominate graduate students and postdocs from China, Africa, India, Central America, South America, Russia, and Ukraine. Funding is available on a first-come, first-served basis.

Apply now for 2022 conferences!
By completing a brief form, eligible applicants can obtain additional information about these funding opportunities!

Visit www.grc.org today!

162 Adam, Eve, and the evolution of humankind
A theologian looks to science for clues about the Bible's first couple *By S. Schaffner*

LETTERS

163 Ensuring recovery for the Przewalski's gazelle
By D. Liang et al.

163 Australia's plan to reduce plastic waste falls short
By T. M. Adyel and P. I. Macreadie

164 Threatened freshwater fish need protection
By W. Chen et al.

RESEARCH

IN BRIEF

165 From *Science* and other journals

REVIEWS

168 Nanophysics
Levitodynamics: Levitation and control of microscopic objects in vacuum
C. Gonzalez-Ballestero et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG3027

169 Ecology
Evasion of tipping in complex systems through spatial pattern formation
M. Rietkerk et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ0359

CREDITS: (PHOTO) KEVIN BRITLAND/UIG/GETTY IMAGES; (ILLUSTRATION) ROBERT NEUBECKER

RESEARCH ARTICLES

170 Circadian rhythms
Reconstitution of an intact clock reveals mechanisms of circadian timekeeping
A. G. Chavan et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD4453

171 Coronavirus
Cross-reactive $CD4^+$ T cells enhance SARS-CoV-2 immune responses upon infection and vaccination
L. Loyal et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABH1823

172 Batteries
Solvation sheath reorganization enables divalent metal batteries with fast interfacial charge transfer kinetics *S. Hou* et al.
PERSPECTIVE P. 156

178 Chemical physics
Conformer-specific photochemistry imaged in real space and time
E. G. Champenois et al.

182 Paleogenomics
Ten millennia of hepatitis B virus evolution
A. Kocher et al.
PODCAST

REPORTS

189 Metallurgy
Grain boundary velocity and curvature are not correlated in Ni polycrystals
A. Bhattacharya et al.

193 Polymer science
Toughening hydrogels through force-triggered chemical reactions that lengthen polymer strands *Z. Wang* et al.
PERSPECTIVE p. 150; REPORT p. 212

197 Cancer
The G protein signaling regulator RGS3 enhances the GTPase activity of KRAS
C. Li et al.
PERSPECTIVE P. 152

201 Neuroscience
Population-specific neuromodulation prolongs therapeutic benefits of deep brain stimulation *T. A. Spix* et al.
PERSPECTIVE P. 153

206 Metals in biology
Microfluidic-like fabrication of metal ion–cured bioadhesives by mussels *T. Priemel* et al.
PERSPECTIVE p. 148

212 Polymer science
Fracture, fatigue, and friction of polymers in which entanglements greatly outnumber cross-links *J. Kim* et al.
PERSPECTIVE P. 150; REPORT P. 193

216 Cancer microbiome
Commensal bacteria promote endocrine resistance in prostate cancer through androgen biosynthesis *N. Pernigoni* et al.
PERSPECTIVE P. 154

225 Topological optics
Topological phonon-polariton funneling in midinfrared metasurfaces
S. Guddala et al.

DEPARTMENTS

125 Editorial
A Decade for restoring Earth
By Bernardo B. N. Strassburg

230 Working Life
The confidence to question
By Samantha J. Butler

ON THE COVER

Artist's rendering of battery electrolytes using amine-based chelants to promote intercalation and deintercalation of magnesium (Mg) ions (yellow) in high-voltage manganese oxide cathodes. The chelants also facilitate highly reversible metal plating and stripping at the Mg anode. Solvation sheath reorganization and electrode-electrolyte interface modification led to an energy density in the Mg ion batteries of 412 watt-hours per kilogram at electrode level. See pages 156 and 172. *Illustration: V. Altounian/*Science*; Data: Xiao Ji, Singyuk Hou, Oleg Borodin*

Science Careers 228

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.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

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

A Decade for restoring Earth

Next week, the United Nations Biodiversity Conference (COP 15) convenes in Kunming, China, to discuss a new global plan to conserve nature and improve the sharing of its benefits across humankind. If such a plan is successfully implemented, it will mark a turning point in the millennia-long decline of the natural world due to human activity. This prospect has been boosted by the United Nations General Assembly decision to declare 2021–2030 as the Decade of Ecosystem Restoration (the "Decade"), which aims to prevent, halt, and reverse the degradation of ecosystems across the planet.

As the world reels from the COVID-19 pandemic and its associated socioeconomic tragedies, it is confronted by worsening climate change and the extinction of life forms on Earth. Both are ranked in the top four threats to the global economy by the World Economic Forum, compound global health and social crises, and threaten vulnerable people the most. A message of hope and the promise of healing Earth could not come at a better time.

"A message of hope and the promise of healing Earth could not come at a better time."

The massive scale of action envisioned over the Decade is key to this promise. The multiple targets it aims to support amount to 1 billion hectares of restoration on land—equivalent to the area of the United States—while calling for similar efforts in marine areas. If achieved, this recovery would amount to the fastest reshaping of Earth's surface by humans ever achieved—this time, in a direction of harmony with the rest of life on the planet. These actions can deliver one-third of the climate change mitigation needed by 2030 while preventing two-thirds of projected global species extinctions. It makes economic sense too, because each US dollar invested in ecosystem restoration generates $30 in economic benefits. Implemented with appropriate social participation—including but not limited to the free, prior, and informed consent of local communities—the conservation and recovery of nature provides multiple social and cultural benefits.

These goals are realistic. Scientific studies have shown that with proper planning and implementation, we can reconcile agricultural production alongside substantial increases in the area and quality of natural ecosystems. Indeed, these are increasingly synergistic goals, with conserved and restored natural lands enhancing the resiliency of lands for farming and urban areas (for example, by improving water security and pollination and reducing disaster risks). This is a welcomed benefit given the predicted increase in extreme climate events in the coming decades.

Fortunately, the science and practice of ecological restoration has matured, providing a growing body of evidence and practical techniques tailored to local conditions. We have a better understanding of how to tap into nature's ability to heal itself through natural regeneration, which slashes recovery costs. New arrangements can combine the restoration of ecosystems with the provision of marketable goods and services, making such recovery in certain conditions more financially attractive than alternative uses, such as low-yield pasturelands.

We need, however, to give the Decade attention and investment commensurate with its potential. The fastest anthropogenic reshaping of Earth must have at least the same amount of cross-sectoral mobilization as the processes that led—and still lead—to a reshaping in the other direction. The 12,000-years-long process of converting natural into anthropic lands was made possible by combined economic, political, cultural, and innovation forces. We need to redirect such resources away from degradation-inducing activities toward restorative ones. Steered correctly, the massive impending public investment in post–COVID-19 recovery could serve as the catalyst for such transformative change. Broad political and societal mobilization; policies that foster enabling conditions; financial support, including innovative mechanisms or subsidies when needed; and targeted investments in science and technology are among the needs if we are to fulfill the promises of this Decade.

As I gaze out of my window in Rio into the forest I grew up in, I can see this promise materialized. Few of Tijuca National Park's 3 million annual visitors, more than any other Brazilian national park, will know that these lands were partially coffee plantations until 1850. These 4000 hectares now provide multiple ecosystem services to its surrounding 12 million people while hosting 2000 animal and plant species. That a 19th-century restoration project could deliver such local impact makes us all optimistic as to what this defining 21st-century global movement can achieve.

–Bernardo B. N. Strassburg

Bernardo B. N. Strassburg is an assistant professor in the Department of Geography and the Environment at Pontifical Catholic University of Rio de Janeiro, Brazil, and executive director of the International Institute for Sustainability, Rio de Janeiro, Brazil. b.strassburg@iis-rio.org

PHOTO: AGNIESZKA LATAWIEC

10.1126/science.abm6556

KAUST: An oasis for tackling global challenges

King Abdullah University of Science and Technology (KAUST), the newest university in Saudi Arabia, when viewed on a Google Maps satellite image, appears as a seemingly nondescript checkerboard of edifices and green gardens perched on the eastern edge of the Red Sea. But a 2D map doesn't do this institution justice. Here at this novel oasis of innovation, nestled between Unity and Discovery Boulevards, is a force for transformation, service, knowledge, and inquiry. On a map, you can't hear the crackling of creativity that is taking place in this desert metropolis. You can't see the movement of minds and bodies as scientists from all corners of the globe and diverse fields converge to solve humanity's grand challenges. You can't feel the excitement or enthusiasm or joy that emanates as scholars, researchers, engineers, and tech experts collaborate, explore the universe, and expose its precious secrets.

Despite only 12 years of existence, KAUST has already set a standard for fostering high-impact, revolutionary science and engineering research. It has been recognized as one of the fastest-growing research institutions in the world. Its star faculty, recruited from prestigious international universities, publish in top-tier journals, and the KAUST faculty citation count is among the highest in scholarship. There are no departments here. Instead, the university is clustered around research centers correlated with KAUST's strategic initiatives, including energy, water, climate and the environment, AI, and data science. Its core labs are robust and advanced, decked out with the latest, most powerful technologies and capabilities to enable cutting-edge investigations into high-risk/high-reward realms of science, technology, engineering, and mathematics (STEM).

Led by President Tony Chan, a mathematician formerly of the University of California, Los Angeles, and the U.S. National Science Foundation, KAUST has blossomed as a beacon of innovation both in the Middle East and on the global stage. "We want to transform Saudi Arabia and the region through world-class

IMAGES PROVIDED BY KAUST

science and technology," says Provost Lawrence Carin, previously of Duke University in the United States. "That mission cannot be achieved in an ivory tower alone. It is important to create the ideas, but we need to bring those ideas to the world. We have a very robust commercialization and entrepreneurship program at KAUST, and our faculty have access to early-stage funding." With 2,697 patent applications, including 161 U.S. patents granted, the university has invested heavily in innovation mechanisms, including an incubator and entrepreneurial education.

KAUST's DNA engenders unique contributions across diverse STEM fields. It is a place where recruitment efforts focus more on attracting the best minds and less on filling a particular disciplinary need, a strategy that has delivered exceptional return on investment. For example, Mark Tester, a professor of plant sciences and associate director of the Center for Desert Agriculture, has demonstrated that tomato plants can grow in salt water. If desalination is not needed for the growth of other crops, this monumental breakthrough has significant ramifications for global food security. To advance and commercialize this research, Tester and his team at KAUST launched a spinoff called Red Sea Farms, with support from the university. Jorge Gascon, a chemical engineering professor and director of the KAUST Catalysis Center, is spearheading an effort to make the Kingdom the world leader in hydrogen energy, and has developed technology that quite literally sucks excess carbon from the air. "Energy is the business of Saudi Arabia," says Carin. "We are trying to be global leaders to mitigate the adverse effects of fossil fuels and lead the future of alternative energy sources."

There are advancements in the education space as well. KAUST is the first university in the Middle East to offer postdoctoral appointments and currently has 400 of these trainees on campus. This decision to initiate a postdoc program is "a consequence of the nature of KAUST and our ambition to produce a world-class STEM workforce," says Carin. "It was transformative for the Kingdom and an entirely different model. It changes the landscape and creates a workforce of people to be hired into domestic companies and universities. Those postdocs are vital to the transformation we are trying to accomplish in the region."

KAUST provost Lawrence Carin: On a mission to support research in service of humanity

In January 2021, Lawrence Carin, an internationally recognized electrical engineer, expert in AI and machine learning, and university leader, became provost of KAUST, a research-focused university based in Thuwal, Saudi Arabia. With a dedication to academic excellence and a flair for facilitating unique, interdisciplinary research programs and collaborations, Carin set to work to shine a light on this unique institution and the outstanding science its faculty are pursuing in service of humanity.

Carin was previously vice president for research at Duke University and a successful serial entrepreneur, but the choice to come to KAUST was an easy one. "The mission of KAUST is inspiring," he shares. "Our goal is to help transform the Middle East and Saudi Arabia through research and education. King Abdullah, for whom the university is named, has a very inspiring vision for the university." That vision has catalyzed a hiring frenzy in recent years, with world-renowned scholars choosing to build or expand their research programs here.

KAUST leverages data to evaluate faculty excellence, says Carin. "Data shows that in almost every case, when faculty were first recruited to KAUST, they were very fine scientists but had not achieved the full impact of their work. But look at their time here and you see their rate of productivity and creativity accelerated dramatically. We can show quantitatively that the model of KAUST is not just attracting really talented people, it is helping to develop these people."

KAUST's service-oriented mission, which ranges from advancing human health and society to elucidating the enigmas of the universe, makes it a wellspring of innovation. There are many aspects of its core tenets that propel discovery. The first principle is how research is executed here, one that aligns with incentivizing high-risk/high-reward endeavors. Rather than have its faculty rely on traditional granting schemes, the university provides baseline funding for all professors. "Grant writing is a time-consuming process, and more often, the proposals that are funded are those that are less risky and more mainstream," notes Carin. "At KAUST, every investigator is provided a substantial amount of funding with no strings attached, with which to take risks. There is tremendous freedom."

Additionally, KAUST takes a unique approach to research structure and personnel organization. While KAUST does have academic programs to fulfill its educational objectives, "there are no departments in the usual sense, so there are fewer walls and this, by its nature, encourages interdisciplinary work," Carin explains. There are 11 research centers that provide thematic focus and coalesce teams to pursue goal-oriented investigations in water desalination, solar energy, advanced membranes and porous materials, and eight other areas critical to the Middle East. And then there are the core facilities, which were "built and designed commensurate with the finest research labs in the world," Carin notes. One example is KAUST's supercomputer, which when first installed was one of the top 10 most powerful supercomputers in world. "Our labs are among the very finest. They have incredible tools to push the frontiers of science and to undertake work that would be unable to be done at any other university."

KAUST takes support for its people very seriously. Active mentorship alliances are prioritized to shape the next generation of scientific leaders. Empowerment is also a treasured principle at KAUST, says Carin, especially for women, as the first multigendered university in the Kingdom. Carin recognizes this awesome opportunity: "33% of our research community are women, and this is high globally. But we want to do better. We are working very hard to make KAUST a place that is very welcoming for women. We have great women leaders. We champion women and showcase how they make an impact, which comes back to the transformation of the region. This empowerment is vital."

"We are tackling some of the biggest challenges," says Carin. "And between the unique funding model, tremendous freedom, no departmental boundaries, strategic centers that foster creativity and interdisciplinarity, and some of the finest core facilities in the world, we can do science not just for science's sake–we can make it transformational, for all."

IMAGES PROVIDED BY KAUST

Transforming ideas into reality

Below are just some of the researchers who are carrying out KAUST's mission of advancing science and technology and catalyzing innovation, economic development, and social prosperity at home and around the world.

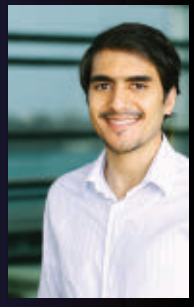
Husam Alshareef

Professor Alshareef is developing nanomaterials for energy storage, electronics, and sensors. Applications of his group's current projects include mobile ion batteries (Li, Na, Zn, and K), electrochemical supercapacitors and microsupercapacitors, wearable electronics and sensors, transparent electronics, and thermoelectric materials. He is veteran of the semiconductor industry who has put materials in volume production at companies such as Texas Instruments.

Mohamed-Slim Alouini

Professor Alouini's research addresses inequities in global distribution and how this impacts access to and use of information and communication technologies. His group has developed aerial and space networks to provide connectivity to remote geographical areas with low populations and uses diverse technologies such as multihop/cooperative communications systems, optical wireless communication systems, cognitive radio systems, and green communication systems.

Osman Bakr

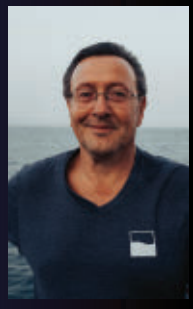


Professor Bakr's research is focused on the physics and chemistry of hybrid materials. His group studies the synthesis and assembly of organic–inorganic hybrid materials and nanomaterials with novel optical and electronic properties. These materials can be used to fabricate advanced material building blocks for solar cells and optoelectronic devices.

Stefaan De Wolf

Professor De Wolf's expertise lies in photovoltaics for terrestrial applications. His research focuses on the fabrication of high-efficiency silicon-based solar cells, especially the fundamental understanding of interface structures and electrical contact formation relevant to solar cells and electronic devices. He is also interested in new device architectures and applications, such as perovskite–silicon tandem solar cells, to capture the full solar spectrum for electricity generation.

Carlos Duarte



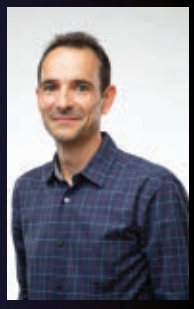
Professor Duarte is a world leader in multiple branches of biological oceanography and marine ecology. His work ranges from the tropics to polar ecosystems, from macrophytes to microbes, and from coastal systems to open ocean gyres, using all types of approaches. His research has stoked discussion on the heterotrophic nature of oligotrophic systems and identified the role of hypoxia thresholds for marine biodiversity.

Mohamed Eddaoudi

Professor Eddaoudi is developing new strategies for the design and synthesis of functional solid-state materials from molecular building blocks and advancing the understanding of the structure–property relationship of materials. His group works toward developing made-to-order metal–organic frameworks (MOFs) that address some of our most challenging and enduring societal needs, such as clean energy alternatives, reduced greenhouse gas emissions, and remediation of chemical and biological threats.

IMAGES PROVIDED BY KAUST

Jorge Gascon



Professor Gascon's research interests fall at the intersection of chemical engineering and materials science, including the development and validation of new nanostructured materials and composites. Applications of his work include multifunctional catalysis, with special emphasis in C1 chemistry.

He has received the ExxonMobil Chemical European Science and Engineering Award for his work on separation techniques and separation modeling in petrochemistry and engineering of metal–organic frameworks for gas separation.

Niveen Khashab

Inspired by the model of evolutionary biological systems, Professor Khashab's research group is designing and synthesizing supramolecular assemblies at the nanoscale, employing non-covalent and coordination interactions. The main advantage of these molecularly defined entities is their superior solution processability, allowing them to be easily and reproducibly integrated in future commercial fabrication processes. These systems are designed with an emphasis on hierarchical assembly (evolution), porosity (ordered self-assembly), and stimuli responsiveness (smart materials).

Yu Han

Professor Han's research interests at KAUST include the synthesis of nanoporous and nanostructured materials, the resolution of their complicated structures and the development of novel applications. These structures are tailored for use in heterogeneous catalysis, gas separation and capture, water harvesting, and energy storage and conversion. His group is also developing low-dose electron microscopy techniques for real-space direct imaging of extremely sensitive crystalline materials.

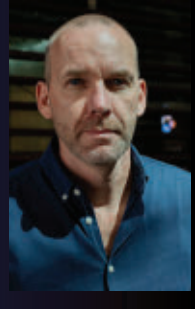
Pierre Magistretti

Professor Magistretti has made significant contributions to the field of brain energy metabolism, discovering cellular and molecular mechanisms that underlie the coupling between neuronal activity and energy consumption by the brain. This work has considerable ramifications for understanding the origin of the signals detected by functional brain imaging techniques used in neurologic and psychiatric research.

Peiying Hong

Professor Hong investigates molecular microbiology and microbial aspects in water and wastewater ecosystems. Her research aims to understand the roles and interactions of microorganisms in these ecosystems and to apply these insights to solving issues related to water quality and water reuse. Another focus of her research is biotic contaminants present in natural and engineered environments, including antibiotic resistance genes, mobile genetic elements, and pathogens.

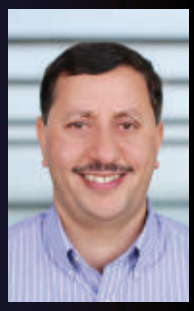
Matthew McCabe



Professor McCabe is focused on water and food security, climate change impacts, precision agriculture, water resources monitoring and modeling, and the novel use of technologies for enhanced Earth system observation. Using modeling and observational approaches, his group investigates the distribution, variability, and exchange of water at local, regional, and global scales, as well as interactions with vegetation.

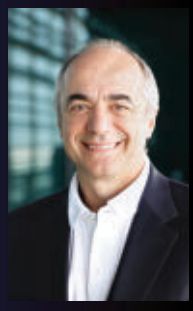
IMAGES PROVIDED BY KAUST

Omar Mohammed



Professor Mohammed's research is directed toward the fundamental understanding of carrier dynamics in a variety of solar-cell systems, including semiconductor quantum dots, polymers and perovskite solar cells (with the aid of cutting-edge nanotechnology), ultrafast laser spectroscopy, and four-dimensional electron imaging. These investigations will provide valuable input for the engineering and optimization of solar cell devices.

J. Carlos Santamarina



Professor Santamarina explores the scientific foundations of geomaterial behavior and subsurface processes using particle-level and pore-scale experiments, combined with numerical methods and high-resolution process-monitoring systems. This conceptual and experimental framework is advancing the study of phenomena and the development of solutions in energy geotechnology for resource recovery, energy geostorage, efficiency and conservation, and energy waste.

Boon Ooi

Professor Ooi studies semiconductor lasers and photonic integrated circuits. He has developed practical technologies for semiconductor photonics integrated circuits, broadband semiconductor lasers, multiple-wavelength lasers, and superluminescent diodes. His recent interest is areas of gallium nitride (GaN)-based nanostructures and lasers for applications such as solid-state lighting and visible light communications.

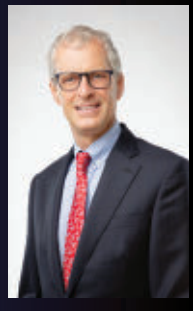
Mani Sarathy

With a passion for developing sustainable energy technologies with decreased net environmental impact, Professor Sarathy is focusing on simulating the combustion chemistry of transportation fuels. He is developing fundamental chemical kinetic models to simulate fuel combustion and pollutant formation in energy systems. The models can be used by engine designers to achieve various performance targets and can illuminate how the chemical structure of a fuel affects pollutant formation.

Håvard Rue

Professor Rue works on computational Bayesian statistics and Bayesian methodology such as priors, sensitivity, and robustness. His main body of research is built around the R-INLA (R-integrated nested Laplace approximation) project (www.r-inla.org), a practical tool to approximate Bayesian analysis of latent Gaussian models, often at extreme data scales. His work on stochastic partial differential equations enables representation of Gaussian fields for use in spatial statistics.

Mark Tester



Professor Tester is elucidating the molecular genetic mechanisms that enable certain plants to thrive in suboptimal conditions, such as high salinity or temperature, and then deliver the outputs in economically viable systems. His group aims to increase the global productivity of crops in conditions of challenging abiotic stress and to unlock seawater by developing an agricultural system where salt-tolerant crops are irrigated with partially desalinized seawater or brackish groundwater.

Magnus Rueping

Professor Rueping's research is directed toward the development and simplification of synthetic catalytic methodologies and technologies and their application in the rapid synthesis of diverse functional molecules. Current projects focus on metal- and organocatalysis as well as photo- and flow catalysis including metal catalysis and biocatalysis, new reaction methodology and technology, smart materials, photochemistry, and flow chemistry.

Sponsored by

جامعة الملك عبدالله للعلوم والتقنية

King Abdullah University of Science and Technology

www.kaust.edu.sa

IMAGES PROVIDED BY KAUST

NEWS

> “Long periods of heavy restrictions [have] not got us to zero cases.”
>
> **Jacinda Ardern,** New Zealand’s prime minister, on her government’s easing of COVID-19 lockdowns, which had kept case numbers low until the Delta variant spread.

IN BRIEF

Edited by **Jeffrey Brainard**

CLINICAL RESEARCH

Merck unveils first pill to treat COVID-19

An experimental antiviral drug with a controversial history may become the first oral medication to help combat the pandemic. Last week, Merck & Co. announced that treating people recently infected with SARS-CoV-2 twice a day for 5 days with a pill called molnupiravir reduces their risk of hospitalization and death by 50%. The compound, which prevents RNA viruses like SARS-CoV-2 from copying their genes, was in clinical trials for flu before the pandemic and emerged from efforts to treat a virus that causes encephalitis. Last year, a whistleblower argued that its original biotech backer inappropriately sought federal money for COVID-19 clinical trials. And in an earlier study, molnupiravir did not help patients already hospitalized. But it proved so effective in recently infected people that an independent oversight board stopped Merck’s clinical trial early. The study included 775 nonhospitalized people who enrolled within 5 days of their first COVID-19 symptoms and had at least one risk factor for developing severe disease. The results have not yet been peer reviewed, and Merck has not announced a price. Molnupiravir joins two other antiviral options already available: remdesivir, which some studies have shown to reduce hospital stays in some patients, and monoclonal antibodies. Unlike molnupiravir, both must be given intravenously or by injection.

NASA keeps Webb name on scope

ASTRONOMY | NASA last week decided not to rename a flagship astronomy mission, the James Webb Space Telescope, after an investigation into claims that Webb, NASA’s second administrator in the 1960s, was complicit in efforts to root out and fire gay and lesbian government employees. In the early 2000s, former NASA Administrator Sean O’Keefe chose to memorialize Webb for his leadership of the Apollo program and promotion of the agency’s science. But earlier this year, four astronomers argued that documents in the U.S. National Archives indicate Webb, while at the Department of State, passed homophobic material to senators during the “lavender scare” that began in the 1940s. Hundreds of astronomers signed a petition calling for a renaming; it noted that while Webb was NASA administrator, the agency fired an employee suspected of being gay. The petition calls for NASA to choose “a name that stands for a future in which we are all free.” In a statement, NASA Administrator Bill Nelson said: “We have found no evidence at this time that warrants changing the name of the James Webb Space Telescope.”

Medical group hushes ouster

SCIENCE ACADEMIES | When the U.S. National Academy of Medicine (NAM) told its 2235 members last year that it had ousted a physician for scientific misconduct, it added a plea: “Please keep this information highly confidential, only for members.” The notice, recently leaked to *Science*, appeared in the members-only section of NAM’s website in December 2020. It said that 2 months earlier, NAM rescinded the membership of Eric Noji for offenses including “fabrication, falsification, and plagiarism.” (Under NAM’s 2019 Code of Conduct, members can also be expelled for sexual harassment or bullying.) Critics say NAM’s public silence may chill reporting of complaints to the academy and ignores its own landmark 2018 report, published with the other U.S. national science academies, which called for research institutions to increase transparency and accountability around behavioral standards and sexual harassment. NAM President Victor Dzau

PHOTO: © 2009-2021 MERCK SHARP & DOHME CORP., A SUBSIDIARY OF MERCK & CO., INC., KENILWORTH, N.J.

An exhibit about neurons explores topics such as vision and addiction.

PHOTO: ARTECHOUSE

SCIENCE AND ART

'Thinking cells' star in 'immersive' multimedia show

A 20-minute, kaleidoscopic, floor-to-ceiling, multimedia presentation about the birth and death of a human neuron is the centerpiece of an exhibit that debuted in Washington, D.C., last week. *Life of a Neuron*, created with the Society for Neuroscience, takes visitors on a tour of the vital cells inside their skulls. It concludes with a dangling neuron, stripped of most of its appendages, accompanied by the haunting sound of a hospital flatline. Other interactive parts of the exhibit capture approaching visitors with cameras and other sensors to mimic how the brain processes visual information, responds to stress, and changes as a result of addiction. The paid event, part of a trend in museums offering large-scale, "immersive" exhibits, runs through 28 November at ARTECHOUSE, an exhibition space focusing on art, science, and technology.

defended the public silence about Noji to *Science*: Announcing an ouster to the members, he said, is sufficient to deter misdeeds as well as discourage the nomination of others with track records of bad behavior for membership in the prestigious academy.

Scientists slam Mexico arrest bid

JUSTICE | Academics in Mexico and elsewhere last week decried a request by Mexico's attorney general for arrest warrants for 31 scientists and administrators on corruption charges even though a judge turned down two previous such requests. The federal Attorney General's Office alleges that the 31, affiliated with the independent Scientific and Technology Advisory Forum (FCCyT), committed organized crime while embezzling and misusing about $12 million they received from Mexico's national research agency. The International Human Rights Network of Academies and Scholarly Societies issued a statement calling the accusations unfounded harassment. Some suspect the prosecution is political in nature: FCCyT has strongly criticized a series of science policy decisions by the administration of President Andrés Manuel López Obrador, including a draft law that FCCyT says would reduce academic freedom, and cuts to funding and scholarships.

New flood maps for world, U.S.

HYDROLOGY | The United Nations last week unveiled a new, public tool that creates street-level maps of floods worldwide since 1985, which may help planners reduce deaths and property damage from future deluges. Detailed maps such as those in the World Flood Mapping Tool have been rare, especially in the Global South where devastating floods have struck in recent years. Worldwide, an estimated 1.5 billion people face hazards from intense floods, but most national flood forecasting centers lack the ability to run complex predictive models, a U.N. study found. In a related development, the U.S. Federal Emergency Management Agency (FEMA) said last week it is raising premiums for its National Flood Insurance Program to more accurately reflect actual risks, including those posed by floods worsened by climate change. The higher rates were shaped by a scientific analysis released last year that used new techniques to create an accurate flood-risk map of the contiguous 48 states. The analysis revealed that 14 million U.S. homes are at high risk of flooding; FEMA's old map only identified 8 million at high risk.

CRISPR eases a genetic blindness

CLINICAL RESEARCH | The first attempt to use the CRISPR gene-editing tool to treat a genetic form of blindness is showing hints of success. Two of six people who received the experimental therapy were able to better perceive light. One can now even navigate a maze, researchers reported last week at the International Symposium on Retinal Degeneration in Nashville, Tennessee. The ongoing phase 1/2 clinical trial, sponsored by Editas Medicine, began in March 2020 as the first to inject genetic instructions

THE NOBEL PRIZES

Medicine prize awarded for research on sensing hot, cold

Discoveries that have helped explain why chili peppers taste hot and mint tastes cool—and revealed proteins that let us sense a gentle breeze or a sharp poke—have earned two scientists this year's Nobel Prize in Physiology or Medicine. David Julius at the University of California, San Francisco, and Ardem Patapoutian at Scripps Research and the Howard Hughes Medical Institute shared the award, unveiled this week. In the 1990s, Julius led research that discovered a sensor in the skin's nerve endings that detects hot temperatures and spicy food. A few years later, working independently, both Julius and Patapoutian found a cell surface receptor that reacts to cold—and to menthol, which makes mint taste minty cool. Patapoutian also led work that discovered key proteins that enable nerve cells to sense pressure. Based on this work, they and other researchers are now working to develop therapies that target the receptors that lead to pain.

Work on climate, atoms' behavior wins physics prize

Work to understand how carbon dioxide drives global warming and research into the seemingly chaotic behavior of atoms in metallic lattices are honored with this year's Nobel Prize in Physics. Half the prize went to Syukuro Manabe of Princeton University and Klaus Hasselmann of the Max Planck Institute for Meteorology for developing models that, despite the variability of weather, made concrete, long-term predictions about Earth's climate and showed how human activity is causing global heating. The other half of the award went to Giorgio Parisi of the Sapienza University of Rome, for finding order in disordered materials such as metal alloys. Despite the apparent randomness of the atoms in the materials, he showed it was possible to identify underlying patterns, work that impacted math, biology, and neuroscience. Members of the Nobel Committee for Physics said the climate focus was a message to world leaders that climate modeling is "solidly based" in physics.

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

for CRISPR's two molecular components directly into the body, instead of modifying patient cells in a dish and returning them. Trial participants have Leber congenital amaurosis type 10, caused by a mutated gene that the trial aimed to fix.

Heat, humidity's health risks grow

CLIMATE SCIENCE | The threat of exposure to extreme heat in cities around the world has tripled since the 1980s, a study this week in the *Proceedings of the National Academy of Sciences* reports. U.S. scientists combined weather and satellite data to estimate the number of days when heat reaches a danger zone—the equivalent of 30°C with 100% humidity—in which people can become sick and die. As temperatures rose during the study period, so did city populations. India had more than half the global increase in extreme heat exposure since 1983.

Lacks family sues over cell line

BIOMEDICINE | The family of Henrietta Lacks, a Black woman whose cervical cancer cells became a widely used research tool without her knowledge, last week sued Thermo Fisher Scientific, a company that commercialized them. Lacks's cells, collected at Johns Hopkins Hospital in 1951, became the "HeLa" cell line, used in fields ranging from infectious disease to genomics. The lawsuit, filed in federal court in Baltimore on 4 October, charges Thermo Fisher with "unjust enrichment," saying the company profited from her tissue despite knowing it was taken without consent. The suit asks the company to transfer "millions of dollars" in profits from the HeLa cell line to her estate. The suit also seeks to block sales of the cell line without the estate's permission.

Wisconsin approves wolf hunt

CONSERVATION | Wisconsin's Department of Natural Resources this week authorized hunting of 130 gray wolves in November. That's less than the quota of 300 backed by a state board of political appointees, but scientists had urged little or no hunting to allow for studies of the population's sustainability—a position that the neighboring states of Michigan and Minnesota have also taken for their gray wolves. Wisconsin's green light for hunting, and the lifting of protections under the federal Endangered Species Act nationwide in January, have contributed to reducing the state's gray wolf (*Canis lupus*) population by 30% from 2020 to 2021, to about 700.

Artist John James Audubon depicted the ivory-billed woodpecker.

ILLUSTRATION: JOHN JAMES AUDUBON/WIKIMEDIA COMMONS

CONSERVATION

Woodpecker among 23 species deemed extinct

U.S. wildlife officials last week said they plan to declare the iconic ivory-billed woodpecker (*Campephilus principalis*) and 22 other species extinct, removing protection under the federal Endangered Species Act (ESA). Previously, the Fish and Wildlife Service had removed just 11 species because of extinction over the ESA's nearly 50-year history. In the 2000s, birdwatchers reported sighting the woodpecker, once the United States's largest, for the first time in decades, but the reports were never fully confirmed. Other species slated for removal include 10 additional birds (most native to Hawaii), eight mussels, two fish species, a bat, and a plant. Many were likely already extinct when the ESA was enacted in 1973, highlighting the importance of protecting wildlife from threats like habitat loss while there is still time, conservationists argue.

A crowd gathers at a Copenhagen nightclub. Denmark dropped all COVID-19 restrictions on 10 September, hoping the virus will remain a tamed threat.

PHOTO: OLAFUR STEINAR GESTSSON/AP IMAGES

COVID-19

Pandemic enters transition phase—but to what?

High vaccination rates in some countries may soon make COVID-19 an "endemic" disease

By **Kai Kupferschmidt**

Life in Denmark now feels so much like it did before the pandemic that it can put visitors on edge, says Lone Simonsen, an epidemiologist at Roskilde University. The country lifted all of its remaining coronavirus restrictions on 10 September. Copenhagen clubs are buzzing, music lovers flock to festivals, and buses are packed with unmasked commuters. The government has given up its power to close schools and shut down the country. "When we have guests now, they feel uncomfortable about how normal everything is," Simonsen says.

Denmark is a pioneer. As the second coronavirus winter approaches in the Northern Hemisphere, Denmark and a few other countries where vaccines now protect a large percentage of the population from severe disease are entering a momentous transition: from pandemic to endemic COVID-19, when the virus is still present but now faces a population that is mostly immune to it. Researchers are closely watching what happens next, because it could yield valuable information about what lies ahead for the rest of the world.

There are many unknowns: how best to manage the transition out of the public health crisis, how it might go wrong, and exactly what endemicity will look like once it arrives. "Going into a pandemic is hard enough, coming out of it is even harder," says Jeremy Farrar, an infectious disease researcher who leads the Wellcome Trust. "We don't just go from a no-vaccine state and horror to a status quo. There's a transition phase, and I think that will be this winter."

Denmark has fully vaccinated more than 88% of people older than 18 and an astonishing 97% of those over age 60, the group most vulnerable to serious COVID-19. That allows the country to try to treat SARS-CoV-2 like influenza and other infectious diseases instead of a threat to the entire health system. "We're thinking of this virus now as a sort of defanged version of the original one. It has gotten its teeth pulled out by the vaccine," Simonsen says. "What's left is not much worse than diseases that we're used to and that we don't close schools for, like seasonal flu or maybe the 2009 influenza pandemic."

That does not mean Denmark is out of the woods. There are still susceptible people, including children, adolescents who aren't yet fully vaccinated, and people in whom the vaccines have not elicited sufficient immunity. Two weeks after the measures were dropped, case numbers in Denmark started to slowly rise, after a monthlong decline; they now stand at about 450 a day in the country of 5.8 million. Simonsen is not alarmed. The continuing spread will result in more natural immunity in unvaccinated children and adults, and vaccine coverage will keep inching upward, she says. "All of this together makes me think Denmark will reach herd immunity in the spring."

Four recently published projections from the Statens Serum Institute (SSI), a governmental public health and research agency, provide some idea of what might happen between now and then. The best-case scenario assumes the country will reach 90% vaccination coverage among those over age 12 and social activities, still below pre-pandemic levels, will not increase further. In that case no surge in cases and hospitalizations is expected, says SSI modeler Camilla Holten Møller. In the worst-case scenario—no further increase in vaccine uptake and a 10% rise in social activities—Denmark could see a wave in cases and hospitalizations as bad as those in December 2020 and January. "We do not expect [that] to happen," Holten Møller says, but if it did the government would likely bring back some restrictions. "The key is to be ready to change one's mind as the data comes out," says Rustom Antia, an immunologist at Emory University.

None of the models included a waning of immunity; Denmark started to give booster shots to vulnerable groups in September and expects to give the entire population a third dose eventually. The models also didn't include a key wild card: the emergence of nasty new variants of the virus.

Norway, with a similarly high vaccine uptake, followed Denmark's lead and lifted all restrictions on 25 September. But many countries around the world do not have access to enough vaccine for even the most vulnerable groups. And even in Europe, awash in vaccine, few countries are as ready as Denmark is to attempt this transition. "If you have a lower coverage in the elderly, you should not try to do what Denmark is trying right now," Simonsen says. Spain has vaccinated 97% of residents over age 60, for example, France 81%, and Greece 77%. Israel, which fully reopened on 1 June, has double-vaccinated about 90% of its population older than 60, meaning that per capita, it has several times as many unvaccinated people in that vulnerable group as Denmark does. It is now struggling to contain a big new wave of cases.

Countries with lower coverage will have to accept a higher level of restrictions during the transition phase to keep the virus at bay, Farrar says. "I think that politicians across the world are sort of pretending that you can have your cake and eat it: You can have zero deaths, no control measures, vaccinate if you want or not vaccinate—and it will all end. I just don't think that's realistic."

> **"We're thinking of this virus now as a sort of defanged version of the original one."**
>
> **Lone Simonsen**, Roskilde University

Farrar has seen that happen in England, which celebrated "Freedom Day" on 19 July, when only about half the population was fully vaccinated. Cases began to trend upward in early August, to well over 30,000 per day now; deaths have risen to an average of more than 100 per day. Keeping in place some public health measures such as wearing masks would have helped keep the burden of disease down, says Farrar, who is part of the U.K. government's scientific advisory group on emergencies. The plan to abolish free testing later this year is mistaken as well, he says. "I don't really understand the tremendous rush to get rid of light touch things like masks," adds David Fisman, an epidemiologist at the University of Toronto.

Meanwhile, scientists are trying to figure out how endemic COVID-19 will behave. "The models are all over the map," Fisman says. "I know lots of smart people who are projecting very different things." The endemic stage probably won't be like endemic measles, which is highly infectious but leaves people immune from infection for life. That results in a childhood disease with a wave every few years, once a large enough new cohort has been born.

With SARS-CoV-2, immunity clearly wanes, Antia says. As a result, it might follow the trajectory of the four endemic coronaviruses—which all cause the common cold—in which protection from infection erodes over time, but protection from severe disease does not. That leads to a pattern of a first infection in early childhood, followed by recurring mild infections—like a case of the sniffles—later in life.

The outlook is different if immunity against severe COVID-19 declines as well. In that case, paradoxically, it might be best if the virus keeps infecting people frequently, Antia and Elizabeth Halloran of the University of Washington, Seattle, argue in a recent paper in *Immunity*. With a lot of virus circulating, those vulnerable to infection but not yet to severe disease are more likely to be infected and have their immunity boosted; otherwise, the population would slowly become vulnerable again. Avoiding infection at all costs might not be the best strategy for vaccinated people, Simonsen says: "I just have this gut feeling for myself that maybe it's not so bad if I meet that virus naturally now that my immune system is strong."

Influenza is a more sobering analogy. It, too, causes infections throughout life, but reinfections are often more severe than those caused by the four endemic coronaviruses because the virus evolves faster and escapes host immunity. And Trevor Bedford of the Fred Hutchinson Cancer Research Center notes that so far, SARS-CoV-2 has evolved about five times faster than H3N2, one of the three influenza types circulating in humans. That rate is likely to slow over time to something comparable to flu, Bedford says. But given SARS-CoV-2's high transmissibility and the low vaccine uptake in the United States, he thinks it might well infect one-third of the U.S. population every year, causing 50,000 to 100,000 deaths, after society fully reopens.

It's a grim vision of endemic COVID-19. Measures like better ventilation and even continuing to test, trace, and isolate could reduce the toll, but Bedford questions the appetite for prolonging those measures, given that the United States routinely accepts 30,000 deaths each year from flu. "We never really did anything about it. People still show up to work sick and so forth," he says. "I don't know if 50,000 people dying a year of COVID would somehow be different." ■

PLANETARY SCIENCE

Asteroid tour could reveal Solar System's chaotic origin

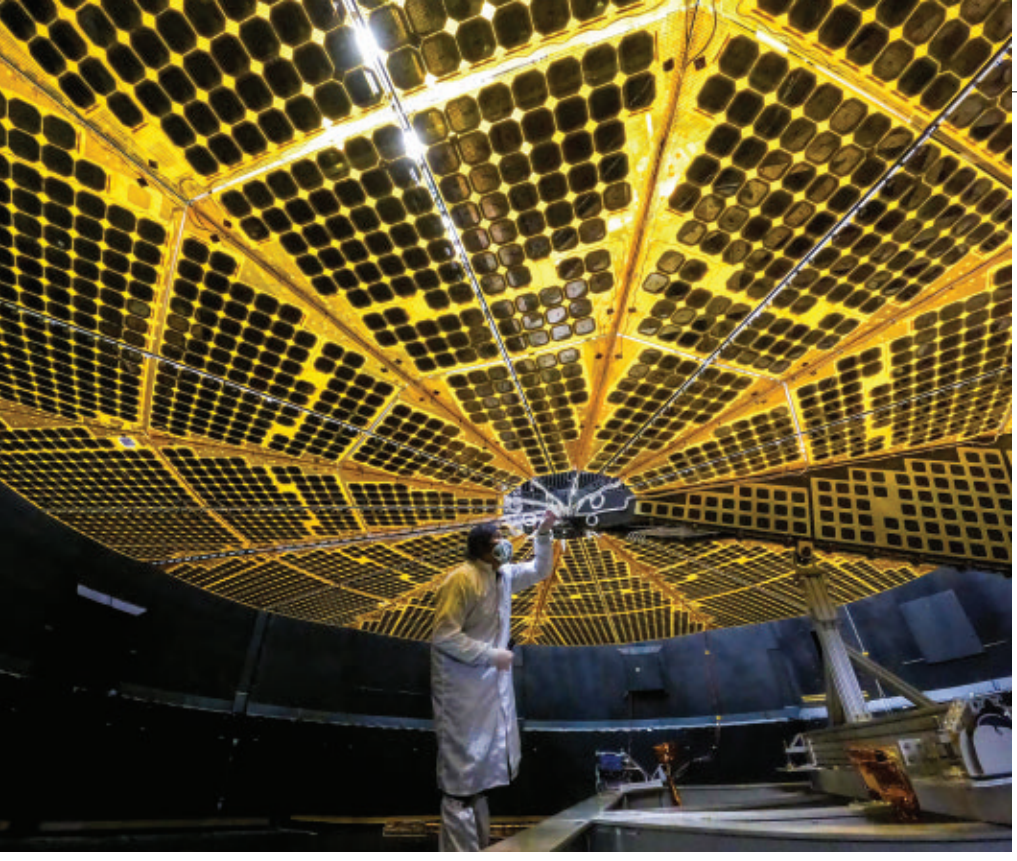
NASA's Lucy mission will survey Jupiter's Trojans, thought to have been swept up by migrating gas giants

By **Paul Voosen**

Jupiter is a planetary bully, tossing other bodies aside with its outsize gravitational tug. But two clusters of asteroids precede and follow Jupiter at a respectful distance, like cowering children, held in a stable place by the pull of the gas giant and the Sun. Astronomers have long been puzzled by these rocks, called the Trojans, which look very different from the objects in the main asteroid belt and settled in the oddest of places, 374 million kilometers in front of and behind Jupiter.

Now, a $981 million robotic NASA mission called Lucy, to be launched on 16 October from Cape Canaveral, Florida, will explore six of these Trojans. During a 12-year tour, Lucy—named after the famed human-ancestor fossil—will test whether the Trojans themselves are fossils of the early Solar System that were dragged inward from the frigid far reaches beyond Neptune during a period of gravitational turmoil. Details of the Trojans' surfaces, shapes, and composition could flesh out not only how this ancient upheaval occurred, but when. Hal Levison, Lucy's principal investigator (PI) at the Southwest Research Institute (SwRI) in Boulder, Colorado, likens their promise to clues in a murder mystery. "Sometimes blood splatter on the walls can tell you more than blood on the corpse. [Asteroids] are really the splatter on the walls."

Decades ago, researchers pictured the formation of the planets as a gentle affair, with Jupiter and the Trojans coalescing near their present location from a disk of gas and dust surrounding the Sun. But telescopes have shown that the Trojans are not uniform in color, as would be expected if they formed from the same local raw material. And rather



Lucy's 7-meter-wide solar panels are needed to collect the dim light available in Jupiter's orbit.

than sitting perfectly in the plane of the planetary disk, as the old model would have predicted, their orbits are slightly inclined.

In 2005, using computer simulations of how newborn planets would have tugged on one another, Levison and others proposed the Trojans were a legacy of early Solar System turmoil. As the idea now goes, gas from the protoplanetary disk initially dampened the effects of the gravitational tugs of the gas giants and kept them in stable orbits much closer to one another than today. Once the gas dissipated, chaos erupted, launching Jupiter inward and the other giants outward. Pluto and other distant icy bodies were flung even farther out into what's called the Kuiper belt, a region beyond Neptune. But the model also predicts that some of those icy bodies would have been scattered inward instead and captured as Trojans by Jupiter.

Although the idea of an instability can explain much, it hasn't really been observationally tested, says Mike Brown, an astronomer at the California Institute of Technology and member of the Lucy team. "The Trojans are where those concrete tests are," Brown says. If they have the same composition as Kuiper belt objects, "there's no way the Solar System was put together without dynamical instability."

To take a close look, Lucy will travel for 6 years, making two Earth flybys to build speed before reaching the leading group of Trojans in 2027. Two circular solar arrays will power the spacecraft, each spanning 7 meters to collect the dim light available in Jupiter's orbit—farther from the Sun than any other solar-powered mission has ventured.

Lucy's first target will be an oddball even for the Trojans: the 64-kilometer-wide Eurybates and its small moonlet, which resemble from afar the dark, carbon-rich asteroids found in the main asteroid belt. Within 1 year, it will then pass Polymele, Leucus, and Orus, which seem similar to icy Kuiper belt objects. Next, Lucy will swoop past Earth for a third gravity assist before flying back to the trailing Trojans. In 2033, it will survey Patroclus and Menoetius, 100-kilometer-wide asteroids orbiting each other. Binaries are oddities in the asteroid belt but common farther out, in the Kuiper belt, says SwRI's Cathy Olkin, the mission's deputy PI. "Could that pair have been captured from the Kuiper belt?" she asks.

A first task will be determining whether the Trojans contain substances that can only form out in the cold twilight of the outer Solar System, like a glassy form of water ice that lacks a crystalline structure. Brown also wants to understand the Trojans' reddish color, a subdued version of the hue seen on many Kuiper belt objects. Brown's best guess for the source of this redness is hydrogen sulfide, which grows red when irradiated by cosmic rays or other sources, but can also evaporate away if brought too close to the Sun. Some Kuiper belt objects on elongated orbits that bring them closer to the Sun "look a lot like Trojans," Brown says.

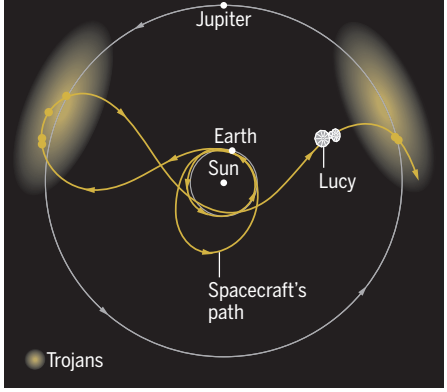
Another task will be counting the number and size of craters on the Trojan surfaces and comparing them with those seen on Pluto's moon Charon and Arrokoth, Kuiper belt objects visited by the New Horizons mission in 2015 and 2019. A similar distribution will suggest they formed in the same spot, and small differences could reveal when the Trojans were torn from their remote homeland, Levison says. Scientists once thought the instability shook up the Solar System some 700 million years after its formation, but some now think it could have happened after only 100 million years, or even earlier (*Science*, 24 January 2020, p. 350).

The Trojans might also provide clues about an even earlier time, when the planets coalesced from primordial gas and dust. Arrokoth's lobed shape fits a model suggesting planetesimals formed when clouds of pebbles in the gassy disk collapsed to form city-size lumps; those might then have merged with others into progressively larger structures. If the Trojans are similarly lumpy, that would be an indication that cloud collapse is the standard way to assemble planetesimals, says Jessica Sunshine, a planetary scientist at the University of Maryland, College Park. The Trojans' relative proximity—just one-sixth as far from the Sun as the Kuiper belt—will make them a ripe target for further exploration, including by NASA's James Webb Space Telescope, which will be launched late this year.

Levison is hoping for surprises. "It'd be damn boring if we just saw what we expected," he says. Lucy's sheer endurance and number of targets will give it plenty of time to follow up on any surprises, Sunshine adds. "We have the opportunity to not only come up with new questions, but maybe even answers." ■

Whistle-stop tour

After launch this month, Lucy will perform two Earth flybys before touring its first four Trojan targets in 2027–28. After swooping past Earth again, it will visit a pair of Trojans in the trailing group in 2033.



CREDITS: (PHOTO) LOCKHEED MARTIN; (GRAPHIC) SOUTHWEST RESEARCH INSTITUTE, ADAPTED BY N. DESAI/*SCIENCE*

Francis Collins directed the National Institutes of Health under three U.S. presidents.

PHOTO: GRAEME JENNINGS/AP IMAGES

SCIENCE POLICY

Stepping down, NIH director says it's time for 'new vision'

Navigating rough political waters and a pandemic, Francis Collins greatly expanded biomedical research agency

By **Jocelyn Kaiser**

After steering the largest biomedical research institute in the world under three presidents, Francis Collins is about to ride his Harley-Davidson into the sunset. On 5 October, the physician-geneticist announced he will step down after 12 years as director of the U.S. National Institutes of Health (NIH) by the end of this year.

"Institutions need new vision and leadership on a regular basis," Collins tells *Science*. And with COVID-19 vaccines now available to stem the pandemic, "I don't think my role is as essential."

Collins, the longest serving NIH director in more than 50 years, says he would have stepped down in January if incoming President Joe Biden had not asked him to stay on, in part to maintain NIH's role in countering the coronavirus pandemic. "He's presided over the brightest part of our COVID-19 response," says Elias Zerhouni, Collins's predecessor as NIH director.

Appointed by then-President Barack Obama in 2009, Collins took the helm of NIH after leading the completion of the human genome project in 2003 as director of NIH's National Human Genome Research Institute (NHGRI). He brought a firm hand to his position, launching "big science" projects that he considers his greatest legacy including a brain mapping effort, then–Vice President Biden's Cancer Moonshot, and the All of Us precision medicine study, as well as a new center focused on translating basic discoveries into treatments.

With folksy charm that disguised his political savvy, the guitar-playing, motorcycle-riding scientist helped convince Congress to expand NIH's budget after 10 years of stagnation. It has grown from $30 billion to $41.3 billion during his tenure. Former President Donald Trump's decision to keep Collins on came as a relief to a research community fearful of whom Trump might choose as director.

Despite qualms from some scientists that his background as an evangelical Christian would influence his decisions, Collins has supported studies using human embryonic stem cells and fetal tissue—although fetal tissue work faced restrictions under Trump. Collins launched NIH efforts to combat sexual harassment, support young researchers, and boost the number of Black investigators funded by the agency. He also led NIH through a controversial probe of hundreds of scientists, mostly of Chinese descent, who allegedly failed to report foreign ties.

During the COVID-19 pandemic, Collins has overseen NIH's role in helping Moderna develop its messenger RNA vaccine for SARS-CoV-2, provided support for coronavirus diagnostics and treatments, and spoken out about the need for people in the United States to wear masks and get vaccinated. He is "incredibly grateful" for NIH's contributions yet "deeply disappointed" by distrust of COVID-19 vaccines, and that some fraction of the 700,000 Americans lost to the pandemic "didn't need to die."

Along with Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, Collins has also defended NIH against claims, including from some in Congress, that coronavirus research funded by the agency at the Wuhan Institute of Virology in China was risky, and might have sparked the pandemic by leading to a lab escape of SARS-CoV-2. At the same time, scientists assailed him for bowing to the Trump administration's wishes to terminate the source of that support, an NIH grant to the EcoHealth Alliance, a nonprofit research organization that had a long-running collaboration with the Wuhan lab. (NIH later reinstated the grant under strict conditions that EcoHealth says it can't fulfill.)

In recent months, Collins has helped convince the White House to steer a new, goal-oriented health research agency proposed by Biden, the Advanced Research Projects Agency for Health, to NIH instead of creating it as a standalone agency.

"Francis has been an enormously successful NIH director on many levels," from winning congressional support for budget increases to supporting young researchers, says Princeton University molecular biologist Shirley Tilghman.

"He's been extremely effective," Zerhouni agrees. "I give him an A plus."

Collins even won grudging respect from some of his fiercest critics. Michael Eisen, a University of California, Berkeley, evolutionary biologist and *eLife* editor-in-chief, feels Collins's focus on big science has cut into stable support for individual labs. But he commented to *Science* in 2019, "He is trying to do the right thing for the institution. You also appreciate that he's probably a bulwark against worse things for science."

Collins plans to return to his NHGRI lab, which focuses on diabetes genomics and gene therapy for progeria, a rare premature aging disorder, and will think about his "next calling." The biomedical community will be closely watching to see whom Biden will nominate to step into Collins's long shadow. "There are a lot of exciting potential candidates," Collins says, but "I would love it if the next NIH director was a woman." ■

NEUROSCIENCE

Megastudy yields brain cell inventory

Census of motor cortex aims to standardize definitions of cell types

By **Kelly Servick**

To deconstruct a thinking machine made of tens of billions of neurons, it helps to have a parts list—an inventory of the brain's cell types. But neuroscientists have struggled to standardize a list across labs and experiments. Now, a network of more than 400 researchers has released the most comprehensive inventory to date: an analysis of millions of human, marmoset, and mouse cells extracted from a brain region involved in coordinating movement.

The results, described in 17 papers this week in *Nature*, collate genetic features of cells along with their shapes, locations, and electrical activity patterns to identify more than 100 cell types in the human brain. The catalog could help researchers define the types of cells affected by brain diseases, identify corresponding cells in animal models, and better target those cells with treatments. The cell atlas "is like the Rosetta Stone for neuroscience," says Jens Hjerling-Leffler, a neuroscientist at the Karolinska Institute who was not involved in the project.

Categories to describe brain cells have emerged over decades as researchers teased apart differences in cells' location, shape, and function (*Science*, 7 August 2015, p. 575). Do they sit in a superficial layer of the cortex, or a deeper one? What is the structure of their branches? Do the neurotransmitters they release excite other cells or inhibit them? But cells often defy tidy definitions. "We've been trying to bucket these cells in ways that are probably oversimplistic," says Kristen Brennand, a stem cell biologist who studies brain disorders at Yale University and is not involved in the new project. "Biology doesn't like black and white."

The National Institutes of Health's BRAIN Initiative Cell Census Network (BICCN), a $250 million program begun in 2017, promises to add some gray. Many studies it supports use genomic sequencing to delineate cell types. A cell's RNA, for example, reveals the set of genes it has recently transcribed to make proteins—its transcriptome. Other sequencing methods describe a cell's epigenome, the set of molecules studding its DNA that influence which genes get expressed.

"These are powerful approaches," says BICCN investigator Xiaowei Zhuang, a biophysicist at Harvard University. But sequencing alone doesn't tell the whole story, Zhuang adds. "We also need to know where [cells] are in the brain, what are their neighbor cells, how do they interact."

To efficiently gather such information, BICCN collaborators agreed to focus on a single region: a strip of tissue across the top of the brain called the primary motor cortex that orchestrates muscle movements. Some techniques could capture multiple features at once. For example, a method developed in Zhuang's lab allows researchers to image hundreds or thousands of RNA sequences in a slice of brain tissue, revealing both cells' transcriptomes and their relative locations. Another method, Patch-seq, records electrical activity from cells, stains them to reveal their shapes, and then sucks out their innards to sequence their RNA.

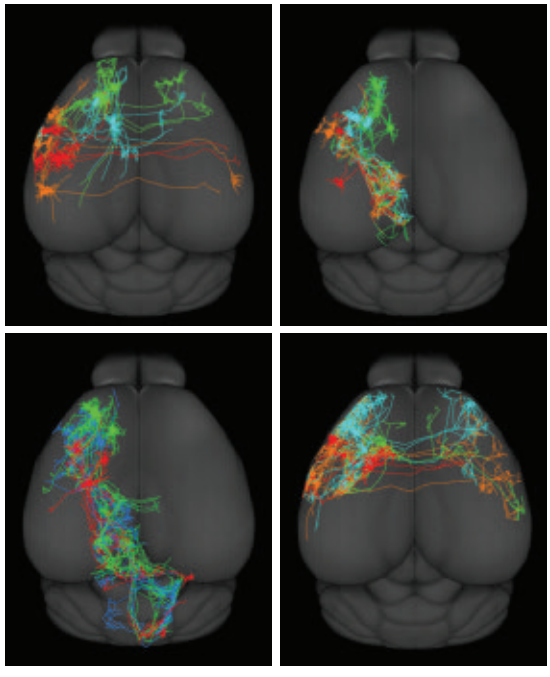
In BICCN's data, cells grouped according to their transcriptomes tend to share other features, such as location, shape, and electrical activity. That finding "provides strong validation to the molecularly defined cell types," the BICCN authors write in a paper summarizing the work. For the human motor cortex, 127 such types emerged.

The exact number depends on the criteria used to group cells, says BICCN investigator Hongkui Zeng, director of the Allen Institute for Brain Science. "Is it 127? Is it 130? Is it 100? … It's not clear cut like that." What's more important, she says, is the hierarchy of cell types that emerged, with a few basic cell types at the top and more nuanced divisions in lower branches. Cells in the same top-level class tend to be similar across various types of measurements—and across the three species studied. Further down, Zeng says, distinctions between cell types "become a lot fuzzier."

The new census may shape the ways researchers model brain disease, says Mina Ryten, a clinical geneticist at University College London (UCL) who studies neurogenetic disorders. "It's actually very hard to predict what a gene mutation is going to do to a person," she says, partly "because you just don't have a framework for understanding … which cell types [it's expressed] in."

The BICCN data could help scientists pin down which human cell types are most affected by specific mutations. Mouse and marmoset data could then help them identify and study comparable cell types in lab animals. The result, researchers hope, will be findings that translate better to humans. "We've cured a lot of diseases in mice," says Michelle Gray, a neuroscientist at the University of Alabama, Birmingham. "It's always a question when we're trying to model disease using rodents: Are we really recapitulating the changes that are observed in humans?"

Having created a partial parts list, researchers still need to study how the cell types behave and interact in a functioning brain. UCL neuroscientist Kenneth Harris is among those imaging genetically labeled cells in living mice to relate gene expression to electrical firing patterns. "We're going to have to learn what all these cell types are and try to figure out how they all work together," he says. "It's going to be difficult." But with the new cell census, "we're entering this stage where we know how much we don't know—and that's progress." ■



Researchers mapped the locations and projections of different genetically defined cell types in the mouse brain.

IMAGES: ALLEN INSTITUTE FOR BRAIN SCIENCE

A child in Burkina Faso receives MenAfriVac, a hugely successful meningitis vaccine introduced in 2010.

PHOTO: WORLD HEALTH ORGANIZATION

GLOBAL HEALTH

Plan aims to slash toll from bacterial meningitis

New five-in-one vaccine could help end epidemics in Africa and cut deaths worldwide by 70%, WHO says

By **Leslie Roberts**

When the sudden, terrifying deaths hit in early June in Banalia, a small mining community in northeastern Democratic Republic of the Congo (DRC), some people suspected witchcraft. Many of the victims were young men living in overcrowded barracks who had come to mine gold and other riches, and rather than seek treatment, they fled, carrying the mysterious ailment with them.

Health authorities were slow to recognize the cause. Banalia is in Tshopo province, which sits in the African meningitis belt, a band stretching from Ethiopia in the east to Senegal in the west, but the province had not seen meningitis outbreaks since 2009. And the puzzling affliction occurred outside the normal meningitis season, which starts in December, when the fierce Harmattan winds whip up the desert dust across large parts of the belt, and ends when the rains begin in June. This year, however, the rains were late.

The DRC suspected meningitis by July and started to provide antibiotic treatment in August, but it wasn't until 7 September that the Pasteur Institute in Paris identified a familiar cause of meningitis, a bacterium named *Neisseria meningitidis* serotype W. The country declared an outbreak and applied for vaccines from an international stockpile, but by then the disease had spread around Tshopo province. As of 3 October, 1349 suspected cases and 189 deaths had been reported. Initially, the fatality rate was a shocking 84%; "Many people didn't arrive at health centers until they were in coma," says André Bita Fouda, meningitis lead for the World Health Organization (WHO) Africa region.

A new "global road map," launched by WHO and many partners on 28 September, could help prevent such tragedies in the future. With the help of a new vaccine targeting five serotypes of *N. meningitidis*, including W, it aims to eliminate epidemics of bacterial meningitis, which kill an estimated 250,000 a year in Africa, by 2030. It would also step up the fight against sporadic cases and small clusters of the disease that occur around the world. Cases worldwide—now some 5 million per year—would be halved by 2030 and deaths reduced by 70%.

The road map, which does not yet have a price tag, is "very ambitious," says WHO's Marie-Pierre Préziosi, who led its development, "but I do think it is feasible"—provided extra money comes through. Still, "How many road maps have been launched before but lost their way?" Mike Ryan, head of health emergencies at WHO, cautioned at the plan's launch. "We must see we don't lose our way on this one."

Meningitis, an inflammation of the membranes shielding the brain and spinal cord, can also be caused by viruses and fungi. But only bacteria spawn the epidemics that sweep across the meningitis belt every 5 to 12 years. Spread by respiratory droplets, bacterial meningitis kills one in 10 affected, often within 24 hours, and leaves one in five with lifelong disabilities such as deafness, cognitive impairment, and loss of limbs. Some people "carry" the bacteria in their nose and throat harmlessly. Trouble strikes, the leading theory goes, when dust and dry weather aggravate the mucous membranes, giving the bacteria a route to invade the bloodstream.

Vaccines already exist for three of the four bacterial species, but they don't cover all serotypes. Many were developed for high-income countries; their price puts them out of reach in Africa, and some are in short supply. That's why the Geneva-based International Coordinating Group on Vaccine Provision (ICG) doles them out only after an outbreak has begun, which is little more than a "Band-Aid," says Mark Alderson, who heads the bacterial meningitis vaccine effort at PATH in Seattle.

The hugely successful 2010 introduction of a long-lasting vaccine against *N. meningitidis* group A, which at the time caused 80% to 90% of all epidemics in Africa, shows the promise of vaccines. Developed by a collaboration between PATH and the Serum Institute of India, it consists of a polysaccharide from the bacterium's surface linked to a protein, tetanus toxoid, that makes the vaccine more powerful. The final price tag was a mere 60 cents per dose, making it cheap enough to use in mass vaccination campaigns (*Science*, 10 December 2010, p. 1466).

So far, 24 of the 26 countries in the meningitis belt have introduced the vaccine, MenAfriVac, in mass campaigns; 11 have also incorporated it into routine child-

hood immunization. The impact has been stunning: The last outbreak of meningitis A occurred in 2014, and the last known case in 2017. (Other *N. meningitidis* serotypes and *Streptococcus* partially replaced the strain, however.)

Now, PATH and Serum are trying to duplicate that success with a "pentavalent" conjugate vaccine that protects against *N. meningitidis* serotypes A, C, Y, W, and X. "If we can switch from MenAfriVac to pentavalent, it could be the end of epidemics in Africa," Préziosi says.

It's a tall order, essentially five vaccines in one. Results from phase 3 clinical trials conducted in Mali and Gambia are not public yet but look encouraging, Alderson says. The partners expect the vaccine to be licensed and receive WHO prequalification—the official blessing for use in poor countries—in 2022. "There is a real sense of urgency coming from WHO," Alderson says.

The price, up to $3 a dose, could still pose a problem. WHO and Gavi, the Vaccine Alliance, which funds vaccines for poor countries, are discussing ways to keep the cost down. MenAfriVac is given to 1- to 29-year-olds, who are most vulnerable to bacterial meningitis; one possibility is to lower the upper age limit for the pentavalent vaccine. Another option is to use it both in mass vaccination campaigns and routine immunization in high-burden countries, but drop the mass campaigns where epidemics occur less often.

The road map also aims for wider use of existing meningitis vaccines and the development of affordable new ones, including against group B *Streptococcus*. The only one of the four species for which no vaccines exist, it is a leading cause of newborn meningitis worldwide.

Meeting WHO's targets will also require stepped up surveillance and affordable, rapid tests so that infected people can be identified and treated with antibiotics. (Now, diagnosis requires a lumbar puncture, which health workers in poor countries don't always perform.) People living with the disease's long-term consequences also need better care and rehabilitation.

In the DRC, weekly case numbers are still climbing, Bita Fouda says. But antibiotics have lowered the death rate to about 10%, and a team of neurologists is helping diagnose and treat complications. ICG approved the DRC's request for 187,460 vaccine doses on 17 September, and the shots arrived in the country on 2 October. Once vaccinations begin, Bita Fouda says, "we should see a drop in cases in 4 to 6 weeks." ■

Leslie Roberts is a science journalist in Washington, D.C.

PHOTO: IMAGEBROKER/ALAMY STOCK PHOTO

ARCHAEOLOGY

Vikings in paradise: Did the Norse settle the Azores?

Seafarers may have landed on lush archipelago more than 1000 years ago, sediment cores suggest

By **Michael Price**

In 1427, the Portuguese navigator Diogo de Silves first set foot on an uninhabited, Sun-kissed island with white sand beaches and crystal blue bays, proclaiming it Santa Maria Island. Later arrivals revealed it to be the southeasternmost island of the Azores archipelago, 1368 kilometers west of Portugal's coast.

But a new study of lake sediment cores suggests de Silves and his crew may not have been the first to reach the island paradise: Viking seafarers may have beaten them by some 700 years. Any Vikings were long gone when Portuguese sailors arrived, the authors note, but some Norse rodent stowaways may have left a lasting genetic mark on the island.

The paper is a welcome addition to the sparse data on the prehistory of the Azores, says evolutionary biologist Jeremy Searle of Cornell University. His team proposed a Norse link to the island in 2015 based on genetic similarities between Azorean and northern European mice. "To actually have firm data supporting that is obviously pretty gratifying."

Biomarkers in lake sediment from the Azores's Pico Island suggest the presence of livestock as early as 700 C.E.

Most archaeological evidence from the Azores dates back to only the 15th century, with a few studies hinting at even earlier occupation. Then about 10 years ago, Pedro Raposeiro, an ecologist at the University of the Azores, Ponta Delgada, and colleagues set out to collect cores of sediment from five lakebeds around the archipelago to help chart the region's climate history. The researchers suspected they would find signs of human disturbance—pollen from nonnative crops, spores from fungi that grow on livestock dung—dating back to the early 1400s. And they did.

But the researchers were surprised to find these signals extended even further back in time. In a layer dating to between 700 C.E. and 850 C.E. taken from Peixinho Lake on the Azores's Pico Island, they saw a sudden uptick of an organic compound found in the feces of ruminants such as cows and sheep. They also saw an increase in charcoal and a dip in the abundance of native tree pollens, perhaps pointing to humans cutting down and burning trees to clear space for livestock to graze, Raposeiro says.

A similar signal shows up in cores from Caldeirão Lake on the Azores's Corvo Island dating to about 100 years later. Pollen from a nonnative ryegrass shows up in layers from Pico Island dating to about 1150, and at 1300 on São Miguel Island, also part of the archipelago.

The results suggest humans were exploiting natural resources on the Azores at least 700 years earlier than thought, the team reports this week in the *Proceedings of the National Academy of Sciences*.

Who were those ancient mariners? "Our best guess is the Norse," who were accomplished and adventurous seafarers, Raposeiro says. Vikings were plundering the coasts of Europe as early as 789 C.E.

Climate models for this time suggest the dominant winds blew from the northeast, so Viking ships sailing southwest from Scandinavia would have fair winds for the Azores, Raposeiro says. And, as Searle and colleagues documented, Azorean house mice share substantial DNA with mice from northern Europe. The rodents could have hitched a ride on the Viking ships; today they are like "living artifacts" of a Viking presence, Searle says.

But others are cautious. Given widespread trade routes, a mouse from Scandinavia could have boarded a ship in what today is Portugal and sailed to the Azores, says Simon Connor, a geographer at the Australian National University. And determined Portuguese sailors might have reached the archipelago despite unfavorable winds, "It's possible it was the Norse," Connor says. "The Vikings were great, great mariners. … But it could have been anyone." ■

THE FAULT IN OUR STARS

Satellite swarms are threatening the night sky. Is low-Earth orbit the next great crucible of environmental conflict?

By **Joshua Sokol**

On 19 December 2019, Tony Tyson, an experimental physicist at the University of California (UC), Davis, joined a conference call with billionaire Elon Musk that helped shape the fate of starry nights on Earth.

The call was cordial but tense. Seven months earlier, Musk's company SpaceX had livestreamed a feed of 60 satellites drifting off into space from the bay of one of the firm's rockets. The satellites, the triumphant first wave of a project called Starlink, were built to beam down broadband internet to every corner of the globe. But as the satellites began to do laps around Earth, people looking up at night saw a string of glinting pearls as bright as the stars in the Big Dipper.

The first flotilla of 60 Starlink satellites was released in May 2019 from a SpaceX rocket.

Those five dozen satellites were just the beginning. Starlink already had permission to launch 12,000 satellites, roughly six times the number of active satellites then in orbit. The next year, the company added another 30,000. Other billionaire-backed companies—Jeff Bezos's Project Kuiper, and OneWeb, funded in part by Richard Branson—were planning comparable space internet swarms, leading to industry forecasts of more than 100,000 satellites in orbit by the end of the 2020s. In the best stargazing conditions, human eyes can perceive about 3000 twinkling stars overhead; if the planned satellites ended up as bright as the first Starlinks, they would fill future summer nights with a comparable number of creeping dots.

Almost overnight, a new arena of environmental conflict opened up. Astronomers

PHOTOS: (TOP TO BOTTOM) © MATT DIETERICH; SPACEX

In August 2020, Starlink satellite tracks were seen next to comet NEOWISE at Idaho's Craters of the Moon, a national monument. Fifty time-lapse shots, each 4 seconds long, were stacked to make the image.

weren't the only ones who saw an existential threat. Environmentalists, amateur stargazers, and Indigenous leaders working to revive astronomical traditions saw an affront to the planet's dwindling dark skies, an act as vandalistic as carving initials into a tree trunk—in front of the whole world. "From a cultural point of view, it is a desecration," says Rangi Mātāmua, a Māori cultural astronomer at Massey University, Manawatū.

For Tyson, the stakes were narrower and more personal—but no less dire. In 1996, when he worked at Bell Labs, he conceived a wide-field telescope capable of soaking in the entire sky overhead out to the edge of the visible universe every single night. That endeavor became a $700 million project supported by the National Science Foundation (NSF), one that would push charge-coupled device (CCD) cameras—a Nobel Prize–winning technology invented down the hall from Tyson's Bell Labs office—to their maximum potential. As the first Starlinks circled the planet, a construction crew in the Chilean Andes had just delivered the main mirror for what's now called the Vera C. Rubin Observatory, scheduled to open its shutters to the universe in the next couple of years. "All was happy and nice with our heads buried underneath the sand," Tyson says.

But Tyson and colleagues quickly apprehended the threat. In the lab, they simulated satellite streaks over the Rubin's Volkswagen-size, 3.2-gigapixel camera. Each photobombing satellite, the researchers found, would cause not just one useless streak of pixels, but a series of ghostly echoes. Yes, code could be written to avert the telescope's gaze when a satellite crossed over—but as the number of satellites in the sky ticked past 10,000 they would start to intrude into every exposure. The swarms would interfere not just with the Rubin, but any other wide-field studies of dark matter, dark energy, interstellar comets, and flaring transients such as supernovae. The satellites would also confound efforts to find potentially hazardous asteroids. Those searches need to be conducted around twilight—exactly when satellites are most visible.

Soon after the launches began, lawyers and activists got involved, asking pointed questions about whether near-Earth space—once the domain of nations and militaries, now open to businesses and billionaires—was an environmental commons in need of protection, like a forest or river. Executives such as Musk pushed back, arguing global broadband would be a boon for poor, isolated communities desperate for online opportunities. But Tyson, for the moment at least, wanted to avoid public conflict. He emailed a former contact at SpaceX, hoping to speak engineer-to-engineer as he would have in his Bell Labs days.

And so it was that Tyson and several colleagues found themselves on the phone with Musk himself, negotiating the first chapter in an ongoing environmental dispute over who exactly decides the future of the sky. According to other people on the call, Musk said that if astronomers were clever, they would just put their telescopes into space, where the seeing was better anyway. But doing that would cost 10 or 100 times more, the astronomers replied—and if SpaceX was clever, their engineers could just solve the satellite brightness problem cheaply.

Tyson, explaining the technical needs of the Rubin detector, pushed for dimming the satellites by a factor of 30 or even 100—with Musk balking at the latter figure. But at the end of the conversation, Musk turned to some SpaceX engineers on the call. "Fix this," he said.

THE DREAM OF satellite internet has come and gone before. In the 1990s, companies such as Globalstar and Iridium started to sell satellite phones that worked virtually anywhere on Earth, and Teledesic raised more than $1 billion for global satellite broadband. But one by one the companies went bankrupt, and by the time Teledesic shuttered, in 2002, its single prototype had burned up in the atmosphere.

More than a decade later, the cost of

building and lofting a satellite had fallen drastically—inspiring a spree of space internet schemes backed not just by billionaires, but also by the Chinese government. Unlike traditional internet satellites—large, expensive beacons perched in deep geostationary orbit—the satellites in these "constellations" would zip past at much lower altitudes. Many more satellites would be needed to ensure one was always above a given spot on the ground. But lower orbits would also mean shorter lag times, with signals whisking up and down in thousandths of a second—better for needs as diverse as online classes, emergency response, competitive video games, stock trading, and military communications.

On a Friday in January 2015, Musk stepped on stage at an invite-only gala in Seattle to announce Starlink's new headquarters, soon to open in nearby Redmond, Washington. There, engineers would re-create internet infrastructure from the top down. It was a first step, Musk said, toward a grander personal goal: making humanity a multiplanetary species. When someone asked how much the service would cost, he said it couldn't be free, because the goal was also to generate enough revenue to pay for a city on Mars.

Only one gatekeeper stood in Musk's way. Satellites need licenses from national agencies, such as the U.S. Federal Communications Commission (FCC), which regulate radio transmissions and coordinate internationally. That system governs the radio sky as a global commons, but imposes few other constraints. The FCC, for example, only requires that defunct satellites get out of low-Earth orbit within 25 years of a mission's end—through either natural or intentional deorbiting into the atmosphere, where they burn up.

Regulators do start to revoke licenses if a company doesn't deploy its approved satellites on schedule, however. Critics say the practice incentivizes corporations to follow a Silicon Valley–style, "fail fast, fail often" strategy. At Starlink, "There was a lot of talk about how, financially, we could possibly make this happen, and how the hell do you even launch that many things," a former team member says. "There was very little talk about it as a moral or ethical or even just a collision problem."

Those collision problems soon came into focus. In 2017, Hugh Lewis, an astronautics researcher at the University of Southampton, reported simulations showing that if even a small fraction of the satellites lost their ability to maneuver while in orbit, collision risks would soar, eventually causing hailstorms of new debris. To address those concerns, Starlink pushed its constellation below 600 kilometers, where spent satellites would fall out

The gathering swarm

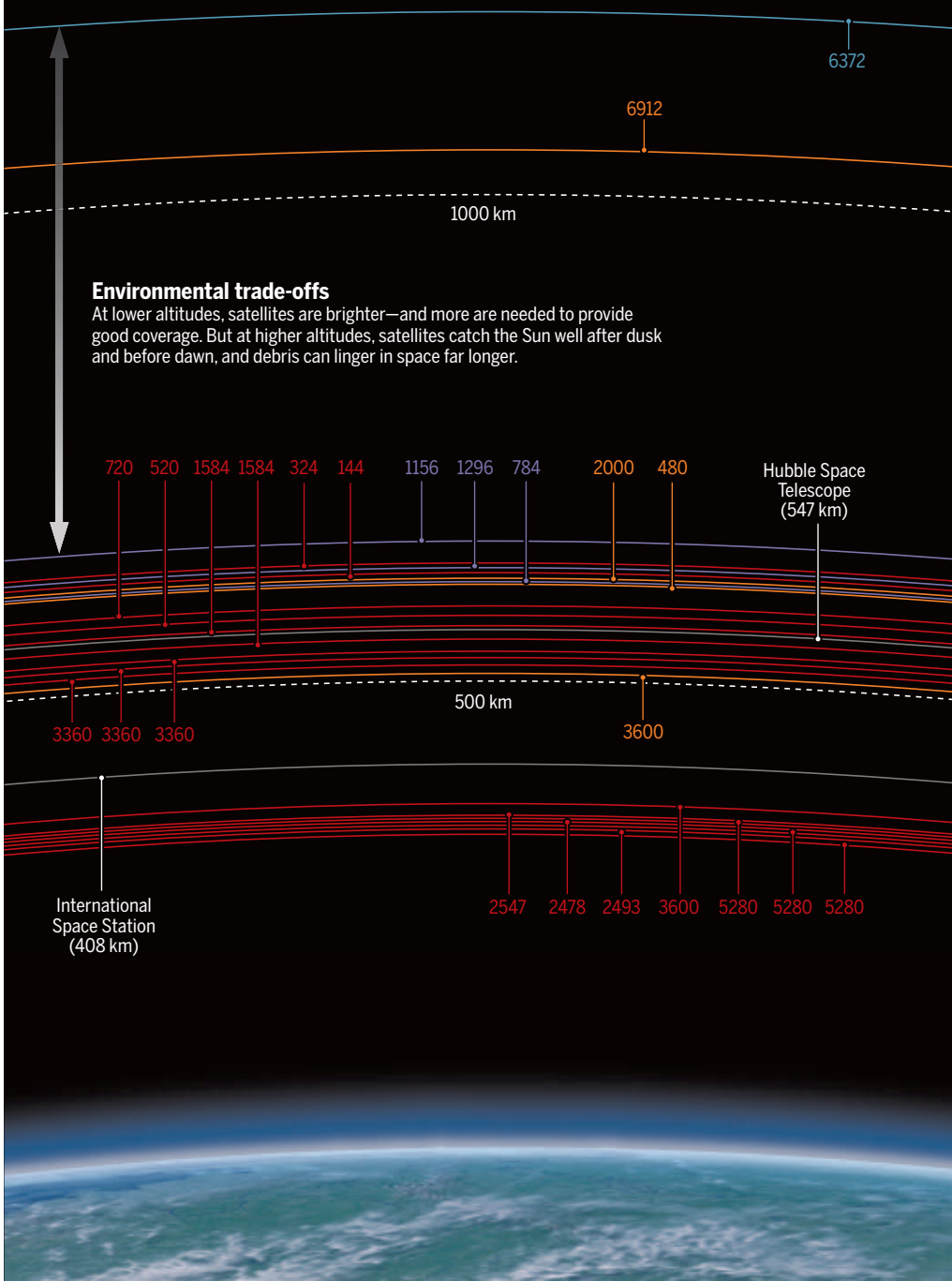
A slew of companies is planning to launch tens of thousands of internet satellites into low-Earth orbit, where they pose threats to astronomers and casual stargazers alike. Here's a 2D representation of what the four largest swarm builders have so far submitted to regulators. *Graphic by* **Chris Bickel**

Starlink (41,914 satellites)
The furthest along, Starlink internet service now reaches more than 90,000 users. Future satellites will be larger but designed from the start to limit visibility, the company says.

OneWeb (6372 satellites)
OneWeb satellites are fainter than Starlink's but are lit up by the Sun for more of the night. In 2020, the U.K. government and other investors bought the company out of bankruptcy.

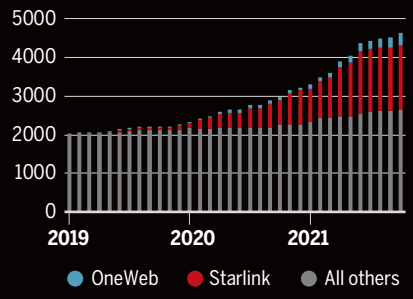
China Star Network (12,992 satellites)
In September 2020, China submitted plans to international regulators for a satellite broadband constellation. This year, a state-backed company was formed to develop it.

Project Kuiper (3236 satellites)
Amazon's project, headquartered a short drive from Starlink, says it will spend $10 billion to launch Kuiper in the 2020s using Atlas V rockets.



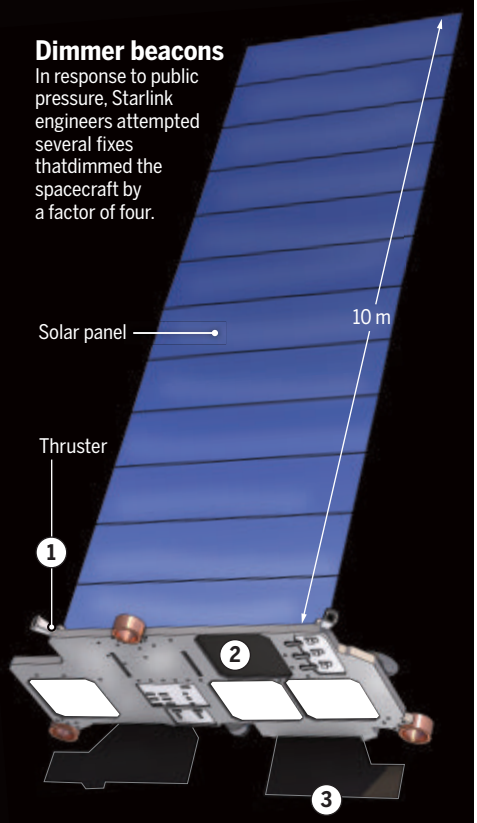
Space race

In less than 3 years, Starlink and OneWeb have launched more than 2000 satellites. Those constellations represent more than 40% of all functioning satellites in orbit.



Dimmer beacons

In response to public pressure, Starlink engineers attempted several fixes thatdimmed the spacecraft by a factor of four.



1 Orbital maneuvers
During the brightest period, days after launch, operators roll the satellite on its side to reduce reflections from the large solar panel. Once raised to operational orbit, the satellite's panel is angled like a shark fin to minimize glare.

2 Dark coating
Engineers tested painting underside surfaces black, but abandoned the practice.

3 Sunshades
The brightness of an operating satellite depends mainly on the flat antenna surfaces pointed down to Earth. Visors block some sunlight from ever hitting those surfaces.

DATA: JONATHAN MCDOWELL/HARVARD-SMITHSONIAN

of orbit quickly. But an accident there could still choke low-Earth orbit with dangerous debris for years afterward, Lewis says. "If you get it wrong, the space environment is not going to forget."

Few optical astronomers were paying attention to the looming threat. In the preceding century, they had coped with the glare of ever-brightening city lights by siting big telescopes on remote mountaintops. Light pollution had cost four-fifths of North Americans and one-third of the world their view of the Milky Way. But the problem was local, when it was a problem at all. Astronomers negotiated it with adjacent communities, not tech companies divvying up the planet's radio frequencies. "I don't read the FCC filings for entertainment purposes," Tyson says. "On balance, my bad."

EVEN BEFORE MUSK'S EDICT to fix the problem, SpaceX engineers had begun a collaboration with the Rubin group. "All they wanted from me at that point was one number," Tyson says. How much dimmer would the satellites have to get?

Astronomers measure brightness by using a system called apparent magnitudes, credited to the ancient Greek polymath Hipparchus, in which larger numbers correspond to fainter objects. The first Starlinks started out at about first or second magnitude, visible even from a city. As they settled into their long-term orbits, higher up, they faded to fourth or fifth magnitude, still visible from some suburbs. By contrast, a 15-second Rubin exposure could pick up distant galaxies at 24th magnitude, fainter by a factor of about 40 million.

Tyson and his team didn't ask for that much dimming. At seventh magnitude, Tyson says, the errant electronic signals caused by a Starlink pass over the Rubin would be manageable. And seventh magnitude mattered for another kind of detector, one with global implications, concluded an American Astronomical Society (AAS) committee on light pollution. "By gosh, that's the limit of human vision," Tyson says. "That's pure coincidence."

While the astronomers converged on that all-important number, Starlink engineers were at work. As sunlight falls on the planet, Earth casts into space a tapering shadow in the shape of a cone. Anything big flitting past the edge of that cone of night can scatter sunbeams down toward people or observatories already in darkness. And internet satellites, which broadcast powerful signals from flat, reflective radio antenna surfaces, tend to be large. The first Starlinks were especially bad, with 10-meter-long solar panels and off-white surfaces.

For a first test, Starlink engineers launched one satellite, painted with a less reflective coating. That darker surface dimmed the satellite, but also absorbed too much heat. In April 2020 they announced alternate tweaks. For the brightest period after launch, all the satellites would now roll on their sides to present a less reflective surface. Each satellite would also be built with a Sun visor, further cutting down on reflections.

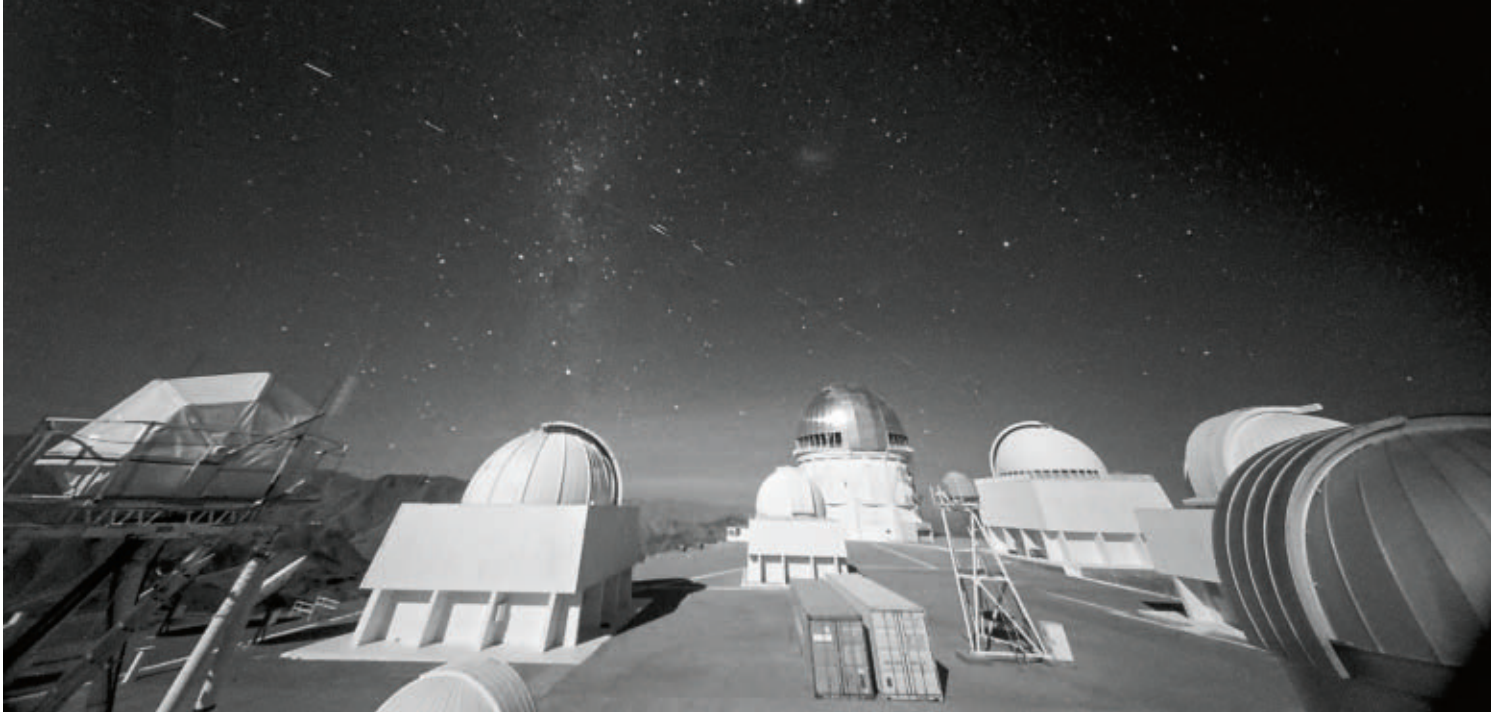
The visors helped. Of the nearly 1700 Starlinks now in orbit, more than 1000 have sunshades. Pat Seitzer, a member of AAS's light pollution committee, estimates that most appear between magnitude six or 6.5, not quite at the seventh-magnitude target, but about one-third to one-quarter as bright as the first Starlinks. "[SpaceX] really have set the standard," he says. Seitzer is relieved, although he says some astronomers and dark-sky advocates did not like the committee's decision to collaborate with SpaceX. "Several of us were compared to Neville Chamberlain, for having sold out the astronomical community."

Astronomers are still struggling to quantify how much the completed swarms will affect wide-field telescopes—assuming Starlink's competitors dim their satellites, too. Starlink's full flock might end up spoiling as little as 1% of Rubin data taken in twilight, and algorithms can scrub damaged pixels to remove some bright trails. But algorithmically corrected images could in turn bias studies of dark matter and dark energy—two unknown entities dominating the large-scale universe—that depend on measuring minute fluctuations across entire images. And Rubin astronomers worry even more that streaks and flashes from passing satellites might mimic the mysterious transient events Rubin was designed to discover: unanticipated flares lasting seconds or minutes.

"I don't mind my career being somewhat eaten by this," says Meredith Rawls, an astronomer at the University of Washington, Seattle, who has led efforts to mitigate satellites' effects on data. "But I would really love it if future astronomy grad students didn't have to have this rabbit hole."

The outlook worsens if future Starlinks trend brighter or if the other companies don't try to limit brightness. None has had comparable discussions with astronomers. "My hope was that it would be so obvious that it would shame all the other industries into complying," Tyson says of his work with Starlink. Lately, though, his hopes are fading. "You realize that it's really just the Wild West, and we astronomers are just a nuisance."

WHILE ASTRONOMERS worked to save their science, other factions began to voice a broader environmentalism: a sense of low-Earth orbit as a shared part of nature—and an unusually vulnerable one. There are no national or international laws to stop anyone



from launching an even lower, undimmed Starlink-size constellation, says Jonathan McDowell, an astronomer and space analyst at the Harvard-Smithsonian Center for Astrophysics. "My feeling now is a mix of relief that we were heard," McDowell says, "and nervousness that we are depending on the kindness of corporations."

One extant law might have teeth. In 2020, Ramon Ryan, a Vanderbilt University law student and former union organizer, read press coverage of the satellite light pollution issue and wondered whether opponents might have grounds to sue. He concluded they do, under the 1969 National Environmental Policy Act (NEPA). The law requires U.S. agencies to assess the environmental consequences of projects before approving them. Satellite licenses have been exempted since 1986 because they were assumed to have no environmental impacts—an assumption that could now be challenged.

The NEPA argument, published in a law school journal, surprised both SpaceX and astronomers. Outside lawyers bombarded AAS with offers to work on a case. But the organization's leadership declined, unsure whether it had the legal expertise and wanting to avoid a confrontation that could damage the SpaceX partnership.

Viasat, which operates geostationary internet satellites, was less skittish about challenging Starlink. In May, the company filed a petition in the U.S. Court of Appeals for the District of Columbia Circuit, arguing that the law requires the FCC to consider Starlink's environmental impacts on space debris and light pollution, but a decision is still pending.

The FCC has been successfully pressured to consider light pollution before, says Travis Longcore, an ecologist and light pollution researcher at UC Los Angeles—but it took a scrappy, grassroots campaign. In the early 2000s, conservation groups cited Longcore's research and leveraged NEPA to push the FCC to change how it regulated the height and illumination of communication towers that birds collide with. "[The FCC] had to get dragged kicking and screaming," he says.

Michelle Hanlon, a space lawyer at the University of Mississippi, says the Viasat lawsuit deserves a wider hearing. "It would be a shame for something this big to be decided by two private companies arguing before a judge." The NEPA debate turns on whether low-Earth orbit can be considered an extension of the human, earthly environment, she says. This summer, a panel of space lawyers—including Hanlon and lawyers from Amazon and SpaceX—considered the question. Even if space is not part of the human environment, they agreed, other U.S. laws—including those that recognize the importance of dark skies—serve as compelling reasons for the FCC to regulate constellation brightness.

Casual stargazers, dark-sky activists, and photographers also have a stake in the outcome, says Aparna Venkatesan, a cosmologist at the University of San Francisco. "Our constituents are humanity, not just professional astronomers," she says. An altered night sky would disproportionately impact Indigenous groups, she says, because it would damage spiritual and cultural practices involving the stars—traditions that have already suffered losses through policies such as residential boarding schools in the United States and Canada, which for decades aimed to assimilate Indigenous children far from home.

Hilding Neilson, an astronomer at the University of Toronto and a member of the Qalipu Mi'kmaq First Nation, thinks it's ironic that astronomers are speaking up for Indigenous cultures. For years, he notes, astronomers and Indigenous protesters in Hawaii have battled over building new telescopes on Mauna Kea, which Native Hawaiians consider a sacred mountain. But he agrees that large satellite constellations are a cultural threat and says launching them should require consent from sovereign groups down below. "If light pollution is erasing Indigenous stories, these satellites are rewriting them," Neilson says. "Did those nations get a choice?"

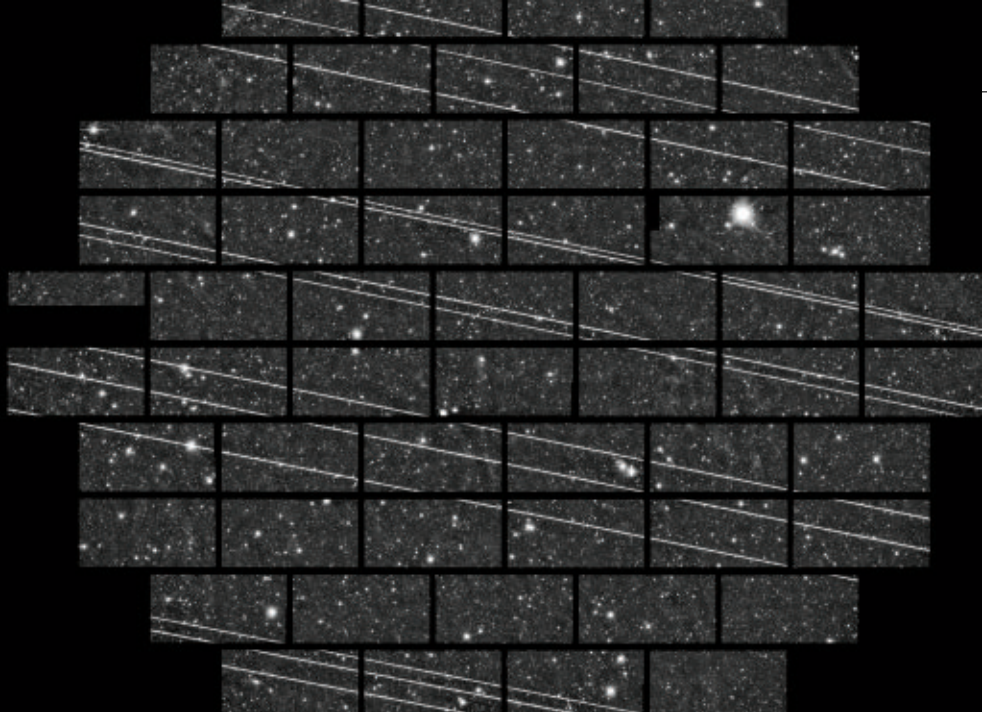
SPACEX EXECUTIVES, responding to those criticisms, have pointed to their own efforts to bring Starlink service to rural Indigenous communities. On a Wednesday morning in June, I visited one on Washington's Olympic Peninsula.

At a checkpoint in front of the Hoh Reservation, I met Melvinjohn Ashue, the tribe's economic development officer, who was wearing a Baltimore Ravens hat. I held out my wrist for a temperature check and confirmed I had no COVID-19 symptoms.

Then we drove in separate cars past moss-blanketed trees and into a small neighborhood of modest single-story houses, each with a white Starlink dish the size of a small pizza gleaming down from slanted roofs. Ashue said he had helped mount them, weighting them down with cinder blocks against the area's battering storms.

At the pandemic's beginning, 15 of the tiny reservation's 28 homes were still on a waiting list for internet service. When school went online, the district equipped families with mobile hot spots, but cell service is also spotty. Ceciliajean Ashue, Melvinjohn's sister, said she was missing online seminars at work

PHOTO: CTIO/NOIRLAB/NSF/AURA



In November 2019, a Starlink satellite train was visible over Cerro Tololo in Chile (left). It left streaks on the Blanco telescope's wide-field camera (right).

PHOTO: CTIO/NOIRLAB/NSF/AURA/DECAM DELVE SURVEY

while trying to juggle the class schedules of three school-age children—only one of whom could go online at a time.

Just then Starlink was searching for a pilot community at the right latitude. The Washington Statewide Broadband Office suggested the Hoh, relegated to the far northwest corner of the state by an 1855 land-grabbing treaty. In September 2020, Starlink team members arrived to help set up the dishes, making the tribe among the first people outside SpaceX to get a glimpse of the service. Today, Melvinjohn Ashue says, the tribe has a 3-year deal to buy Starlink terminals and subscriptions—services that retail for $500 for the dish and $99 monthly payments. Eventually, he adds, the Hoh hope to switch from Starlink to a ground-based fiber connection, one that would let the tribe be its own internet service provider—and perhaps reap some of the federal incentive money that went to Starlink for offering rural broadband access.

Connectivity through Starlink has helped with education and a vaccine clinic and allows ceremonies to be livestreamed. Coverage drops out when it rains but quickly returns. And the satellites themselves may have even inspired careers in aerospace. "At night, my kids like to watch for the blinking lights," Ceciliajean Ashue says. "We get the Navy flying over us constantly. A satellite going by doesn't bug us as much."

Juan-Carlos Chavez, a Seattle-based information scientist affiliated with the Yaqui tribe of the Sonoran desert, is more concerned. "Telehealth, tele-education, we need that too," says Chavez, who advises Pacific Northwest tribes on securing broadband access. "But deploying a bunch of satellites up there to quote-unquote solve the problem is something that we really gotta consult communities with."

Later in the summer, in the state's northeast corner, Chavez waited for wildfire smoke to clear and the stars to rise over his campsite at Colville National Forest. As he concentrated that evening, gazing up to connect with the grandfather who first taught him to seek wisdom from his elders in the stars, a moving dot crossed by. "When you're in deep ceremony and you see something moving that's not supposed to be there, it detracts from being in the place of harmony with your elders and ancestors," Chavez says.

LAWS MAY NOT protect that view, but diplomats can change norms on the international stage. "There's no enforcement, but there is that sort of 'name and shame,'" Hanlon says. "That carries weight."

The International Astronomical Union (IAU) has long prodded UNESCO to declare the night sky a protected World Heritage Site, to no avail. But in 2017, international astronomers approached a different U.N. committee in Vienna that might have more jurisdiction over the sky—the Committee on the Peaceful Uses of Outer Space (COPUOS).

Three years later, a group of astronomers, satellite engineers, and other experts gathered by IAU submitted their final report to COPUOS summarizing threats to the night sky: ground-based light pollution, radio interference, and then satellites, an issue they hadn't even anticipated when they started. The report recommends satellite constellations orbit below 600 kilometers—a tradeoff that makes them visible only around twilight for much of the year, but brighter when they are visible—and that they never outshine seventh magnitude, the original threshold Tyson negotiated.

Astronomers hope COPUOS will pass the recommendations to the U.N. General Assembly and that they will spur diplomats to acknowledge the importance of dark, quiet skies. But building consensus takes time. "It's even slower than you might think it would be," says Connie Walker, an astronomer at NSF's National Optical-Infrared Astronomy Research Laboratory, who helped present the report.

Timiebi Aganaba, a space governance expert at Arizona State University, Tempe, predicts such developments will eventually shift space from an inaccessible, specialized domain to a shared one, governed by laws and norms. "These are age-old problems," she says. "Space is not that unique—it's just another domain of human activity now."

For now, the future is uncertain. Before leaving Washington in June, I drove to the center of the state and then rattled down several gravel and dirt roads toward a bare lot with a sign out front: Goldendale Sky Village.

In the past 3 years, a neurosurgeon named Christopher Smythies has persuaded 46 professionals from Seattle and Portland, Oregon—lawyers, accountants, firefighters—with a shared loved of amateur astronomy to buy into a limited liability company that lets them set up camp here, under skies that are darker than the surrounding areas and almost the entire eastern United States.

True to Seattle stereotype, my entire week in Washington had been cloudy. This night's forecast was little better, but a smattering of hobbyists had come out in recreational vehicles anyway, for the company and fresh air. Smythies says he was here in May 2019, when the first string of 60 Starlink satellites blazed across the sky. "We thought we were being invaded by space aliens or something," he says. "It's not a problem for us right now, but if they put up tens of thousands of satellites it'll become one." Meanwhile the village has grown—and gained an internet connection through a Starlink dish.

As the clock ticked past 10 p.m., then 11 p.m., and clouds still obscured the sky, I set off for a motel and stopped at a remote, unlit intersection a few minutes away. Head craning up, I watched the Big Dipper and a spill of fainter stars finally peek from a gap. And within 1 minute the dim speck of a satellite traced a line through the clearing, sliding across the constellation, until it, too, passed behind the clouds. ■

Joshua Sokol's journalism on the night sky is supported by the Alicia Patterson Foundation.

INSIGHTS

PERSPECTIVES

BIOCHEMISTRY

Rare metal, precious adhesion

Mussels use microfluidic assembly and vanadium to create an adhesive that anchors them in their natural environment

By **Jonathan J. Wilker**[1,2]

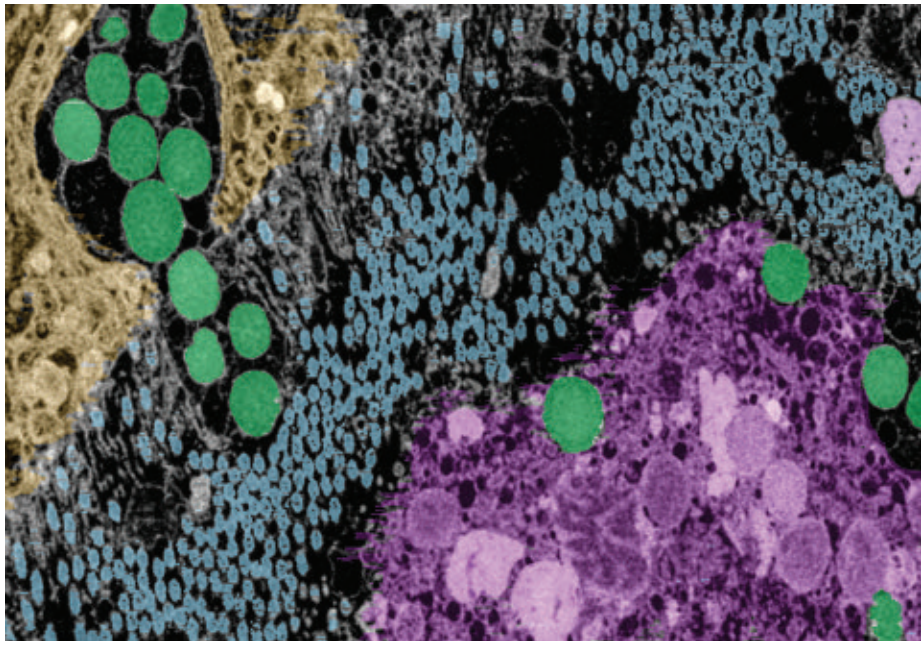
Metals are essential for life. Usually, however, it is the cheaper elements that give rise to function. Iron is essential for producing blood, and zinc is vital for stabilizing protein structures. Vanadium, by contrast, is not seen too often in biology (*1*). On page 206 of this issue, Priemel *et al.* (*2*) report that marine mussels use vanadium, along with iron, to construct their protein-based adhesive. Their study also provides a clearer picture of the machinery that assembles these components into a high-performance material.

For a mussel, survival is no vacation at the shore. Pounding waves, predating birds, and mating struggles are constant. In the face of such challenges, many species of shellfish find safety by sticking to rocks and to each other. They have developed adhesives that function in wet environments.

The mussel bonding apparatus is protein-based and composed of plaques (or disks) of glue that grab the surface and threads that connect this adhesive to the animal's tissues. Two tricks transform these proteins into adhesives. The proteins are posttranslationally modified to generate catechol side chains (*3*). These 3,4-dihydroxyphenylalanine (DOPA) residues

PHOTO: KEVIN BRITLAND/UIG/GETTY IMAGES

Blue mussels cling to rocks at Godrevy Beach, Cornwall, England. They fabricate a metallopolymeric adhesive that anchors them to the seashore.



Mussels secrete protein-containing vesicles (green) and metal storage particles into the lumen of microchannels where they are moved along by cilia (blue) and mixed into the nascent adhesive plaque (purple).

give rise to cross-linking by several types of chemical reactions and bonds (*3*). In addition, these DOPA proteins are combined with metal ions. What has been unclear is how such metal-protein systems are formed and how organic-inorganic interactions influence bulk properties (*4*).

Consider a double-barrel syringe of epoxy from the hardware store. The components are only mixed together right before they are needed for bonding. Likewise, mussels cannot just generate a protein with metals for later use. The resulting combination would be a cross-linked, hardened mess. There must be a delivery and assembly system that keeps the components separated until adhesion is required. How do mussels accomplish this feat of materials engineering? This question raises the challenge of mimicking protein-based spider webs. People can make solutions of silk proteins but not achieve the notable mechanical properties displayed by webs (*5*). Spiders have a complex spinneret system to forge webs from proteins.

Priemel *et al*. reveal that mussels also process their proteins in distinctive ways. The DOPA proteins are contained within vesicles (*2*). Metal ions are stored in separate particles. The animal has a microfluidic channel system to mix all parts where and when needed. A trigger, likely a change in pH, opens both the metal storage particles and the protein vesicles. At this point, metals and proteins make first contact and become cross-linked as they pass through the channel on the way to forming an adhesive material.

In the threads that connect adhesive to mussel, iron binding provides the beneficial properties of hardness and flexibility (*6*). The proteinaceous plaques of glue also require iron during curing, when liquid precursors transform into a solid (*7*). Changes to the concentration of iron in water alters the microstructure and adhesion strength of plaques (*8*). Both reversible metal-protein chelation (*6*) as well as more permanent protein-protein covalent bonds are present in the plaques. Binding and oxidizing DOPA groups by iron(III) ions appear to commence the cross-linking chemistry (*7*).

Among the redox-active metals in ocean waters, iron is the most prevalent (*9*). There, iron is usually particulate and easy for mussels to collect in their feeding filter (*8*). Copper and manganese are also present at concentrations higher than that of vanadium. The exact concentrations of metals in sea water vary a great deal with location, but there tends to be 5- to 100-fold more iron than vanadium (*9*). Perhaps not surprisingly, then, vanadium is seldom seen in biology. Some sea squirts, worms, and mushrooms accumulate this metal for reasons that remain unclear (*1*). A small collection of enzymes, including haloperoxidases and nitrogenases, have vanadium-containing active sites (*1*).

How mussels obtain this somewhat rare metal remains an open question. Why are they using vanadium? Mussels accumulate iron and vanadium but largely keep them separated (*10*). Such partitioning suggests that a means of differentiating metal ions is present. Upon mixing with proteins, both metals can colocate (*2*). Addition of vanadium to steel increases strength and decreases weight upon formation of an alloy. Even if alloys are not present in mussels, having more than one metal in this adhesive system could have implications for mechanical performance.

With higher oxidation states come electron deficiency and tighter binding to electron-rich ligands. Vanadium (V) has a +4 oxidation state (and even +5) not readily

[1]Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA. [2]School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA. Email: wilker@purdue.edu

PHOTO: PRIEMEL *ET AL*. (2)

available to iron (Fe). Metal-protein bond strengths and kinetic stability of potential $V^{4+}(DOPA)_3$ complexes are likely greater than the corresponding $Fe^{3+}(DOPA)_3$ cross-links. Binding of three DOPA residues to V^{4+} requires forcing conditions (*11*). An excess of this ligand is present within the mussel adhesive (*12*).

What benefit would there be from using two metals? Biological materials such as bone and wood gain toughness by having breakable, noncovalent bonds (*13*). When subjected to mechanical forces, these bonds fragment at varied loads to relieve stress buildup while maintaining the larger structure intact (*13*). Synthetic polymers containing catechol have shown variations in viscoelastic properties when cross-linked with iron versus vanadium (*14*). There could be an analogy, with mussels appearing to have both stronger V^{4+}-DOPA bonds and also relatively weaker Fe^{3+}-DOPA interactions. In doing so, the animals may be able to dissipate a wider range of mechanical forces from their environment than if only iron was present.

We are beginning to learn how and why sea creatures process low-abundance metals. This knowledge is already being used to create biomimetic materials, including tissue engineering scaffolds, coatings, and adhesives that work underwater (*15*). Perhaps the variety of properties afforded by different metals will play an increasingly prominent role in these design efforts. Much like how gold and platinum are afforded a distinctive status in commerce, rare metals appear to be saved for special purposes in biology. ■

REFERENCES AND NOTES

1. D. Rehder, *Metallomics* **7**, 730 (2015).
2. T. Priemel *et al.*, *Science* **374**, 206 (2021).
3. J. H. Waite, *J. Exp. Biol.* **220**, 517 (2017).
4. A. Narayanan *et al.*, *Macromolecules* **54**, 5417 (2021).
5. T. Scheibel, *Microb. Cell Fact.* **3**, 14 (2004).
6. M. J. Harrington, A. Masic, N. Holten-Andersen, J. H. Waite, P. Fratzl, *Science* **328**, 216 (2010).
7. M. J. Sever, J. T. Weisser, J. Monahan, S. Srinivasan, J. J. Wilker, *Angew. Chem. Int. Ed.* **43**, 448 (2004).
8. N. A. Hamada, C. Gilpin, J. J. Wilker, *Environ. Sci. Technol.* **54**, 10254 (2020).
9. J. R. Donat, K. W. Bruland, in *Trace Elements in Natural Waters*, B. Salbu, E. Steinnes, Eds. (CRC Press, 1995), pp. 247–281.
10. F. Jehle *et al.*, *Nat. Commun.* **11**, 862 (2020).
11. M. J. Sever, J. J. Wilker, *Dalton Trans. 2006*, 813 (2006).
12. C. N. Z. Schmitt *et al.*, *J. R. Soc. Interface* **12**, 0466 (2015).
13. U. G. K. Wegst, H. Bai, E. Saiz, A. P. Tomsia, R. O. Ritchie, *Nat. Mater.* **14**, 23 (2015).
14. N. Holten-Andersen *et al.*, *J. Mater. Chem. B Mater. Biol. Med.* **2**, 2467 (2014).
15. A. H. Hofman, I. A. van Hees, J. Yang, M. Kamperman, *Adv. Mater.* **30**, 1704640 (2018).

ACKNOWLEDGMENTS

Support provided by the Office of Naval Research (grants N00014-19-1-2342 and N00014-21-1-2178) and the National Science Foundation (2104783) is appreciated.

10.1126/science.abm2298

POLYMER SCIENCE

Pathways to tough yet soft materials

Chemical and physical strategies allow polymer chains in hydrogels to extend or slip

By **Nikola Bosnjak** *and* **Meredith N. Silberstein**

Various engineering applications, ranging from biomedical devices to soft robotics, have increased the demand for functional soft materials. In many of these applications, materials must undergo large deformations without failure. Soft polymeric materials (elastomers) can have the required mechanical and chemical properties but often lack the necessary toughness (the ability to absorb energy and deform without fracturing) to maintain performance when undergoing large deformations. Two papers in this issue showcase distinctly different yet effective paths to create tough soft polymers (see the figure). On page 193, Wang *et al.* (*1*) use force-driven chemical reactions to extend the length of the polymer chains. On page 212, Kim *et al.* (*2*) rely on the physical interactions, in this case, sliding of individual polymer chains. In both cases, the authors report a substantial increase in material toughness that could provide a route for reliable implementation of soft materials in emerging technologies.

"Physics-based models could greatly facilitate strategies for combining different soft material tailoring approaches for target mechanical properties."

The ability of soft polymeric materials to sustain large deformation is directly related to their underlying molecular structure. Elastomers consist of long polymer chains typically coiled up like cooked spaghetti. The chains are interconnected, either by chemical cross-links or physical entanglements, and form a network. Pulling on the material causes the chains to straighten, and interchain connections distribute the applied mechanical load. These connections prevent the chains from slipping past one another and give rise to a solid-like response. Many polymeric networks can absorb water and form a hydrogel that preserves these solid-like features. The chemical interaction of the polymer network with water or solvent allows additional tuning of mechanical and chemical properties, which already depend on the length of the chain segment between cross-links, the physical properties of chain-to-chain connections, and monomers that form the chains.

Early work on elastomer mechanics focused on manipulating properties through cross-link density, entanglement density, and chain flexibility. Toughening could be achieved by introducing filler particles. Detachment of the polymer from the filler dissipates energy but also leads to irreversible damage and is only effective the first time a particular stretch occurs. Many later innovations relied on sacrificial bonds within the material that dissipated energy through bond breakage, a mechanism widely found in biological tissue.

In addition to these sacrificial bonds, reversible bonds were implemented that could reform after breaking, thus avoiding toughness-to-damage coupling and providing both good stiffness and extensibility. Approaches have included ionic bonds, hydrogen bonds, metal coordination bonds, and reversible covalent bonds (*3*–*5*). However, the replacement of permanent covalent bonds with these reversible bonds can lead to a gradual shape change that is undesirable in many applications.

To help address these problems, Gong *et al.* (*6*) developed double-network (DN) gels that combined two polymer networks with complementary properties. Typically, one network is stiffer, more brittle, and has reversible bonds, and the second network is more compliant and stretchable. The second, stretchable network holds the material together and maintains a memory of the initial shape as the first network breaks and dissipates energy locally to pro-

Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14850, USA. Email: meredith.silberstein@cornell.edu

vide toughness. This innovation led to huge improvements in toughness, particularly for hydrogels, whose brittleness severely limited applications. However, the DN approach introduces large hysteresis. The material is much softer as it unloads than during loading. Although this energy dissipation, or lossy behavior, can sometimes be advantageous, it can also be a drawback.

The approach proposed by Wang *et al.* can drastically increase hydrogel toughness through force-driven chemical reactions occurring within the polymeric chains. The limitation of hydrogels caused by network failure after exceeding the maximum end-to-end length of polymer segments is overcome by combining a DN structure with force-sensitive groups. In contrast to prior materials, the brittle network has length-extending, force-sensitive chemical groups, or mechanophores (*7*), integrated into the polymer chains. Previous studies (*8*, *9*) have shown that DN structures can enhance mechanophore response to external forces compared with single networks. The mechanophores of Wang *et al.* open up through cycloreversion just before the chain breakage, which increases the overall end-to-end length of the polymer chain segment by up to 40%. Although force-driven chain unfolding has been demonstrated for single polymer chains and is widespread in biological materials, it had not been reported for bulk synthetic polymers.

An exciting aspect of this approach is that the unfolding inherently takes place where the force is highest. The local chain extension delays failure, allows for greater load to be transferred onto the stretchable network, and redistributes load across the material. Energy is dissipated only when needed to avert fracture, which minimizes lossy response. Experimental tests revealed that the resulting hydrogel is twice as tough against tearing as the regular DN hydrogel in the absence of force-activated unfolding groups.

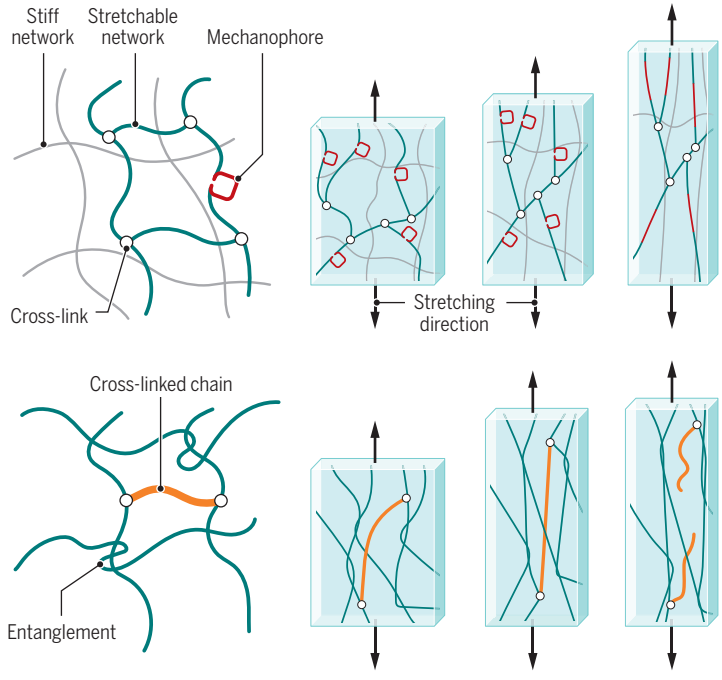
Kim *et al.* realized tough elastomers and hydrogels by modifying the chain-to-chain connections to increase the entanglement-to-cross-link ratio. They identified a regime in which entanglements add substantial stiffness and strength beyond what is provided by the covalent cross-links. However, the entanglements are not cross-links, so applied forces cause the chains to slip rather than fracture. They argue that this approach leads to a more fabric-like rather than net-like material topology. This highly entangled elastomer still follows the theoretical findings of Lake and Thomas from 1967 (*10*) of inverse scaling of toughness with the square root of cross-link density. The energy stored by the full-chain segment length between two cross-links can be dissipated during stretching when one of the bonds in the polymer chain segment breaks.

Although the approach of Kim *et al.* seems quite different from that of Wang *et al.*, it also results in high toughness, allowing three times as much stretching as a regular brittle hydrogel prior to failure, and has even less lossy behavior than the mechanophore approach. Additionally, the added toughness combined with an increased lubrication layer resulting from the low cross-link density leads to ultralow wear. Both studies present fundamental concepts that can be implemented with different polymer chemistries and combined with other state-of-the-art techniques for tailoring polymers. For example, reversible bonds could be incorporated into an entanglement-dominated material to help restore the stiffness lost by cross-links breaking at large strain. Unfolding mechanophores could be reversed through photostimulation (*11*) or even used as cross-link initiation sites (*12*).

Physics-based models could greatly facilitate strategies for combining different soft material tailoring approaches for target mechanical properties. These models will also enable the performance of the material under a particular application to be simulated and lead to a rapid combined material and device design cycle. Tailoring the mechanical properties can enable implementation of soft elastomers in advanced technologies. The delicate restructuring of materials on a molecular level will also further our understanding of polymer physics. By doing so, not only can soft materials be designed and implemented in new devices, but profound knowledge of the mechanisms underlying the macroscopically observed material behavior can also be obtained. ■

Two pathways to tougher hydrogels

Toughness in polymers has been achieved with two different strategies. Wang *et al*. use unfolding molecules in stiff networks, and Kim *et al*. use high ratios of chain entanglements to cross-links.



Lengthening chains
The approach by Wang *et al*. relies on unfolding of mechanophores, or cyclic groups (red) in the cross-linked stiff network. The chains of this network can elongate by 40% during stretching.

Sliding chains
Kim *et al.* use a high ratio of entanglement to cross-links. Applied forces cause the entangled chains to slip and elongate. Even if a cross-link (orange) breaks, the hydrogel does not fracture.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. Z. Wang *et al*., *Science* **374**, 193 (2021).
2. J. Kim, G. Zhang, M. Shi, Z. Suo, *Science* **374**, 212 (2021).
3. Z. P. Zhang, M. Z. Rong, M. Q. Zhang, *Prog. Polym. Sci.* **80**, 39 (2018).
4. E. Khare, N. Holten-Andersen, M. J. Buehler, *Nat. Rev. Mater.* **6**, 421 (2021).
5. S. Utrera-Barrios, R. Verdejo, M. A. López-Manchado, M. Hernández Santana, *Mater. Horiz.* **7**, 2882 (2020).
6. J. P. Gong, Y. Katsuyama, T. Kurokawa, Y. Osada, *Adv. Mater.* **15**, 1155 (2003).
7. C. R. Hickenboth *et al*., *Nature* **446**, 423 (2007).
8. E. Ducrot, Y. Chen, M. Bulters, R. P. Sijbesma, C. Creton, *Science* **344**, 186 (2014).
9. P. Jayathilaka *et al*., *Chem. Commun.* **57**, 8484 (2021).
10. G. J. Lake, A. G. Thomas, *Proc. R. Soc. London Ser. A* **300**, 108 (1967).
11. H. Chen, F. Yang, Q. Chen, J. Zheng, *Adv. Mater.* **29**, 1606900 (2017).
12. A. L. B. Ramirez *et al*., *Nat. Chem.* **5**, 757 (2013).

10.1126/science.abl6358

CANCER

Filling in the GAPs in understanding RAS

A newly identified regulator increases the efficacy of a new class of targeted anti-RAS drugs

By **Adrienne D. Cox**[1,2,3] *and* **Channing J. Der**[1,3]

Since the discovery of mutationally activated RAS in human cancers, small-molecule inhibitors have been long sought but disappointingly elusive. RAS is in an inactive, "off" state when bound to guanosine diphosphate (GDP) and in an active, "on" state when bound to guanosine triphosphate (GTP). According to dogma, mutationally activated RAS is insensitive to GTPase–activating proteins (GAPs) that accelerate conversion of active GTP-bound RAS to inactive, GDP-bound RAS. Yet, a small molecule blocks the growth of cancer cells that express a KRAS Gly^{12}→Cys (KRAS^{G12C}) mutant, despite binding only GDP-bound KRAS^{G12C} (*1*). The US Food and Drug Administration recently approved sotorasib, a KRAS^{G12C} off inhibitor, and more are in the pipeline. How can there be sufficient GDP-bound KRAS^{G12C} to enable inhibitor efficacy? On page 197 of this issue, Li *et al.* (*2*) find that a GAP for heterotrimeric Gα subunits can also promote GTP hydrolysis to stimulate formation of GDP-bound KRAS^{G12C}.

Initial efforts to establish the structural and biochemical consequences of activating RAS point mutations demonstrated that wild-type RAS was largely GDP-bound, whereas mutant RAS was largely GTP-bound. A major advance in understanding the GDP-GTP regulation of RAS came from the seminal discovery in 1987 of a cytosolic GAP that could accelerate the hydrolysis of GTP to GDP on wild-type but not mutant RAS proteins (*3*). This was followed by the identification of p120 RASGAP and neurofibromin as RAS-selective GAPs. From the findings that mutant RAS remained active because it was insensitive to GAP-stimulated hydrolysis of GTP to GDP was born the concept of exploiting GAP activity for anti-RAS drug development. Despite efforts to identify a small molecule that could reawaken GAP activity toward mutant RAS, none were discovered. Numerous other failed approaches earned RAS the unfortunate designation of an "undruggable" cancer target (*4*).

This designation changed with the breakthrough discovery of a compound (*1*) that directly bound to KRAS^{G12C}, impaired downstream effector engagement, and displayed activity against KRAS^{G12C}-mutant cancer cells in vitro. Building on these principles, a flurry of new potent and clinically effective KRAS^{G12C} inhibitors marked an exciting point in anti-RAS drug discovery. However, in the face of dogma that mutant KRAS proteins are GAP-insensitive, these compounds also posed a continuing paradox because they selectively recognized only inactive, GDP-bound KRAS^{G12C}.

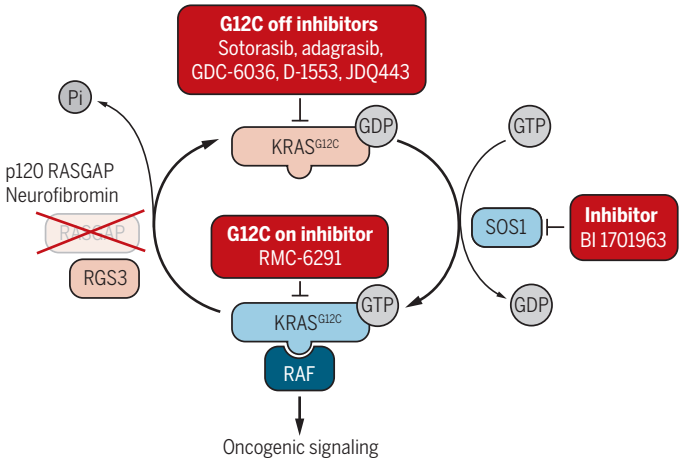
Li *et al.* speculated that there must exist a cellular factor that stimulates formation of GDP-bound KRAS^{G12C}, enabling its vulnerability to KRAS^{G12C} off inhibitors. Their search led them to identify regulator of G protein signaling 3 (RGS3), a GAP for heterotrimeric GαI and Gαq (*5*), as an unexpected GAP for KRAS^{G12C}. Their determination that KRAS^{G12C} is sensitive to the GAP activity of RGS3 marks the second revelation of an unexpected biochemical feature of this particular KRAS mutant. An earlier study (*6*) showed that, also contrary to dogma, the fully activated, GTP-bound KRAS^{G12C} is dependent on receptor tyrosine kinase–mediated stimulation of the RAS guanine nucleotide exchange factor (RASGEF) SOS1 (son of sevenless homolog 1) to promote exchange of GDP for GTP. Consequently, whereas SOS1 inhibitors were once thought to have no therapeutic value against persistently on, mutationally activated RAS, one has now entered clinical evaluation for KRAS-mutant cancers, in combination with either an inhibitor of the downstream kinase MEK (MAPK-ERK kinase) or the KRAS^{G12C} off inhibitor adagrasib (*7*).

The discovery of RGS3 as a weakly activating GAP for mutant KRAS^{G12C} may have implications for the clinical application of KRAS^{G12C} off inhibitors. A subset of non–small-cell lung cancers (NSCLCs) harboring KRAS^{G12C} are resistant de novo to adagrasib (*8*, *9*). Li *et al.* showed that RGS3 expression correlated with sensitivity to KRAS^{G12C} off inhibitors in patient-derived NSCLC xenograft models (*2*). RGS3 may thus serve as a biomarker to identify patients who will be most responsive to these inhibitors. Of the genetic alterations associated with treatment-induced acquired resistance, most are the logical suspects: mutational activation of upstream receptor tyrosine kinases, emergence of second-site RAS mutations that are refractory to drug binding and/or inhibition, and mutational activation of downstream effector signaling (*8*, *9*). However, in many cancers that are resistant to KRAS^{G12C} inhibitors, the mechanisms remain to be identified—perhaps decreased expression of RGS3 is one contributor.

Although the discovery of RGS3 GAP activity for KRAS^{G12C} provides one explanation for the effectiveness of sotorasib, adagrasib, and other KRAS^{G12C} off inhibitors under clinical evaluation (see the figure), a debate

Inhibitors of mutant KRAS

The KRAS^{G12C} (Gly^{12}→Cys) mutant is refractory to canonical GAPs, p120 RASGAP and neurofibromin, but not RGS3, which promotes GTP hydrolysis. The resulting GDP-bound KRAS^{G12C} is an anticancer drug target. Combination with inhibitors of SOS1 or with inhibitors of downstream signaling may further improve efficacy.



GAP, GTPase activating protein; GDP, guanosine diphosphate; GTP, guanosine triphosphate; Pi, inorganic phosphate; RGS3, regulator of G protein signaling 3; SOS1, son of sevenless homolog 1.

GRAPHIC: C. BICKEL/*SCIENCE*

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, NC, USA. [2]Department of Radiation Oncology, University of North Carolina at Chapel Hill, NC, USA. [3]Department of Pharmacology, University of North Carolina at Chapel Hill, NC, USA. Email: adrienne_cox@med.unc.edu; cjder@med.unc.edu

remains whether a KRAS^{G12C} on inhibitor that binds directly to GTP-bound KRAS^{G12C} will be superior, with or without SOS1 inhibition. One such KRAS^{G12C} on inhibitor, RMC-6291, is currently under advanced preclinical development.

RGS3 lacks the catalytic arginine finger through which canonical RASGAPs promote GTP hydrolysis and to which KRAS G12 mutants are insensitive. Instead, Li *et al.* show that the RASGAP activity of RGS3 relies on an asparagine residue that fosters the transition state from GTP to GDP and that is oriented in a manner that renders RGS3 capable of still acting on conventionally GAP-insensitive KRAS^{G12C} and other mutants. Perhaps this provides a clue to revisit the concept of a small-molecule drug that can accelerate the impaired GTPase activity of mutant RAS proteins in cancer. The RASGAP activity of RGS3 on KRAS^{G12C} was stronger on KRAS^{G12C} than that of the canonical RASGAPs but was one to two orders of magnitude weaker than that of the canonical RASGAPs on wild-type KRAS. The anticancer efficacy of such a small molecule would be enhanced by targeting it selectively to RAS rather than G proteins. Additionally, because RGS3 was also active against the most prevalent KRAS mutations (G12D, G12V, and G13D), an RGS3 mimetic may serve as an effective pan-mutant RAS inhibitor.

RGS3 is a member of a larger family of more than 20 RGS proteins, many of which—including RGS3—have multiple splice variants. Still other proteins possess an RGS domain, which is where the GAP activity of RGS3 resides (*5*). The critical asparagine residue of RGS3 is conserved among RGS isoforms, suggesting that more RGS RASGAPs remain to be identified. Indeed, Li *et al.* showed that RGS4 has RASGAP activity on KRAS^{G12C}. This asparagine residue is also catalytically important in several GAPs for RAS-related small GTPases, some of which may also have previously unappreciated RASGAP activity toward specific RAS mutants. Perhaps GAP, a once discarded target, will reenter the stage for anti-RAS cancer drug discovery. ■

REFERENCES AND NOTES

1. J. M. Ostrem *et al.*, *Nature* **503**, 548 (2013).
2. C. Li *et al.*, *Science* **374**, 197 (2021).
3. M. Trahey, F. McCormick, *Science* **238**, 542 (1987).
4. B. Papke, C. J. Der, *Science* **355**, 1158 (2017).
5. J. B. O'Brien *et al.*, *J. Biol. Chem.* **294**, 18571 (2019).
6. P. Lito, M. Solomon, L. S. Li, R. Hansen, N. Rosen, *Science* **351**, 604 (2016).
7. J. Hallin *et al.*, *Cancer Discov.* **10**, 54 (2020).
8. N. Tanaka *et al.*, *Cancer Discov.* **11**, 1913 (2021).
9. M. M. Awad *et al.*, *N. Engl. J. Med.* **384**, 2382 (2021).

ACKNOWLEDGMENTS

We thank C. Stalnecker for help with the figure. The authors are supported by grants from the National Cancer Institute (CA203657, CA224428, and CA232113). C.J.D. is a scientific advisory board member for, and receives research funding from, Mirati Therapeutics.

10.1126/science.abl3639

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

NEUROSCIENCE

Revisiting brain stimulation in Parkinson's disease

Electrical burst stimulation rescues movement effectively in parkinsonian mice

By **Carola A. Haas**

Parkinson's disease (PD) is a progressive neurological disorder characterized by tremor and slowness of movement as well as debilitating nonmotor symptoms, including cognitive decline and depression (*1*). PD is caused by the progressive degeneration of dopaminergic neurons in the substantia nigra (SN), located in the midbrain. The death of SN neurons causes a deficiency in the neurotransmitter dopamine in the basal ganglia circuitry—a brain region that is important for movement control. The main symptomatic treatment for PD consists of dopamine replacement therapy; however, this becomes ineffective over time, and patients inevitably develop motor complications. At this stage, deep brain stimulation (DBS) is applied at high frequencies to rescue motor function. Unfortunately, DBS also has strong side effects and is only effective as long as stimulation is maintained (*2*). On page 201 of this issue, Spix *et al.* (*3*) describe a highly precise DBS stimulation protocol with long-term efficacy in mice.

Department of Neurosurgery, Medical Center–University of Freiburg, Faculty of Medicine, Freiburg, Germany. Email: carola.haas@uniklinik-freiburg.de

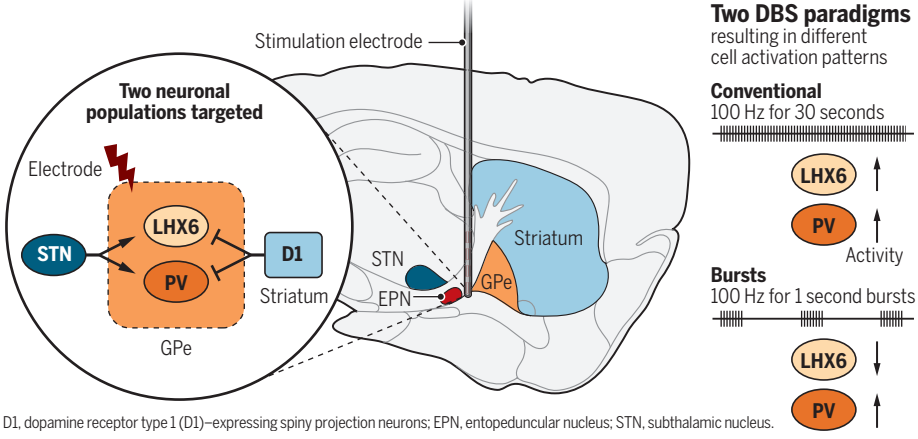
In PD, the loss of dopaminergic neurons disrupts the function of the basal ganglia—a group of interconnected subcortical nuclei. It is unclear why these neurons die, but an enhanced vulnerability to environmental toxins has been proposed. The dopamine loss causes an imbalance in the synaptic connections between the different nuclei, which leads to an over-inhibition of the thalamus and cortex, thereby hampering movement execution (*4*).

The therapeutic gold standard for PD is treatment with the dopamine precursor levodopa. However, there are side effects of levodopa, such as involuntary movements (dyskinesias), particularly after several years of treatment. Dopamine receptor agonists represent an alternative option early in the course of the disease because they offer antiparkinsonian effects with a lower risk of developing troublesome dyskinesias. Yet, the use of dopamine receptor agonists is not devoid of motor complications, and these compounds often cause severe nonmotor side effects, such as psychiatric disorders and autonomic malfunction (*5*).

Here, DBS comes into play for relieving PD motor symptoms. In 1989, DBS was successfully applied for the first time to reduce tremor in PD patients (*6*). Since then, DBS has been developed further. The two most

Electrical burst stimulation in parkinsonian mice

Deep brain stimulation (DBS) electrode placement in the globus pallidus externus (GPe) in parkinsonian mice can activate parvalbumin (PV)–GPe neurons and repress LIM homeobox protein 6 (LHX6)–GPe neurons, which results in lasting motor improvement in parkinsonian mice.



D1, dopamine receptor type 1 (D1)–expressing spiny projection neurons; EPN, entopeduncular nucleus; STN, subthalamic nucleus.

common targets for PD-DBS are the internal pallidal segment [internal globus pallidus (GPi)] and the subthalamic nucleus (STN), both nodes of the basal ganglia–thalamocortical motor circuit. Although both targets are effective for reducing levodopa-induced dyskinesias and improving mobility, STN-DBS has the additional advantage of reducing the daily levodopa requirement. However, there are considerable nonmotor side effects of STN-DBS, such as depression, psychosis, and impulse-control disorders (*2*). Additionally, the effect of DBS decays rapidly once stimulation is switched off, and the mechanism underlying DBS efficacy is largely unknown.

These issues were addressed by Spix *et al.*, who targeted specific neuronal subtypes in the globus pallidus externus (GPe) of mice with the aim to develop a cell type–specific DBS protocol. The GPe is a key contributor to motor-suppressing pathways in the basal ganglia by means of γ-aminobutyric acid (GABA)–ergic inhibitory synapses on thalamic neurons. Moreover, the GPe is a larger structure than the STN or the GPi in humans and mice and is therefore easier to target and stimulate. Spix *et al.* built on previous findings showing that optogenetic manipulation of distinct, spatially intermixed neuronal GPe subpopulations provides long-lasting therapeutic effects in dopamine-depleted mice (*7*). Specifically, they found that parvalbumin-expressing (PV)-GPe neurons must be excited, whereas LIM homeobox protein 6 (LHX6)-GPe neurons must be suppressed. Because optogenetics is not feasible in human patients, Spix *et al.* aimed to identify parameters of electrical stimulation tailored for cell-type specificity to improve the therapeutic efficacy of DBS, which can be transferred to human patients.

To identify the responses of PV-GPe and LHX6-GPe neurons to electrical stimulation, Spix *et al.* recorded from these neuron populations in brain slices obtained from transgenic mice in which both GPe neuron types were endogenously labeled. Both neuron types were equally excited by continuous electrical stimulation (100 Hz/30 s). These results led to the hypothesis that delivering stimulation in short bursts, rather than continuously, could achieve population-specific neuromodulation in the GPe. Therefore, Spix *et al.* delivered repeated bursts (100 Hz/1 s) to pairs of LHX6-GPe and PV-GPe neurons. This pattern caused robust and reproducible suppression of LHX6-GPe neurons and simultaneous excitation of PV-GPe neurons—a result similar to that obtained using optogenetic stimulation (*7*).

To investigate the parameter space for optimal stimulation Spix *et al.* used a machine learning approach. They found that shorter, higher-frequency bursts (175 Hz/200 ms)—but not lower-frequency bursts (50 Hz/200 ms)—were effective. A similar approach was applied recently to determine efficient stimulation parameters for the STN, which normalized pathological hyperactivity of pyramidal cells in the motor cortex of parkinsonian mice (*8*).

Spix *et al.* also investigated the mechanisms underlying the cell type–specific inhibitory synaptic transmission in the GPe. To this end, they optogenetically stimulated incoming fibers from either the STN or the striatum at high frequencies using a new channelrhodopsin variant called *Chronos*, which responds rapidly to light activation (*9*). However, these stimulations did not lead to selective inhibition of LHX6-GPe neurons. By contrast, expression of *Chronos* in striatal dopamine receptor 1–positive neurons and stimulation at high frequencies led to robust inhibition of LHX6-GPe but not of PV-GPe neurons, showing that striatal afferents from dopamine receptor type 1 (D1)-expressing spiny projection neurons (SPNs) preferentially inhibit LHX6-GPe neurons.

Spix *et al.* also tested the relevance of their so-called circuit-inspired burst stimulation in vivo in a PD mouse model by alternating the application of conventional and burst DBS. The burst mode resulted in restoration of movement during and, most notably, for 2 hours after DBS stopped, showing long-lasting therapeutic benefits that far exceeded those induced by conventional DBS. In the future, this burst stimulation protocol could be tested in PD patients because the frequencies are in a clinically approved range.

There is a growing amount of research in which cell type–specific, optogenetic stimulation is applied in mice to unravel the cellular and circuit mechanisms of DBS (*7*, *8*, *10*, *11*) and to optimize it with a translational perspective. The study by Spix *et al.* is an excellent example of "optogenetics-inspired" DBS and may pave the way to more-robust DBS approaches that ultimately can be translated to humans. ■

REFERENCES AND NOTES

1. W. Poewe *et al.*, *Nat. Rev. Dis. Primers* **3**, 17013 (2017).
2. M. Paff *et al.*, *J. Mov. Disord.* **13**, 185 (2020).
3. T. A. Spix *et al.*, *Science* **374**, 201 (2021).
4. M. M. McGregor, A. B. Nelson, *Neuron* **101**, 1042 (2019).
5. Y. Smith, T. Wichmann, S. A. Factor, M. R. DeLong, *Neuropsychopharmacology* **37**, 213 (2012).
6. A. L. Benabid *et al.*, *Rev. Neurol.* **145**, 320 (1989).
7. K. J. Mastro *et al.*, *Nat. Neurosci.* **20**, 815 (2017).
8. S. Valverde *et al.*, *Nat. Commun.* **11**, 2388 (2020).
9. N. C. Klapoetke *et al.*, *Nat. Methods* **11**, 338 (2014).
10. E. Paschen *et al.*, *eLife* **9**, e54518 (2020).
11. G. D. R. Watson *et al.*, *Sci. Adv.* **7**, eabe9192 (2021).

ACKNOWLEDGMENTS

The authors thank M. Kirsch for critical reading of the manuscript and helpful discussions and P. Kleis for figure design.

10.1126/science.abl9915

CANCER

Gut bacteria enable prostate cancer growth

Testosterone-synthetizing gut bacteria drive resistance to therapy

By **John A. McCulloch**[1,2] *and* **Giorgio Trinchieri**[3]

Prostate cancer is driven by androgens (particularly testosterone), so androgen deprivation therapy (ADT) is the standard of care for patients with advanced prostate cancer. After an initial response, hormone-sensitive prostate cancer (HSPC) usually progresses to castration-resistant prostate cancer (CRPC) owing to various mechanisms that may include altered androgen receptor (AR) signaling and intratumor accumulation of androgens even during ADT. On page 216 of this issue, Pernigoni *et al.* (*1*) show that gut commensal bacteria that are able to synthesize the sex hormone precursor dehydroepiandrosterone (DHEA) and testosterone increase in patients and mice with CRPC and prevent complete suppression of androgens during ADT. In mice, these androgen-producing bacteria can be targeted to slow progression from HSPC to CRPC.

Commensal bacteria and other microbes, called the microbiota, are present on all epithelial interfaces of higher organisms. The gut microbiota have a profound influence on host physiology and pathology, and in particular affect all stages of cancer (*2*). Additionally, the microbiota can modulate the responsiveness to cancer chemotherapy and immunotherapy (*3*). Clinical proof of concept for this hypothesis was provided by showing that fecal microbiota transplantation (FMT) from responsive patients overcomes resistance to immunotherapy in some melanoma patients (*4*).

Sex influences microbiota composition, and the microbiota determines the concentration of androgens; the sexual dimorphism

[1]Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD, USA. [2]Genetics and Microbiome Core, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. [3]Cancer Immunobiology Section, Laboratory of Integrative Cancer Immunology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. Email: john.mcculloch@nih.gov; trinchig@mail.nih.gov

in testosterone is almost completely abolished in germ-free mice (*5, 6*). Male FMT into female mice increases circulating testosterone and attenuates susceptibility to autoimmune diseases (*5*). Thus, sex differences are modulated through cross-talk between microbiota, hormones, and immunity that can alter susceptibility to diseases (*7*). Although gut bacteria can modulate the amount of circulating sex hormones by affecting host cells, they can also directly biotransform or synthesize them. Androgens are glucuronidated in the liver for solubilization and excretion, but a fraction of these conjugated androgens are carried with the bile into the intestine. However, unconjugated androgens, testosterone and dihydrotestosterone (DHT), are present in the bacteria-rich large intestine of young men at concentrations much higher than in the serum and from there they can be reabsorbed (*8*). This deglucuronidation of androgens does not happen in germ-free mice and is dependent on bacterial β-glucuronidase (*8*). Gut bacteria, such as *Clostridium scindens*, catalyze the side chain of the steroid hormone cortisol by steroid-17,20-desmolases, forming 11β-hydroxyandrostenedione, primarily produced in humans from androstenedione in the adrenal gland. 11β-Hydroxyandrostenedione gives rise to the potent 11-oxyandrogens that participate in the intratumor androgen pool that favors tumor progression in CRPC (*9, 10*).

The gut microbiota of ADT-treated prostate cancer patients showed an increased abundance of species such as *Akkermansia muciniphila* and *Oscillospiraceae (Ruminoccocaceae)* spp., as well as an enriched representation of bacterial gene pathways involved in steroid hormone biosynthesis (*11*). Pernigoni *et al.* show that androgen biosynthesis by commensal gut bacteria increases in a mouse model of CRPC and in prostate cancer patients. Using different mouse models of progression from HSPC to CRPC, they observed that tumor progression following castration (androgen deprivation) was significantly slowed by depletion of gut microbiota with antibiotics. The composition of the gut microbiota was different in HSPC versus CRPC mice, with an enrichment in CRPC of *Ruminococcus (Mediterraneibacter) gnavus* and the *Bacteroides acidifaciens*. This alteration of the gut microbiota composition and its ability to sustain CRPC growth were not due to modulation of immunity.

Metabolic analyses revealed that testosterone and DHEA, but not their precursor pregnenolone, were decreased in antibiotic-treated mice but increased in mice receiving FMT from CRPC mice or *R. gnavus*. In the presence of pregnenolone and hydroxypregnenolone, *R. gnavus*, *B. acidifaciens*, and *C. scindens* (but not other bacteria) produced DHEA and testosterone. These are normally synthesized from pregnenolone in the host, catalyzed by 17α-hydroxylase/17,20-desmolase (CYP17A1). The selective CYP17A1 inhibitor abiraterone inhibits androgen synthesis in cultured bacteria. *R. gnavus* exposed to pregnenolone up-regulates the expression of 22 genes that could be potential candidates for the CYP17A1-like activity.

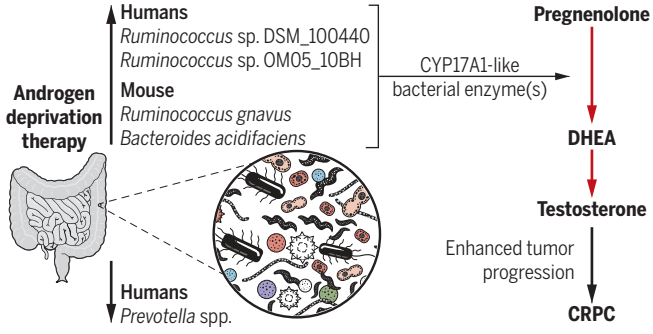
The production of 11-oxyandrogens from cortisol as described for *C. scindens* was not detected by the metabolomic analysis and not investigated. Genes homologous to *C. scindens desA* and *desB* that encode the steroid-17,20-desmolase (which initiates 11-oxyandrogen synthesis) are present in many *Lachnospiraceae* and *Oscillospiraceae* family members, which are related but not identical to the gut bacteria enriched in CRPC mice and patients. This suggests that 11-oxyandrogen synthesis may be worth investigating in CRPC (*9*). Also, several *Lachnospiraceae*, including *R. gnavus*, express β-glucuronidase, raising the possibility that in vivo deglucuronidation of testosterone and DHT by bacteria in the large intestine may play a role in increasing androgen bioavailability in tumors (*8, 12*).

Pernigoni *et al.* found that CRPC patients on ADT had increased *Oscillospiraceae* family members *Ruminococcus* spp. DSM_100440 and OM05_10BH and decrease of *Prevotellaceae* spp. in the gut. *Ruminococcus* sp. DSM_100440 synthesized testosterone and DHEA from pregnenolone in culture and when instilled in castrated and antibiotic-treated mice, enhanced the progression of CRPC tumors (see the figure). Conversely, like antibiotics, *Prevotella stercorea* decreased the growth of CRPC tumors in castrated mice, likely by competing for the expansion of androgen-producing bacteria. The presence of such beneficial species in the gut microbiota correlated with improved survival of patients with either HSPC or CRPC, whereas the presence of the *Oscillospiraceae* species correlated with decreased survival. Moreover, residual plasma testosterone in patients treated with ADT correlated with the presence in their gut microbiota of *Ruminococcus* sp. DSM_100440.

The work of Pernigoni *et al.* reveals that gut commensal bacteria that become more abundant in ADT-treated animals and patients synthesize androgens that affect CRPC tumor progression and, for the first time, describes testosterone biosynthesis by bacteria. Different pathways of androgen synthesis and biotransformation could be shared by the same bacterial species (e.g., *C. scindens*) or mediated by distinct species. In addition to the gut commensals, whether the microbiota within a tumor or the urinary tract could be involved in production of androgens should be investigated. How the microbiota senses the reduction of androgen concentrations and responds with expansion of species that are able to synthesize androgens remains unclear. It is possible that metabolites downstream of AR signaling, such as fatty acids, may regulate the composition of the microbiota (*13*). Targeting the gut microbiota or the metabolic pathways involved in bacterial androgen biosynthesis or the cross-talk between the host and bacteria may open new therapeutic options for advanced prostate cancer patients. ■

Gut bacteria can enhance prostate cancer

Advanced prostate cancer patients on androgen deprivation therapy have altered gut microbiota composition, with an accumulation of bacterial species that can synthesize dehydroepiandrosterone (DHEA) and testosterone from pregnenolone [through 17α-hydroxylase/17,20-desmolase (CYP17A1)–like activity]. The presence of these species enhances tumor growth and progression to castration-resistant prostate cancer (CRPC).



GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. N. Pernigoni *et al.*, *Science* **374**, 216 (2021).
2. A. Dzutsev *et al.*, *Annu. Rev. Immunol.* **35**, 199 (2017).
3. D. Davar *et al.*, *Science* **371**, 595 (2021).
4. N. Iida *et al.*, *Science* **342**, 967 (2013).
5. J. G. Markle *et al.*, *Science* **339**, 1084 (2013).
6. G. Falony *et al.*, *Science* **352**, 560 (2016).
7. R. Vemuri *et al.*, *Semin. Immunopathol.* **41**, 265 (2019).
8. H. Colldén *et al.*, *Am. J. Physiol. Endocrinol. Metab.* **317**, E1182 (2019).
9. J. M. Ridlon, *Appl. Environ. Microbiol.* **86**, e00641 (2020).
10. M. Barnard, E. A. Mostaghel, R. J. Auchus, K. H. Storbeck, *J. Steroid Biochem. Mol. Biol.* **197**, 105506 (2020).
11. K. S. Sfanos *et al.*, *Prostate Cancer Prostatic Dis.* **21**, 539 (2018).
12. R. M. Pollet *et al.*, *Structure* **25**, 967 (2017).
13. K. Mittelstrass *et al.*, *PLOS Genet.* **7**, e1002215 (2011).

10.1126/science.abl7070

BATTERIES

Chelated electrolytes for divalent metal ions

Chelating electrolytes reorganize ion solvation and enable reversible magnesium batteries

By **Pengjian Zuo** and **Geping Yin**

Divalent metal batteries, particularly those based on magnesium (Mg), could offer lower cost and higher safety and energy density than lithium (Li) batteries. Demonstration of a prototype Mg-based battery almost two decades ago (*1*) led to efforts to find suitable electrolytes and Mg^{2+}-ion insertion compounds and to modify the metal anode (*2*). However, no Mg-based electrochemical device has been practical for energy storage up to now. Drawbacks include low working voltages and poor cycling performance that result from the lack of an electrolyte that avoids formation of a passivation interface layer on the Mg anode or that allows fast Mg^{2+} migration in high-voltage metal-oxide cathodes (*3*, *4*). On page 172 of this issue, Hou *et al.* (*5*) propose a versatile electrolyte design strategy for divalent metal batteries in which strong chelating agents interact with the cations. This approach greatly increased the reversibility of the divalent metal battery and the charge-transfer kinetics.

High-quality electrode-electrolyte interfaces are the key to realizing practical batteries with high energy density. In Li^+-ion batteries, strategies such as using high-concentration electrolyte (*6*, *7*), using weakly solvating electrolyte (*8*), and introducing additional salt (*9*) were applied to improve the electrode-electrolyte interfacial chemistry and promote the formation of high-quality solid-electrolyte interface (SEI) films by introducing anions in the solvation sheath of Li^+ ions. However, in Mg-based batteries, the SEI that is formed on the Mg anode by the decomposition of solvents and salts inhibits Mg^{2+}-ion migration because of strong electrostatic interactions.

Hou *et al.* found that the solvation sheath of Mg^{2+} is expected to be stable and anion-free, unlike Li^+ ions. In the traditional Mg-battery electrolyte that contains chloride ions (Cl^-), Mg^{2+} is tetra- or hexacoordinated by a combination of easily reduced tetrahydrofuran and Cl^- (*10*, *11*). In $Mg(TFSI)_2$-ether electrolytes, where TFSI is bis(trifluoromethanesulfonimide), Mg^{2+} is usually solvated with unstable ether solvent molecules (*12*), and $TFSI^-$ will also enter the solvation sheath in the presence of cosolvent (*13*). At present, there are no reports suggesting that a solvent, instead of ether and anions, can reorganize the solvation sheath of Mg^{2+}.

Considering the diversity of electrolyte structure and the complexity of electrochemical reactions at the electrode interface, an effective electrochemical reaction model is urgent for the development of a Mg metal battery. Hou *et al.* developed a simple and high-accuracy theoretical model that combines molecular dynamics simulations and Marcus-Hush-Chidsey dynamics theory to determine the effect of the electrolyte coordination environment on the charge transfer kinetics. This model provides a powerful tool to understand the nature of electrochemical reactions and of electrode reactions by tuning the solvation sheath.

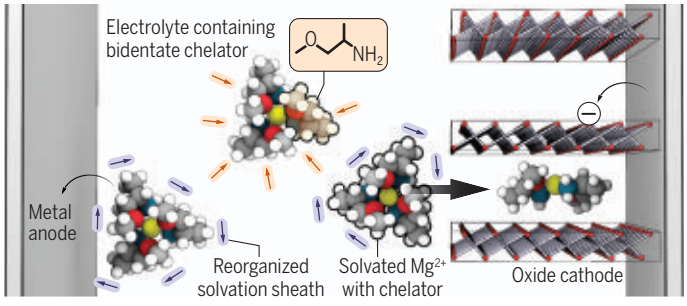
The method of solvation sheath reorganization effectively inhibits the continuous decomposition of dimethyl ether solvents on the Mg electrode and avoids the formation of a passivated SEI. It creates an alternative approach compared with the suitable SEI film that was formed on the surface of the Mg electrode in previous studies (*14*). The chelate-rich solvation sheath can alleviate the electrostatic interaction of naked Mg^{2+} ions with the framework constituents, which results in fast Mg^{2+} diffusion (see the figure). In aqueous electrolytes, Mg^{2+} ions can bypass dehydration during intercalation and extraction (*15*). The method of solvation sheath reorganization by Hou *et al.* pioneers the co-intercalation of solvents and Mg^{2+} ions in organic electrolytes.

This approach of Hou *et al.* could enable relatively high concentrations of Mg^{2+} ions in battery electrolytes. The high affinity between methoxyethyl amine–based chelate and Mg^{2+} could allow it to prevail over the anions during solvation and avoid the formation of an undesired SEI layer through anion decomposition. It may enable the use of Mg salts and carbonate-based electrolytes in Mg^{2+} ion batteries. A chelant could prevent the components that easily passivate the Mg electrode—such as perchlorate, hexafluorophosphate, and propylene carbonate—from entering the first solvation sheath. This proposed method will reduce dependence on Grignard reagent–based electrolytes in the Mg deposition and dissolution electrochemistry and enable simple Mg-salt electrolytes in Mg batteries.

The proposed strategy of sheath reorganization provides guidance for achieving partial desolvation to match the active material frame structure. The application of solvent sheath reorganization should be extended to other high–specifc energy cathode materials. Greater attention to the divalent metal cation solvation sheath should lead to the development of new electrolytes, and the establishment of electrode–solvated metal cation interface models will be extensively researched. ■

Improving both ends of a magnesium battery

Hou *et al.* show that chelating electrolytes reduce the magnesium overpotential at the metal anode. This approach also achieves fast intercalation and deintercalation at layered oxide cathodes through the solvation sheath reorganization.



MIIT Key Laboratory of Critical Materials Technology for New Energy Conversion and Storage, School of Chemistry and Chemical Engineering, Harbin Institute of Technology, Harbin 150001, China. Email: zuopj@hit.edu.cn; yingphit@hit.edu.cn

REFERENCES AND NOTES

1. D. Aurbach *et al.*, *Nature* **407**, 724 (2000).
2. F. Liu, T. Wang, X. Liu, L. Z. Fan, *Adv. Energy Mater.* **10**, 2000787 (2020).
3. Z. Rong *et al.*, *Chem. Mater.* **27**, 6016 (2015).
4. R. Attias, M. Salama, B. Hirsch, Y. Goffer, D. Aurbach, *Joule* **3**, 27 (2019).
5. S. Hou *et al.*, *Science* **374**, 172 (2021).
6. L. L. Jiang *et al.*, *Angew. Chem.* **60**, 3402 (2021).
7. Z. Chang *et al.*, *Energy Environ. Sci.* **13**, 4122 (2020).
8. Y. X. Yao *et al.*, *Angew. Chem.* **60**, 40907 (2021).
9. S. H. Lee *et al.*, *Adv. Energy Mater.* **10**, 2000567 (2020).
10. L. F. Wan, D. Prendergast, *J. Am. Chem. Soc.* **136**, 14456 (2014).
11. P. Canepa *et al.*, *Energy Environ. Sci.* **8**, 3718 (2015).
12. T. Kimura, K. Fujii, Y. Sato, M. Morita, N. Yoshimoto, *J. Phys. Chem. C* **119**, 18911 (2015).
13. S. Fan *et al.*, *ACS Appl. Mater. Interfaces* **12**, 10252 (2020).
14. Y. Li *et al.*, *J. Energy Chem.* **37**, 215 (2019).
15. S. Yagi *et al.*, *J. Electrochem. Soc.* **162**, A2356 (2015).

10.1126/science.abi6643

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

RETROSPECTIVE

Edmond Fischer (1920–2021)

A biochemical virtuoso

By **John D. Scott[1], Trisha N. Davis[2], Rachel E. Klevit[2], William A. Catterall[1]**

Visionary scientist Edmond "Eddy" Henri Fischer was a virtuoso in multiple fields. He transformed our understanding of biochemistry and physiology by discovering how reversible phosphorylation of proteins works as a switch to regulate cellular processes. He was also an exceptional pianist who regularly delighted friends, family, and colleagues with his renderings of Chopin, Schumann, or Bach. When he died on 27 August at the age of 101, the Swiss American biochemist was the oldest living Nobel laureate.

Born on 6 April 1920 in the consular section of Shanghai to a French mother and an Austrian father, Fischer was introduced to the pioneering work of the renowned bacteriologist Louis Pasteur at La Châtaigneraie International School in Geneva. This stirred in him a passion for science, and the young Fischer expressed a desire to "cure the ills of the world." Admitted to the Geneva Conservatory of Music as a pianist, he contemplated becoming a professional musician, but science was a stronger pull. Instead, he attended the University of Geneva, where he excelled in both organic chemistry and biology, foreshadowing his entry into the burgeoning field of biochemistry.

Under the supervision of Kurt Heinrich Meyer, Fischer completed his PhD in 1947 on the structure of polysaccharides and α-amylases, enzymes that catalyze hydrolysis of starch into sugars. In 1953, Fischer crossed the Atlantic with the intention of starting a postdoctoral fellowship at Caltech, but instead he joined the faculty of the Department of Biochemistry at the University of Washington (UW) in Seattle, where he remained active as professor emeritus until 2020, long after his retirement. This providential recruitment was orchestrated by Hans Neurath, the inaugural chair of biochemistry in the newly formed UW School of Medicine. Shortly after arriving in Seattle, Fischer met UW biochemist Edwin G. Krebs, thus beginning a lifelong collaboration and friendship.

Fischer and Krebs were a formidable research team, focusing on biochemical aspects of hormone action and metabolism. Universally known to colleagues and other scientists as Eddy and Ed, they liked to explain their success by quipping that "two Eds are better than one." Together, they investigated how endocrine factors influence blood glucose levels and trigger the adrenal response, noting intriguing parallels between plant enzymes called phosphorylases that metabolize starch into glucose-1-phosphate and glycogen metabolism in rabbit skeletal muscle. In 1955, they found that adenosine triphosphate (ATP) is required for phosphorylase activation. In a serendipitous experiment, they discovered that calcium, leaching from filter paper used to clarify the extract, is an important cofactor in the process. By using γ-^{32}P–labeled ATP, they demonstrated a seminal finding: Phosphate is incorporated into a specific serine residue of phosphorylase.

PHOTO: RAPHAEL GAILLARDE/GAMMA-RAPHO/GETTY IMAGES

Subsequently, they confirmed that serine phosphorylation is mediated by phosphorylase b kinase, which is itself controlled by a cyclic adenosine monophosphate–responsive kinase, leading to the notion of a kinase phosphorylation cascade. In 1992, Fischer and Krebs shared the Nobel Prize in Physiology or Medicine for their discovery of "reversible protein phosphorylation as a biological regulatory mechanism," arguably the most important discovery in cell biology in the 20th century. By then, their work had sparked an area of research that remains one of the most wide-ranging and active in biomedical science. Their work paved the way for the development of drugs that combat cancer, diabetes, and other diseases.

Fischer was appropriately recognized for his contributions to science. He won the Prix Jaubert from the University of Geneva and was elected a member of the US National Academy of Sciences and a foreign fellow of the Royal Society, London. However, his scientific legacy reaches far beyond published works and international recognition. He was a consummate mentor firmly committed to the career development and well-being of the next generation of researchers. His guidance nourished an extensive scientific tree, including many leading figures in the signal transduction community. One branch includes former postdoctoral fellows Sir Philip Cohen (University of Dundee) and Nicholas Tonks (Cold Spring Harbor Laboratory). Fischer was proud that this field of science now also includes his granddaughter Elyse. His commitment to the next generation was never more evident than during his regular participation in the Lindau Nobel Laureate Meetings, a forum for exchange between scientists of different generations, cultures, and disciplines. Thus, the Fischer dynasty will be in good hands for years to come.

We remember Eddy for his penetrating questions in seminars, no matter what the topic; his wry humor and good spirits, no matter what the situation; and his constant generosity of time and insights to colleagues. During the search to fill the Edmond H. Fischer–Washington Research Foundation Endowed Chair in Biochemistry, Eddy, then in his 90s, insisted on holding the dinners for candidates in his home and cooking the food himself, despite the trepidation expressed by the search committee. He served his favorite Swiss dishes and sweet Swiss white wine. Not wanting to take credit he was not due, he announced that he had not made the dessert and emphasized, "I don't do desserts." One candidate remarked that, regardless of whether an offer was made, a dinner at Eddy's would stand as one of the highlights of their life.

Unfortunately, Eddy's 100th birthday coincided with the first wave of the COVID-19 pandemic, thwarting plans for an in-person celebration. Instead, we convened a few months later in a virtual symposium to honor him. The socially distanced event attracted several hundred participants from all over the world. Former trainees (including J.D.S.), family members, faculty colleagues (including T.N.D., R.E.K., and W.A.C.), and friends celebrated his magnificent life and paid tribute to the warmth, humanity, and charisma of this most beloved virtuoso of biochemistry. ■

[1]Department of Pharmacology, University of Washington School of Medicine, Seattle, WA 98195, USA. [2]Department of Biochemistry, University of Washington School of Medicine, Seattle, WA 98195, USA. Email: wcatt@uw.edu

10.1126/science.abm3289

POLICY FORUM

CLIMATE CHANGE

Intergenerational inequities in exposure to climate extremes

Young generations are severely threatened by climate change

By **Wim Thiery, Stefan Lange, Joeri Rogelj, Carl-Friedrich Schleussner, Lukas Gudmundsson, Sonia I. Seneviratne, Marina Andrijevic, Katja Frieler, Kerry Emanuel, Tobias Geiger, David N. Bresch, Fang Zhao, Sven N. Willner, Matthias Büchner, Jan Volkholz, Nico Bauer, Jinfeng Chang, Philippe Ciais, Marie Dury, Louis François, Manolis Grillakis, Simon N. Gosling, Naota Hanasaki, Thomas Hickler, Veronika Huber, Akihiko Ito, Jonas Jägermeyr, Nikolay Khabarov, Aristeidis Koutroulis, Wenfeng Liu, Wolfgang Lutz, Matthias Mengel, Christoph Müller, Sebastian Ostberg, Christopher P. O. Reyer, Tobias Stacke, Yoshihide Wada**

Under continued global warming, extreme events such as heat waves will continue to rise in frequency, intensity, duration, and spatial extent over the next decades (*1*–*4*). Younger generations are therefore expected to face more such events across their lifetimes compared with older generations. This raises important issues of solidarity and fairness across generations (*5*, *6*) that have fueled a surge of climate protests led by young people in recent years and that underpin issues of intergenerational equity raised in recent climate litigation. However, the standard scientific paradigm is to assess climate change in discrete time windows or at discrete levels of warming (*7*), a "period" approach that inhibits quantification of how much more extreme events a particular generation will experience over its lifetime compared with another. By developing a "cohort" perspective to quantify changes in lifetime exposure to climate extremes and compare across generations (see the first figure), we estimate that children born in 2020 will experience a two- to sevenfold increase in extreme events, particularly heat waves, compared with people born in 1960, under current climate policy pledges. Our results highlight a severe threat to the safety of young generations and call for drastic emission reductions to safeguard their future.

Meteorological extremes, hazards, or climate change impacts are mostly studied as they evolve over time under varying emission scenarios and socioeconomic pathways (*2*, *4*, *8*). For example, applying a heat wave indicator (see table S1) (*9*) to four bias-adjusted global climate models indicates that the land area annually affected by such heat waves will increase from ~15% around 2020 to ~22% by 2100 under a scenario compatible with limiting global warming to 1.5°C, and to ~46% under a scenario in line with current emission reduction pledges (see the first figure). Recent studies extended this approach, studying aspects of climate change as a function of global mean temperature (GMT) increments, highlighting the scenario-independence of several extreme event indicators (*1*, *3*, *8*) but remaining, in essence, a comparison of time windows.

By contrast, we performed a birth cohort analysis by combining a collection of multimodel extreme event projections (*3*) with country-scale life expectancy information (*10*), gridded population data (*11*), and future global temperature trajectories (*12*) from the Intergovernmental Panel on Climate Change (IPCC) Special Report on Global Warming of 1.5°C (see supplementary materials). By integrating the exposure of an average person in a country or region to extreme events across their lifetime, we encapsulate spatiotemporal changes in climate hazards, population density, cohort size, and life expectancy (see the first figure).

EXTREME EVENT EXPOSURE

Our results allow for comparing lifetime exposure to climate extremes across birth cohorts globally. For example, a person born in 1960 will on average experience around 4 ± 2 (1σ) heat waves across their lifetime according to our extreme heat wave definition (see the first figure). The lifetime heat wave exposure of this cohort is largely insensitive to the three future temperature scenarios considered here. By contrast, a child born in 2020 will experience 30 ± 9 heat waves under a scenario that follows current climate pledges, which could be reduced to 22 ± 7 heat waves if warming is limited to 2°C or 18 ± 8 heat waves if it is limited to 1.5°C. In any case, that is seven, six, or four times more, respectively, compared with that of a person born in 1960. Repeating this analysis for all cohorts born between 1960 and 2020 highlights clear differences in lifetime exposure to heat waves between older and younger cohorts globally (see the first figure). The effect of alternative future temperature trajectories on the lifetime exposure multiplication factor becomes discernible only for cohorts younger than 40 years in 2020, with the largest differences for the youngest cohorts.

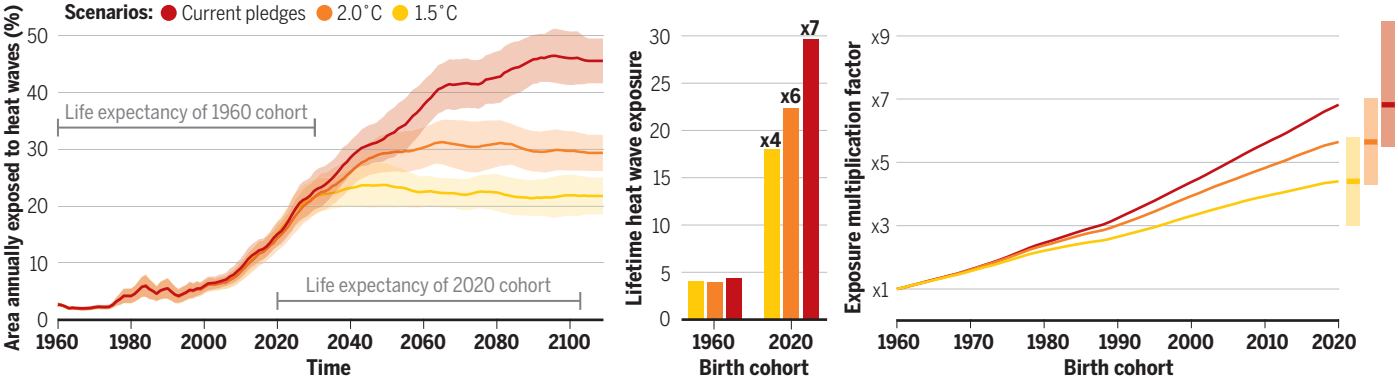
The previous example only uses one hazard indicator and a subset of all possible future temperature pathways. We expanded this approach and considered six extreme event categories: wildfires, crop failures, droughts, river floods, heat waves, and tropical cyclones (see table S1), which we analyzed under a wide range of temperature pathways that resulted in future warming that ranges from constant present-day levels up to 3.5°C by 2100 (see materials and methods and fig. S1). To this end, we generated a total of 273 global-scale projections with 15 impact models forced by four bias-adjusted global climate models (see table S2). Inspired by the IPCC's Reasons for Concern Framework (*7*), we visualized the exposure multiplication factors, relative to a hypothetical reference person living under preindustrial climate conditions, as a function of the 2100 GMT anomaly and cohort (see the second figure and fig. S2). Life expectancy varies with the cohort, whereas the hypothetical reference person is given the same life expectancy as that of the 1960 birth cohort in our figures. Therefore, in contrast to the previous comparison of lifetime exposure across generations given historical and climate conditions (see the first figure), we assessed how projected lifetime exposure of birth cohorts is affected by climate change since the preindustrial era and by increased life expectancy since 1960.

Our results highlight that lifetime exposure to each of the considered extreme events consistently increases for higher warming levels and younger cohorts. Changes in extreme event frequencies have relatively little effect on lifetime exposure for cohorts above age 55 in 2020, but this rapidly changes for younger cohorts as they experience increasing extreme events in the coming years and decades (see the second figure and fig. S2). For a 3°C global warming pathway, a 6-year-old in 2020 will experience twice as many wildfires and tropical cyclones, three times more river floods, four times more crop failures, five times more droughts, and 36 times more

Author affiliations are available in the supplementary materials. Email: wim.thiery@vub.be

From a period to a cohort perspective on extreme event exposure

(Left) Global land area annually exposed to heat waves under three scenarios is shown. Lines represent multimodel means of a heat wave metric calculated from four global climate models. Lines were smoothed by using a 10-point moving average. Uncertainty bands span 1 standard deviation across the model ensemble. (Middle) Lifetime heat wave exposure for the 1960 and 2020 birth cohorts under the three scenarios is shown. Numbers above bars indicate exposure multiplication factors relative to the 1960 cohort. (Right) Shown are multiplication factors for lifetime heat wave exposure across birth cohorts relative to the 1960 cohort. Uncertainty bands represent the interquartile range for the 2020 cohort exposure relative to the multi-model mean exposure of the 1960 cohort.



heat waves relative to the reference person.

Although qualitatively consistent, quantitative exposure changes differ among categories: For wildfires and tropical cyclones, increases in exposure remain limited relative to the other categories, whereas heat wave exposure increases much more strongly, up to a factor of 44 for the 2020 birth cohort under a scenario with 3.5°C global warming. Aggregating the exposure multiplication factors across the six categories shows that people younger than 10 years in 2020 will experience about a fourfold increase in extreme events if global warming is limited to 1.5°C, an increase that older cohorts will never experience, even if a scenario toward 3.5°C warming is followed (see fig. S3A). Under a 3°C global warming pathway, children under 8 years will face an almost fivefold increase in extreme event exposure. These exposure multiplication factors scale robustly with the warming pathway and cohort across a range of aggregation methods, despite some variation in the factor values (see fig. S3).

We then calculated the probability of each person's lifetime exposure occurring under preindustrial climate conditions. Lives with an accumulated exposure that would occur with less than 0.01% probability under preindustrial climate (that is, with less than a 1 in 10,000 chance) are classified as "unprecedented." We found that cohorts above age 55 years in 2020 will on average live an unprecedented life only for heat waves and crop failures, whereas cohorts aged 0 to 40 years in 2020 will additionally face unprecedented exposure to droughts and flooding above 1.5°C warming (see the second figure). Aggregated across all the event categories, lifetime exposure to extremes is unprecedented at all warming levels and cohorts (see fig. S3A).

REGIONAL PATTERNS

Behind this global average picture, there are important spatial variations. Repeating the analysis for a selection of world regions (see fig. S4) reveals marked differences between regions (see figs. S5 and S6), whereas a country-level assessment highlights even stronger spatial disparities (see supplementary text 1 and figs. S7 and S8). We found a particularly strong increase in lifetime exposure across the Middle East and North Africa, with on average at least seven times higher exposure for all cohorts younger than 25 years in 2020 under current emission reduction pledges (see fig. S9A). In sub-Saharan Africa, the 2020 birth cohort will on average experience 5.9 times more extreme events compared with a reference person living under preindustrial climate, whereas in other regions this cohort will on average experience 3.7 to 5.3 times more extremes. This burden is substantially reduced when limiting global warming to 1.5°C: the strongest reductions in exposure are found in the Middle East and North Africa (–39%), Europe and Central Asia (–28%), and North America (–26%), whereas benefits in sub-Saharan Africa, East Asia, and the Pacific roughly correspond to the global average (–24%).

Grouping countries by income category instead of by region highlights that young generations in low-income countries will face the strongest increases in lifetime exposure, with a more than fivefold increase for the 2020 birth cohort under current pledges (see fig. S9B). High-income countries face the smallest increases for younger cohorts and the smallest variation across generations. In 2020, 22% of all 60-year-old people lived in high-income countries, whereas only 5% lived in low-income countries. By contrast, only 10% of all newborns lived in high-income countries, whereas 18% lived in low-income countries (see figs. S10 and S11). Thus, children born in the present and future are much more likely to be born in regions facing the highest increase in lifetime extreme event exposure. For example, 64 million children born in Europe and Central Asia between 2015 and 2020 will experience 3.8 to 4.0 times more extreme events under current pledges, but 205 million children of the same age in sub-Saharan Africa face a factor of 5.4 to 5.9 increase in lifetime extreme event exposure, including a factor of 49 to 54 increase in lifetime heat wave exposure (see figs. S10 and S11). This combined rapid growth in cohort size and extreme event exposure (see figs. S10 to S13) highlights a disproportionate climate change burden for young generations in the Global South.

Improvements in life expectancy (see the first figure and fig. S14) represent a confounding factor in the signal of increasing exposure to extreme events over a person's lifetime. However, we found that globally, climate change explains 98% of the exposure increase for the 2020 birth cohort under the current pledges scenario (see fig. S15 and materials and methods). In high-income countries, the enhanced exposure of this cohort is almost entirely attributable to climate change (99%), whereas in low, lower-middle, and upper-middle income countries, climate change contributes 98% of the total exposure change (see figs. S15 and S16).

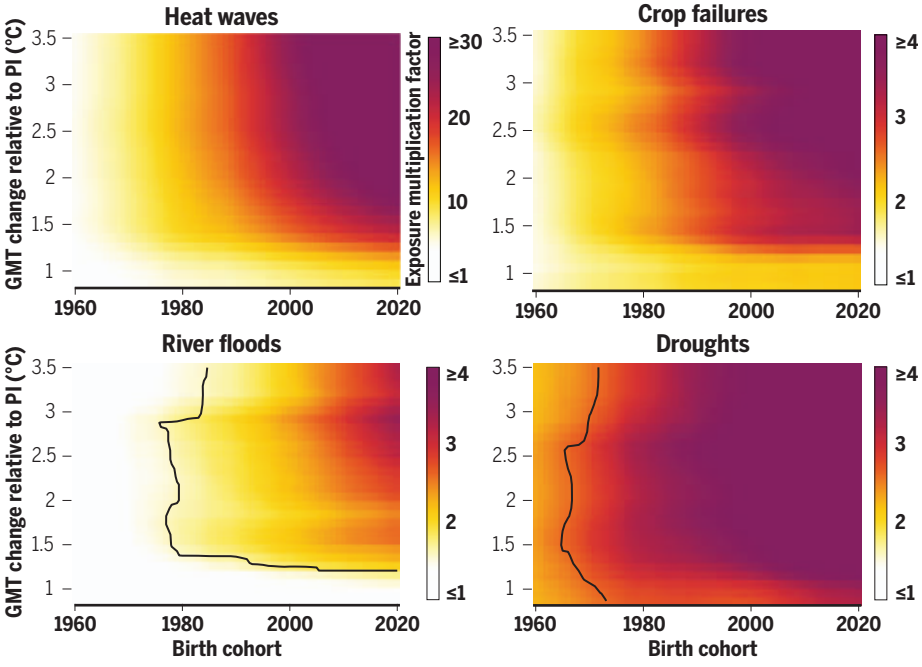
DISCUSSION

Although we comprehensively account for hazards and exposure in our modeling (*3*, *11*) and attention to within-country population density variability, there are reasons that our approach may underestimate intergenerational differences in exposure (see supplementary text 2). For example, we treated

GRAPHIC: K. FRANKLIN/*SCIENCE*

Lifetime exposure to extreme events on the rise

Exposure multiplication factors across birth cohorts under a range of global warming trajectories (fig. S1) reaching 0.87°C to 3.5°C global mean temperature (GMT) anomalies in 2100 relative to the preindustrial (PI) reference period are shown. Factors are computed relative to the mean exposure of a hypothetical reference person living under PI climate conditions with 1960 cohort life expectancy. The black contours delineate lifetime exposure with 0.01% probability of occurrence under PI climate conditions; absence of the contour indicates that this probability is lower for all cases covered.



multiple extremes within a year as one, neglected compound events (*13*), and ignored changes in event duration and intensity.

In this study, we selected six extreme event categories that have a near-immediate response to warming (*3*), which we can adequately simulate (*11*) and which are known to trigger major impacts (*1*–*9*). We omitted, for methodological reasons, slow-onset events such as coastal flooding (see supplementary text 2). Lifetime exposure to other types of extremes may increase, remain unchanged, or even decrease, such as we found for cold extremes (see supplementary text 3).

Further work could aim at extending this approach to include further demographic dimensions and vulnerability (*14*). Vulnerability to extreme events depends on a range of socioeconomic and demographic factors, such as income or gender, but may also evolve over the course of a lifetime. For example, a young person may experience few health impacts from a heat wave compared with those of older people (*15*), but schooling infrastructure destroyed by a tropical cyclone may have a disproportionate detrimental effect on children's education that could persist throughout their lifetime. And although communities may possibly become less vulnerable to extreme events over time, limits to adaptive capacity remain even under optimistic pathways beyond mid-century (see supplementary text 4) (*14*). Extending our exposure analysis to a comprehensive lifetime risk framework may ultimately inform the further development of climate change adaptation strategies.

Climate change impacts may also engender migration and ultimately even affect life expectancy through increased mortality, two aspects that are not considered in this study. Even though climate change increases mortality (*15*), it is currently not included in life expectancy estimates and population projections such as the ones we used in this work. Likewise, migration triggered by environmental degradation may change both exposure and vulnerability to extreme events. Further analysis should therefore aim at systematically integrating population dynamics and climate risk assessments to better understand the long-term impacts of extreme climate events and to improve socioeconomic scenario development.

Our results highlight the strong benefits of aligning policies with the Paris Agreement for safeguarding the future of current young generations. Lifetime exposure to extreme events drastically increases for younger generations as global warming progresses, especially in low-income countries where strongly rising extremes (*8*) affect a rapidly growing young population. Limiting global warming to 1.5°C instead of following the current pledges scenario nearly halves (–40%) the additional exposure of newborns to extreme heat waves and substantially reduces the burden for wildfires (–11%), crop failures (–27%), droughts (–28%), tropical cyclones (–29%), and river floods (–34%) but still leaves younger generations with unprecedented extreme event exposure. These findings have implications for climate litigation and call for ambitious mitigation to improve intergenerational and international justice. ■

REFERENCES AND NOTES

1. R. A. Betts *et al.*, *Philos. Trans. - Royal Soc., Math. Phys. Eng. Sci.* **376**, 20160452 (2018).
2. J. Sillmann, V. V. Kharin, F. W. Zwiers, X. Zhang, D. Bronaugh, *J. Geophys. Res. Atmos.* **118**, 2473 (2013).
3. S. Lange *et al.*, *Earths Futur.* **8**, 1 (2020).
4. L. J. Harrington *et al.*, *Environ. Res. Lett.* **11**, 055007 (2016).
5. G. Hagedorn *et al.*, *Science* **364**, 139 (2019).
6. J. Hansen *et al.*, *Earth Syst. Dyn.* **8**, 577 (2017).
7. IPCC, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, C. Field *et al.*, Eds. (IPCC, 2014), pp. 1–32.
8. N. W. Arnell *et al.*, *Environ. Res. Lett.* **14**, 084046 (2019).
9. S. Russo, J. Sillmann, E. M. Fischer, *Environ. Res. Lett.* **10**, 124003 (2015).
10. United Nations Department of Economic and Social Affairs Population Division, *World Population Prospects 2019* (United Nations, 2019).
11. K. Frieler *et al.*, *Geoscientif. Mod. Dev.* **10**, 4321 (2017).
12. J. Rogelj *et al.*, *Mitigation pathways compatible with 1.5°C in the context of sustainable development, in Global Warming of 1.5°C*, V. Masson-Delmotte *et al.*, Eds. (IPCC, 2018), pp. 93–174.
13. J. Zscheischler *et al.*, *Nat. Rev. Earth Environ.* **1**, 333 (2020).
14. M. Andrijevic, J. Crespo Cuaresma, R. Muttarak, C. F. Schleussner, *Nat. Sustain.* **3**, 35 (2020).
15. N. Watts *et al.*, *Lancet* **394**, 1836 (2019).

ACKNOWLEDGMENTS

For their roles in producing, coordinating, and making available the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; www.isimip.org) input data and impact model output, we acknowledge the modeling groups, the ISIMIP sector coordinators, and the ISIMIP cross-sectoral science team. W.T. thanks J. Steyaert and V. Leroy for sparking the idea that led to this study. U. Beyerle and the ETH Zurich cluster team are thanked for support with the CLM4.5 simulations. We also thank the National Center for Atmospheric Research (NCAR) for maintaining CLM (Community Land Model) and making the source code publicly available. T.G. and S.N.W. acknowledge support by the German Federal Ministry of Education and Research (BMBF) under the research projects SLICE (FKZ: 01LA1829A) and CLIC (FKZ: 01LA1817C), respectively. C.-F.S. acknowledges support by the German Federal Ministry of Education and Research (FKZ: 01LS1905A). N.H. was supported by the Environment Research and Technology Development Fund (JPMEERF20182R02). S.I.S. acknowledges partial support from the European Research Council (ERC) through the ERC Proof-of-concept grant MESMER-X under H2020-EU.1.1 (grant 964013). A.K. acknowledges support through the FP7 project HELIX (grant 603864). S.L., K.F., T.G., M.B., J.V., M.M., and C.P.O.R. acknowledge funding from BMBF under the ERA4CS project ISIPedia (FKZ: 01LS1711A). F.Z. thanks the National Key R&D Program of China (2017YFC1503001). V.H. received support from the Spanish Ministry of Economy, Industry and Competitiveness (grant PCIN-2017-046). C.P.O.R. acknowledges funding from the EU Horizon 2020 research and innovation program under grant agreement 821010 (CASCADES) and J.R. under number 820829 (CONSTRAIN).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi7339

Published online 26 September 2021
10.1126/science.abi7339

GRAPHIC: K. FRANKLIN/*SCIENCE*

Bats can host many viruses without succumbing to them and are often implicated in disease outbreaks.

PHOTO: MILLETSTUDIO/SHUTTERSTOCK

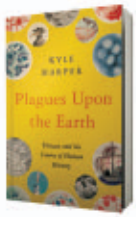
EPIDEMIOLOGY

The infectious diseases that shaped us

An ambitious history traces humanity's entanglements with communicable illness

By **Gregory J. Morgan**

In the mid-19th century, the polymath William Whewell coined the term "consilience" to describe the progress of science through time (*1*). According to Whewell, inductive knowledge should "jump together" with evidence from different disciplines of science supporting more and more general, unified knowledge claims. The historian Kyle Harper explicitly adopts and enlarges this goal—which he attributes to E. O. Wilson—beyond science proper, incorporating insights from history, economics, anthropology, paleogenomics, parasitology, ecology, and phylogenetics, to write an ambitious, engaging, and unified history of humanity's interaction with infectious disease. His focus on consilience largely succeeds but remains a promissory note in at least one arena.



Plagues Upon the Earth
Kyle Harper
Princeton University Press, 2021. 704 pp.

Harper begins by arguing that the discovery of fire by our ancestors in the Pleistocene allowed them to disperse beyond Africa, where they were exposed to a large variety of geographically diverse wild animals and their pathogens, including bacteria, viruses, and helminths. Levels of typhoid, malaria, yellow fever, Chagas disease, and dengue fever were disproportionately high in the Pleistocene, he notes.

Farming emerged about 10,000 years ago, exposing humans to domesticated animals and their pathogens and increasing the amount of excrement that accumulated in one place as humans relinquished nomadic lifestyles. Human fecal-oral life cycles, often with flies as vectors, became possible for many microbes during this period. Harper presents genomic data that undercuts Jared Diamond's thesis that the bulk of new human disease came from domesticated animals during this period, suggesting instead that domesticated animals formed a bridge to wild animals and their pathogens.

Harper suggests that the Neolithic period was a catastrophe for human health because of diseases and pathogens mostly understudied by historians, including shigella, campylobacter, rotavirus, dysentery, and typhoid fever. He speculates that, as in the Pleistocene, Holocene human populations experienced cycles of boom and bust likely driven by infectious diseases.

Respiratory pathogens such as measles and smallpox came later, with the rise of the first higher-density cities. Genomic data suggest that measles came to humans from cattle, which in turn caught the virus from bats. Here, Harper highlights the "paradox of progress": Technological advances can diminish human health by improving living conditions for our parasites. As we began to travel more, first overland and then overseas, the human disease pool expanded.

We cannot understand the Black Death (1346–1353 CE) without insights from both the humanities and natural sciences, argues Harper. He details how the *Yersinia pestis* bacteria build biofilms in the alimentary canals of fleas, blocking the ingestion of blood and causing the fleas to ravenously search for more blood, consequently spreading the bacteria. The plague cycle involves the transfer of *Y. pestis* living in marmots to the black rat and then to humans via fleas. Harper summarizes the work of historian Monica Green, who integrated the branching pattern of the phylogenetic tree of *Y. pestis* into histories of the plague, arguing that it was the growth of the Mongol Empire that allowed for the spread of the plague from marmots in the Tian Shan mountains (where Kyrgyzstan meets Kazakhstan) to Europe and perhaps China (*2*).

Harper questions the "virgin-soil hypothesis," the idea that Europeans were able to colonize the Americas because their infectious diseases ravaged the region's Indigenous peoples, who had no previous immunity. He pushes for a more nuanced theory with input from genetics and immunology, dividing the Americas into 13 regions to be considered separately, arguing that the biological exchanges that occurred are inseparable from the dynamics of slavery, settlement, and commerce. However, while science has settled some debates—syphilis was in Europe before the Columbian Exchange, for instance—it has not radically changed the view that emerged from historians who did not have access to genetic data. In this case, the prospect of enlightening history through paleogenetics remains largely promissory. ■

The reviewer is at the College of Arts and Letters, Stevens Institute of Technology, Hoboken, NJ 07030, USA. Email: gmorgan@stevens.edu

REFERENCES AND NOTES

1. W. Whewell, *The Philosophy of Inductive Sciences* (John W. Parker, 1847).
2. M. H. Green, *Am. Hist. Rev.* **125**, 1601 (2020).

10.1126/science.abl7417

HUMAN ORIGINS

Adam, Eve, and the evolution of humankind

A theologian looks to science for clues about the Bible's first couple

By **Stephen Schaffner**

Adam and Eve are not figures that turn up often in serious scientific discourse. Introducing them as historical characters in, say, an evolutionary biology lecture would raise more than a few eyebrows. William Lane Craig, a widely published philosopher, theologian, and Christian apologist, is not one to shy away from such a provocative pairing, however. In his latest book, *In Quest of the Historical Adam*, Craig sets out to bring academic and scientific rigor to bear on the famous first couple of Genesis. He seeks to answer two questions: whether his theological commitments as a Christian necessitate believing in a historical Adam and Eve and, if so, what science can tell us about that couple.

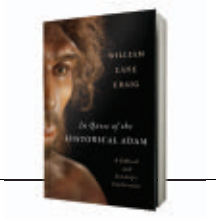
The first question occupies more than half the book, and to answer it, Craig draws on a range of academic fields, including myth studies, the study of ancient Near Eastern religion, and biblical criticism. The conclusions of this largely nonscientific portion of the book are twofold. As far as the Genesis story of Adam and Eve is concerned, Craig concludes that it was written in mythological terms and, while intended to teach theological beliefs, was not meant to be taken literally (a view that will not endear him to creationists). By contrast, he concludes that some statements about Adam in the New Testament, specifically ones in Paul's letter to the Romans, do entail a real, historical Adam, one who was the first human.

This reasoning, which Craig ultimately finds convincing, is the crux of the book and the basis for the scientific investigation that follows. However, it is also the least-developed part of Craig's argument. In his section on Genesis, for example, he probes the authors' understanding of the historicity of what they wrote, but he does not similarly probe Paul's own understanding of his interpretation of the same text. Rather than ask how exegetes within Paul's cultural milieu, that of Second Temple Judaism, viewed the historicity of their interpretations, he assumes that Paul, by treating Adam as a historical figure whose actions had real-world consequences, is asserting that Adam really lived.

Having established that he should believe in Adam's existence, Craig sets out to locate him. He does so in the form of a question amenable to scientific analysis: When did hominins acquire the cognitive capacity for abstract thought, symbolic behavior, and the like, such that they should be considered human? The relevant subject matter is large and touches on evolutionary biology, paleontology, paleoneurology, archaeology, and genetics; the data are often scanty and contentious. Nevertheless, he does a more than creditable job of synthesizing both the conclusions and the uncertainties offered by these various fields, often drawing on primary scientific literature to do so.

Craig's theory of human origins allows for admixture with other early hominin lineages.

Craig argues, for example, on the basis of brain size, that the first humans could not have lived before the time of *Homo heidelbergensis* and late *Homo erectus*. A number of facts about Neanderthals—symbolic behavior, ability to cooperate and plan, probable linguistic capacity, possession of human-specific genetic modifiers of brain development—convince him that they qualify as human. He therefore concludes that humanness was a trait inherited by Neanderthals, Denisovans, and *Homo sapiens* from their common ancestral populations and that Adam must have lived circa 700,000 years ago.



In Quest of the Historical Adam: A Biblical and Scientific Exploration
William Lane Craig
Eerdmans, 2021. 439 pp.

Out of necessity, given scant scientific interest in the topic, Craig abandons the formal scientific literature for a discussion of the limits that human genetic diversity places on the occurrence of a population bottleneck of size two in our ancestry. Relying on less-formal studies, he concludes that population genetics can rule out a pair of sole genetic progenitors for humans within roughly the past 500,000 to 700,000 years. This conclusion, while reasonable, turns out to have little relevance to Craig's overall argument; we eventually learn that his model allows for an unspecified amount of admixture from other hominin lineages into the descendants of Adam, eliminating the need for such a tight bottleneck. A more clearly stated hypothesis in this section would have saved the reader time and frustration.

Craig's goal in writing this book, of course, is not a scientific one, and it cannot be judged on scientific grounds. I suspect that for many scientists, including religious ones, the exercise will be seen as misguided or simply incomprehensible. Even leaving aside the religious motivations, biologists are likely to be highly skeptical of the idea that humanness is a binary condition that can be induced by a change in a single pair of ancestors—declaring the change to be miraculous and to incorporate an immaterial soul, as Craig proposes, will not make it more appealing.

While my own reaction is along similar lines, I very much welcome the book. I think that it is entirely a good thing that an individual with Craig's theological commitments and credentials turns to science to answer questions about the physical world, takes evolution as a given, and puts in the hard work to understand scientific findings. I can only hope that others who work at the intersection of science with philosophy or religion emulate his efforts. ■

The reviewer is at the Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. Email: sfs@broadinstitute.org

10.1126/science.abl8547

IMAGE: IAN DAGNALL/ALAMY STOCK PHOTO

The Przewalski's gazelle (*Procapra przewalskii*) population has come back from the brink of extinction but still faces challenges.

PHOTO: STAFFAN WIDSTRAND/WILD WONDERS OF CHINA

Edited by **Jennifer Sills**

Ensuring recovery for the Przewalski's gazelle

When the Przewalski's gazelle (*Procapra przewalskii*) was first identified 150 years ago (*1*), the species was distributed across western China (*2*). By the 1990s, the gazelle had dwindled to just 300 individuals (*2*), all of which lived near the Qinghai Lake Basin on the Qinghai-Tibetan Plateau (*2*, *3*). In 2000, the International Union for Conservation of Nature Red List listed the species as Critically Endangered. The gazelle's habitat had been engulfed by livestock and human activities and fragmented by grassland fences and roads, pushing the vulnerable ungulate to the brink of extinction (*3*–*7*). Recently, the Przewalski's gazelle population has grown, but decreasing protection efforts in response would sabotage the species' recovery.

In the past two decades, the Przewalski's gazelle has attracted wide attention. In 2001, it was listed as one of the 15 species in urgent need of rescue in the National Wildlife Conservation and Nature Reserve Construction Project as well as the key species in the Biodiversity Survey and Assessment Project of the Ministry of Ecology and Environment of China (*3*). In 2004, at the International Workshop on Saving the Przewalski's Gazelle, experts worldwide discussed how to save the species (*8*). Since then, conservation measures backed by scientific research have been implemented to protect the Przewalski's gazelle and its habitat. As a result of these efforts, the Przewalski's gazelle achieved a population size of about 2700 individuals in 2020 (*9*). Reflecting the positive change, the species' threatened status was downgraded from Critical Endangered to Endangered in the most recent version of *China's Red List of Biodiversity: Vertebrates* (*10*).

However, the Przewalski's gazelle still faces severe threats. The species reached high density in very limited ranges, where it must compete with livestock, habitat degradation and fragmentation due to anthropogenic disturbance, and low genetic exchange among isolated populations. The Przewalski's gazelle is the only species that has no nature reserve specifically dedicated to its success (*3*).

In July, the first corridor for the conservation of the Przewalski's gazelle was established; the safe passage across a highway will allow isolated populations to access each other (*11*). To ensure the species' continued growth, we should also eliminate grassland fences, road traffic, and human activities in its core habitats. Providing compensation to local herdsmen to mitigate the impact of livestock-wildlife conflict would also be a step forward. Finally, conservation projects should be explored to restore the Przewalski's gazelle populations in the species' historical ranges in Gansu and Inner Mongolia of western China.

Dongni Liang[1,2], Yaxin Liu[1,2], Xiaoge Ping[1,3], Zhigang Jiang[1,2], Chunwang Li[1,2]*
[1]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]Endangered Species Scientific Commission of the People's Republic of China, Beijing 100101, China.
*Corresponding author. Email: licw@ioz.ac.cn

REFERENCES AND NOTES

1. G. M. Allen, *The Mammals of China and Mongolia, Part 2* (American Museum of Natural History, New York, 1940).
2. Z. Jiang *et al.*, *Oryx* **34**, 129 (2000).
3. Z. Jiang, *Przewalski's Gazelle* (China Forestry Publishing House, Beijing, 2004) [in Chinese with English summary].
4. Z. You *et al.*, *Chin. Sci. Bull.* **58**, 2262 (2013).
5. C. Li *et al.*, *Wildl. Res.* **36**, 379 (2009).
6. J. Yang *et al.*, *PLOS One* **6**, e20144 (2011).
7. D. P. Mallon, S. C. Kingswood, *Global Survey and Regional Action Plans on Antelope* (IUCN, 2001).
8. Y. Zeng, C. Li, *Biodivers. Sci.* **12**, 645 (2004) [in Chinese].
9. L. Li, H. Geng, "The number of Przewalski's gazelle population exceeds to 2700" (2020); www.gov.cn/xinwen/2020-08/25/content_5537304.htm [in Chinese].
10. Z. Jiang *et al.*, *China's Red List of Biodiversity: Vertebrates* (Science Press, Beijing, 2021) [in Chinese].
11. M. Wan, "The first corridor for the conservation of the Przewalski's gazelle was established in Qinghai" (2021); epaper.gmw.cn/gmrb/html/2021-07/14/nw.D110000gmrb_20210714_5-08.htm [in Chinese].

10.1126/science.abm2865

Australia's plan to reduce plastic waste falls short

The Australian government recently launched the National Plastics Plan 2021 (NPP) to manage plastic waste, reduce its environmental impacts, and promote a circular economy (*1*), with the goal of making substantive changes by 2025. This initiative is an important step forward in combating Australia's plastic problem; the country consumes 59 kg of single-use plastic per person per year (second only to Singapore) (*2*) and recycles only 18% of discarded plastics (*3*). The NPP aims to reduce plastic waste by preventing its creation at the source, doubling recycling capacity, encouraging responsible stewardship by both industry and consumers, providing better information about disposal, and streamlining waste collection procedures (*1*). However, the current plan may not succeed without more specific targets and consequences for inaction.

The NPP largely relies on industries' voluntary actions to take custodial responsibility of their plastic waste (*4*). Voluntary action has a history of yielding poor outcomes when applied to plastic packaging and waste elimination goals (*5*). Without stronger legislation, states and territories have no way to take legal action against plastic polluters (*6*). Furthermore, the NPP does not clearly address commingled collection of recyclables (*7*), provide incentives for plastic alternative market initiatives, or define measurable indicators to evaluate targeted action plans.

To make the NPP effective, Australia should add a specific timeline with measurable interim goals and penalties for inaction to ensure that industrial plastic producers meet the proposed stewardship and abatement targets. The country should provide support to existing plastic and packaging businesses to enable them to adopt responsible practices. Financing should be made available to entrepreneurs working on alternatives to plastic. Urgent and clear guidelines and strategies should be coordinated across all jurisdictions, allowing them to work together to collect, sort, and dispose of recyclable plastics. Australia should also work to foster a culture of limited plastic use, which will facilitate behavioral changes among citizens.

Tanveer M. Adyel* and Peter I. Macreadie
Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Burwood, VIC 3125, Australia.
*Corresponding author.
Email: t.adyel@deakin.edu.au

REFERENCES AND NOTES

1. Commonwealth of Australia, Australian Government, National Plastics Plan (2021); www.environment.gov.au/system/files/resources/a327406c-79f5-47f1-b71b-7388407c35a0/files/national-plastics-plan-2021.pdf.
2. D. Charles, L. Kimman, N. Saran, "The plastic waste makers index," Minderoo Foundation (2021); https://cdn.minderoo.org/content/uploads/2021/05/06123145/20210518-Plastic-Waste-Makers-Index.pdf.
3. World Wildlife Fund, Boston Consulting Group, "Plastic revolution to reality: A roadmap to halve Australia's single-use plastic litter" (2020); www.bcg.com/en-au/plastic-revolution-to-reality.
4. K. Lindsay, "Commonwealth plastic plan announced," Boomerang Alliance (2021); www.boomerangalliance.org.au/commonwealth_plastic_plan_announced.
5. G. Readfearn, "Australia to miss plastic reduction targets without tougher enforcement, waste industry says," *The Guardian* (2021).
6. World Wildlife Fund Australia, "The state of Australia's recycling—how did we get into this mess?" (2021); www.wwf.org.au/news/blogs/the-state-of-australias-recycling-how-did-we-get-into-this-mess.
7. A. Schiavo, "Breaking down Australia's National Plastics Plan 2021," Lux Research (2021); www.luxresearchinc.com/blog/breaking-down-australias-national-plastics-plan-2021.

10.1126/science.abm4271

Threatened freshwater fish need protection

Fish species in freshwater habitats are under threat worldwide due to overexploitation, pollution, species invasion, dam construction, and climate change (*1*–*3*). About one-third of freshwater fish species are threatened with extinction; 80 species were declared extinct in 2020 (*2*). Action plans such as the 2020 Yangtze River fishing ban (*4*) and the 2021 US Recovering America's Wildlife Act (*5*) have been adopted to restore and conserve biodiversity in freshwater habitats, but they are not enough. The 15th meeting of the Conference of the Parties to the Convention on Biological Diversity (COP 15) will be held in Kunming, China, in October to consider the post-2020 global biodiversity framework (*6*). Conference attendees should include improvements in monitoring and protecting freshwater fish in their discussions.

In the context of global change, monitoring species communities is critical for the evaluation of management strategies. However, traditional fish survey methods are inefficient, selective, destructive, or dependent on declining taxonomic expertise (*7*). Furthermore, the COVID-19 crisis has resulted in disruptions to routine environmental monitoring and affected fisheries management and stock assessment across the globe (*8*). Thus, there is an urgent need for sensitive, noninvasive, nondestructive, and non–labor-intensive approaches to document species and assess biodiversity trends in nature reserves and other areas, especially after fishery bans.

We urge scientists, nongovernmental organizations, natural resource managers, and governments to work together to protect freshwater fish diversity from severe and growing threats. Environmental DNA (eDNA) metabarcoding is a promising conservation tool for fish monitoring. By sequencing a small sample of water or sediment, eDNA technology can efficiently track rare, endangered, or alien fish species across space and time as well as assess phylogenetic and functional diversity and reproductive ecology (*9*–*11*).

In addition to better monitoring, freshwater fish populations would benefit from habitat restoration to help improve natural habitat conditions, fishing bans to provide fish populations with sufficient time for natural recovery, and stock enhancement with the release of hatchery-reared juveniles to make up for natural reproduction deficits. These measures should be combined to ensure comprehensive, adaptive, and effective implementation; not all solutions are suitable for all situations, and no single solution can solve the freshwater fish crisis. International attention and collaboration will be critical to prevent further extinctions and rebuild a future in harmony with nature.

Wenjian Chen[1], Li Liu[1], Junjie Wang[2], Lei Zhou[1]*
[1]College of Marine Sciences, South China Agricultural University, Guangzhou 510642, China.
[2]School of Life Science, South China Normal University, Guangzhou 510631, China.
*Corresponding author. Email: zhoulei@scau.edu.cn

REFERENCES AND NOTES

1. G. Su, M. Logez, J. Xu, S. Tao, S. Brosse, *Science* **371**, 835 (2021).
2. World Wildlife Fund, "The World's Forgotten Fishes" (2021); https://wwf.panda.org/discover/our_focus/freshwater_practice/the_world_s_forgotten_fishes/.
3. C. P. Carmona, R. Tamme, M. Prtel, F. D. Bello, A. Toussaint, *Sci. Adv.* **7**, eabf2675 (2021).
4. "Notice of the Ministry of Agriculture and Rural Affairs on the scope and time of arrests in key waters of the Yangtze River Basin" (2019); www.cjyzbgs.moa.gov.cn/tzgg/201912/t20191227_6334009.htm [in Chinese].
5. American Fisheries Society, "Fish and wildlife scientists commend introduction of Recovering America's Wildlife Act" (2021); https://fisheries.org/2021/07/fish-and-wildlife-scientists-commend-introduction-of-recovering-americas-wildlife-act/.
6. 2021 UN Biodiversity Conference (COP 15) (2021); www.cbd.int/meetings/COP-15.
7. D. Causey *et al.*, *Science* **305**, 1103 (2004).
8. S. J. Cooke *et al.*, *Biol. Conserv.* **253**, 108932 (2020).
9. V. Marques *et al.*, *Conserv. Biol.* 10.1111/cobi.13802 (2021).
10. K. Deiner *et al.*, *Mol. Ecol.* **26**, 5872 (2017).
11. J. Bylemans *et al.*, *Meth. Ecol. Evol.* **8**, 646 (2017).

10.1126/science.abm1123

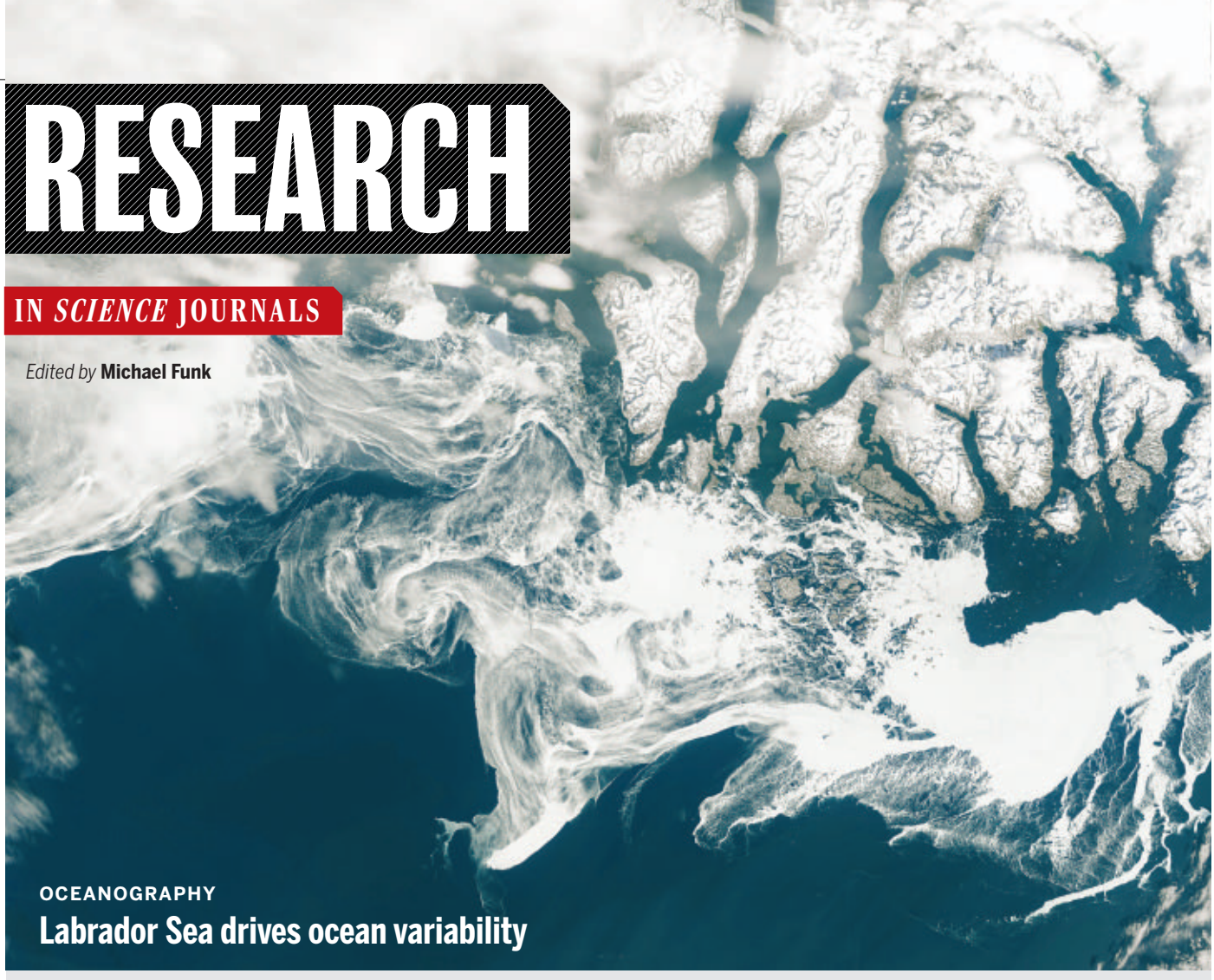
Australia's plastic waste production per capita ranks second in the world.

PHOTO: WATERFRAME/ALAMY STOCK PHOTO

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



OCEANOGRAPHY

Labrador Sea drives ocean variability

The Atlantic meridional overturning circulation (AMOC) transports warm surface water into the subpolar North Atlantic, where the cooled, dense water sinks below 1000 meters and flows southward and across the equator. Recent observations call into question the fidelity of climate models that place the sinking branch in the Labrador Sea. Yeager *et al.* inspected a high-resolution climate model and found that by resolving ocean eddies, this model simulates the mean circulation realistically. The twist is that even though the Labrador Sea is not the main region of sinking in the North Atlantic, it does drive a significant component of multidecadal variability in the AMOC and it does so with a time lead of a few years. This finding suggests the possibility of predicting ocean variability based on Labrador Sea conditions. —SPX

Sci. Adv. 10.1126/sciadv.abh3592 (2021).

High-resolution simulations confirm an important role for the Labrador Sea in multidecadal variability of ocean thermohaline circulation.

IMAGE: LAUREN DAUPHIN/NASA EARTH OBSERVATORY (DATA) LANDSAT/U.S. GEOLOGICAL SURVEY

SPECTROSCOPY

Conformer-specific dynamics

Conformation-dependent dynamics play an important role in organic chemistry syntheses such as electrocyclic reactions, as well as in biological processes such as protein folding. However, current time-resolved experimental methods struggle to distinguish conformers from each other, and conformational isomerism is usually analyzed through reactant and product distributions. Using a combination of mega–electron volt ultrafast electron diffraction and quantum wave packet simulations, Champenois *et al*. directly followed the photochemical electrocyclic ring opening of the molecule α-phellandrene with femtosecond time resolution and confirmed that the transformation of a specific molecular conformer follows the famous Woodward-Hoffmann rules. The proposed method is potentially a powerful tool to follow conformer specificity in various organic and biological systems in real time. —YS

Science, abk3132, this issue p. 178

PALEOGENOMICS

Ancient DNA traces the history of hepatitis B

Hepatitis B virus (HBV) infections represent a worldwide human health concern. To study the history of this pathogen, Kocher *et al*. identified 137 human remains with detectable levels of virus dating between 400 and 10,000 years ago. Sequencing and analyses of these ancient viruses suggested a common ancestor between 12,000 and 20,000 years ago. There is no evidence indicating that HBV was present in the earliest humans as they spread out of Africa; however, HBV was likely present in human populations before farming. Furthermore, the virus was present in the Americas by about 9000 years ago, representing a lineage sister to the viral strains found in Eurasia

that diverged about 20,000 years ago. —LMZ

Science, abi5658, this issue p. 182

METALLURGY

Revising grain growth

Polycrystalline materials will often coarsen during annealing, a process that can modify the physical properties. A generally accepted theory ties the grain boundary curvature to the velocity at which grain boundaries migrate. Bhattacharya *et al*. measured this relationship for 52,000 grains in a nickel sample. They did not observe any relationship between curvature and velocity, suggesting that this accepted theory is not robust for polycrystalline samples. The observations require grain-coarsening models to be updated to more accurately predict the effects of annealing. —BG

Science, abj3210, this issue p. 189

TOPOLOGICAL OPTICS

Coupling light and heat

Understanding of the topological features of bandgaps has provided a route for engineering optical structures that exhibit directional propagation of light and are robust to defects. Guddala *et al*. combined a silicon-based topological photonic crystal with an atomic monolayer of hexagonal boron nitride (hBN). The topological features of the photonic crystal are coupled to the lattice vibrations of the hBN through the formation of phonon-polaritons. Funneling of helical infrared phonons along arbitrary pathways and across sharp bends provides the possibility of realizing directional heat dissipation along topologically resilient heat sinks. —ISO

Science, abj5488, this issue p. 225

TUMOR IMMUNOLOGY

A niche in the node

The presence of "stem-like" TCF1$^+$ CD8$^+$ T (T_{SL}) cells in tumors correlates with better patient outcomes after immunotherapy. T_{SL} cells undergo terminal differentiation and exhaustion in the tumor microenvironment, but how a stem-like population is maintained throughout tumor development is unclear. Using a mouse model of lung adenocarcinoma programmed to express neoantigens, Connolly *et al.* identified a stable reservoir of tumor-specific T_{SL} cells in the tumor-draining lymph node (dLN). T_{SL} cells migrated from the dLN to the tumor, where they became terminally differentiated as the tumor microenvironment shifted from T cell inflamed to noninflamed. These findings suggest that therapeutic strategies directed at tumor-specific T cells in the dLN may be a promising approach to treating patients with so-called "cold" tumors. —HMI

Sci. Immunol. **6**, eabg7836 (2021).

POLYMER SCIENCE

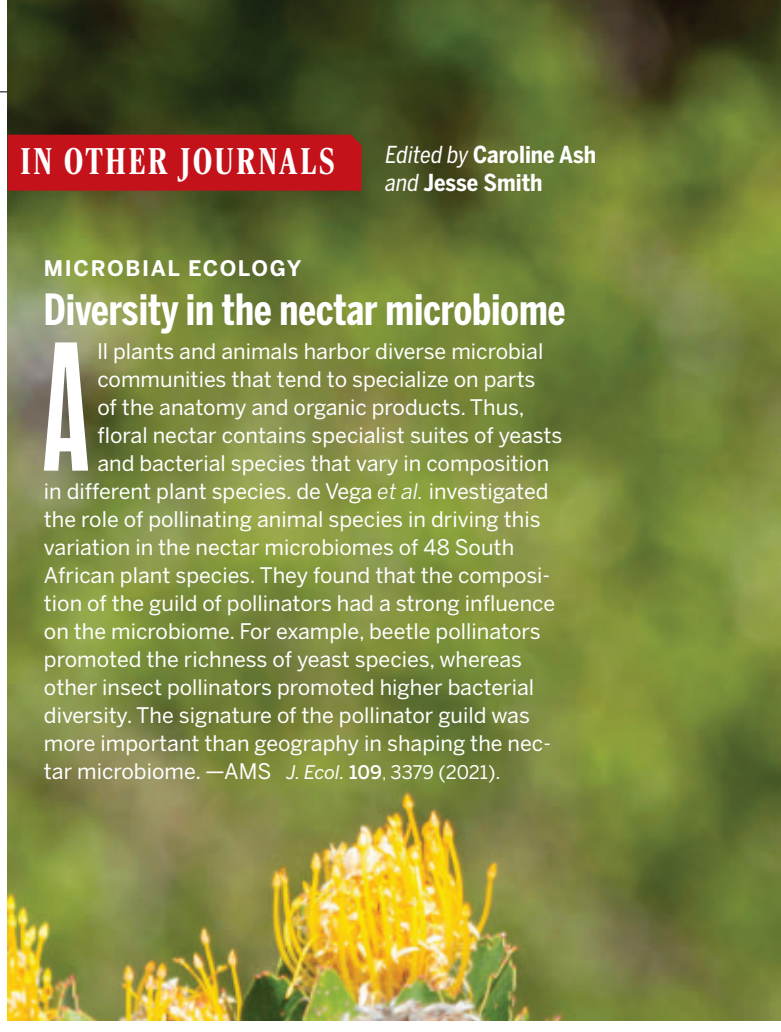
Longer and stronger; stiff but not brittle

Hydrogels are highly water-swollen, cross-linked polymers. Although they can be highly deformed, they tend to be weak, and methods to strengthen or toughen them tend to reduce stretchability. Two papers now report strategies to create tough but deformable hydrogels (see the Perspective by Bosnjak and Silberstein). Wang *et al*. introduced a toughening mechanism by storing releasable extra chain length in the stiff part of a double-network hydrogel. A high applied force triggered the opening of cycling strands that were only activated at high chain extension. Kim *et al*. synthesized acrylamide gels in which dense entanglements could be achieved by using unusually low amounts of water, cross-linker, and initiator to cause crowding during the synthesis. This approach improves the mechanical strength in solid form while also improving the wear resistance once swollen as a hydrogel. —MSL

Science, abg2689, this issue p. 193; see also abl6358, p. 150 see also abl6358, p. xxxx

IN OTHER JOURNALS

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



MICROBIAL ECOLOGY

Diversity in the nectar microbiome

All plants and animals harbor diverse microbial communities that tend to specialize on parts of the anatomy and organic products. Thus, floral nectar contains specialist suites of yeasts and bacterial species that vary in composition in different plant species. de Vega *et al.* investigated the role of pollinating animal species in driving this variation in the nectar microbiomes of 48 South African plant species. They found that the composition of the guild of pollinators had a strong influence on the microbiome. For example, beetle pollinators promoted the richness of yeast species, whereas other insect pollinators promoted higher bacterial diversity. The signature of the pollinator guild was more important than geography in shaping the nectar microbiome. —AMS *J. Ecol.* **109**, 3379 (2021).

EXOPLANETS

A multiplanet system with a sub-Earth

Small terrestrial planets are easiest to detect if they orbit nearby red dwarf stars such as L 98-59, which was already known to host three transiting terrestrial planets. Demangeon *et al.* collected additional radial velocity observations of L 98-59 and further constrained the orbits of its planets by requiring the system to be dynamically stable. They detected an additional fourth planet, identified a candidate fifth, and improved the parameter measurements for the other three. The smallest planet, L 98-59 b, has a mass of about 40% of Earth, whereas the other planets are between two and three Earth masses. The system provides an opportunity to compare multiple terrestrial planets subjected to the same stellar activity. —KTS

Astron. Astrophys. **653**, A41 (2021).

METABOLISM

How taste might drive obesity

Umami, or savory taste, results from monosodium glutamate (MSG) and other nucleotides. Andres-Hernando *et al*. show that MSG induces obesity and metabolic syndrome in mice through the induction of adenosine monophosphate deaminase 2 and purine degradation in hepatocytes and neurons. This cascade causes inflammation and insensitivity to leptin, a hormone that modulates satiety, food intake, and metabolic syndrome. Inosine monophosphate, another umami taste–associated nucleotide, also induced purine degradation and metabolic syndrome in mice that were resistant to the effects of MSG. Similar pathways were activated by the consumption of fructose and salt, suggesting that excess exposure to these tastes can drive caloric intake, metabolic syndrome, and obesity. —GKA

Nat. Metab. **3**, 1189 (2021).

PHOTO: CHRISTIAN DIETZ/ALAMY STOCK PHOTO

Pollinating animals, such as this Cape sugarbird, that feed on Protea flowers influence microbial communities in plant nectar.

IMAGE: IMAGEBROKER/ALAMY STOCK PHOTO

LUNG INFLAMMATION

LRRK2 to restrict pulmonary fibrosis

In pulmonary fibrosis, the lungs become scarred over time. This is a serious and irreversible disease and new treatments are greatly needed. Tian *et al.* uncovered a critical role for leucine-rich repeat kinase 2 (LRRK2) in mediating innate immune responses to lung injury and safeguarding against pulmonary fibrosis. The researchers found that the expression of the LRRK2 gene was decreased in human and mouse fibrotic lungs compared with healthy controls. Mice deficient for LRRK2 showed more rapid progression of experimentally induced pulmonary fibrosis, and this was accompanied by exacerbated immune responses. The alveolar type II epithelial cells of these mice had an increased capacity to recruit macrophages to the lung and to mediate profibrotic inflammatory responses by CCL2/CCR2 chemokine signaling. —PNK

Proc. Natl. Acad. Sci. U.S.A. **118**, e2106685118 (2021).

SIGNAL TRANSDUCTON

Informing antiviral responses

Receptor proteins in the cell membrane are key to the flow of information between cells. Interferon receptors alert cells to viral infection. Binder *et al.* explored how information processing works in such receptors, one of which recognizes at least 17 different ligands. The authors used information theory and mathematical modeling to explore how the interferon I receptor (IFNAR) intercepts noisy signals from multiple ligands that are encoded in the chemical properties of the ligand rather than the ligand's concentration. The characteristics of IFNAR, which forms heterodimers, allow discrimination between different ligands even when concentrations of ligand vary. Removal of receptor from the surface after ligand binding further improves sensitivity. —LBR

Proc. Natl. Acad. Sci. U.S.A. **118**, e2103939118 (2021).

MICROPLASTIC

Fraying ropes foul the ocean

Synthetic rope, which is typically made of polypropylene and/or polyethylene plastic fiber, has obvious cost and durability benefits for marine uses over natural or biodegradable alternatives. However, it does degrade with time. Napper *et al.* investigated how such ropes release microplastic fibers and particles as they age. They found that during abrasion, older ropes released large quantities of small fragments rather than fibers. These results suggest a reevaluation of sources of microplastic particles and reconsideration of when to replace older ropes to avoid such pollution. —MAF

Sci. Total Environ. **804**, 150155 (2021).

MEMBRANES

A crowning choice for lithium

The extraction of lithium, whether through mining or recycling, is critical for the continuing expansion of energy-related applications, but it requires energy-intensive processes. One possible option is the use of membranes, which tend to have poor ion selectivity that is further hampered by the greater solubility of sodium over lithium in water. Warnock *et al.* polymerized norborene molecules functionalized with a 12-crown-4 ligand, a polyethylene oxide (PEO) side chain, or a cross-linker. The PEO helps to adjust the water content of the membranes, and the cross-linker provides some rigidity to the membranes. The key to extracting lithium is the 12-crown-4, which binds sodium more strongly, allowing for more lithium to diffuse through, even in mixed-ion conditions. —MSL

Proc. Natl. Acad. Sci. U.S.A. **118**, e2022197118 (2021).

Abrasion causes plastic rope used in marine settings to release large numbers of small fragments, rather than fibers, as it ages.

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Cross-reactive T cells aid in the fight

There is mounting evidence that immunological memory after infection with seasonal human coronaviruses (hCoVs) contributes to cross-protection against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Loyal *et al*. identified a universal immunodominant coronavirus peptide found within the fusion peptide domain of the coronavirus spike protein. This peptide is recognized by $CD4^+$ T cells in 20% of unexposed individuals, more than 50% of SARS-CoV-2 convalescents, and 97% of subjects treated with the Pfizer–BioNTech COVID-19 vaccine. Although ubiquitous, these coronavirus-reactive T cells decreased with age, which may explain in part the increased susceptibility of elderly people to COVID-19. —STS

Science, abh1823, this issue p. 171

BATTERIES

Efficient, rechargeable Mg and Ca batteries

Divalent rechargeable metal batteries such as those based on magnesium and calcium are of interest because of the abundance of these elements and their lower tendency to form dendrites, but practical demonstrations are lacking. Hou *et al*. used methoxyethyl amine chelants in which the ligands attach to the metal atom in more than one place, modulating the solvation structure of the metal ions to enable a facile charge-transfer reaction (see the Perspective by Zuo and Yin). In full battery cells, these components lead to high efficiency and energy density. Theoretical calculations were used to understand the solvation structures. —MSL

Science, abg3954, this issue p. 172; see also abi6643, p. 156

CANCER MICROBIOME

Microbes hijack prostate cancer therapy

Androgens such as testosterone and dihydrotestosterone are essential for male reproduction and sexual function. Androgens can also influence the growth of prostate tumor cells, and androgen deprivation therapy (ADT) either by surgical means (castration) or pharmacological approaches (hormone suppression), is the cornerstone of current prostate cancer treatments. Pernigoni *et al.* found that when the body was deprived of androgens during ADT, the gut microbiome could produce androgens from androgen precursors (see the Perspective by McCulloch and Trinchieri). Gut commensal microbiota in ADT-treated patients or castrated mice produced androgens that were absorbed into the systemic circulation. These microbe-derived androgens appeared to favor the growth of prostate cancer and helped to facilitate development into a castration- or endocrine therapy–resistant state. —PNK

Science, abf8403, this issue p. 216; see also abl7070, p. 154

METALS IN BIOLOGY

Mussel mooring made mighty by metals

Mussels produce an exceptional proteinaceous adhesive to withstand waves and currents. Metal ions bound to modified tyrosine residues play an important role in reinforcing the adhesive. Priemel *et al*. brought together a variety of spectroscopy and microscopy techniques to study the cellular mechanisms involved in adhesive fabrication in mussels (see the Perspective by Wilker). They found that metal ion–rich vesicles are secreted alongside vesicles containing the adhesive protein and mix in a microfluidic-like process within interconnected microchannels found in the lateral duct of the mussel foot to create porous, adhesive plaque filaments. —MAF

Science, abi9702, this issue p. 207; see also abm2298, p. 148

GPCR SIGNALING

Location matters for GPCR signaling

Parathyroid hormone receptor (PTHR) controls vitamin D, calcium ion, and bone homeostasis by stimulating cyclic AMP (cAMP) production at the plasma membrane or endosomes after β-arrestin–dependent internalization. White *et al*. generated a synthetic PTH derivative, PTH^{7d}, that inhibited β-arrestin coupling to PTHR, thereby leading to sustained cAMP production at the cell surface. Comparison of cellular and mouse responses with wild-type PTH, PTH^{7d}, or a PTH analog that elicits sustained endosomal signaling showed that the biological outcome was determined by the subcellular location of PTHR activity. —AMV

Sci. Signal. **14**, eabc5944 (2021).

TISSUE REGENERATION

The mechanics of muscle regeneration

Immune cells play an important role in skeletal muscle response to injury. Seo *et al.* studied the effects of mechanical loading on hindlimb muscle regeneration after intramuscular myotoxin injection and ischemia in mice. Compressive force applied by a soft robotic system facilitated neutrophil clearance, reduced proinflammatory cytokines and chemokines, and improved muscle fiber composition and function. These results support the development of therapeutic mechanical stimulation regimens to accelerate muscle regeneration. —CC

Sci. Transl. Med. **13**, eabe8868 (2021).

CIRCADIAN RHYTHMS

A biological clock in a test tube

The biological clock of cyanobacteria, which remarkably requires just three proteins, has been reconstituted in vitro in a system that allows detailed study of its inputs and outputs, bringing new understanding of how environmental signals can influence a biological oscillator and how the clock controls cellular events such as gene transcription. Chavan *et al*. extended the known in vitro function of the core clock components to include output signals to transcriptional regulation and allow monitoring through fluorescence measurements in real time. The authors combined crystallography, mutagenesis, and quantitative modeling to further explore the clock mechanism, which may enable future synthetic biology applications. —LBR

Science, abd4453, this issue p. 170

NEUROSCIENCE

How to optimize deep-brain stimulation

Deep-brain stimulation as presently used in the clinical setting, for example, to treat Parkinson's disease, does not differentiate between different neural circuitries. Considerable improvements could thus be achieved with selective stimulation that targets particular neuronal populations. Spix *et al*. used optogenetics to develop a clever electrical stimulation protocol that enhances cell-type specificity (see the Perspective by Haas). The authors managed to drive population-specific neuromodulation in a brain region called the external globus pallidus with brief bursts of electrical stimulation, which then yielded a long-lasting effect in a mouse model of Parkinson's disease. —PRS

Science, abi7852, this issue p. 201; see also abl9915, p. 153

CANCER

Paving the way for KRAS inhibitors

KRAS is a key oncogene in multiple cancer types, but existing inhibitors target only a mutant form of KRAS containing the G12C mutation, and their function presents a mechanistic conundrum. It is known that $KRAS^{G12C}$ inhibitors bind to the oncoprotein in its inactive form; however, KRAS mutations such as G12C interfere with the action of proteins that normally help it hydrolyze GTP to achieve the inactive state. Li *et al*. have now identified a protein that enhances GTP hydrolysis by mutant KRAS, helping to explain the clinical activity of current drugs targeting this oncoprotein (see the Perspective by Cox and Der). —YN

Science, abf1730, this issue p. 197; see also abl3639, p. 152

NANOPHYSICS

To have and to hold

The ability to couple the motion of levitated systems to internal degrees of freedom—in some cases down to their motional ground state—as well as to external forces and systems provides opportunities for fundamental science and technology. Gonzalez-Ballestero *et al*. provide an overview of the status, challenges, and prospects of levitodynamics, the mutidisciplinary research area that focuses on the understanding and control of optical trapping of microparticles and nanoparticles in vacuum, with the goal of eventually using such levitated particles as ultrasensitive probes for sensing applications. —ISO

Science, abg3027, this issue p. 168

ECOLOGY

Resilience to tipping points in ecosystems

Spatial pattern formation has been proposed as an early warning signal for dangerous tipping points and imminent critical transitions in complex systems, including ecosystems. Rietkerk *et al*. review how ecosystems and Earth system components can actually evade catastrophic tipping through various pathways of spatial pattern formation. With mathematical and real-world examples, they argue that evading tipping and enhancing resilience could be relevant for many ecosystems and Earth system components that until now were known as tipping prone. Many of these complex systems may be more resilient than currently thought because of overlooked spatial dynamics and multiple stable states, and may thus not undergo critical or catastrophic transitions with global change. —AMS

Science, abj0359, this issue p. 169

REVIEW SUMMARY

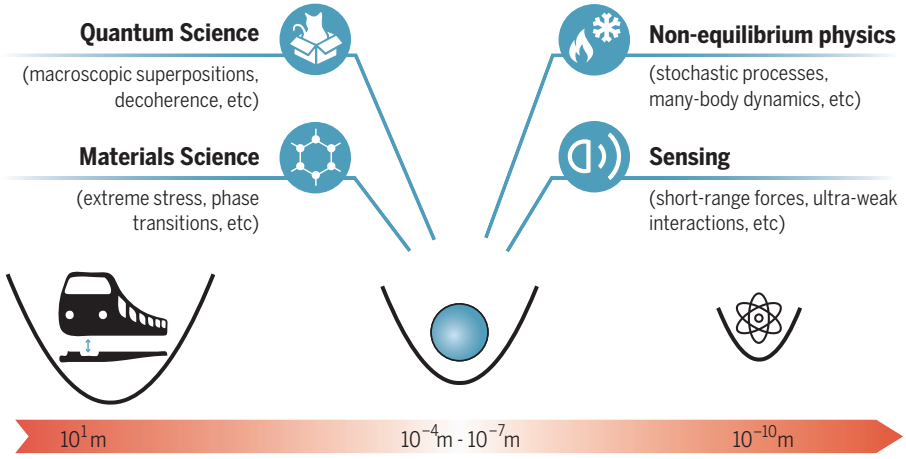
NANOPHYSICS

Levitodynamics: Levitation and control of microscopic objects in vacuum

C. Gonzalez-Ballestero, M. Aspelmeyer, L. Novotny, R. Quidant, O. Romero-Isart*

BACKGROUND: Levitation of large objects has become a widely used technique in science and engineering, and the control of levitated nano- and micro-objects in vacuum has gained considerable momentum. The ability to manipulate and measure translation, rotation, and free dynamics of these objects with high precision has led to a new experimental platform with opportunities for fundamental and applied research. For example, the high acceleration and force sensitivities of levitated objects are fueling both sensor development and searches for new physics, and the underdamped dynamics in complex potential landscapes enables the testing of stochastic thermodynamic models. Moreover, damping and noise can be reduced to a fundamental minimum in ultrahigh vacuum, paving the way for quantum-limited sensing and detection, as well as for exploration of macroscopic quantum physics in a hitherto unexplored regime of large masses.

ADVANCES: Levitation of micrometer-sized objects in vacuum was first demonstrated in the pioneering work of Arthur Ashkin in the 1970s. In the following decades, his work inspired the development of trapping and cooling of atoms and contributed to the rise of modern atomic physics. Levitation of microscopic solids focused on applications in air and liquids, with substantial breakthroughs, particularly in biophysics. In 2010, quantum optical control techniques were first proposed for levitation in vacuum, with the possibility of cooling the motion of a levitated nanoparticle in an optical cavity to the quantum regime. These proposals have been complemented by experimental research of control mechanisms based on optical, electrical, and magnetic forces. By now, both optical cavity–based and active-feedback cooling schemes have successfully cooled the motion of a dielectric levitated nanoparticle into the quantum ground state. Aside from translation, rotational degrees of freedom can also be controlled. Nanoparticle rotation frequencies in the gigahertz range are limited only by the tensile strength of the material. Furthermore, both translation and rotation of levitated particles have been coupled to single solid-state spins, providing a means of engineering the quantum motional and rotational states of a levitated object.



Levitation across length scales

Controlled levitation of macroscopic- and atomic-scale objects has advanced technology and fundamental science, respectively. Levitodynamics—the levitation and control of microscopic solids in vacuum—targets objects in the intermediate-size regime. This research area provides access to unexplored research directions in quantum science, materials science, nonequilibrium physics, and sensing.

Levitation allows particles to evolve in complex potential landscapes that can be actively shaped or switched off to include free dynamics. Furthermore, the coupling of the particle to its environment, and thus its dissipation and noise, can be tuned via gas pressure over several orders of magnitude. This flexibility has become the basis for studying nonequilibrium physics with levitated particles in both the classical and quantum regimes.

OUTLOOK: Levitation of nano- and micro-objects in high vacuum offers new opportunities for research and applications by providing previously unattainable isolation from the environment. The current toolbox includes magnets, metals, diamonds containing color centers, graphene, liquid droplets, and even superfluid helium—all of which can be levitated and controlled by means of optical, electrical, and magnetic interactions. These interactions provide a means to couple the internal degrees of freedom (e.g., phonons, magnons, excitons) to the well-controlled external degrees of freedom (translation, rotation). Levitated systems are clean test beds for materials science, where matter in extreme conditions (e.g. high internal stress or out-of-equilibrium behavior) can be probed and even engineered. Furthermore, levitated systems are an ideal platform for the study of nonequilibrium physics, such as the thermalization of mesoscopic objects in vacuum as well as complex many-particle physics (e.g. interparticle interactions induced by short-range electromagnetic fields or by tunneling of internal excitations between particles). Extending control to all degrees of freedom of a levitated particle is key to reduce sources of noise and decoherence. This step opens the door to a new regime of macroscopic quantum physics (e.g., the preparation of macroscopic quantum superpositions of objects composed of billions of atoms) and the probing of weak forces in unexplored regimes, including non-Newtonian gravity and searches for new physics beyond the standard model. Finally, the use of levitated systems for ultrasensitive detection of forces brings opportunities for commercial sensing applications, including gravimeters, pressure sensors, inertial force (acceleration, rotation) sensors, and electric and magnetic field sensors. Exploration of levitated sensors has already begun, but their performance has yet to be benchmarked against those of existing sensing technologies. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: oriol.romero-isart@uibk.ac.at
Cite this article as C. Gonzalez-Ballestero *et al.*, *Science* **374**, eabg3027 (2021). DOI: 10.1126/science.abg3027

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg3027

REVIEW

NANOPHYSICS

Levitodynamics: Levitation and control of microscopic objects in vacuum

C. Gonzalez-Ballestero[1,2], M. Aspelmeyer[3,4], L. Novotny[5,6], R. Quidant[6,7], O. Romero-Isart[1,2]*

The control of levitated nano- and micro-objects in vacuum—which capitalizes on scientific achievements in the fields of atomic physics, control theory, and optomechanics—is of considerable interest. The ability to couple the motion of levitated systems to internal degrees of freedom, as well as to external forces and systems, provides opportunities for science and technology. Attractive research directions, ranging from fundamental quantum physics to commercial sensors, have been unlocked by the many recent experimental achievements, including motional ground-state cooling of an optically levitated nanoparticle. Here we review the status, challenges, and prospects of levitodynamics, the multidisciplinary research area devoted to understanding, controlling, and using levitated nano- and micro-objects in vacuum.

Levitation has been a topic of fascination for centuries. Although it was initially a subject reserved solely for science fiction, it became a reality by means of recent advances in science and technology (*1*). Well-known examples include magnetically levitated trains used for high-speed transportation and magnetically levitated centimeter-sized superconductors used as accelerometers (*2*). Further reducing the size of the levitated object enables a previously unattainable degree of control over its dynamics in vacuum. Indeed, in vacuum, the plethora of external (motion, rotation, libration) and internal (e.g., phonons, excitons, magnons, quantum impurities) degrees of freedom of a levitated micro- or nanoparticle can freely evolve or interact with each other in conditions of extreme environmental isolation. High-precision control control over these degrees of freedom can be achieved by borrowing and building on techniques developed in atomic physics and biophysics (*3*–*8*), such as optical tweezers, dynamical control of trapping potentials, and motional cooling.

Levitodynamics—the trapping and control of nano- and micro-objects in vacuum—has established itself as an exciting research field by combining several ingredients from other fields in a single platform. When compared with other levitation platforms, two features stand out: (i) the large mass and density (and hence complexity) of the levitated object, as compared with the mass and density of trapped atoms; (ii) the high degree of control over both conservative dynamics and coupling to the environment, as compared with the degree of control for levitated meter-scale objects. The combination of these two features provides a springboard to many applications and branches of physics. For instance, large masses provide high inertial sensitivities for acceleration sensors. Moreover, the possibility of dynamically tuning both external potentials and dissipation provides a platform to study nonequilibrium physics and thermodynamics. This tunability also enables a top-down approach to quantum physics—namely, to bring a mesoscopic object of billions of atoms from the classical to the quantum regime to address long-standing questions about the nature of quantum mechanics at large scales. The possibilities of levitating particles of any material and of externally accessing their multitude of internal degrees of freedom also enable studies of condensed matter in unexplored regimes (e.g., in the absence of any substrate). Applications range from studying exotic material properties under extreme conditions to answering fundamental questions about the nature of equilibration in de facto isolated complex systems. Levitation also provides new opportunities for established fields—for instance, the study of liquid droplet levitation is useful for model development of aerosol processes in the atmosphere (*9*). In essence, levitodynamics offers the possibility to levitate microscale objects of arbitrary shape and type in high vacuum and to study its motional and rotational dynamics, its internal properties, and the interplay between them.

In this Review, we detail advances in levitodynamics that have brought maturity to the field, as well as research directions currently under exploration. We discuss near-future prospects enabled by the untapped potential of levitodynamics and conclude with an overview of the main challenges that lie ahead.

[1]Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria. [2]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences A-6020 Innsbruck, Austria. [3]Vienna Center for Quantum Science and Technology, Faculty of Physics, University of Vienna, A-1090 Vienna, Austria. [4]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, A-1090 Vienna, Austria. [5]Photonics Laboratory, ETH Zürich, 8093 Zürich, Switzerland. [6]Quantum Center, ETH Zürich, 8093 Zürich, Switzerland. [7]Nanophotonic Systems Laboratory, Department of Mechanical and Process Engineering, ETH Zürich, 8092 Zürich, Switzerland.
*Corresponding author. Email: oriol.romero-isart@uibk.ac.at

Advances

Trapping of polarizable objects by means of optical forces was pioneered in the 1970s (*10*). The subsequent development of optical tweezers revolutionized the field of biophysics (*7*), whereas the levitation of microparticles (*11*, *12*) and, later, atoms in vacuum set the basis for modern atomic physics (*4*, *6*). Building on these advances, several theoretical works in the 2010s proposed applying quantum optical control techniques to levitated objects in vacuum (*13*–*16*). In this section, we review the advances sparked by the aforementioned works, which established levitodynamics as a vibrant research field.

Levitation schemes

The physical mechanisms involved in the levitation of nano- and microparticles are summarized in Box 1, and a comparison between different levitation methods is provided in Table 1. Early levitation experiments made use of optical potentials and weakly absorbing dielectric polarizable particles. In the subsequent years, the toolbox expanded to include techniques borrowed from the atom-trapping community. The development of electrostatic and magnetic levitation made it possible to overcome excessive photoheating of the trapped specimen (*17*, *18*) and extended levitation to magnets (*19*–*22*), metals (*23*), diamond with quantum emitters (*24*–*27*), liquid droplets (*28*), graphene flakes (*29*), and even superfluid helium droplets (*30*).

Loading in high vacuum

Regardless of the trapping mechanism, one of the main challenges for trapping is the controlled loading of a particle into the trap in vacuum. Simple loading techniques such as spraying with a nebulizer rely on gas damping, which requires loading at higher pressures and subsequent pumping to desired vacuum conditions. More refined approaches enable particle loading in high vacuum. These methods make use of particle loading from a charged tip (*27*) or particle launching with piezoelectric transducers (*11*), electrospray (*31*), laser-induced acoustic desorption (*32*–*34*), or transfer through a load-lock (*35*, *36*).

Measurement and cooling

Another key advance has been the implementation of feedback control (cooling; see Fig. 1) of levitated particle motion, which is essential for mitigating and controlling the impact of external noise. Freezing the motion of a levitated particle into the quantum ground state sets the stage for macroscopic quantum mechanics and has thus been a major objective

Box 1. Physical mechanisms for levitation of nano- and microparticles.

Levitation is the stable suspension against gravity of an otherwise free particle, detached from any substrate, in the minimum of a potential $V(\boldsymbol{R})$, where $\boldsymbol{R}$ is the center-of-mass position of the particle. Near a local minimum, the trapping potential is approximately parabolic and the particle behaves as a 3D damped harmonic oscillator. Its motion along each axis $j = x, y, z$ obeys the equation

$$\ddot{R}_j + \gamma_j \dot{R}_j + \Omega_j^2 R_j = -g\delta_{jz} + F_j^{\text{fluct}}(t)/m$$

Here, Ω_x, Ω_y, and Ω_z are the mechanical frequencies along each axis, which can depend on the particle mass m (also, a single dot represents a first-order time derivative, a double dot represents a second-order time derivative, and δ_{ij} is the Kronecker delta). The particle is subject to the acceleration of gravity g, to dampings at a rate γ_j, and to a fluctuating force $\mathbf{F}^{\text{fluct}}(t)$. Both the damping and the fluctuating force describe the effect of the environmental noise and can stem from many different sources (Fig. 4). Stable trapping requires a potential that is able to compensate for the force of gravity and is deep enough to prevent the particle from escaping as a result of fluctuating forces. The trap depth along axis j is given by

$$U_j \propto m\Omega_j^2 L_j^2$$

where L_j is a typical length-scale characteristic of each trapping mechanism. As trap depths increase with particle mass, the same trap is much deeper for a particle than for an atom or an ion, assuming mass-independent frequencies. The trap frequencies and depths depend strongly on the trapping mechanism, which can be mediated by optical, electrical, and/or magnetic forces.

Optical trapping

Dipole forces induced by the 3D intensity gradients at the focus of a laser beam can be used to trap dielectric particles. Mechanical frequencies Ω_l and Ω_t (motion parallel or transverse, respectively, to the optical axis) for subwavelength particles are typically in the range of $\Omega_j/2\pi \sim 10$ to 300 kHz. They are mass independent and scale as

$$\Omega_t \sim \sqrt{P}/w^2; \Omega_l \sim \sqrt{P}\lambda/w^3$$

where P is the laser power, w is the beam waist at the focus, and λ is the wavelength. They also depend on the particle density.

Electrical trapping

Radiofrequency traps (also known as Paul traps) use an inhomogeneous ac electric potential with amplitude V_0 and frequency ω to trap charged particles. For small particles, mechanical frequencies, typically in the $\Omega_j/2\pi \sim 1$- to 10-kHz range, scale with their charge-to-mass (q/m) ratio

$$\Omega_j \propto V_0/\omega \times q/m$$

Magnetic trapping

Many workarounds to Earnshaw's theorem facilitate stable levitation of magnetic particles without external time-dependent fields. An example is diamagnetic levitation (typical frequencies $\Omega_j/2\pi \sim 0.1$ to 1 kHz) of permanent magnets above type I or type II superconductors (*19*–*22*), via the Meissner effect and magnetic flux trapping, respectively. Diamagnetism also allows trapping of diamagnetic [including superconducting (*179*)] particles in magneto-gravitational and quadrupole traps. The coupling between magnetization and rotations (e.g., Einstein–de Haas effect) (*187*) enables, according to theoretical predictions, stabilization of levitated magnets through rotation or quantum spin (*188*, *189*).

in the field of levitodynamics. This quest has been pursued by means of both active- and passive-feedback cooling.

Active-feedback cooling relies on an actual measurement of the particle, the result of which is used to counteract particle motion by means of a control loop (*37*–*42*). The efficiency of active feedback relies on three factors: (i) quantum-limited measurement (*43*, *44*), (ii) optimal control (*45*–*47*), and (iii) high detection efficiency (*48*). The trade-off between measurement imprecision and measurement backaction is fundamentally constrained by the Heisenberg limit, which sets a lower bound on the performance of active-feedback cooling. Optimization on all three parameters has led to measurement near the Heisenberg limit, observation of motional sideband asymmetry (*49*), and demonstration of quantum ground-state cooling for an optically levitated dielectric particle in both room temperature (*50*) and cryogenic (*51*) environments.

Passive-feedback cooling makes use of the autonomous feedback provided by an optical cavity (*52*–*54*). In the so-called resolved sideband limit, photons in the cavity most effectively counteract, and thereby slow, the motion of the particle by means of radiation pressure (*13*–*16*). Furthermore, as known from research into cavity cooling of atoms and molecules (*55*, *56*), laser noise can be minimized if the cavity is populated solely by the photons scattered from the particle (coherent scattering) (*57*–*59*). This configuration led to the first report on ground-state cooling of an optically levitated dielectric particle (*60*).

Motional and rotational control

One of the main advantages of levitation is that the particles are free to move when presented with external influences, allowing for precise control of their external degrees of freedom (motion, rotation). Control of conservative particle motion is achieved via engineering of the trapping potential. On the one hand, complex static potentials can be synthesized using optical trapping beams with desired intensity profiles (*61*–*64*) or by following a hybrid approach (i.e., combining optical and electric fields that act on charged dielectric particles) (*53*, *65*, *66*). On the other hand, potentials can also be dynamically modified, either externally (*24*, *33*, *67*, *68*)—including fast control of a complex optical potential (*69*)—or by the backaction induced by the particle itself on the trapping field (*70*, *71*). Furthermore, the equilibrium position of a levitated particle can be controlled with high precision, enabling its positioning and translation with nanometer-scale accuracy. Such control has been implemented with moving single-beam traps (*35*, *36*, *72*), standing-wave traps (*73*–*75*), and electric Paul traps (*76*).

Precise control has also been achieved over the rotational degrees of freedom of levitated particles. On the one hand, tuning the polarization direction of linearly polarized light on an optical trap allows external control of the equilibrium angular orientation of an anisotropic particle (*23*, *70*, *77*). On the other hand, using circularly polarized light, particles of different materials and shapes can be set into sustained rotation in a controlled fashion (*29*, *32*, *70*, *78*–*82*). Record spinning velocities in the gigahertz range have been reported for nanodumbbells (*83*–*85*) and are limited only by particle disintegration when the elastic tensile limit is reached. Furthermore, rotational degrees of freedom can be cooled (*86*), even simultaneously with the translational motion (*87*–*89*). Enhanced understanding of rotational noise (*90*), especially noise induced by particle motion (*91*), has made it possible to reach the shot-noise limit, defined by radiation pressure torque (*88*). This paves the way for bringing rotational degrees of freedom to the quantum regime.

Coupling to external systems

Experiments focused on coupling a levitated particle to external systems have been driven by the goals of (i) preventing photodamage in absorptive particles and (ii) exploiting nonlinear interactions (e.g., via coupling to a two-level system). Two prominent cases are levitated magnets coupled to other magnetic systems and diamond particles that contain color centers. Diamagnetically levitated magnets generate a position-dependent magnetic field that

Table 1. Advantages (✓) and drawbacks (✗) of optical, electrical, and magnetic interactions in terms of levitation, control, and measurement of the particle.

	Optical	Electrical	Magnetic
Levitation	✓ High trapping frequencies (>100 kHz) ✓ Charged or neutral particles ✓/✗ Large distance from surfaces ✗ Laser recoil heating ✗ Internal heating due to absorption ✗ Limited to ultralow-absorption, high-temperature–resistant particles	✓ No internal heating ✓ Applicable to a wide range of charged particles ✓ Large potential depth ✓ On-chip integration ✓ Compatible with optical manipulation ✗ Low trapping frequencies (<10 kHz) ✗ Highly charged particles (sensitive to decoherence) ✗ Poor spatial confinement	✓ No internal heating ✓ Potentials created by static fields ✓ Applicable to very large particles ✓ On-chip integration ✓/✗ Advanced cryogenics required for superconducting traps or particles ✗ Low trapping frequencies (<1 kHz) ✗ Dissipation and decoherence mechanisms not fully understood
Control	✓ Cooling via active and passive feedback ✓ Potentials can be easily engineered and reconfigured ✓ Control of rotational degrees of freedom ✗ Coherent control limited by decoherence due to light scattering	✓ Cooling via active and passive feedback ✓ Simple implementation of cold damping ✓ Particles with internal degrees of freedom (quantum emitters) ✗ Proximity to surfaces (sensitive to stray fields)	✓ Cooling via active and passive feedback ✓ Easy access to other degrees of freedom (e.g., magnetization) ✗ Static potentials
Measurement	✓ Shot-noise–limited measurements ✗ Laser recoil heating	✓ No need for bulk optical elements ✗ Requires the detection of very weak currents	✓ Sensitive readout with magnetometers (e.g., SQUIDs) or superconducting circuits (cavities, qubits) ✗ Low compatibility with optical measurements

has been read out with both a superconducting quantum interference device (SQUID), which offers high sensitivity and minimal power dissipation at cryogenic temperatures (*20*, *92*), and a single electronic spin in a nitrogen-vacancy color center (NV center) embedded in a nearby diamond slab (*19*). NV centers can also be embedded in a levitated nanodiamond (*26*), thereby equipping the particle with an internal energy-level structure. These quantum emitters can be optically manipulated and coherently controlled (*25*, *27*), even at the single-electron level (*23*, *24*). Moreover, in the presence of a magnetic field gradient, electronic spins couple to the motion and rotation of the particle, as the Zeeman splitting of the electronic levels depends on the position and orientation of the particle. This spin-mechanical coupling in an ensemble of electronic spins has been used to control the orientation of a levitated diamond (*77*), exert a torque (*93*), and cool its librational motion via optical polarization (*86*). Hosting quantum emitters in a levitated particle also provides a local probe for its internal dynamics.

Stochastic thermodynamics

The advances outlined above have enabled controlled levitation of micro- and nanosolids in high vacuum, advancing the use of levitodynamics in the pursuit of several research directions. The first example is the study of single-particle stochastic thermodynamics. Levitated particles provide a convenient platform for this goal, as motional and rotational degrees of freedom can be controlled, and their coupling to the environment can be tuned by many orders of magnitude. This tuning, a distinctive feature of levitation, can be performed by controlling gas pressure or by engineering artificial reservoirs (e.g., using an optical cavity) (*94*). The controllable isolation makes it possible to study equilibration (*17*) or underdamped Brownian motion in both linear (*95*) and nonlinear (*96*) regimes. As a specific example, the tunability of the damping rate facilitates exploration of thermally activated transitions between two potential minima separated by a barrier (*61*, *64*). It has been shown that the transition probability is maximized at intermediate damping rates, a phenomenon predicted by Kramers in 1940. Dynamical modulation of trapping potentials is also a powerful tool. Specifically, it provides a means to study the equilibration from nonequilibrium initial states, enabling tests of fundamental fluctuation theorems (*97*, *98*), and to implement shortcuts to adiabaticity for heat engine cycles (*99*). Along with feedback control, this technique also enables assessment of fluctuations under nonequilibrium thermal driving (*100*) and of a generalized second law in non-Markovian information thermodynamics (*101*). The combination of dynamical potential modulation and tunable dissipation provides a window into the thermodynamics of small systems, where fluctuations are relevant. The many open questions pertaining to this regime are explored, for example, by devising microscopic levitated heat engines (*95*, *99*) that can operate near the Curzon-Ahlborn efficiency bound (*94*, *102*). In this realm, another quantum system, such as an optical cavity, enables gradual substitution of the thermal reservoir to experimentally study the extension of the concepts of classical stochastic thermodynamics into the quantum regime (*103*).

Investigation of materials

Another feature of levitated nano- and microparticles is the ability to control and investigate mesoscopic solids composed of billions of atoms in conditions of extreme isolation. Important applications are measurement and reduction of a particle's internal temperature. This is relevant for preventing particle loss at low pressures, reducing blackbody decoherence for fundamental tests of quantum mechanics, or reducing decoherence of embedded color centers for use in quantum transduction. One way to measure the internal temperature of optically levitated particles is via the mechanical frequency shift induced by a temperature-dependent refractive index (*104*). Another way is to use the fluorescence of embedded quantum emitters (*25*, *105*), which also allows for

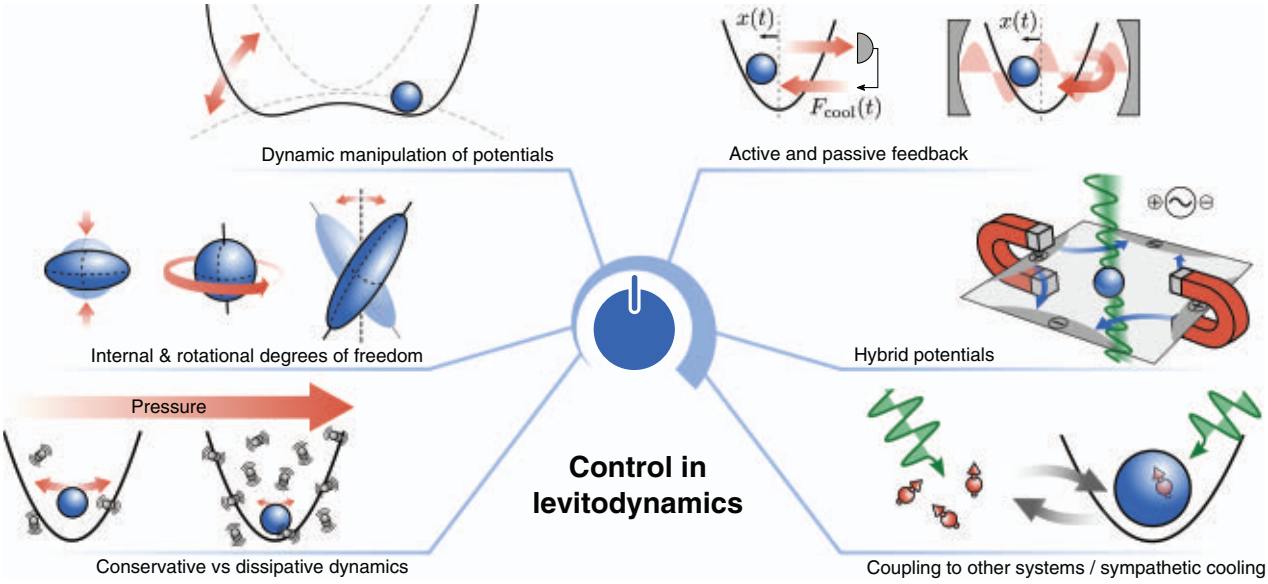


Fig. 1. Control in levitodynamics. The control toolbox of levitodynamics makes it possible to tailor every degree of freedom of a levitated particle in multiple ways. Both the conservative dynamics and the dissipation of motional, rotational, and internal (e.g., acoustic) degrees of freedom can be tailored via dynamic manipulation of potentials and modification of gas pressure, respectively. They can also be cooled and parametrically driven via both active and passive feedback. Additional control can be achieved through hybrid traps and coupling to external systems, which also provide controlled nonlinearity. $F_{cool}(t)$, cooling force.

internal refrigeration of the particle (*106*). Additionally, surface temperature can be estimated by tracking particle motion in the background gas (*17*).

A related research direction is the study of matter in controlled conditions (i.e., without the perturbing effects of any substrate). The absence of clamping allows particles to be rotated at frequencies that exceed 1 GHz (*79*, *83*, *84*), which, for particles of 100-nm radius, results in surface velocities of 600 m/s and centrifugal accelerations of $4 \times 10^{11} g$ (where g is Earth's gravitational acceleration). The extreme surface forces and material stresses provide information about material susceptibilities, their modification as a result of large centrifugal stress (*29*), and tensile failure (*107*). The absence of a substrate has received particular attention in the study of liquid particle (droplet) physics, which is important for modeling a wide range of systems, from atomic nuclei to black holes (*28*). Moreover, at low temperatures, levitation of liquid helium droplets provides insight into superfluid physics (*108*)—e.g., into the complex interplay between fast droplet rotation and the formation of superfluid vortices. In this scenario, by providing controlled and stable trapping, levitation offers a definite advantage over current methods based on x-ray diffraction of liquid helium jets (*109*).

Sensing

Levitated particles also excel as force sensors (Fig. 2). Their three-dimensional (3D) motion and rotation allows for independent sensing of forces and torques along three dimensions (6D sensing) with shot-noise–limited sensitivity. Moreover, their large motional amplitude provides a much larger dynamic range than that afforded by clamped resonators. In addition, the spectral response of levitated sensors can be largely tuned by potential engineering, thus modifying their oscillation frequencies. Force sensitivities in the sub-femtonewton regime have been demonstrated (*37*, *38*, *110*–*114*). The high force sensitivity can be exploited for the sensing of different fields. For example, if the levitated particle carries a charge, it responds to electric fields and thus constitutes an electric field sensor (*40*, *110*, *111*, *115*, *116*). Likewise, if the particle carries a magnetic moment it responds to magnetic fields and acts as a magnetic field sensor (*82*, *117*). Furthermore, the high mass of a levitated particle (as opposed to, e.g., a trapped atom) makes it a natural sensor for gravitational and inertial forces, including acceleration and rotation (gyroscopes) (*21*, *68*, *118*–*120*). Anisotropic particles, such as rods, dumbbells, or other particles composed of anisotropic materials, can be set into rotation and used for torque sensing with sensitivities exceeding 10^{-27} $N{\cdot}m/Hz^{-1/2}$ (*81*, *82*, *84*, *85*, *91*). In addition, a polarizable particle also responds to optical forces, such as radiation pressure, which can be measured with photon-recoil–limited precision (*43*). Finally, the lack of a clamping mechanism allows levitated particles to be released from their trap and recaptured at a later time and different location. This feature opens up the possibility of measuring forces "in the dark" [e.g., the gravitational force in a free-fall experiment (*68*)]. Aside from sensing, this protocol is of interest for quantum matter-wave interferometry with massive, solid-state particles (*121*–*127*).

Future research directions

The potential of levitodynamics is largely untapped. However, enabled by the aforementioned advances, a large variety of research directions and applications will become attainable in the near future. We classify these directions in three broad, interconnected areas: (i) sensing and metrology, from commercial navigation sensors to searches for new physics such as dark matter; (ii) novel quantum physics, e.g., generation of macroscopic quantum superpositions; and (iii) physics of complex systems, from nonequilibrium physics to materials engineering.

Integrated sensors

On the applied side, levitation-based sensing may find uses in areas ranging from seismometry to inertial navigation systems. Development of commercial sensors, however, will require integration of trapping, control, and readout on a single platform to reduce noise and achieve long-term stability. The prospect of compact and robust integrated sensors motivates the ongoing effort toward on-chip levitodynamics. Controlled levitation near structured surfaces, a key first step, has already been achieved in ambient conditions and low vacuum. Dielectric nanoparticles have been levitated in planar Paul traps (*128*), near dielectric photonic structures (*129*) and metasurfaces (*130*),

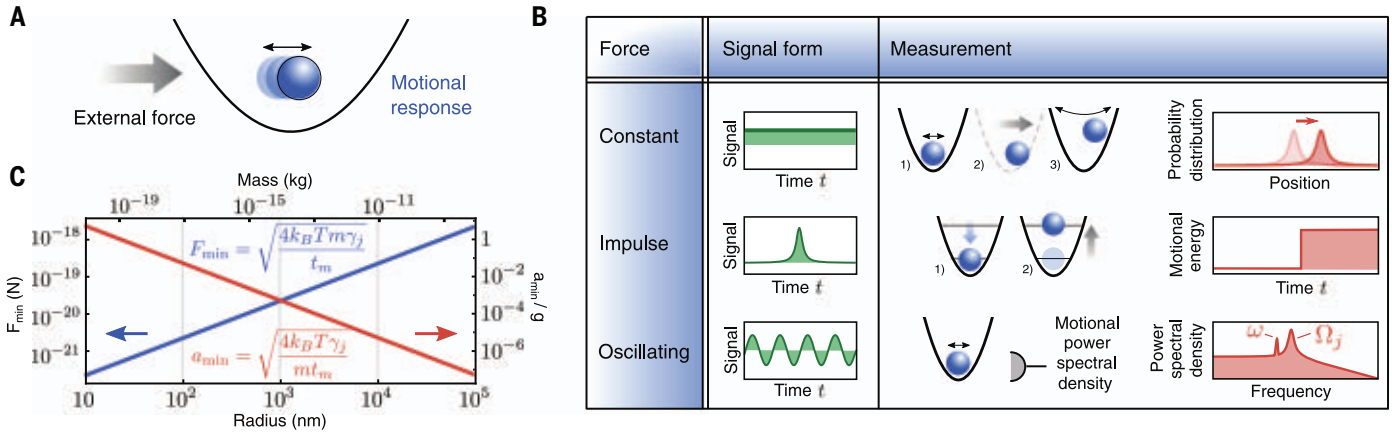


Fig. 2. Sensing. (**A**) Owing to its distinctive mechanical properties, a levitating particle in ultrahigh vacuum is very sensitive to external forces and torques. (**B**) External signal types: (i) Constant (e.g., Earth gravity): measured through the shift in phase-space probability distribution after letting the particle motion dynamically evolve (e.g., free-fall expansion). This requires sharp-featured phase-space distributions (e.g., lower initial temperatures, interference fringes) and many identical experimental repetitions. (ii) Impulse (e.g., inertia in navigation or scattering of a gas molecule): measured by monitoring the energy of a particle with low initial energy (achieved, e.g., via motional cooling). (iii) Oscillating: ~$F\cos(\omega t)$ (e.g., gravity generated by an oscillating mass), detected as additional peaks in the position power spectral density. (**C**) Minimum detectable force (F_{min}; blue) and acceleration (a_{min}; red) for oscillating signals, using a SiO_2 particle at motional thermal equilibrium at T = 300 K. We assume gas-limited damping ($\gamma_j \propto$ pressure) at 10^{-8} mbar and an integration time t_m = 1 s. Lower dampings improve sensitivity. Motional cooling does not, as $\gamma_j T$ is constant, but does increase stability and decrease ring-down time after perturbations. Small (or large) particles are advantageous for force (or acceleration) sensing. For $T \to 0$, measurements become limited by quantum backaction (e.g., photon recoil). The ultimate sensitivity is bound by the Heisenberg limit.

and near ultrathin membranes (*131*). Levitation of magnetic particles near the surface of superconductors (*92*), which has applications in high-sensitivity magnetometry (*117*), has also been reported (*19*–*22*). Current efforts focus on reducing particle-surface separations, increasing vacuum conditions, and using the same nearby surfaces or structures for particle control and readout (*132*, *133*). On this front, particles have already been levitated at subwavelength distances from surfaces (*129*, *134*), even in high vacuum with the aid of feedback cooling (*131*). Moreover, exploitation of the strong mode confinement of a photonic crystal cavity has enabled enhanced readout of the particle dynamics (*129*). Controlling and mitigating the still poorly understood surface-induced noise, also known as anomalous heating in the trapped-ion community (*135*), remains a major challenge for on-chip levitodynamics.

Rotation-based sensing and metrology

Owing to their complex dynamics, the rotational degrees of freedom of a levitated particle can be exploited in many areas of sensing and metrology. Fast-rotating nanorotors can act as local magnetic fields (*117*) or pressure sensors (*80*). Moreover, the frequency of optically levitated nanorotors can be locked to external clocks and tuned across a wide frequency range. The frequency stability of these nanorotors (*91*) is limited only by clock stability, even at room temperature (*80*). This will enable their use as frequency standards for metrology. Finally, rotation can be used for sensing fundamental forces. Fast-rotating nanorotors have been proposed as torsion balances for measuring processes such as Casimir torques or vacuum friction (*84*, *85*, *136*, *137*). Measuring such electromagnetic dispersion forces could elucidate the mechanisms behind surface-induced heating, opening the door for compact levitation platforms. Furthermore, controlled rotation will facilitate cancellation of limiting noise in fundamental sensing experiments, such as the measurement of Newton's gravitational constant (*138*) or the search for exotic particles beyond the standard model (*110*). A major challenge for rotational sensing and metrology is to reach the fundamental sensitivity limit imposed by the quantization of angular momentum and associated radiation torque shot noise (*88*).

New physics

The possibility of bringing particles to the quantum regime provides a novel sensing platform for fundamental science. Specifically, the substantial degree of isolation and control required to reach the quantum regime yields laboratory-scale experiments that are extremely sensitive to forces or phenomena that couple to mass. A first possible use is the search for new physics beyond the standard model of particle physics. As an alternative approach to colliders pushing the energy frontier, these compact levitodynamics experiments push the sensitivity frontier (*139*). They can exploit (i) the high isolation and mass density of levitated particles and (ii) levitation near structured surfaces (*131*) to detect tiny deviations arising from high-energy physics at short distances (*140*). This will enable tests of short-distance corrections to Coulomb's law to exclude dark matter models in the hidden sector, as well as searches for fundamental non-neutrality of matter (*141*). Experiments along this route have already been performed using optically levitated microparticles (*110*). Moreover, these experiments will allow researchers to test potential corrections to Newton's gravitational law at the unexplored micrometer and submicrometer scale (*140*, *142*, *143*), an order of magnitude beyond the 50-μm scale where strong constraints to deviations have been reported (*144*). A second related prospect for these sensitive experiments is the detection of astrophysical signals, including high-frequency gravitational waves (*145*) and signatures of dark matter and dark energy (*110*, *114*, *146*–*148*). Finally, levitated particles could shed light on the quantum nature of gravity by entangling two levitated massive objects via their mutual gravitational interaction (*149*, *150*). As discussed below, exploring all of these research directions requires devising advanced techniques to suppress quantum decoherence.

Quantum sensing and transduction

Bringing a levitated particle to the quantum regime also enables a variety of genuine quantum applications (Fig. 3), such as quantum-enhanced sensing. These applications require

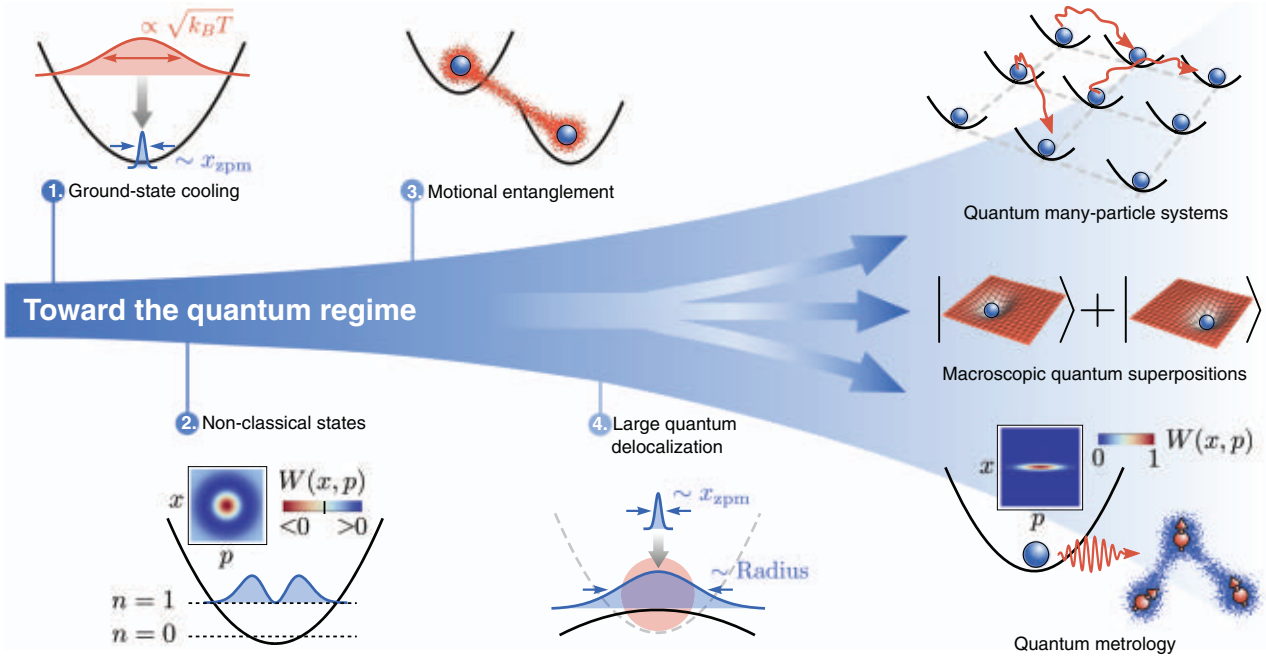


Fig. 3. Quantum levitodynamics. Levitated particles in the quantum regime open new research directions in fundamental quantum mechanics, quantum sensing, and many-particle physics, among other areas. Reaching this regime is a challenging effort that must be undertaken sequentially. Some milestones along the way are (1) ground-state cooling, which reduces the thermal motional amplitude to the zero-point motion x_{zpm} (k_B, Boltzmann constant); (2) generation of nonclassical motional states at length scales ~x_{zpm} (e.g., Fock states); (3) demonstration of entanglement between the motional degrees of freedom of two levitated particles; and (4) delocalization of the particle wave function to an extension orders of magnitude larger than the zero-point motion, comparable to its size.

cooling of the particle to the quantum regime, as well as the ability to coherently couple its motion to other quantum systems—that is, to establish a coherent coupling at a rate higher than the decoherence rate of both the particle and the external quantum system. Although such strong coupling has been demonstrated with dielectric particles coupled to an optical cavity (*151*), establishing coherent coupling to other quantum systems remains to be achieved. Two-level systems are particularly relevant for current research, owing to their large intrinsic nonlinearity (*152*). As in clamped quantum nanomechanical systems, coherent coupling to such nonlinear systems will enable the preparation of levitated particles in non-Gaussian quantum states—that is, states characterized by a phase-space quasi–probability density with negative values. These states can be used to unequivocally certify the quantum nature of the particle's motional state and to identify decoherence sources, to which they are extremely sensitive. Moreover, coherent coupling to an ensemble of entangled two-level quantum systems will improve the displacement readout of levitated particles, with potential uses in inertial or—in the case of levitated magnets—magnetic quantum sensing. More specific to levitodynamics are applications in which an interface with internal solid-state quantum excitations, such as acoustic phonons, could be established. Theoretical proposals (*153*) indicate that, owing to their nearly perfect isolation, these long-lived quantum excitations could be used, for instance, as quantum memories or quantum transducers.

Macroscopic quantum superpositions

One of the most appealing features of quantum physics is the ability to prepare macroscopic superpositions, in which an object behaves as if it were in two locations at once, separated by a distance comparable to or larger than its size. Experimental confirmations started as early as 1927, using electrons, and have today reached the size of organic molecules that contain thousands of atoms (*154*). Levitodynamics will enable the preparation of macroscopic quantum superpositions of nanoparticles that contain billions of atoms and, in principle, far beyond. These superpositions will enable quantum matter-wave interferometry with mesoscopic particles (*121*–*124*, *155*) and will corroborate the quantum superposition principle in a hitherto inaccessible parameter regime where collapse models predict the breakdown of quantum mechanics (*122*, *156*). Moreover, they will serve as ultraprecise sensors able to detect tiny signals (*125*, *155*) and will thus promote enhanced understanding and testing of theoretical models of environmental decoherence. This is made possible by the extreme sensitivity of these superpositions to decoherence. Finally, superpositions will open the door for exploring the interplay between quantum mechanics and gravity—for example, by elucidating the gravitational field generated by a massive object in a macroscopic quantum superposition state (*157*).

Macroscopic quantum superpositions require sufficiently pure states that are coherently delocalized over large scales. These conditions could be achieved by ground-state motional cooling, followed by large-scale expansion of the wave function through dynamical control of the trapping potential (*127*, *158*, *159*) (Fig. 3). This expansion increases the spatial extent of the wave function by several orders of magnitude, from the zero-point motion in the ground state (typically around 10^{-12} m) to, ideally, a size comparable to that of the particle itself (10^{-7} m). Several theoretical works have proposed and analyzed these ideas and have determined the requirements to prevent the detrimental action of decoherence from the environment (*121*–*124*, *126*). After the achievement of motional ground-state cooling (*50*, *51*, *60*), the next steps consist of reaching the low decoherence level required for unitary expansion of the wave function and devising detection schemes to certify the successful preparation of a macroscopic superposition (*160*).

Nonequilibrium physics

Levitated particles offer a distinctive channel to probe and engineer complex systems across multiple scales. These encompass arrays of levitated particles for many-body physics, single

particles for motional quantum heat engines, and solid-state physics inside a single particle. At the largest scale, ensembles of levitated particles can be used to explore many-body physics as done with atomic ensembles, with the addition of rotational degrees of freedom, rich internal solid-state structure, tunable coupling to the environment, and gravitational interactions, among others. On the one hand, ensembles of nanoparticles levitated in a single wide and deep trap will enable the exploration of nonequilibrium self-assembly, liquid crystal phases, and the transitions between them (*161*, *162*). Conversely, arrays of physically separated particles with tunable interactions can be tailored by multiplexing several traps into an array (*163*). In these systems, the ability to tune the dissipation enables exploration of the transition between closed and open systems. This may contribute to previous studies (*164*) of elusive phenomena, such as prethermalization (*165*), and tests of long-standing hypotheses, such as the eigenstate thermalization hypothesis.

At the single-particle level, levitodynamics has turned out to be an excellent approach to implement far-from-equilibrium processes and understand irreversibility, similar to the complementary and long-established approach of colloids in liquids. This will certainly not be limited to optical approaches in the future. With the first demonstrations of levitated systems in the quantum regime, new ground can be explored at the interface of classical, information, and quantum thermodynamics. Focusing on the refinement of levitated microheat engines as a paradigmatic example, the rapid control enabled (e.g., by all-optical approaches) theoretically designed optimal protocols (*102*) to become a reality. Generally, substituting the thermal environment by engineered (quantum) reservoirs (*166*)—e.g., via cavity optomechanical methods (*167*–*169*)—opens the door for investigating extended schemes with nonthermal or nonclassical baths, making it possible to bypass the Carnot efficiency limit (*170*, *171*). More importantly, these engineered reservoirs will enable researchers to understand and minimize entropy production in the finite time domain, which is paramount, in engines and elsewhere, to optimize efficiency and minimize wasted energy. In the quantum domain, coherence, entanglement, and measurement backaction, which are now starting to become observable in optical levitodynamics, will play key roles in engine design and performance, addressing open questions in quantum thermodynamics (*99*, *172*). By helping researchers to understand these concepts and providing flexible possibilities for spatiotemporal design, levitodynamics may inspire an entirely new generation of classical and quantum nanoscale mechanical engines.

Materials science

A defining advantage of levitated particles is their high internal complexity as a mesoscopic solid-state system. The ability to investigate and control this isolated solid will open new avenues for materials science. Specifically, the absence of surrounding gas and substrate will enable precise measurements of physi- and chemisorption rates in unusually clean environments (*173*). Moreover, the enhanced control provided by levitation can be exploited for the growth of novel crystalline heterostructures (*174*). In the field of droplet levitation (*175*), the control techniques developed in levitodynamics will make it possible to scale down current experiments with microdroplets to the nanoscale. The preparation of nanodroplets with different chemical composition, viscosity, refractive index, and surface tension will facilitate the study of processes such as phase transitions, coagulation, and photochemical reactions, among others, as a function of external stimuli. In the future, levitodynamics could provide a nanolab for engineering the material properties of levitated nanoobjects.

Condensed matter physics

Levitodynamics also enables researchers to study previously unexplored regimes of condensed matter where solid-state excitations (e.g., phonons, electrons, magnons) exist in extreme isolation and confinement (*176*). Acoustic phonons have received the most attention, owing to their ubiquity and highest-recorded quality factor. This provides acoustic modes with extremely long coherence times but also makes them difficult to probe experimentally. Future detection will be possible by means of Brillouin scattering or, in rapidly rotating spheres, by the change in polarizability stemming from centrifugal elastic deformation (*177*). In optical traps, this deformation would additionally give rise to a complex dynamical coupling between rotation, motion, and acoustic field, which could shed light on the motional decoherence induced by internal vibrations. Once detection schemes become available, acoustic phonons will represent a powerful resource. According to theoretical works, a tunable and strong coupling between acoustic phonons and the particle motion can be engineered by using mediating degrees of freedom (e.g., magnons in levitated magnets) (*178*). This could enable the use of internal acoustic resonances for motional cooling without the need for external feedback, or for engineering internal and external states via motional control (*153*), such as squeezed magnonic, acoustic, or motional states. Because the novel properties of acoustic phonons stem entirely from confinement and isolation, similar exotic properties can be expected for any other condensed matter excitations (e.g., semiconductor excitons in levitated quantum dots or supercurrents in levitated superconducting particles) (*179*). Despite the potential of levitated particles, both for fundamental solid-state science and as tools to control external dynamics, their internal physics remains largely unexplored.

Open challenges

The prospects outlined above have established levitodynamics as a vibrant research field. However, several challenges must be addressed before such research directions can be fully explored.

On-chip levitodynamics

The implementation of levitation-based sensing platforms faces three substantial obstacles. The first one concerns the development of chip-scale levitation systems operating in high isolation. On-chip integration is key to interface levitodynamics with other existing technologies, to increase platform robustness, and to devise autonomous and portable sensors. Despite achievements in this direction, the transition from current bulky proof-of-principle experiments to compact on-chip platforms remains challenging. On the technical side, it requires the engineering of optical, magnetic, and electric fields, or combinations of them, at the chip surface to create microscale traps. These traps must include on-chip feedback-control schemes to stabilize traps in ultrahigh vacuum and/or reduce the ring-down time of trapped high–quality factor oscillators. Developing an autonomous integrated platform will also require miniaturization of key surrounding components, from vacuum equipment to cryogenic setup. On the fundamental side, particles in microscale traps experience surface-induced forces and decoherence, which are relevant at short particle-surface separations. These short-range interactions affect the trapping potential, and the dynamics they induce depend on the internal degrees of freedom of the trapped object and their properties (e.g., temperature, internal equilibration times). Because these degrees of freedom can be largely out of equilibrium, the forces may differ from typical predictions on the basis of quasi-equilibrium assumptions (*176*). Second, in analogy to anomalous heating in trapped ions (*135*), surface-induced dissipative forces create additional heating. These forces can have multiple origins, such as Johnson noise or patch potentials in Paul traps (*135*, *180*) or eddy currents or hysteresis losses for magnets trapped above superconductors (*19*, *20*, *92*). In most cases, these mechanisms are not well understood for levitated particles. To implement on-chip levitation in high vacuum, future research must focus on developing a fundamental understanding of surface-induced forces and implementing strategies to mitigate them.

Sensor performance

The second challenge ahead for levitation-based sensors is to establish their true sensing potential. Most studies are at the proof-of-concept level and focus on the minimum detectable signal, commonly referred to as sensitivity. However, for industrial and field applications, other attributes—such as robustness, reliability, and stability—are more important. The stability of an instrument is characterized by the Allan minimum, corresponding to the signal integration time that yields the lowest variance (*181*). Undesired effects, such as laser pointing noise (*118*), eddy currents in magnetic levitation (*21*), or technical noise (*120*), limit the stability of levitated sensors. Stability against long-term drifts is especially critical for low-frequency signal detection using high–quality factor levitated sensors. Technical noise is a particular obstacle for torque sensing with fast nanorotors because it linearly increases with rotational frequency and makes shot-noise–limited operation difficult (*91*). Other challenges include calibration uncertainties (e.g., particle mass) (*116*, *181*–*183*) and the temperature dependence of system parameters (e.g., index of refraction of particle). Although some of these noise sources can be suppressed by differential measurements, the potential of levitation-based sensors for commercial applications requires thorough evaluation and optimization procedures.

Enhancing sensitivity

The third important challenge pertains to boosting the sensitivity of levitation-based sensors, especially for uses in fundamental science (e.g., detection of weak forces or of quantum features). A fundamental sensitivity limit is reached when all thermal and technical noise sources are eliminated and the measurement becomes shot-noise limited. State estimation based on Kalman filtering (*50*, *184*) in combination with quantum control can be used to compensate for drifts and stabilize operation. Stretching the measurement precision beyond the standard quantum limit will require implementation of quantum-enhanced sensing protocols (e.g., squeezed states of light and/or motion). In the future, completely new sensing strategies will open up once it is possible to harness the "wave properties" of levitated particles and to capitalize on progress achieved in the field of matter-wave interferometry.

Decoherence

Levitated objects are macroscopic and thus very sensitive to quantum decoherence (Fig. 4). To perform quantum experiments, strategies to mitigate or circumvent such decoherence must be developed. Decoherence due to collisions with molecules of the surrounding gas can be suppressed by operating in ultrahigh vacuum, such that the probability to scatter a single gas molecule in each experimental run becomes negligible. The use of laser light also generates motional decoherence in two ways: by direct scattering of photons and by raising the internal energy of the particle through absorption, thereby increasing the emission rate and frequency of the thermal electromagnetic field radiated by the particle. In both cases, the photons (scattered and thermally emitted, respectively) carry away information about the center-of-mass position and orientation of the particle, thus generating decoherence. Controlling levitated particles without light does not necessarily reduce decoherence, as it requires either highly charged particles, which are more sensitive to stray electric fields, or magnetic or superconducting particles levitated using magnetic fields, which typically need on-chip schemes that are prone to surface-induced decoherence. The decoherence challenge is amplified as soon as the extension of the particle wave function grows, a required step toward exploration of macroscopic quantum physics. In addition, adding nonlinearities to the quantum motional control toolbox of a levitated particle, as required for exploration of non-Gaussian physics, typically involves either strong coupling to a two-level quantum system or the use of strongly nonharmonic potentials. In both cases, a levitated particle must be placed in a region of large field gradients, which unavoidably require the proximity of the particle to other materials and are hence accompanied by surface-induced decoherence. To tackle these challenges, quantum experiments with levitated particles, and particularly those exploring macroscopic quantum physics, will have to involve fast experimental runs and minimal use of laser light while avoiding large surfaces and time-dependent electromagnetic fields. A deeper understanding of decoherence in levitodynamics, attainable by the progress of experiments and theory, will most likely lead to innovative ideas for controlling levitated particles in the quantum regime.

Repeatability

In addition to minimizing decoherence, quantum experiments must be robust and repeatable. A quantum tomography protocol involves (i) preparation of the initial state, (ii) state evolution, and (iii) measurement of the final state. This sequence must be repeated many times, with the same particle and under identical conditions, to extract the underlying quantum statistics. Because protocols likely involve sequences of free fall and changes in trapping potential, it is difficult to retrap the particle and bring it back to its initial state. Optimal control protocols, methods for low-noise measurements and filtering, and ultrafast data acquisition and processing techniques will thus be needed.

Scalability

Scaling up the size and number of levitated particles (to enhance inertial sensitivities or to study many-particle physics, respectively) also poses several challenges. The more massive the particle, the more difficult it is to levitate

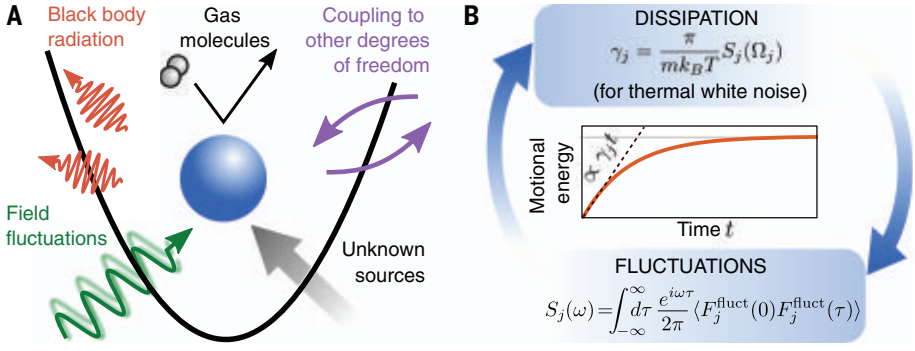


Fig. 4. Noise and decoherence in levitodynamics. (**A**) Uncontrolled interactions with the environment and noise in the trapping potentials generate fluctuations and dissipation in particle motion. In the quantum regime, these generate decoherence on the motional quantum state by providing information to the environment about the position of the particle. The main noise sources in levitodynamics are (i) collisions with molecules of the surrounding gas; (ii) emission, absorption, and scattering of thermal electromagnetic radiation; (iii) noise in the trap center from vibrations and/or stray fields; (iv) noise in the trap frequency from fluctuations of the electromagnetic fields that generate the trapping potentials (e.g., laser intensity noise in optical tweezers); and (v) cross-coupling to other thermal degrees of freedom in the particle—either motion along orthogonal directions, rotation, or internal excitations (e.g., acoustic phonons). (**B**) Noise sources manifest as a damping and a fluctuating force (Box 1), related by the fluctuation-dissipation theorem. The energy of the trapped particle initially grows linearly in time until it thermalizes with the environment. For stochastic forces proportional to the mass of the particle, the heating rate is linear with the mass. S_j, power spectral density of the fluctuating force.

in a high-frequency trap, the stronger its coupling to the environment, and the smaller its zero-point motion. This introduces additional noise and hinders motional ground-state cooling and all related applications (e.g., metrology of impulses or transient signals, tests of quantum mechanics) (*120*). Levitating several particles is also challenging, owing to the difficulty of simultaneously multiplexing different traps (*185*) and controlling the resulting interactions between the particles contained in them [e.g., coupling that stems from optical binding (*186*)]. Once these complex traps are engineered, practical questions, such as those pertaining to controlled loading of several particles and the cooling required for applications, will have to be addressed. A step ahead in terms of complexity are hybrid multiplexed traps, which would enable levitation of particles of different sizes and types (e.g., dielectric, charged, magnetic, superconducting) for exploring, e.g., multispecies statistical mechanics, including complex (magnetic, electrostatic) interactions, or reconfigurable levitated particle lattices to model solid-state physics. Developing such hybrid multiplexed traps will likely require a combined effort toward hybrid trap design and on-chip integration.

Outlook

Levitodynamics brings together appealing features from different fields—examples include levitation techniques from atomic physics and biosciences, control techniques from atomic and molecular optics, the richness of solid-state physics, and the potential to couple motion to various degrees of freedom and/or external systems. It is the combination of all of these properties in a single platform that brings forth unique opportunities for science and technology. The path toward attractive research directions, from commercial sensing devices to fundamental quantum physics, has been unlocked by the many recent experimental achievements. Exploring the multiple future prospects for science and technology is a multidisciplinary endeavor, which will establish levitated objects in high vacuum as highly controllable nanolabs for use in areas as diverse as nonequilibrium physics, condensed matter physics, particle physics, and the foundations of quantum mechanics.

REFERENCES AND NOTES

1. E. H. Brandt, Levitation in physics. *Science* **243**, 349–355 (1989). doi: 10.1126/science.243.4889.349; pmid: 17787252
2. J. M. Goodkind, The superconducting gravimeter. *Rev. Sci. Instrum.* **70**, 4131–4152 (1999). doi: 10.1063/1.1150092
3. S. Arnold, in *Optical Effects Associated with Small Particles* (World Scientific, 1988), pp. 63–137.
4. S. Chu, J. E. Bjorkholm, A. Ashkin, A. Cable, Experimental observation of optically trapped atoms. *Phys. Rev. Lett.* **57**, 314–317 (1986). doi: 10.1103/PhysRevLett.57.314; pmid: 10034028
5. K. Frye *et al*., The Bose-Einstein Condensate and Cold Atom Laboratory. *EPJ Quantum Technol.* **8**, 1 (2021). doi: 10.1140/epjqt/s40507-020-00090-8
6. W. D. Phillips, Nobel Lecture: Laser cooling and trapping of neutral atoms. *Rev. Mod. Phys.* **70**, 721–741 (1998). doi: 10.1103/RevModPhys.70.721
7. C. J. Bustamante, Y. R. Chemla, S. Liu, M. D. Wang, Optical tweezers in single-molecule biophysics. *Nat. Rev. Methods Primers* **1**, 25 (2021). doi: 10.1038/s43586-021-00021-6
8. K. Dholakia, T. Čižmár, Shaping the future of manipulation. *Nat. Photonics* **5**, 335–342 (2011). doi: 10.1038/nphoton.2011.80
9. O. Reich, G. David, K. Esat, R. Signorell, Weighing picogram aerosol droplets with an optical balance. *Commun. Phys.* **3**, 223 (2020). doi: 10.1038/s42005-020-00496-x
10. A. Ashkin, Acceleration and Trapping of Particles by Radiation Pressure. *Phys. Rev. Lett.* **24**, 156–159 (1970). doi: 10.1103/PhysRevLett.24.156
11. A. Ashkin, J. Dziedzic, Optical Levitation by Radiation Pressure. *Appl. Phys. Lett.* **19**, 283–285 (1971). doi: 10.1063/1.1653919
12. A. Ashkin, J. Dziedzic, Optical levitation in high vacuum. *Appl. Phys. Lett.* **28**, 333–335 (1976). doi: 10.1063/1.88748
13. O. Romero-Isart, M. L. Juan, R. Quidant, J. I. Cirac, Toward quantum superposition of living organisms. *New J. Phys.* **12**, 033015 (2010). doi: 10.1088/1367-2630/12/3/033015
14. D. E. Chang *et al*., Cavity opto-mechanics using an optically levitated nanosphere. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1005–1010 (2010). doi: 10.1073/pnas.0912969107; pmid: 20080573
15. P. F. Barker, M. N. Shneider, Cavity cooling of an optically trapped nanoparticle. *Phys. Rev. A* **81**, 023826 (2010). doi: 10.1103/PhysRevA.81.023826
16. O. Romero-Isart *et al*., Optically levitating dielectrics in the quantum regime: Theory and protocols. *Phys. Rev. A* **83**, 013803 (2011). doi: 10.1103/PhysRevA.83.013803
17. J. Millen, T. Deesuwan, P. Barker, J. Anders, Nanoscale temperature measurements using non-equilibrium Brownian dynamics of a levitated nanosphere. *Nat. Nanotechnol.* **9**, 425–429 (2014). doi: 10.1038/nnano.2014.82; pmid: 24793558
18. A. T. M. A. Rahman *et al*., Burning and graphitization of optically levitated nanodiamonds in vacuum. *Sci. Rep.* **6**, 21633 (2016). doi: 10.1038/s41598-016-0001-8; pmid: 28442746
19. J. Gieseler *et al*., Single-Spin Magnetomechanics with Levitated Micromagnets. *Phys. Rev. Lett.* **124**, 163604 (2020). doi: 10.1103/PhysRevLett.124.163604; pmid: 32383959
20. A. Vinante *et al*., Ultralow Mechanical Damping with Meissner-Levitated Ferromagnetic Microparticles. *Phys. Rev. Appl.* **13**, 064027 (2020). doi: 10.1103/PhysRevApplied.13.064027
21. C. Timberlake, G. Gasbarri, A. Vinante, A. Setter, H. Ulbricht, Acceleration sensing with magnetically levitated oscillators above a superconductor. *Appl. Phys. Lett.* **115**, 224101 (2019). doi: 10.1063/1.5129145
22. T. Wang *et al*., Dynamics of a Ferromagnetic Particle Levitated over a Superconductor. *Phys. Rev. Appl.* **11**, 044041 (2019). doi: 10.1103/PhysRevApplied.11.044041
23. A. W. Schell, A. Kuhlicke, G. Kewes, O. Benson, "Flying Plasmons": Fabry-Pérot Resonances in Levitated Silver Nanowires. *ACS Photonics* **4**, 2719–2725 (2017). doi: 10.1021/acsphotonics.7b00526
24. L. P. Neukirch, E. Von Haartman, J. M. Rosenholm, A. N. Vamivakas, Multi-dimensional single-spin nano-optomechanics with a levitated nanodiamond. *Nat. Photonics* **9**, 653–657 (2015). doi: 10.1038/nphoton.2015.162
25. T. M. Hoang, J. Ahn, J. Bang, T. Li, Electron spin control of optically levitated nanodiamonds in vacuum. *Nat. Commun.* **7**, 12250 (2016). doi: 10.1038/ncomms12250; pmid: 27432560
26. G. P. Conangla, A. W. Schell, R. A. Rica, R. Quidant, Motion Control and Optical Interrogation of a Levitating Single Nitrogen Vacancy in Vacuum. *Nano Lett.* **18**, 3956–3961 (2018). doi: 10.1021/acs.nanolett.8b01414; pmid: 29772171
27. T. Delord *et al*., Ramsey Interferences and Spin Echoes from Electron Spins Inside a Levitating Macroscopic Particle. *Phys. Rev. Lett.* **121**, 053602 (2018). doi: 10.1103/PhysRevLett.121.053602; pmid: 30118282
28. R. J. A. Hill, L. Eaves, Nonaxisymmetric shapes of a magnetically levitated and spinning water droplet. *Phys. Rev. Lett.* **101**, 234501 (2008). doi: 10.1103/PhysRevLett.101.234501; pmid: 19113558
29. P. Nagornykh, J. E. Coppock, J. P. J. Murphy, B. E. Kane, Optical and magnetic measurements of gyroscopically stabilized graphene nanoplatelets levitated in an ion trap. *Phys. Rev. B* **96**, 035402 (2017). doi: 10.1103/PhysRevB.96.035402
30. C. D. Brown *et al*., Characterization of Levitated Superfluid Helium Drops in High Vacuum. arXiv:2109.05618 [cond-mat.other] (2021).
31. P. Nagornykh, J. E. Coppock, B. Kane, Cooling of levitated graphene nanoplatelets in high vacuum. *Appl. Phys. Lett.* **106**, 244102 (2015). doi: 10.1063/1.4922705
32. S. Kuhn *et al*., Cavity-Assisted Manipulation of Freely Rotating Silicon Nanorods in High Vacuum. *Nano Lett.* **15**, 5604–5608 (2015). doi: 10.1021/acs.nanolett.5b02302; pmid: 26167662
33. D. S. Bykov, P. Mestres, L. Dania, L. Schmöger, T. E. Northup, Direct loading of nanoparticles under high vacuum into a Paul trap for levitodynamical experiments. *Appl. Phys. Lett.* **115**, 034101 (2019). doi: 10.1063/1.5109645
34. J. Flajšmanová *et al*., Using the transient trajectories of an optically levitated nanoparticle to characterize a stochastic Duffing oscillator. *Sci. Rep.* **10**, 14436 (2020). doi: 10.1038/s41598-020-70908-z; pmid: 32879371
35. P. Mestres *et al*., Cooling and manipulation of a levitated nanoparticle with an optical fiber trap. *Appl. Phys. Lett.* **107**, 151102 (2015). doi: 10.1063/1.4933180
36. M. Calamai, A. Ranfagni, F. Marin, Transfer of a levitating nanoparticle between optical tweezers. *AIP Adv.* **11**, 025246 (2021). doi: 10.1063/5.0024432
37. T. Li, S. Kheifets, M. G. Raizen, Millikelvin cooling of an optically trapped microsphere in vacuum. *Nat. Phys.* **7**, 527–530 (2011). doi: 10.1038/nphys1952
38. J. Gieseler, B. Deutsch, R. Quidant, L. Novotny, Subkelvin parametric feedback cooling of a laser-trapped nanoparticle. *Phys. Rev. Lett.* **109**, 103603 (2012). doi: 10.1103/PhysRevLett.109.103603; pmid: 23005289
39. J.-F. Hsu, P. Ji, C. W. Lewandowski, B. D'Urso, Cooling the Motion of Diamond Nanocrystals in a Magneto-Gravitational Trap in High Vacuum. *Sci. Rep.* **6**, 30125 (2016). doi: 10.1038/s41598-016-0001-8; pmid: 28442746
40. B. R. Slezak, C. W. Lewandowski, J.-F. Hsu, B. D'Urso, Cooling the motion of a silica microsphere in a magneto-gravitational trap in ultra-high vacuum. *New J. Phys.* **20**, 063028 (2018). doi: 10.1088/1367-2630/aacac1
41. D. Goldwater *et al*., Levitated electromechanics: All-electrical cooling of charged nano- and micro-particles. *Quantum Sci. Technol.* **4**, 024003 (2019). doi: 10.1088/2058-9565/aaf5f3
42. L. Dania, D. S. Bykov, M. Knoll, P. Mestres, T. E. Northup, Optical and electrical feedback cooling of a silica nanoparticle levitated in a Paul trap. *Phys. Rev. Res.* **3**, 013018 (2021). doi: 10.1103/PhysRevResearch.3.013018
43. V. Jain *et al*., Direct Measurement of Photon Recoil from a Levitated Nanoparticle. *Phys. Rev. Lett.* **116**, 243601 (2016). doi: 10.1103/PhysRevLett.116.243601; pmid: 27367388
44. M. Kamba, H. Kiuchi, T. Yotsuya, K. Aikawa, Recoil-limited feedback cooling of single nanoparticles near the ground state in an optical lattice. *Phys. Rev. A* **103**, L051701 (2021). doi: 10.1103/PhysRevA.103.L051701
45. F. Tebbenjohanns, M. Frimmer, A. Militaru, V. Jain, L. Novotny, Cold Damping of an Optically Levitated Nanoparticle to Microkelvin Temperatures. *Phys. Rev. Lett.* **122**, 223601 (2019). doi: 10.1103/PhysRevLett.122.223601; pmid: 31283294
46. G. P. Conangla *et al*., Optimal Feedback Cooling of a Charged Levitated Nanoparticle with Adaptive Control. *Phys. Rev. Lett.* **122**, 223602 (2019). doi: 10.1103/PhysRevLett.122.223602; pmid: 31283263
47. M. Iwasaki *et al*., Electric feedback cooling of single charged nanoparticles in an optical trap. *Phys. Rev. A* **99**, 051401 (2019). doi: 10.1103/PhysRevA.99.051401
48. F. Tebbenjohanns, M. Frimmer, L. Novotny, Optimal position detection of a dipolar scatterer in a focused field. *Phys. Rev. A* **100**, 043821 (2019). doi: 10.1103/PhysRevA.100.043821
49. F. Tebbenjohanns, M. Frimmer, V. Jain, D. Windey, L. Novotny, Motional Sideband Asymmetry of a Nanoparticle Optically Levitated in Free Space. *Phys. Rev. Lett.* **124**, 013603 (2020). doi: 10.1103/PhysRevLett.124.013603; pmid: 31976693
50. L. Magrini *et al*., Real-time optimal quantum control of mechanical motion at room temperature. *Nature* **595**, 373–377 (2021). doi: 10.1038/s41586-021-03602-3; pmid: 34262213
51. F. Tebbenjohanns, M. L. Mattana, M. Rossi, M. Frimmer, L. Novotny, Quantum control of a nanoparticle optically levitated in cryogenic free space. *Nature* **595**, 378–382 (2021). doi: 10.1038/s41586-021-03617-w; pmid: 34262214
52. N. Kiesel *et al*., Cavity cooling of an optically levitated submicron particle. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14180–14185 (2013). doi: 10.1073/pnas.1309167110; pmid: 23940352
53. J. Millen, P. Z. G. Fonseca, T. Mavrogordatos, T. S. Monteiro, P. F. Barker, Cavity cooling a single charged levitated nanosphere. *Phys. Rev. Lett.* **114**, 123602 (2015). doi: 10.1103/PhysRevLett.114.123602; pmid: 25860743

54. N. Meyer *et al*., Resolved-Sideband Cooling of a Levitated Nanoparticle in the Presence of Laser Phase Noise. *Phys. Rev. Lett.* **123**, 153601 (2019). doi: 10.1103/PhysRevLett.123.153601; pmid: 31702279
55. P. Horak, G. Hechenblaikner, K. M. Gheri, H. Stecher, H. Ritsch, Cavity-Induced Atom Cooling in the Strong Coupling Regime. *Phys. Rev. Lett.* **79**, 4974–4977 (1997). doi: 10.1103/PhysRevLett.79.4974
56. V. Vuletić, S. Chu, Laser cooling of atoms, ions, or molecules by coherent scattering. *Phys. Rev. Lett.* **84**, 3787–3790 (2000). doi: 10.1103/PhysRevLett.84.3787; pmid: 11019206
57. D. Windey *et al*., Cavity-Based 3D Cooling of a Levitated Nanoparticle via Coherent Scattering. *Phys. Rev. Lett.* **122**, 123601 (2019). doi: 10.1103/PhysRevLett.122.123601; pmid: 30978044
58. U. Delić *et al*., Cavity Cooling of a Levitated Nanosphere by Coherent Scattering. *Phys. Rev. Lett.* **122**, 123602 (2019). doi: 10.1103/PhysRevLett.122.123602; pmid: 30978033
59. C. Gonzalez-Ballestero *et al*., Theory for cavity cooling of levitated nanoparticles via coherent scattering: Master equation approach. *Phys. Rev. A* **100**, 013805 (2019). doi: 10.1103/PhysRevA.100.013805
60. U. Delić *et al*., Cooling of a levitated nanoparticle to the motional quantum ground state. *Science* **367**, 892–895 (2020). doi: 10.1126/science.aba3993; pmid: 32001522
61. F. Ricci *et al*., Optically levitated nanoparticle as a model system for stochastic bistable dynamics. *Nat. Commun.* **8**, 15141 (2017). doi: 10.1038/ncomms15141; pmid: 28485372
62. Y. Arita, M. Chen, E. M. Wright, K. Dholakia, Dynamics of a levitated microparticle in vacuum trapped by a perfect vortex beam: Three-dimensional motion around a complex optical potential. *J. Opt. Soc. Am. B* **34**, C14 (2017). doi: 10.1364/JOSAB.34.000C14
63. L.-M. Zhou, K.-W. Xiao, J. Chen, N. Zhao, Optical levitation of nanodiamonds by doughnut beams in vacuum. *Laser Photonics Rev.* **11**, 1600284 (2017). doi: 10.1002/lpor.201600284
64. L. Rondin *et al*., Direct measurement of Kramers turnover with a levitated nanoparticle. *Nat. Nanotechnol.* **12**, 1130–1133 (2017). doi: 10.1038/nnano.2017.198; pmid: 29209016
65. P. Z. G. Fonseca, E. B. Aranas, J. Millen, T. S. Monteiro, P. F. Barker, Nonlinear Dynamics and Strong Cavity Cooling of Levitated Nanoparticles. *Phys. Rev. Lett.* **117**, 173602 (2016). doi: 10.1103/PhysRevLett.117.173602; pmid: 27824467
66. G. P. Conangla, R. A. Rica, R. Quidant, Extending Vacuum Trapping to Absorbing Objects with Hybrid Paul-Optical Traps. *Nano Lett.* **20**, 6018–6023 (2020). doi: 10.1021/acs.nanolett.0c02025; pmid: 32692184
67. M. Rashid *et al*., Experimental Realization of a Thermal Squeezed State of Levitated Optomechanics. *Phys. Rev. Lett.* **117**, 273601 (2016). doi: 10.1103/PhysRevLett.117.273601; pmid: 28084746
68. E. Hebestreit, M. Frimmer, R. Reimann, L. Novotny, Sensing Static Forces with Free-Falling Nanoparticles. *Phys. Rev. Lett.* **121**, 063602 (2018). doi: 10.1103/PhysRevLett.121.063602; pmid: 30141659
69. M. A. Ciampini *et al*., Experimental nonequilibrium memory erasure beyond Landauer's bound. arXiv:2107.04429 [cond-mat.stat-mech] (2021).
70. S. Kuhn *et al*., Full rotational control of levitated silicon nanorods. *Optica* **4**, 356 (2017). doi: 10.1364/OPTICA.4.000356
71. L. Neumeier, R. Quidant, D. E. Chang, Self-induced back-action optical trapping in nanophotonic systems. *New J. Phys.* **17**, 123008 (2015). doi: 10.1088/1367-2630/17/12/123008
72. U. Delić *et al*., Levitated cavity optomechanics in high vacuum. *Quantum Sci. Technol.* **5**, 025006 (2020). doi: 10.1088/2058-9565/ab7989
73. T. Čižmár, V. Garcés-Chávez, K. Dholakia, P. Zemánek, Optical conveyor belt for delivery of submicron objects. *Appl. Phys. Lett.* **86**, 174101 (2005). doi: 10.1063/1.1915543
74. O. Brzobohatý *et al*., Experimental demonstration of optical transport, sorting and self-arrangement using a 'tractor beam'. *Nat. Photonics* **7**, 123–127 (2013). doi: 10.1038/nphoton.2012.332
75. D. Grass, J. Fesel, S. G. Hofer, N. Kiesel, M. Aspelmeyer, Optical trapping and control of nanoparticles inside evacuated hollow core photonic crystal fibers. *Appl. Phys. Lett.* **108**, 221103 (2016). doi: 10.1063/1.4953025
76. T. M. Ostermayr *et al*., A transportable Paul-trap for levitation and accurate positioning of micron-scale particles in vacuum for laser-plasma experiments. *Rev. Sci. Instrum.* **89**, 013302 (2018). doi: 10.1063/1.4995955; pmid: 29390683
77. M. Perdriat, P. Huillery, C. Pellet-Mary, G. Hétet, Angle Locking of a Levitating Diamond using Spin-Diamagnetism. arXiv:2102.13637 [quant-ph] (2021).
78. Y. Arita, M. Mazilu, K. Dholakia, Laser-induced rotation and cooling of a trapped microgyroscope in vacuum. *Nat. Commun.* **4**, 2374 (2013). doi: 10.1038/ncomms3374; pmid: 23982323
79. F. Monteiro, S. Ghosh, E. C. van Assendelft, D. C. Moore, Optical rotation of levitated spheres in high vacuum. *Phys. Rev. A* **97**, 051802 (2018). doi: 10.1103/PhysRevA.97.051802
80. S. Kuhn *et al*., Optically driven ultra-stable nanomechanical rotor. *Nat. Commun.* **8**, 1670 (2017). doi: 10.1038/s41467-016-0009-6; pmid: 28232747
81. M. Rashid, M. Toroš, A. Setter, H. Ulbricht, Precession Motion in Levitated Optomechanics. *Phys. Rev. Lett.* **121**, 253601 (2018). doi: 10.1103/PhysRevLett.121.253601; pmid: 30608788
82. T. M. Hoang *et al*., Torsional Optomechanics of a Levitated Nonspherical Nanoparticle. *Phys. Rev. Lett.* **117**, 123604 (2016). doi: 10.1103/PhysRevLett.117.123604; pmid: 27689273
83. R. Reimann *et al*., GHz Rotation of an Optically Trapped Nanoparticle in Vacuum. *Phys. Rev. Lett.* **121**, 033602 (2018). doi: 10.1103/PhysRevLett.121.033602; pmid: 30085794
84. J. Ahn *et al*., Optically Levitated Nanodumbbell Torsion Balance and GHz Nanomechanical Rotor. *Phys. Rev. Lett.* **121**, 033603 (2018). doi: 10.1103/PhysRevLett.121.033603; pmid: 30085795
85. J. Ahn *et al*., Ultrasensitive torque detection with an optically levitated nanorotor. *Nat. Nanotechnol.* **15**, 89–93 (2020). doi: 10.1038/s41565-019-0605-9; pmid: 31932762
86. T. Delord, P. Huillery, L. Nicolas, G. Hétet, Spin-cooling of the motion of a trapped diamond. *Nature* **580**, 56–59 (2020). doi: 10.1038/s41586-020-2133-z; pmid: 32238930
87. J. Bang *et al*., Five-dimensional cooling and nonlinear dynamics of an optically levitated nanodumbbell. *Phys. Rev. Res.* **2**, 043054 (2020). doi: 10.1103/PhysRevResearch.2.043054
88. F. van der Laan *et al*., Sub-Kelvin feedback cooling and heating dynamics of an optically levitated librator. *Phys. Rev. Lett.* **127**, 123605 (2021). doi: 10.1103/PhysRevLett.127.123605
89. J. Schäfer, H. Rudolph, K. Hornberger, B. A. Stickler, Cooling Nanorotors by Elliptic Coherent Scattering. *Phys. Rev. Lett.* **126**, 163603 (2021). doi: 10.1103/PhysRevLett.126.163603; pmid: 33961470
90. B. A. Stickler, B. Schrinski, K. Hornberger, Rotational Friction and Diffusion of Quantum Rotors. *Phys. Rev. Lett.* **121**, 040401 (2018). doi: 10.1103/PhysRevLett.121.040401; pmid: 30095961
91. F. van der Laan *et al*., Optically levitated rotor at its thermal limit of frequency stability. *Phys. Rev. A* **102**, 013505 (2020). doi: 10.1103/PhysRevA.102.013505
92. J. Prat-Camps, C. Teo, C. C. Rusconi, W. Wieczorek, O. Romero-Isart, Ultrasensitive Inertial and Force Sensors with Diamagnetically Levitated Magnets. *Phys. Rev. Appl.* **8**, 034002 (2017). doi: 10.1103/PhysRevApplied.8.034002
93. C. Pellet-Mary, P. Huillery, M. Perdriat, G. Hétet, Magnetic-Torque Enhanced by Tunable Dipolar interactions. arXiv:2103.00836 [cond-mat.mes-hall] (2021).
94. A. Dechant, N. Kiesel, E. Lutz, All-optical nanomechanical heat engine. *Phys. Rev. Lett.* **114**, 183602 (2015). doi: 10.1103/PhysRevLett.114.183602; pmid: 26001001
95. T. Li, S. Kheifets, D. Medellin, M. G. Raizen, Measurement of the instantaneous velocity of a Brownian particle. *Science* **328**, 1673–1675 (2010). doi: 10.1126/science.1189403; pmid: 20488989
96. J. Gieseler, M. Spasenović, L. Novotny, R. Quidant, Nonlinear mode coupling and synchronization of a vacuum-trapped nanoparticle. *Phys. Rev. Lett.* **112**, 103603 (2014). doi: 10.1103/PhysRevLett.112.103603; pmid: 24679293
97. J. Gieseler, R. Quidant, C. Dellago, L. Novotny, Dynamic relaxation of a levitated nanoparticle from a non-equilibrium steady state. *Nat. Nanotechnol.* **9**, 358–364 (2014). doi: 10.1038/nnano.2014.40; pmid: 24681775
98. T. M. Hoang *et al*., Experimental Test of the Differential Fluctuation Theorem and a Generalized Jarzynski Equality for Arbitrary Initial States. *Phys. Rev. Lett.* **120**, 080602 (2018). doi: 10.1103/PhysRevLett.120.080602; pmid: 29542995
99. J. Gieseler, J. Millen, Levitated Nanoparticles for Microscopic Thermodynamics-A Review. *Entropy* **20**, 326 (2018). doi: 10.3390/e20050326; pmid: 33265416
100. M. Rademacher, M. Konopik, M. Debiossac, D. Grass, E. Lutz, N. Kiesel arXiv:2103.10898 [cond-mat.stat-mech] (2021).
101. M. Debiossac, D. Grass, J. J. Alonso, E. Lutz, N. Kiesel, Thermodynamics of continuous non-Markovian feedback control. *Nat. Commun.* **11**, 1360 (2020). doi: 10.1038/s41467-020-15148-5; pmid: 32170062
102. A. Dechant, N. Kiesel, E. Lutz, Underdamped stochastic heat engine at maximum efficiency. *Europhys. Lett.* **119**, 50003 (2017). doi: 10.1209/0295-5075/119/50003
103. G. S. Agarwal, S. Chaturvedi, Quantum dynamical framework for Brownian heat engines. *Phys. Rev. E* **88**, 012130 (2013). doi: 10.1103/PhysRevE.88.012130; pmid: 23944437
104. E. Hebestreit, R. Reimann, M. Frimmer, L. Novotny, Measuring the internal temperature of a levitated nanoparticle in high vacuum. *Phys. Rev. A* **97**, 043803 (2018). doi: 10.1103/PhysRevA.97.043803
105. T. Delord, L. Nicolas, M. Bodini, G. Hétet, Diamonds levitating in a Paul trap under vacuum: Measurements of laser-induced heating via NV center thermometry. *Appl. Phys. Lett.* **111**, 013101 (2017). doi: 10.1063/1.4991670
106. A. A. Rahman, P. Barker, Laser refrigeration, alignment and rotation of levitated Yb3+:YLF nanocrystals. *Nat. Photonics* **11**, 634–638 (2017). doi: 10.1038/s41566-017-0005-3
107. M. Schuck, D. Steinert, T. Nussbaumer, J. W. Kolar, Ultrafast rotation of magnetically levitated macroscopic steel spheres. *Sci. Adv.* **4**, e1701519 (2018). doi: 10.1126/sciadv.1701519; pmid: 29326976
108. L. Childress *et al*., Cavity optomechanics in a levitated helium drop. *Phys. Rev. A* **96**, 063842 (2017). doi: 10.1103/PhysRevA.96.063842
109. B. Langbehn *et al*., Three-Dimensional Shapes of Spinning Helium Nanodroplets. *Phys. Rev. Lett.* **121**, 255301 (2018). doi: 10.1103/PhysRevLett.121.255301; pmid: 30608832
110. D. C. Moore, A. D. Rider, G. Gratta, Search for millicharged particles using optically levitated microspheres. *Phys. Rev. Lett.* **113**, 251801 (2014). doi: 10.1103/PhysRevLett.113.251801; pmid: 25554874
111. G. Ranjit, M. Cunningham, K. Casey, A. A. Geraci, Zeptonewton force sensing with nanospheres in an optical lattice. *Phys. Rev. A* **93**, 053801 (2016). doi: 10.1103/PhysRevA.93.053801
112. D. Hempston *et al*., Force sensing with an optically levitated charged nanoparticle. *Appl. Phys. Lett.* **111**, 133111 (2017). doi: 10.1063/1.4993555
113. J. Gieseler, L. Novotny, R. Quidant, Thermal nonlinearities in a nanomechanical oscillator. *Nat. Phys.* **9**, 806–810 (2013). doi: 10.1038/nphys2798
114. A. D. Rider *et al*., Search for Screened Interactions Associated with Dark Energy below the 100 μm Length Scale. *Phys. Rev. Lett.* **117**, 101101 (2016). doi: 10.1103/PhysRevLett.117.101101; pmid: 27636465
115. M. Frimmer *et al*., Controlling the net charge on a nanoparticle optically levitated in vacuum. *Phys. Rev. A* **95**, 061801 (2017). doi: 10.1103/PhysRevA.95.061801
116. F. Ricci, M. T. Cuairan, G. P. Conangla, A. W. Schell, R. Quidant, Accurate Mass Measurement of a Levitated Nanomechanical Resonator for Precision Force-Sensing. *Nano Lett.* **19**, 6711–6715 (2019). doi: 10.1021/acs.nanolett.9b00082; pmid: 30888180
117. D. F. Jackson Kimball, A. O. Sushkov, D. Budker, Precessing Ferromagnetic Needle Magnetometer. *Phys. Rev. Lett.* **116**, 190801 (2016). doi: 10.1103/PhysRevLett.116.190801; pmid: 27232012
118. F. Monteiro, S. Ghosh, A. G. Fine, D. C. Moore, Optical levitation of 10-ng spheres with nano- g acceleration sensitivity. *Phys. Rev. A* **96**, 063841 (2017). doi: 10.1103/PhysRevA.96.063841
119. A. D. Rider, C. P. Blakemore, G. Gratta, D. C. Moore, Single-beam dielectric-microsphere trapping with optical heterodyne detection. *Phys. Rev. A* **97**, 013842 (2018). doi: 10.1103/PhysRevA.97.013842
120. F. Monteiro *et al*., Force and acceleration sensing with optically levitated nanogram masses at microkelvin temperatures. *Phys. Rev. A* **101**, 053835 (2020). doi: 10.1103/PhysRevA.101.053835
121. O. Romero-Isart *et al*., Large quantum superpositions and interference of massive nanometer-sized objects. *Phys. Rev. Lett.* **107**, 020405 (2011). doi: 10.1103/PhysRevLett.107.020405; pmid: 21797585
122. O. Romero-Isart, Quantum superposition of massive objects and collapse models. *Phys. Rev. A* **84**, 052121 (2011). doi: 10.1103/PhysRevA.84.052121
123. R. Kaltenbaek *et al*., Macroscopic quantum resonators (MAQRO). *Exp. Astron.* **34**, 123–164 (2012). doi: 10.1007/s10686-012-9292-3

124. J. Bateman, S. Nimmrichter, K. Hornberger, H. Ulbricht, Near-field interferometry of a free-falling nanoparticle from a point-like source. *Nat. Commun.* **5**, 4788 (2014). doi: 10.1038/ncomms5788; pmid: 25179560
125. A. Geraci, H. Goldman, Sensing short range forces with a nanosphere matter-wave interferometer. *Phys. Rev. D* **92**, 062002 (2015). doi: 10.1103/PhysRevD.92.062002
126. C. Wan *et al*., Free Nano-Object Ramsey Interferometry for Large Quantum Superpositions. *Phys. Rev. Lett.* **117**, 143003 (2016). doi: 10.1103/PhysRevLett.117.143003; pmid: 27740804
127. H. Pino, J. Prat-Camps, K. Sinha, B. P. Venkatesh, O. Romero-Isart, On-chip quantum interference of a superconducting microsphere. *Quantum Sci. Technol.* **3**, 025001 (2018). doi: 10.1088/2058-9565/aa9d15
128. I. Alda, J. Berthelot, R. A. Rica, R. Quidant, Trapping and manipulation of individual nanoparticles in a planar Paul trap. *Appl. Phys. Lett.* **109**, 163105 (2016). doi: 10.1063/1.4965859
129. L. Magrini *et al*., Near-field coupling of a levitated nanoparticle to a photonic crystal cavity. *Optica* **5**, 1597 (2018). doi: 10.1364/OPTICA.5.001597
130. K. Shen *et al*., On-chip optical levitation with a metalens in vacuum. arXiv:2107.09659 [physics.optics] (2021).
131. R. Diehl *et al*., Optical levitation and feedback cooling of a nanoparticle at subwavelength distances from a membrane. *Phys. Rev. A* **98**, 013851 (2018). doi: 10.1103/PhysRevA.98.013851
132. S. Kuhn *et al*., Nanoparticle detection in an open-access silicon microcavity. *Appl. Phys. Lett.* **111**, 253107 (2017). doi: 10.1063/1.5008492
133. G. Winstone *et al*., Direct measurement of the electrostatic image force of a levitated charged nanoparticle close to a surface. *Phys. Rev. A* **98**, 053831 (2018). doi: 10.1103/PhysRevA.98.053831
134. C. Montoya *et al*., Scanning force sensing at μm-distances from a conductive surface with nanospheres in an optical lattice. arXiv:2103.03420 [physics.optics] (2021).
135. M. Brownnutt, M. Kumph, P. Rabl, R. Blatt, Ion-trap measurements of electric-field noise near surfaces. *Rev. Mod. Phys.* **87**, 1419–1482 (2015). doi: 10.1103/RevModPhys.87.1419
136. A. Manjavacas, F. J. García de Abajo, Thermal and vacuum friction acting on rotating particles. *Phys. Rev. A* **82**, 063827 (2010). doi: 10.1103/PhysRevA.82.063827
137. R. Zhao, A. Manjavacas, F. J. García de Abajo, J. B. Pendry, Rotational quantum friction. *Phys. Rev. Lett.* **109**, 123604 (2012). doi: 10.1103/PhysRevLett.109.123604; pmid: 23005949
138. T. Westphal, H. Hepach, J. Pfaff, M. Aspelmeyer, Measurement of gravitational coupling between millimetre-sized masses. *Nature* **591**, 225–228 (2021). doi: 10.1038/s41586-021-03250-7; pmid: 33692556
139. D. C. Moore, A. A. Geraci, Searching for new physics using optically levitated sensors. *Quantum Sci. Technol.* **6**, 014008 (2021). doi: 10.1088/2058-9565/abcf8a
140. A. A. Geraci, S. B. Papp, J. Kitching, Short-range force detection using optically cooled levitated microspheres. *Phys. Rev. Lett.* **105**, 101101 (2010). doi: 10.1103/PhysRevLett.105.101101; pmid: 20867507
141. G. Afek *et al*., Limits on the abundance of millicharged particles bound to matter. *Phys. Rev. D* **104**, 012004 (2021). doi: 10.1103/PhysRevD.104.012004
142. A. Kawasaki *et al*., High sensitivity, levitated microsphere apparatus for short-distance force measurements. *Rev. Sci. Instrum.* **91**, 083201 (2020). doi: 10.1063/5.0011759; pmid: 32872897
143. C. P. Blakemore *et al*., Search for non-Newtonian interactions at micrometer scale with a levitated test mass. arXiv:2102.06848 [hep-ex] (2021).
144. J. G. Lee, E. G. Adelberger, T. S. Cook, S. M. Fleischer, B. R. Heckel, New Test of the Gravitational 1/r^2 Law at Separations down to 52 μm. *Phys. Rev. Lett.* **124**, 101101 (2020). doi: 10.1103/PhysRevLett.124.101101; pmid: 32216404
145. N. Aggarwal *et al*., Searching for new physics with a levitated-sensor-based gravitational-wave detector. arXiv:2010.13157 [gr-qc] (2020).
146. J. Bateman, I. McHardy, A. Merle, T. R. Morris, H. Ulbricht, On the existence of low-mass dark matter and its direct detection. *Sci. Rep.* **5**, 8058 (2015). doi: 10.1038/srep08058; pmid: 25622565
147. F. Monteiro *et al*., Search for Composite Dark Matter with Optically Levitated Sensors. *Phys. Rev. Lett.* **125**, 181102 (2020). doi: 10.1103/PhysRevLett.125.181102; pmid: 33196261
148. D. Carney *et al*., Mechanical quantum sensing in the search for dark matter. *Quantum Sci. Technol.* **6**, 024002 (2021). doi: 10.1088/2058-9565/abcfcd
149. C. Marletto, V. Vedral, Gravitationally Induced Entanglement between Two Massive Particles is Sufficient Evidence of Quantum Effects in Gravity. *Phys. Rev. Lett.* **119**, 240402 (2017). doi: 10.1103/PhysRevLett.119.240402; pmid: 29286752
150. S. Bose *et al*., Spin Entanglement Witness for Quantum Gravity. *Phys. Rev. Lett.* **119**, 240401 (2017). doi: 10.1103/PhysRevLett.119.240401; pmid: 29286711
151. A. de Los Ríos Sommer, N. Meyer, R. Quidant, Strong optomechanical coupling at room temperature by coherent scattering. *Nat. Commun.* **12**, 276 (2021). doi: 10.1038/s41467-020-20419-2; pmid: 33436586
152. L. Martinetz, K. Hornberger, J. Millen, M. Kim, B. A. Stickler, Quantum electromechanics with levitated nanoparticles. *NPJ Quantum Inf.* **6**, 101 (2020). doi: 10.1038/s41534-020-00333-7
153. C. Gonzalez-Ballestero, J. Gieseler, O. Romero-Isart, Quantum Acoustomechanics with a Micromagnet. *Phys. Rev. Lett.* **124**, 093602 (2020). doi: 10.1103/PhysRevLett.124.093602; pmid: 32202851
154. Y. Y. Fein *et al*., Quantum superposition of molecules beyond 25 kDa. *Nat. Phys.* **15**, 1242–1245 (2019). doi: 10.1038/s41567-019-0663-9
155. B. A. Stickler *et al*., Probing macroscopic quantum superpositions with nanorotors. *New J. Phys.* **20**, 122001 (2018). doi: 10.1088/1367-2630/aaece4
156. A. Bassi, K. Lochan, S. Satin, T. P. Singh, H. Ulbricht, Models of wave-function collapse, underlying theories, and experimental tests. *Rev. Mod. Phys.* **85**, 471–527 (2013). doi: 10.1103/RevModPhys.85.471
157. C. M. DeWitt, D. Rickles, Eds., *The Role of Gravitation in Physics: Report from the 1957 Chapel Hill Conference*, vol. 5 (Edition Open Access, Max Planck Institute for the History of Science, 2011).
158. O. Romero-Isart, Coherent inflation for large quantum superpositions of levitated microspheres. *New J. Phys.* **19**, 123029 (2017). doi: 10.1088/1367-2630/aa99bf
159. T. Weiss, M. Roda-Llordes, E. Torrontegui, M. Aspelmeyer, O. Romero-Isart, Large Quantum Delocalization of a Levitated Nanoparticle Using Optimal Control: Applications for Force Sensing and Entangling via Weak Forces. *Phys. Rev. Lett.* **127**, 023601 (2021). doi: 10.1103/PhysRevLett.127.023601; pmid: 34296896
160. T. Weiss, O. Romero-Isart, Quantum motional state tomography with nonquadratic potentials and neural networks. *Phys. Rev. Res.* **1**, 033157 (2019). doi: 10.1103/PhysRevResearch.1.033157
161. K. Dholakia, P. Zemánek, *Colloquium*: Gripped by light: Optical binding. *Rev. Mod. Phys.* **82**, 1767–1791 (2010). doi: 10.1103/RevModPhys.82.1767
162. W. Lechner, S. J. M. Habraken, N. Kiesel, M. Aspelmeyer, P. Zoller, Cavity optomechanics of levitated nanodumbbells: Nonequilibrium phases and self-assembly. *Phys. Rev. Lett.* **110**, 143604 (2013). doi: 10.1103/PhysRevLett.110.143604; pmid: 25166988
163. D. R. Burnham, D. McGloin, Holographic optical trapping of aerosol droplets. *Opt. Express* **14**, 4175–4181 (2006). doi: 10.1364/OE.14.004175; pmid: 19516567
164. M. Gring *et al*., Relaxation and prethermalization in an isolated quantum system. *Science* **337**, 1318–1322 (2012). doi: 10.1126/science.1224953; pmid: 22956685
165. S. Liu, Z. Yin, T. Li, Prethermalization and Nonreciprocal Phonon Transport in a Levitated Optomechanical Array. *Adv. Quantum Technol.* **3**, 1900099 (2020). doi: 10.1002/qute.201900099
166. J. Millen, A. Xuereb, Perspective on quantum thermodynamics. *New J. Phys.* **18**, 011002 (2016). doi: 10.1088/1367-2630/18/1/011002
167. E. E. Wollman *et al*., Quantum squeezing of motion in a mechanical resonator. *Science* **349**, 952–955 (2015). doi: 10.1126/science.aac5138; pmid: 26315431
168. J.-M. Pirkkalainen, E. Damskägg, M. Brandt, F. Massel, M. A. Sillanpää, Squeezing of Quantum Noise of Motion in a Micromechanical Resonator. *Phys. Rev. Lett.* **115**, 243601 (2015). doi: 10.1103/PhysRevLett.115.243601; pmid: 26705631
169. F. Lecocq, J. B. Clark, R. W. Simmonds, J. Aumentado, J. D. Teufel, Quantum Nondemolition Measurement of a Nonclassical State of a Massive Object. *Phys. Rev. X* **5**, 041037 (2015). doi: 10.1103/PhysRevX.5.041037; pmid: 27057422
170. J. Roßnagel, O. Abah, F. Schmidt-Kaler, K. Singer, E. Lutz, Nanoscale heat engine beyond the Carnot limit. *Phys. Rev. Lett.* **112**, 030602 (2014). doi: 10.1103/PhysRevLett.112.030602; pmid: 24484127
171. J. Klaers, S. Faelt, A. Imamoglu, E. Togan, Squeezed Thermal Reservoirs as a Resource for a Nanomechanical Engine beyond the Carnot Limit. *Phys. Rev. X* **7**, 031044 (2017). doi: 10.1103/PhysRevX.7.031044
172. P. Talkner, P. Hänggi, Aspects of quantum work. *Phys. Rev. E* **93**, 022131 (2016). doi: 10.1103/PhysRevE.93.022131; pmid: 26986312
173. F. Ricci *et al*., A chemical nano-reactor based on a levitated nanoparticle in vacuum. arXiv:2107.01084 [physics.chem-ph] (2021).
174. B. E. Kane, Levitated spinning graphene flakes in an electric quadrupole ion trap. *Phys. Rev. B* **82**, 115441 (2010). doi: 10.1103/PhysRevB.82.115441
175. U. K. Krieger, C. Marcolli, J. P. Reid, Exploring the complexity of aerosol particle properties and processes using single particle techniques. *Chem. Soc. Rev.* **41**, 6631–6662 (2012). doi: 10.1039/c2cs35082c; pmid: 22739756
176. A. E. Rubio López, C. Gonzalez-Ballestero, O. Romero-Isart, Internal quantum dynamics of a nanoparticle in a thermal electromagnetic field: A minimal model. *Phys. Rev. B* **98**, 155405 (2018). doi: 10.1103/PhysRevB.98.155405
177. D. Hümmer *et al*., Acoustic and optical properties of a fast-spinning dielectric nanoparticle. *Phys. Rev. B* **101**, 205416 (2020). doi: 10.1103/PhysRevB.101.205416
178. C. Gonzalez-Ballestero, D. Hümmer, J. Gieseler, O. Romero-Isart, Theory of quantum acoustomagnonics and acoustomechanics with a micromagnet. *Phys. Rev. B* **101**, 125404 (2020). doi: 10.1103/PhysRevB.101.125404
179. O. Romero-Isart, L. Clemente, C. Navau, A. Sanchez, J. I. Cirac, Quantum magnetomechanics with levitating superconducting microspheres. *Phys. Rev. Lett.* **109**, 147205 (2012). doi: 10.1103/PhysRevLett.109.147205; pmid: 23083277
180. M. Teller *et al*., Heating of a Trapped Ion Induced by Dielectric Materials. *Phys. Rev. Lett.* **126**, 230505 (2021). doi: 10.1103/PhysRevLett.126.230505; pmid: 34170180
181. E. Hebestreit *et al*., Calibration and energy measurement of optically levitated nanoparticle sensors. *Rev. Sci. Instrum.* **89**, 033111 (2018). doi: 10.1063/1.5017119; pmid: 29604723
182. C. P. Blakemore *et al*., Precision Mass and Density Measurement of Individual Optically Levitated Microspheres. *Phys. Rev. Appl.* **12**, 024037 (2019). doi: 10.1103/PhysRevApplied.12.024037
183. Y. Zheng *et al*., Robust Optical-Levitation-Based Metrology of Nanoparticle's Position and Mass. *Phys. Rev. Lett.* **124**, 223603 (2020). doi: 10.1103/PhysRevLett.124.223603; pmid: 32567927
184. W. Wieczorek *et al*., Optimal State Estimation for Cavity Optomechanical Systems. *Phys. Rev. Lett.* **114**, 223601 (2015). doi: 10.1103/PhysRevLett.114.223601; pmid: 26196621
185. J. E. Curtis, B. A. Koss, D. G. Grier, Dynamic holographic optical tweezers. *Opt. Commun.* **207**, 169–175 (2002). doi: 10.1016/S0030-4018(02)01524-9
186. Y. Arita, E. M. Wright, K. Dholakia, Optical binding of two cooled micro-gyroscopes levitated in vacuum. *Optica* **5**, 910 (2018). doi: 10.1364/OPTICA.5.000910
187. C. C. Rusconi, O. Romero-Isart, Magnetic rigid rotor in the quantum regime: Theoretical toolbox. *Phys. Rev. B* **93**, 054427 (2016). doi: 10.1103/PhysRevB.93.054427
188. C. C. Rusconi, V. Pöchhacker, K. Kustura, J. I. Cirac, O. Romero-Isart, Quantum Spin Stabilized Magnetic Levitation. *Phys. Rev. Lett.* **119**, 167202 (2017). doi: 10.1103/PhysRevLett.119.167202
189. C. C. Rusconi, V. Pöchhacker, J. I. Cirac, O. Romero-Isart, Linear stability analysis of a levitated nanomagnet in a static magnetic field: Quantum spin stabilized magnetic levitation. *Phys. Rev. B* **96**, 134419 (2017). doi: 10.1103/PhysRevB.96.134419

ACKNOWLEDGMENTS

The authors acknowledge the valuable feedback from colleagues all over the world and thank C. Dellago in particular for suggesting the term "levitodynamics." **Funding:** This work has received funding from the Q-Xtreme project of the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 951234). **Competing interests:** None declared.

10.1126/science.abg3027

Science
JOURNALS AAAS

Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

REVIEW SUMMARY

ECOLOGY

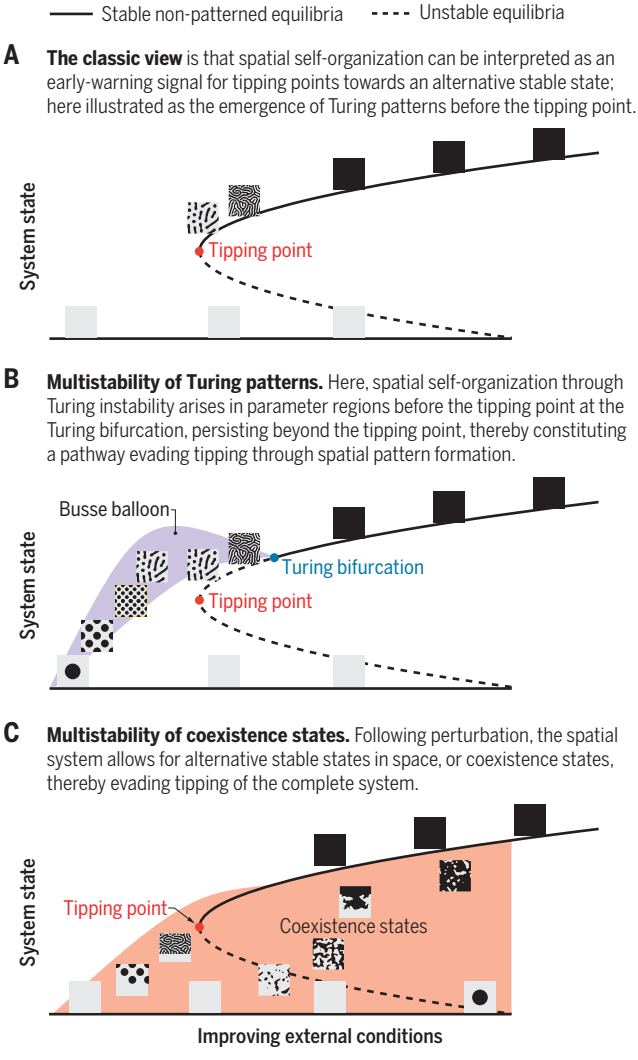
Evasion of tipping in complex systems through spatial pattern formation

Max Rietkerk*, Robbin Bastiaansen, Swarnendu Banerjee, Johan van de Koppel, Mara Baudena, Arjen Doelman

BACKGROUND: In the Anthropocene, there is a need to better understand the catastrophic effects that climate and land-use change may have on ecosystems, Earth system components, and the whole Earth system. The concept of critical transitions, or tipping from one state to another, contributes to this understanding. Tipping occurs in a system when it is forced outside the basin of attraction of the original equilibrium, resulting in a critical transition to an alternative, often less-desirable, stable state. In this context, the search for early warning signals for such imminent critical transitions has become a focus of research. In particular, spatial self-organization in ecosystems, such as the spontaneous formation of regular vegetation patterns—so-called Turing patterns—has been suggested as a prominent early warning signal.

ADVANCES: However, recent findings indicate that such spatial self-organization should not necessarily be interpreted as an early warning signal for critical transitions. Instead, spatial self-organization can cause ecosystems to evade tipping points and can thereby be a signal of resilience. These findings are based on recent mathematical analyses of spatial models and on observations of real ecosystems. Both have revealed multistability, meaning that many different spatial patterns can co-occur under the same environmental conditions, and each of these patterns can stay stable for a wide range of conditions. This enables complex system states to persist beyond tipping points through spatial self-organization. Moreover, if a complex system with tipping properties experiences a perturbation, subsequent change of the system does not necessarily lead to tipping of the complete system. Instead, the change can stay localized because the system allows for alternative states to coexist in space—thus referred to as coexistence states. These spatial patterns can also persist beyond tipping points with worsening conditions through this alternative pathway. We refer to both Turing patterns and coexistence states as spatial pattern formation. Evasion of tipping through these various pathways of spatial pattern formation may be relevant for many ecosystems and Earth system components that were hitherto interpreted as prone to tipping, including for Earth as a whole.

OUTLOOK: To further study how complex systems evade tipping through spatial pattern formation, savanna ecosystems can be considered as a concrete archetypal example because of the alternative states and spatial patterns observed for them. Moreover, universal conditions for evading tipping points in both ecosystems and Earth system components can be derived by mathematical analyses. Scenarios can be revealed by which Turing patterns with small amplitudes can grow and form large-scale localized interacting structures, thereby aiding complex systems to evade tipping. The effects that global change has on the spatial boundaries between coexistence states should be studied, and the impacts of restrictions of spatial domain and localized and nonlocal homogenizing effects by humans should be revealed. This approach will advance our understanding and predictions of critical transitions in nature and reveal how these may be avoided or reversed. ■



Evasion of tipping points. We illustrate the response of complex systems to changes in external conditions (i.e., a bifurcation diagram). Homogeneous dark gray squares depict high density of the system state variable, and homogeneous light gray squares illustrate low density. (**A**) Classic view. (**B**) Multistability of Turing patterns. Recent model analysis has revealed multistability of Turing patterns in Busse balloons, supported by satellite observations of real ecosystems. A Busse balloon is the region in parameter space in mathematical models where multistability of patterned equilibria occurs. (**C**) Multistability of coexistence states. Evading tipping can also be the result of multistability of coexistence states. These spatial patterns originate in the bistability region before the tipping point; the evolving spatial patterns can also persist beyond the tipping point with worsening external conditions, thereby constituting an alternative pathway to evade tipping points.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: m.g.rietkerk@uu.nl
Cite this article as M. Rietkerk *et al.*, *Science* **374**, eabj0359 (2021). DOI: 10.1126/science.abj0359

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj0359

REVIEW

ECOLOGY

Evasion of tipping in complex systems through spatial pattern formation

Max Rietkerk[1]*, Robbin Bastiaansen[2], Swarnendu Banerjee[1,3,4], Johan van de Koppel[5,6], Mara Baudena[1,7], Arjen Doelman[8]

The concept of tipping points and critical transitions helps inform our understanding of the catastrophic effects that global change may have on ecosystems, Earth system components, and the whole Earth system. The search for early warning indicators is ongoing, and spatial self-organization has been interpreted as one such signal. Here, we review how spatial self-organization can aid complex systems to evade tipping points and can therefore be a signal of resilience instead. Evading tipping points through various pathways of spatial pattern formation may be relevant for many ecosystems and Earth system components that hitherto have been identified as tipping prone, including for the entire Earth system. We propose a systematic analysis that may reveal the broad range of conditions under which tipping is evaded and resilience emerges.

The concept of critical transitions and tipping points contributes to our understanding of planetary changes in the Anthropocene (*1*–*3*). This idea is that ecosystems or Earth system components undergoing global change can typically persist and stay in a similar equilibrium state. The impending danger, however, is that this is only the case until a tipping point is reached, at which time this (often desired) stable state disappears, and the system undergoes a critical or catastrophic transition toward an alternative equilibrium (*1*, *2*, *4*). The latter is a state that will also prevail if the external change goes back to its original value, thus displaying hysteresis (*1*, *5*). Classic examples of ecosystems thought to exhibit critical transitions with tipping points between alternative stable states include clear lakes becoming turbid because of nutrient overloading (*6*); barren deserts replacing vegetated areas in dry savannas, or drylands in general, because of drought or overgrazing (*5*, *7*); and savannas replacing tropical forests because of deforestation associated with fire, possibly combined with less rainfall (*8*–*10*).

Mechanisms that cause such tipping dynamics are positive feedbacks—i.e., processes that amplify change imposed on complex systems. In the above-mentioned ecosystems, the positive feedbacks are increased turbidity in lakes leading to less macrophyte plants, which hinders nutrient uptake, leading to even more turbidity (*6*); reduced vegetation cover resulting in less water infiltration into the soil in dry savannas, which in turn leads to even less vegetation (*5*, *7*); and fewer forest trees in tropical forests because of deforestation and fires allowing more grass growth, fueling more fires, and preventing forest tree establishment, thereby leading to even fewer forest trees (*10*–*12*). At tipping points, positive reinforcing feedbacks overwhelm the negative balancing feedback processes that maintain the desired state of ecosystems, often leading to the catastrophic loss of ecosystem services to humans.

There are many possible ways that positive feedbacks may overtake negative feedbacks and tipping can be triggered. Tipping occurs as a result of the loss of resilience, which is interpreted as the amount of environmental change or the strength of perturbation that a system can withstand before it tips to another basin of attraction (*13*). Three of the most prominent possibilities have been classified as follows (*14*): Bifurcation-induced tipping (B-tipping) happens when a parameter shift (e.g., a change in environmental conditions) reduces the basin of attraction of the original stable state to zero; noise-induced tipping (N-tipping) occurs when a perturbation of the system state—e.g., as the result of environmental noise or disturbance—knocks the system outside the basin of attraction of the original state; and rate-induced tipping (R-tipping) arises when an environmental (parameter) change is faster than the restorative attraction to the original state. The restorative attraction to the original state, or the speed after which an equilibrium state is restored, is called engineering resilience, or stability, in the ecological literature (*13*).

This idea of tipping or critical transitions is important not only for ecosystems at a local scale but also for many regional-scale Earth system components (*15*, *16*). Two examples include tipping of the Arctic ice sheets because of warming, inducing changes of surface albedo, leading to more warming (*17*–*19*), and tipping of the Atlantic Ocean circulation induced by changes in surface water fluxes (*20*–*22*). The same notion of tipping points also underlies the concept of hazardous planetary boundaries at the global scale (*23*) and has similarly been applied to Earth's biosphere as a whole, as a response to climate and land-use changes (*3*, *24*). This concept is therefore relevant for all spatial scales, ranging from ecosystems to the entire complex Earth system.

For spatially extended ecosystems, such as drylands, savannas, and peatlands, it has been highlighted that critical transitions are associated with the formation of self-organized spatial patterns of vegetation (*2*, *25*). In these systems, as environmental conditions worsen, a uniform coverage becomes unstable to nonuniform (spatial) disturbances, as a result of the spatial processes, which leads to the formation of regular spatial patterns. Such spatial destabilization of a uniform state is called a Turing instability or a Turing bifurcation after Alan Turing, who first studied this in reaction-diffusion systems (*26*). After a Turing instability, so-called Turing patterns emerge that can have various spatial forms depending on environmental conditions. This has previously been interpreted as preceding a tipping point to an alternative ecosystem state (*25*). Most notably in drylands, the following vegetation patterns are observed (listed here in the order in which they appear with worsening environmental conditions, such as increasing drought or grazing): bare gaps in homogeneous vegetation cover, labyrinthine or striped vegetation cover, and spotty vegetation in homogeneous bare soil (*25*, *27*–*30*).

The mechanistic base of Turing instability is that the positive feedback mentioned earlier is scale dependent in spatially extended systems: The positive feedback dampens and is subsequently replaced by a negative feedback further away in space, generating scale-dependent feedbacks as a result of spatial processes (*31*). The crucial spatial effects of these processes leading to scale-dependent feedbacks are typically neglected when assuming that systems are homogeneous. In drylands, for example, the scale-dependent feedback relates to increased infiltration of water into the soil and larger soil water uptake at places where vegetation is growing. This generates surface and soil water flows toward the vegetation at the cost of available water further

[1]Copernicus Institute of Sustainable Development, Utrecht University, 3508 TC, Utrecht, Netherlands. [2]Department of Physics, Institute for Marine and Atmospheric Research Utrecht, Utrecht University, 3508 TA, Utrecht, Netherlands. [3]The Institute of Mathematical Sciences, CIT Campus, Taramani, Chennai 600113, India. [4]Indian Statistical Institute, Agricultural and Ecological Research Unit, Kolkata 700108, India. [5]Department of Estuarine and Delta Systems, Royal Netherlands Institute for Sea Research, 4400 AC, Yerseke, Netherlands. [6]Groningen Institute for Evolutionary Life Sciences, Conservation Ecology Group, University of Groningen, 9700 CC, Groningen, Netherlands. [7]National Research Council of Italy, Institute of Atmospheric Sciences and Climate (CNR-ISAC), 10133 Torino, Italy. [8]Mathematical Institute, Leiden University, 2300 RA, Leiden, Netherlands.
***Corresponding author. Email: m.g.rietkerk@uu.nl**

away (*28*, *29*), ultimately resulting in spatial patterns of concentrated vegetation in some places and bare soil in others.

Evading tipping points through Turing patterns

The regular spatial patterns resulting from Turing instability have been understood until now as early warning signals for tipping points or critical transitions toward an alternative state (B-tipping) in various ecosystems (*2*, *25*). However, recent mathematical analyses, combined with satellite observations of real ecosystems, have revealed a drastic alternative view (*32*–*34*). These analyses have exposed the existence of pattern-driven multistability described by Busse balloons, after F. H. Busse (*32*, *33*, *35*). The Busse balloon (Fig. 1) indicates a region of the parameter space in mathematical models for which a large range of wavelengths of regular spatial patterns (vegetation in the dry savanna example) are possible and stable. Also, when conditions or parameters in the model are changed, a specific spatial organization can remain stable, which means that the wavelength does not change for a wide range of environmental conditions until the edge of the Busse balloon is reached. Here, the spatial organization and ecosystem state variable, such as ecosystem productivity, may adjust in a noncritical or noncatastrophic way (Fig. 1). This can be at parameter values for which, in the nonspatial model, the homogeneous vegetation state may still exist or may have already collapsed to bare soil.

The alternative view mentioned above stems from the fact that earlier analysis of the ecosystem models only considered the stability of uniform states. Such analysis concealed the many possible responses of spatially patterned states, overlooking the existence of Busse balloons. Studies beyond the conventional analysis (*32*) have also considered the stability of patterned states, exposing the Busse balloons, implying qualitatively new model predictions, and inspiring new observations in real ecosystems.

These model predictions, including the existence of multistability and the notion of Busse balloons, have lately been supported by observations in real ecosystems (*33*). Satellite observations have shown regular spatial vegetation patterns that occur in vast areas in different regions in dry savannas in Africa. These observations have shown that many spatial patterns with different wavelengths co-occur next to each other in one and the same area with similar environmental conditions, and for different areas within the same larger region, supporting the Busse balloon theory. Moreover, the wavelengths of the patterns in specific areas within those regions have remained stable in time for decades, despite changes in environmental conditions. Once again, we refer to these two phenomena combined as multistability, which is illustrated by the Busse balloon (Fig. 1) (*32*, *33*). Also, recent model analyses have revealed that at the edge of the Busse balloon, ecosystems adjust their spatial organization in such a way that they stay within the Busse balloon. In other words, the dominant variable generating the spatial pattern (ecosystem productivity in our example) does not change drastically in the way that it would with a critical or catastrophic transition, but more gradually instead (*34*). Moreover, in the patterned state, vegetation persists for environmental conditions beyond the tipping point (Fig. 1B). This therefore demonstrates a case where the system shows spatial pattern formation at the Turing bifurcation before the tipping point is reached, which then extends beyond the tipping point, essentially constituting a pathway evading it (Fig. 1).

Thus, a phenomenon once considered to be an illustrative early warning signal of imminent critical transitions now appears to be a sign of resilience instead. In this case, the patterns originate from a Turing instability, leading to multistability of patterns and subsequent gradual change of the system. Similarly, in mussel beds, multistability of patterned states has been found in models and real systems (*36*, *37*). It is noteworthy that the predicted size and number of ecosystem shifts can vary; these depend on the magnitude and rate of environmental change (*32*, *34*).

There are now strong indications that evading tipping points through spatial self-organization may be very common for ecosystems and Earth system components. This is based on earlier work on regular pattern formation in real ecosystems (*31*), combined with the recent insights outlined here, and a mathematical analysis (Box 1) of a class of models. This is also underpinned by other observations—tipping points and alternative stable states (bistability) have been suggested on the basis of simple, nonspatial models for at least the following ecosystems: drylands or dry savannas (*5*, *7*), savannas (*10*, *11*, *38*), peatlands (*39*), mussel beds (*40*), intertidal mudflats (*41*), barrier islands (*42*),

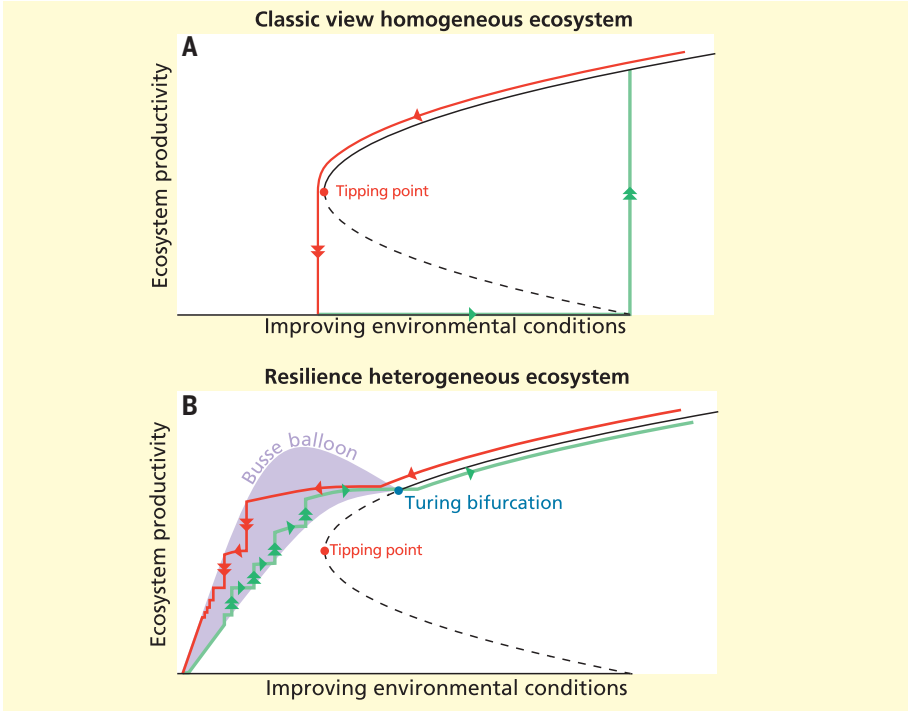


Fig. 1. Resilience in homogeneous and heterogeneous ecosystems. Solid lines denote stable nonpatterned equilibria, and dashed lines indicate unstable equilibria. Double arrows indicate ecosystem shifts, and single arrows indicate minor ecosystem adjustments. Typical trajectories for ecosystem degradation (red) and recovery (green) are given, showing the narrowed hysteresis loop for heterogeneous, spatially self-organized ecosystems. (**A**) Classic view of a homogeneous ecosystem. Classically, worsening environmental conditions correspond to a minor adjustment of the ecosystem state or ecosystem productivity, until these drive the system over a tipping point, and a critical or catastrophic transition (B-tipping) occurs. (**B**) Resilience heterogeneous ecosystem. In multistable spatial ecosystems, every set of environmental conditions beyond the Turing bifurcation allows for multiple stable, spatially self-organized states, indicated by the purple area, named the Busse balloon. Here, instead of one critical transition, multiple smaller ecosystem shifts from one spatially patterned state to another occur, which have minor effects on the function or productivity of the ecosystem as a whole (*32*, *34*).

Box 1. Turing-before-tipping.

In mathematical models, B-tipping takes place when an equilibrium state loses its attraction and disappears in response to a parameter, for example A, passing through a critical value, A^*. At that moment, the system has to shift to an alternative equilibrium state. In the literature (*10*, *38*, *39*), the models in which tipping occurs typically are spatially homogeneous—i.e., not spatially extended. However, the associated modeled systems are almost without exception spatially extended, and it is thus natural and necessary to model the spatial effects. The main idea underlying the concept of Turing-before-tipping is as follows: For tipping, it is necessary that the (initial) equilibrium state stays stable until it tips. Therefore, it is self-evident to ask the question, is the homogeneous equilibrium state that precedes tipping also stable against spatial effects? If not, a spatially extended version of the model will not exhibit tipping; spatial patterns emerge before parameter A has reached its critical (tipping) value A^*. In that case, the modeled ecosystem will not tip but instead will evade the critical transition by forming spatial patterns.

As a general and relatively simple example of how one can determine whether Turing-before-tipping occurs (or not), we consider the dynamics of a biological quantity $b(t)$ in interaction with a (typically limited) resource $r(t)$ and assume that this is described by the model $\frac{db}{dt} = F(b, r)$, and $\frac{dr}{dt} = G(b, r)$, in which $F(b, r)$ and $G(b, r)$ represent various growth, decay, and interaction effects that vary with parameter A. For example, in the (nondimensionalized) model for vegetation dynamics in drylands of Bastiaansen *et al*. (*33*), $b(t)$ is the biomass of the vegetation, $r(t)$ is the available water, and $F(b, r) = -Mb + rb^2$, $G(b, r) = A - r - rb^2$, where M models the vegetation mortality rate and A the rainfall. Tipping occurs when, as a function of parameter A, two equilibrium states merge into one, the threshold state (b^*, r^*) at $A = A^*$, and subsequently disappear. To guarantee that one of these equilibrium states is stable, and thus observable, until it tips, it is necessary that $\frac{\partial F}{\partial b}(b^*, r^*) + \frac{\partial G}{\partial r}(b^*, r^*) < 0$. To consider the question whether Turing-before-tipping may occur, one therefore needs to incorporate spatial effects into the homogeneous model. A simple way to do so is to extend the model for $b(t)$ and $r(t)$ into a system of reaction-diffusion equations for biomass $B(x, t)$ and resource $R(x, t)$: $\frac{dB}{dt} = d_B \Delta B + F(B, R)$, $\frac{dR}{dt} d_R \Delta R + G(B, R)$, in which Δ models (spatial) diffusion, and d_B and d_R are the diffusion coefficients that govern the diffusive spreading speeds of $B(x, t)$ and $R(x, t)$. It can be obtained that only if $d_R \frac{\partial F}{\partial b}(b^*, r^*) + d_B \frac{\partial G}{\partial r}(b^*, r^*) < 0$ the threshold state (b^*, r^*) is stable against spatial perturbations. In other words, Turing-before-tipping occurs if $\frac{\partial F}{\partial b}(b^*, r^*) + \frac{\partial G}{\partial r}(b^*, r^*) < 0$ and $d_R \frac{\partial F}{\partial b}(b^*, r^*) + d_B \frac{\partial G}{\partial r}(b^*, r^*) > 0$. This condition therefore determines whether an ecosystem can evade collapse by forming patterns (or not) and can be checked explicitly in any given model. For example, in the dryland model of Bastiaansen *et al*. (*33*), these conditions are given by $M < 2$ and $eM > 2$ [at $(b^*, r^*) = (1, M)$ with $A^* = 2M$, $d_R = e$, and $d_B = 1$]. In Bastiaansen *et al*. (*33*), the following realistic choices for M and e were made: $M = 0.45$, $e = 500$ (*32*). Thus, Turing-before-tipping takes place as is also exhibited by the observations reported. Explicit conditions for Turing-before-tipping can also be deduced for multicomponent models and/or for models with spatial effects beyond (linear) diffusion.

and sea grasses (*43*). Most notably, for all of these model ecosystems, spatial self-organization, or Turing patterns, have been observed in their real counterpart systems, which are always spatially extended (*31*, *44*–*47*).

The evasion of tipping points may not be restricted to the case of Turing patterns in Busse balloons, but rather may include more comprehensive spatial pattern formation, as we will outline below. Apart from spatial pattern formation, there are also other mechanisms through which tipping may be evaded. For example, sufficiently fast reset of the changing parameters as compared with the rate of the changing state variable may repair the overshooting of a tipping point (that is, put the state variable back into the original basin of attraction) in simple Earth system component models (*48*). We expect that both spatial patterns and time delay of the state variable will not only affect B-tipping but also have a very similar effect on N-tipping and R-tipping—that is, evasion of (complete) tipping of the system. In this context, it is also interesting to note that thresholds related to tipping points are generally rarely detectable from empirical data (*49*).

Evading tipping points through coexistence states

The ranges of conditions and mechanisms for which complex systems can evade tipping through spatial pattern formation are not restricted to those leading to Turing patterns in Busse balloons—they include more comprehensive spatial pattern formation. Although nonspatial complex systems can respond to disturbances or perturbations only with a system-wide response that leads either to complete tipping to an alternative stable state (N-tipping) or to full recovery, this is not the case for real systems, which are always spatially extended. In those systems, localized or random disturbances that are omnipresent can lead to spatial pattern formation in which only part of the spatial domain transitions to the alternative state, and system-wide N-tipping is evaded. Such spatial patterns consist of the coexistence of alternative stable states in space and are herein referred to as coexistence states. Multistability of different spatial patterns of those coexistence states can occur for given environmental conditions (*50*, *51*). Spatial boundaries, or interfaces, necessarily arise between alternative states in space. These boundaries can range from very simple, in the form of a single front, to a rich patterned structure of alternative stable states. For example, presence and absence of ice are two alternative stable states that can occur in the spatial domain of the whole Earth system; these states coexist on a global scale. Ice only appears at the polar latitudes, and no ice occurs in between, with a simple spatial boundary between them named the ice line or grounding line (*52*, *53*). At smaller scales, such spatial boundaries between ice and no-ice states can consist of much more complex structures of alternative stable states (*54*).

The spatial boundaries separating the coexisting states may themselves become unstable, which might lead to one alternative state invading the other, thereby providing a mechanism for evolving such a rich and complex patterned structure of alternative stable states (*51*, *55*). Furthermore, reorganization of spatial patterns may take place as a result of environmental change, without any abrupt change in system characteristics at the global system scale. Therefore, the system can evade critical transitions, and a smoother and much more gradual response can be expected instead (*50*, *51*). In this manner, the spatial patterns can also persist beyond the tipping point with worsening environmental conditions. This constitutes another essential pathway for ecosystems and Earth system components to evade tipping points through spatial pattern formation and multistability. Unexpectedly, such spatial organization could even lead to nonforced reversed transitions in which a state counterinvades, or (in other words) the automatic and spontaneous recovery of the original state from the alternative one (*55*).

These coexistence states may also form in ecosystems because of the aggregation of organisms into self-organized patchiness (*56*), as was observed in the clustering of mussels in mussel beds (*57*). Here, there is an interesting parallel with the physical theory of phase separation. This theory describes the dynamics of spatial interfaces of phases (i.e., spatial boundaries between coexistence states in our terminology) and explains a rich variety of possibly very slow transient spatial patterns. These include maze or labyrinths patterns, spot patterns, and viscous fingering, which may cause very gradual transitions when conditions are changed. Phase separation dynamics leading to coexistence states can be present

in ecosystems and Earth system components as well (*58*). For instance, some of the spatial aggregation of organisms and resources can be interpreted as such. Comparisons with phase segregation mechanisms have been made already for other ecosystems besides mussels (*57*)—e.g., for the aggregation of vegetation in fingering patterns (*55*).

New theories and procedures need to be developed to distinguish spatial pattern formation from patterns originating from preexisting heterogeneity, because real systems are generally a mixture of both (*59*). One example of preexisting heterogeneity that could falsify the mechanism of spatial pattern formation in the specific context of vegetation patterns could be preexisting drainage patterns governed by topography explaining tree distribution patterns.

Toward a theory on spatial pattern formation and multistability

The insights described above highlight how spatial pattern formation in complex systems can lead to evasion of tipping and increased resilience. We highlight a few illustrative examples in Box 2 and Box 3. However, some complex systems may still exhibit tipping even though they are spatially extended. So when does the classical nonspatial framework of tipping points break down and need to be supplemented by a new theory on spatial pattern formation? For what ecosystems and Earth system components, including the complex Earth system as a whole, is this relevant? And for what conditions and at which spatial scales is this the case? Currently, these questions are not fully resolved. In the text below, we provide a synthesis of the current understanding, based on analysis of model and real patterned systems, in which we also explicitly point out missing pieces of knowledge.

An archetypical system: Evading savanna tipping

Savannas are defined by the coexistence of grasses and trees, spread over one-eighth of the land surface worldwide, and are an important source of livelihood for many (*60*). The expected changes in climate and land use can lead to substantial alteration of vegetation characteristics and global savanna distributions (*61*, *62*). Tree cover is highly variable and notoriously difficult to predict; it has been suggested that space could be a main element elucidating this, which is missing in many studies of savannas (*63*). Savannas can be considered as a model ecosystem to derive and validate conditions for evading tipping points because of the following two main reasons. First, observations in models and real ecosystems show multiple alternative states and tipping phenomena. The ecosystem may change between a system with no trees (barren desert or grassland), open savanna with grasses and scattered trees, closed

Box 2. Spatial pattern formation evading tipping in local-scale ecosystems.

Savanna ecosystems are characterized by the coexistence of trees and grasses. Most homogeneous models reproduce this coexistence. They show that, depending on rainfall, fire intensity, and level of herbivory, open savannas (with dispersed trees) may generally switch between multiple alternative homogeneous states (Fig. 2): no trees; either barren desert or grassland, closed tree cover of mainly savanna trees; savanna woodland, or closed tree cover of mainly forest trees; or tropical forest (*5*, *10*, *11*, *38*, *64*, *65*, *75*). In other words, neglecting spatial effects, these models predict that savanna ecosystems can exhibit alternative states and critical transitions with tipping between them.

However, when spatially extended, the models instead show a rich variety of spatial patterns through the emergence of Turing patterns (*71*, *104*) and because of the coexistence of alternative stable states in space (*51*, *73*) (Fig. 2). The system may, or may not, evade tipping by the appearance of spatial patterns, as analysis (Box 1) indicates that both situations may occur. Whether a tropical forest or savanna collapses or increases its resilience by the multistability associated with spatial pattern formation depends on local, but system-wide, conditions—i.e., the parameters in the model.

Box 3. Multistability evading tipping in regional- and global-scale Earth systems.

Tipping behavior and critical transitions are attributed to many Earth system components (*3*, *15*). However, even for these systems, the framework of tipping points may be too limited, and multistability may play a more important role than previously thought.

As a classic example, we consider how spatial effects lead to the introduction of coexistence states and thereby change the tipping behavior related to the ice-albedo feedback in an Earth's global climate model (*52*, *53*, *105*, *106*). Changes in Earth's temperature are directly related to changes in the energy budget, which is computed as incoming solar radiation minus reflected and outgoing (Planck) radiation. The temperature dependency of the albedo is the ice-albedo feedback: As long as temperatures are low, Earth is covered in ice, which reflects much of the solar radiation. However, when temperatures rise, the ice melts and consequently less radiation is reflected, which leads to further temperature increase. If one does not take spatial effects into account, as is commonly done, this feedback mechanism leads to two alternative Earth states in this model: An Earth fully covered in ice ("snowball earth") or an Earth with no ice ("no-ice Earth"). There is a bistable region where both of these states exist, and critical transitions and tipping points between those states occur when one stable state vanishes.

However, in reality, we are clearly not in either of those two Earth states because ice is present only at the polar regions. That is because spatial effects that play an important role in the real Earth's global energy budget are ignored in such a model. For example, incoming solar radiation is latitude dependent, and meridional heat flow forms an integral part of energy distribution. When adding such spatial mechanisms, the models also predict coexisting states of ice and no-ice in addition to the alternative snowball and no-ice Earth states. In the coexistence states, ice is present in only part of Earth—near a pole there is ice, while simultaneously there is no ice at the equator, with a spatial boundary between these alternative states occurring at some intermediate latitude. The presence of these additional states changes the classical tipping properties of the nonspatial system: When a fully (un)covered Earth state disappears, a less-critical transition to a system with coexistence states might happen, and more gradual transitions are possible.

In addition to this example of coexistence states, there is multistability attributed to Earth system components. In fact, the Busse balloon finds its origin in the study of thermal convection (*35*), which is closely related to turbulence in fluid mechanics and thus relevant also for the global atmosphere and the ocean circulation of Earth. However, in the example of the Atlantic meridional overturning circulation mentioned above, when using box models with relatively few boxes, tipping is observed (*107*), whereas multistability of coexistence states is not detected. We suggest that, despite taking into account spatial processes between the boxes, homogenizing within only a restricted number of boxes severely limits the many emergent spatial responses that the system may unfold. Increasing the spatial resolution, by adding more boxes, increases the number and nature of stable states (*108*, *109*) and gradually brings back the multistability. In this context, an interesting line of research is to investigate whether such tipping of Earth system components in general (*15*) persists in higher-resolution models, such as in the state-of-the-art global climate models (*16*), or is replaced by multistability of coexistence states including more gradual transitions. If the latter is the case, this may be a possible explanation why the full-complexity global climate models seem more stable than the simple or intermediate-complexity ones (*110*).

savanna woodland covered with savanna trees and shrubs, and closed tropical forest with forest trees (*5*, *8*–*12*, *38*, *64*–*68*) (Fig. 2). Savanna trees are more fire tolerant and less shade tolerant than forest trees (*69*). This, together with the flammability and fire resistance of grasses that can easily regrow after fires, is the key of the positive feedback mentioned earlier, generating alternative stable states between open savanna and tropical forests. Second, both types of spatial pattern formation as we outlined, namely Turing patterns and coexistence states, have been observed in real and model systems (*12*, *25*, *68*, *70*–*76*) (Fig. 2). Therefore, studying savanna ecosystems is an excellent opportunity to obtain results that are of general interest and applicability.

Toward a unified spatial savanna model

To the best of our knowledge, no unified savanna model exists that explains all of the possible ecosystem states including spatial Turing patterns and coexistence states. To develop a unified spatially explicit savanna model, existing savanna models (*71*, *76*, *77*) could be combined in such a way that the new model consists of (at least) the state variables water, grass biomass, savanna tree biomass, and forest tree biomass. Savanna and forest trees should be distinguished by their different fire responses and shade tolerance (*78*, *79*). Nonspatial models that consider rain, fire, and herbivory may aid in the development of such a unified framework (*64*, *80*), and studies of savanna-forest transitions may also prove useful (*10*, *81*). The new model could include positive feedbacks between water infiltration into the soil and biomass and between fire and grass biomass (*11*). Herbivory (grazing and browsing) could be added as an extra negative term (*82*), or possibly as an extra dynamic equation (*83*). It could then be analyzed to address the following questions along a rainfall gradient (Fig. 2): Under which conditions do sharp or gradual spatial boundaries exist between open savanna on the one hand and tropical forest, savanna woodland, or desert on the other? How are those boundaries evolving, bifurcating, or moving with respect to (changes in) climate, herbivory, and fire? Under which conditions do spatial patterns occur, either Turing patterns or coexistence states? Answers to these questions can be obtained by a combination of simulations, mathematical bifurcation analyses, and numerical continuation (*84*, *85*).

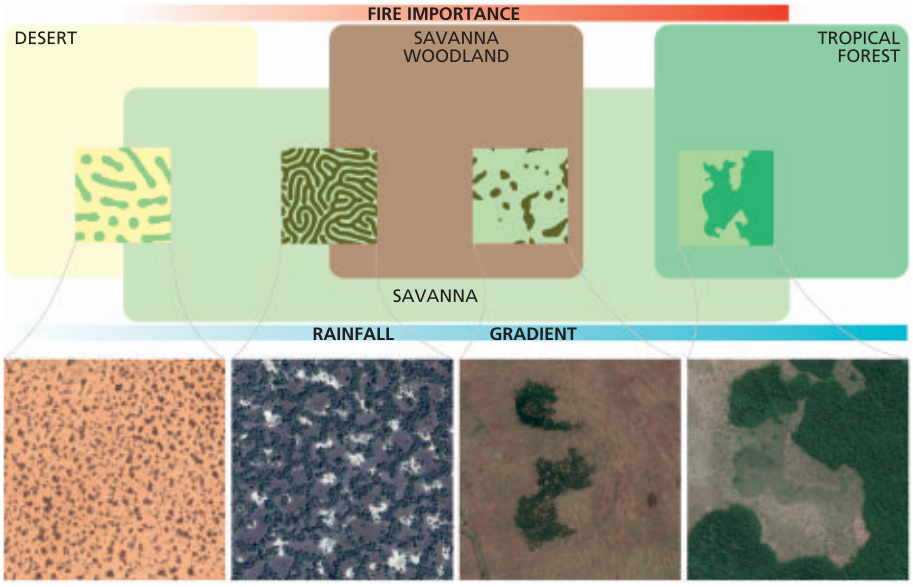


Fig. 2. Savannas along a rainfall gradient. Open savannas with dispersed trees can be classified into three broad ranges according to annual rainfall (*67*): humid, mesic, and dry (blue bar indicates rainfall decreasing from humid to dry, from right to left). The humid range potentially demonstrates bistability with tropical forest. The mesic range shows possible bistability with savanna woodland, and the dry end shows possible bistability with barren desert (or grassland, which is not shown). Fire and herbivory play an important role in savanna dynamics, determining varying tree cover given a certain amount of rainfall. The importance of fire (illustrated by the red shade in the fire bar) decreases from humid toward dry ranges, where water availability is the main limiting factor and driving force. We compare the expected spatial structures with symbolic Google Earth image examples, from right to left with decreasing rainfall, namely: coexistence states between forests and humid savannas (Gabon 1°16′1.06″ S to 13°56′18.63″ E, 2000 m by 2000 m), coexistence states as well as Turing patterns between savanna woodland and mesic open savanna (Kenya 1°28′8.83″ S to 34°53′55.60″ E, 1200 m by 1200 m; Mali 13°03′15.09″ N to 6°40′50.86″ W, 800 m by 800 m), and Turing patterns between dry savannas and desert (Sudan 11°26′47.53″ N to 27°53′30.07″ E, 1500 m by 1500 m). There is no unifying model explaining all savanna states and spatial patterns observed. We suggest that developing and analyzing such a unifying savanna model is needed by incorporating relevant state variables and parameters, making the model spatially explicit, and using a combination of mathematical bifurcation analyses, simulations, and numerical continuation methods (*84*, *85*).

General new predictions based on the theory that could be tested with such a unified model are as follows (Fig. 2): (i) At the humid end of the rainfall gradient, at the tropical forest–savanna boundaries, where fire is generally assumed to drive alternative stable states between tropical forests and open savannas, we expect to discover coexistence states (*68*, *86*) evading tipping. (ii) In the mesic range, where a combination of fire and herbivory drives alternative states between closed savanna woodland and open savanna, we expect to find both coexistence states and Turing patterns (*70*, *73*, *74*) evading tipping. (iii) At the dry end of the gradient, where dryness drives alternative stable states between open savanna on the one hand and desert on the other, we suppose that the occurrence of Turing patterns associated with Busse balloons (*33*, *34*) leads to the evasion of tipping points. (iv) Increased fire intensity will lead to movement of the tropical forest–savanna boundary toward tropical forest (humid savannas replacing forest) (*81*). A decrease in fire or herbivory will lead to savanna woodland invading mesic open savanna, and decreased rainfall will lead to desert replacing dry savanna.

Observations of spatial patterns evading tipping points

Toward the wet end of the rainfall gradient, sharp tropical forest–savanna fronts (*87*) without spatial pattern formation could be interpreted as spatial substitutions of critical transitions (*88*). However, it is not clear for which conditions tropical forest–savanna boundaries are sharp or gradual, whether they are moving, and whether patterned structures of alternative stable states (coexistence states) occur. Therefore, the same questions as those for the model analyses described above can be addressed through observations. After being tested by model analyses, the predictions mentioned above can be validated, focusing on fire and dryness as main drivers. Spatial signatures of sharp vegetation boundaries (*88*) between tropical forests and open savannas can be explored worldwide from satellite data (*87*, *89*). This can be done at multiple spatial resolutions to create time series for multiple years or decades (*90*) to detect vegetation boundaries at both coarser and finer resolutions necessary to screen boundaries of spatial patterned structures of coexistence states (*68*, *86*) and Turing patterns, which could both be easily overlooked by only using coarser

spatial resolutions. The same approach can be followed for open savanna–savanna woodland boundaries for the mesic ranges of the rainfall gradient and for dry savanna–desert boundaries at the dry end of the gradient.

Such findings can be related to the main drivers, such as dryness and fire as mentioned above, through change indicators (*91*, *92*). Metrics of spatial boundaries can also be correlated to annual burned area maps (*93*). This would allow for tests of the predictions of how these drivers determine the sharpness of vegetation boundaries, whether patterned structures of coexistence states and Turing patterns occur, and how moving the boundaries is associated with a change of the main drivers. This is important in the light of expected changes in climate and land use, and it will provide insight into the conditions for which tipping points may be evaded or not.

Universal conditions for spatial pattern formation and multistability

A major part of the mathematical literature on spatial pattern formation focuses on the onset of Turing patterns close to homogeneous equilibria (*94*). However, real ecosystems and Earth system components in their natural state typically are not close to such an onset at which the patterns first emerge, but rather are far from equilibrium, where the spatial patterns are fully developed. There can only be limited mathematical insight into many of the spatial patterns in real systems, unless the governing system shows a sufficient large-scale separation—the condition in which spatial transport of the components of the system plays out on vastly different spatial and temporal scales. A system for which this occurs is called singularly perturbed in the mathematical literature. Notably, the spatially extended models and real systems considered here typically are singularly perturbed, where the singularly perturbed nature stems from observations that pattern formation in ecosystems is in fact driven by counteracting feedback mechanisms on widely different spatial scales (*31*). Therefore, the realistic, far-from-equilibrium patterns considered here can be studied in mathematical detail by the methods of singular perturbation theory (*95*). Moreover, the most relevant patterns commonly exhibited by ecosystems and Earth system components (including the examples of Box 2 and 3) have the nature of interacting localized structures, such as areas with vegetation bands surrounded by areas with bare soil (*33*), the boundaries mentioned above between open savanna and savanna woodlands (*73*), and the grounding line of glaciers (*54*). Resilience by multistability (*33*, *34*) is directly coupled to the very rich variety of patterns associated with interacting localized structures like localized stripes and spatial boundaries.

Therefore, the formation of these fundamental spatial patterns and multistability provide mechanisms by which spatial ecosystems (Box 2) or Earth system components (Box 3), including the complex Earth system as a whole, may evade tipping in general. Whether this occurs or not will depend on local (but system-wide) conditions in real systems, or parameter combinations in models. For various classes of systems, there will be universal conditions that lead to the evasion of tipping points through spatial pattern formation and multistability, such as for savanna ecosystems. These are classes of ecosystems and Earth system components for which tipping points are currently predicted.

Turing-before-tipping

Preliminary analyses of spatially extended ecosystems, such as dry savannas, modeled by activator-inhibitor–type reaction-diffusion equations (*27*, *75*, *76*), show that tipping may or may not be preceded by a pattern-forming Turing bifurcation (Box 1 and Fig. 1). The conditions for which this may or may not happen can be explicitly expressed in terms of model parameters (Box 1). However, the Turing bifurcation only gives insight into the onset of spatial patterns, not into their behavior beyond onset (far from equilibrium), where the patterns become more developed and observable. Moreover, it is the dynamics beyond onset that determine if and how these Turing patterns lead to evasion of tipping. An explicit scenario has been unraveled within the literature on dry savannas (*32*, *34*) where Turing patterns evolved beyond the onset in such a way that tipping is evaded in the system. The crucial question now is whether this scenario may also play a role in other types of systems and whether there are alternative scenarios through which small-amplitude Turing patterns may evolve into large-scale interacting localized structures that enable the system to evade tipping. Because of the singularly perturbed nature of the models, it is possible to study the basic localized structures and their interactions mathematically (*96*–*98*). To make the crucial connection between these localized patterns and those that appear from the Turing bifurcation, a further analysis of the Busse balloon—and especially the nature of its boundary—by a combination of bifurcation analyses, simulations, and numerical continuation (*32*, *84*, *85*) is necessary.

Such analyses will also reveal the dynamics of systems preceding bifurcation points, thereby leading to better early warning signs. The classical theory of tipping, predominantly based on nonspatial models, led to the creation of generic early warning signs before such tipping occurs (*2*, *99*), mainly associated with critical slowing down of recovery after perturbation. In spatial systems, this idea becomes more complicated because this may now crucially depend on the nature of the different spatial perturbations; critical slowing down will only show for perturbations that have a specific spatial structure (*74*), which may be hard or even impossible to detect in spatially averaged data. Simultaneously, the form of the perturbation that does show critical slowing down can help to determine the type of bifurcation before onset—i.e., whether a system will tip or evade tipping by forming spatial patterns. After the system is exposed to some generic perturbation (that might be random or localized) before a bifurcation is reached, the system will restore during a transient restoration period, and temporarily, either a patterned or a nonpatterned state may emerge. The form of this emerging transient state could now be distinctive, and the type of bifurcation it precedes could be identified. Leading up to Turing bifurcation, such emerging transient state may be some spatially periodic pattern. By contrast, tipping bifurcation could be preceded by emerging transient states that are either spatially homogeneous or extremely localized. This would make it possible to distinguish between both types of bifurcations from spatial time series before they occur. Of course, one main challenge is the extraction of these spatial perturbations from data, but existing so-called mode decomposition algorithms (*100*) may be a viable option.

Destabilizations of interfaces between coexistence states

Isolated interfaces between coexistence states, such as the open savanna–savanna woodland boundary, the tropical forest–savanna boundary (Box 2 and Fig. 2), and the interfaces between ocean and ice (Box 3), are necessary for the multistability by which tipping of the whole spatial system through perturbation of the stable state may be evaded. However, for the spatial coexistence patterns to persist beyond the tipping point with worsening external conditions, an additional mechanism is required and likely. Similarly to how homogeneous states can be destabilized by Turing bifurcations, spatial fronts between coexistence states may typically also bifurcate and may thus be the origin of a multitude of evolving localized spatial patterns. These patterns subsequently may provide the ecosystem or Earth system component with further multistability and with various gradual routes it may follow beyond the tipping point—i.e., with evasion of the tipping point when environmental conditions worsen. These bifurcations can be traced numerically (*51*, *55*). But more importantly, the singular perturbed nature of the models allows for the development of a fundamental understanding of the underlying destabilizing mechanisms (*98*, *101*). For instance, conditions can be derived for which an invading front that leaves a homogeneous stable state

behind triggers a counterinvasion of an alternative patterned state, which comprises multistability and thus a gradual route, thereby circumventing tipping points.

The impact of domain, localized, and nonlocal homogenizing effects

Most mathematical studies of spatial pattern formation in spatially extended complex systems take place assuming a highly idealized domain—a sufficiently large open space or volume in which environmental conditions do not change throughout the domain. In real systems, however, such idealized domains do not exist, and it is not evident if and how results carry over from idealized domains to more realistic ones. For example, if the spatial domain in which a system can evolve is too small for spatial pattern formation, evasion of tipping cannot work anymore. A prominent example could be alternative stable states and tipping of the spatially confined and shallow lakes (*6*). This implies there is a minimum domain size for complex systems to form spatial patterns and enhance resilience; this minimum size would depend on the spatial scale of the dominant mechanisms and resulting spatial patterns of the specific system under consideration.

Moreover, localized effects—e.g., those produced by human interventions—can have a strong impact on the formation, stability, and dynamics of patterns (*102*). Examples of this include logging in tropical forests and imposed fixing of sand dunes or building dikes in coastal dune systems. Such human perturbation, albeit localized, may substantially reduce the flexibility and thus the resilience of the patterned system as a whole. Similarly, and probably more importantly, the same can happen if humans homogenize spatial patterns characteristic of pristine systems nonlocally, because the mechanisms outlined here enhancing resilience will not function anymore. Examples of this include large-scale agriculture in terrestrial ecosystems, spatially homogeneous restoration efforts combating desertification, and destructive bottom trawling in marine ecosystems. Therefore, the study of resilience through spatial pattern formation in complex systems should be embedded in a thorough analysis of the impact of spatial restrictions of the domain and the effects of localized and nonlocal homogenizing, human-induced effects. A relevant approach would be a combination of computational and analytical studies to determine the effects of such spatial (in)homogeneities on pattern dynamics and resilience (*103*). This is important for ecosystem restoration and mitigating the effects of land use and climate change.

Conclusions

Here, we have shown how spatial self-organization and multistability resulting from Turing patterns may help complex systems evade tipping points and enhance resilience. Additionally, we have outlined that the ranges of conditions and mechanisms for which tipping is evaded are supposed to be much broader than those leading to Turing patterns, owing to more comprehensive spatial pattern formation and multistability, including the occurrence of coexistence states. We highlighted that both types of spatial pattern formation originate before and can persist beyond tipping points, demonstrating various pathways evading tipping while strongly enhancing the resilience. We emphasized that such spatial pattern formation and multistability have also been observed recently for real ecosystems, and we argued that this may be relevant for many ecosystems and Earth system components, including for the complex Earth system as a whole. Savannas can be considered an archetypal ecosystem to further investigate this because of observations of tipping phenomena together with spatial pattern formation. A better understanding of the dynamics of spatial pattern formation in general is needed to determine how these patterns respond to external changes of various magnitudes and rates and to localized and nonlocal homogenizing perturbations. Such an understanding will help determine which conditions and spatial patterns lead to the evasion of tipping and which do not. We expect that identifying these in the many ecosystems and Earth system components that are supposedly tipping prone will reveal that some are actually much more resilient than currently thought.

REFERENCES AND NOTES

1. M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001). doi: 10.1038/35098000; pmid: 11595939
2. M. Scheffer *et al*., Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009). doi: 10.1038/nature08227; pmid: 19727193
3. W. Steffen *et al*., Trajectories of the Earth System in the Anthropocene. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8252–8259 (2018). doi: 10.1073/pnas.1810141115; pmid: 30082409
4. J. van de Koppel, M. Rietkerk, F. J. Weissing, Catastrophic vegetation shifts and soil degradation in terrestrial grazing systems. *Trends Ecol. Evol.* **12**, 352–356 (1997). doi: 10.1016/S0169-5347(97)01133-6; pmid: 21238102
5. M. Rietkerk, F. van den Bosch, J. van de Koppel, Site-specific properties and irreversible vegetation changes in semi-arid ecosystems. *Oikos* **80**, 241–252 (1997). doi: 10.2307/3546592
6. M. Scheffer, S. H. Hosper, M.-L. Meijer, B. Moss, E. Jeppesen, Alternative equilibria in shallow lakes. *Trends Ecol. Evol.* **8**, 275–279 (1993). doi: 10.1016/0169-5347(93)90254-M; pmid: 21236168
7. M. Rietkerk, J. van de Koppel, Alternate stable states and threshold effects in semi-arid grazing systems. *Oikos* **79**, 69–76 (1997). doi: 10.2307/3546091
8. M. Hirota, M. Holmgren, E. H. Van Nes, M. Scheffer, Global resilience of tropical forest and savanna to critical transitions. *Science* **334**, 232–235 (2011). doi: 10.1126/science.1210657; pmid: 21998390
9. A. C. Staver, S. Archibald, S. A. Levin, The global extent and determinants of savanna and forest as alternative biome states. *Science* **334**, 230–232 (2011). doi: 10.1126/science.1210465; pmid: 21998389
10. A. C. Staver, S. A. Levin, Integrating theoretical climate and fire effects on savanna and forest systems. *Am. Nat.* **180**, 211–224 (2012). doi: 10.1086/666648; pmid: 22766932
11. F. Van Langevelde *et al*., Effects of fire and herbivory on the stability of savanna ecosystems. *Ecology* **84**, 337–350 (2003). doi: 10.1890/0012-9658(2003)084[0337:EOFAHO]2.0.CO;2
12. J. C. Aleman *et al*., Floristic evidence for alternative biome states in tropical Africa. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28183–28190 (2020). doi: 10.1073/pnas.2011515117; pmid: 33109722
13. C. S. Holling, in *Engineering Within Ecological Constraints*, P. E. Schulze, Ed. (National Academy Press, 1996), pp. 31–43.
14. P. Ashwin, S. Wieczorek, R. Vitolo, P. Cox, Tipping points in open systems: Bifurcation, noise-induced and rate-dependent examples in the climate system. *Phil. Trans. R. Soc. A* **370**, 1166–1184 (2012). doi: 10.1098/rsta.2011.0306; pmid: 22291228
15. T. M. Lenton *et al*., Tipping elements in the Earth's climate system. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1786–1793 (2008). doi: 10.1073/pnas.0705414105; pmid: 18258748
16. S. Drijfhout *et al*., Catalogue of abrupt shifts in Intergovernmental Panel on Climate Change climate models. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5777–E5786 (2015). doi: 10.1073/pnas.1511451112; pmid: 26460042
17. P. Huybrechts, J. de Wolde, The dynamic response of the Greenland and Antarctic ice sheets to multiple-century climatic warming. *J. Clim.* **12**, 2169–2188 (1999). doi: 10.1175/1520-0442(1999)012<2169:TDROTG>2.0.CO;2
18. F. Pattyn, M. Morlighem, The uncertain future of the Antarctic ice sheet. *Science* **367**, 1331–1335 (2020). doi: 10.1126/science.aaz5487; pmid: 32193321
19. J. Garbe, T. Albrecht, A. Levermann, J. F. Donges, R. Winkelmann, The hysteresis of the Antarctic ice sheet. *Nature* **585**, 538–544 (2020). doi: 10.1038/s41586-020-2727-5; pmid: 32968257
20. T. F. Stocker, D. G. Wright, Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991). doi: 10.1038/351729a0
21. S. Rahmstorf, A. Ganopolsky, Long-term global warning scenario's computed with an efficient coupled climate model. *Clim. Change* **43**, 353–367 (1999). doi: 10.1023/A:1005474526406
22. J. Lohmann, P. D. Ditlevsen, Risk of tipping the overturning circulation due to increasing rates of ice melt. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e20179891 (2021). doi: 10.1073/pnas.2017989118; pmid: 33619095
23. J. Rockström *et al*., A safe operating space for humanity. *Nature* **461**, 472–475 (2009). doi: 10.1038/461472a; pmid: 19779433
24. A. D. Barnosky *et al*., Approaching a state shift in Earth's biosphere. *Nature* **486**, 52–58 (2012). doi: 10.1038/nature11018; pmid: 22678279
25. M. Rietkerk, S. C. Dekker, P. C. de Ruiter, J. van de Koppel, Self-organized patchiness and catastrophic shifts in ecosystems. *Science* **305**, 1926–1929 (2004). doi: 10.1126/science.1101867; pmid: 15448261
26. A. M. Turing, The chemical basis of morphogenesis. *Bull. Math. Biol.* **52**, 153–197 (1990). doi: 10.1007/BF02459572; pmid: 2185858
27. C. A. Klausmeier, Regular and irregular patterns in semiarid vegetation. *Science* **284**, 1826–1828 (1999). doi: 10.1126/science.284.5421.1826; pmid: 10364553
28. J. von Hardenberg, E. Meron, M. Shachak, Y. Zarmi, Diversity of vegetation patterns and desertification. *Phys. Rev. Lett.* **87**, 198101 (2001). doi: 10.1103/PhysRevLett.87.198101
29. M. Rietkerk *et al*., Self-organization of vegetation in arid ecosystems. *Am. Nat.* **160**, 524–530 (2002). doi: 10.1086/342078; pmid: 18707527
30. K. Gowda, H. Riecke, M. Silber, Transitions between patterned states in vegetation models for semiarid ecosystems. *Phys. Rev. E* **89**, 022701 (2014). doi: 10.1103/PhysRevE.89.022701; pmid: 25353503
31. M. Rietkerk, J. van de Koppel, Regular pattern formation in real ecosystems. *Trends Ecol. Evol.* **23**, 169–175 (2008). doi: 10.1016/j.tree.2007.10.013; pmid: 18255188
32. K. Siteur *et al*., Beyond Turing: The response of patterned ecosystems to environmental change. *Ecol. Complex.* **20**, 81–96 (2014). doi: 10.1016/j.ecocom.2014.09.002
33. R. Bastiaansen *et al*., Multistability of model and real dryland ecosystems through spatial self-organization. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11256–11261 (2018). doi: 10.1073/pnas.1804771115; pmid: 30322906
34. R. Bastiaansen, A. Doelman, M. B. Eppinga, M. Rietkerk, The effect of climate change on the resilience of ecosystems with adaptive spatial pattern formation. *Ecol. Lett.* **23**, 414–429 (2020). doi: 10.1111/ele.13449; pmid: 31912954

35. F. H. Busse, Non-linear properties of thermal convection. *Rep. Prog. Phys.* **41**, 1929–1967 (1978). doi: 10.1088/0034-4885/41/12/003
36. R. H. Wang, Q. X. Liu, G. Q. Sun, Z. Jin, J. van de Koppel, Nonlinear dynamic and pattern bifurcations in a model for spatial patterns in young mussel beds. *J. R. Soc. Interface* **6**, 705–718 (2009). doi: 10.1098/rsif.2008.0439; pmid: 18986965
37. Q. X. Liu *et al*., Pattern formation at multiple spatial scales drives the resilience of mussel bed ecosystems. *Nat. Commun.* **5**, 5234 (2014). doi: 10.1038/ncomms6234; pmid: 25335554
38. E. H. van Nes, M. Hirota, M. Holmgren, M. Scheffer, Tipping points in tropical tree cover: Linking theory to data. *Glob. Change Biol.* **20**, 1016–1021 (2014). doi: 10.1111/gcb.12398; pmid: 24106057
39. J. Pastor, B. Peckham, S. Bridgham, J. Weltzin, J. Chen, Plant community dynamics, nutrient cycling, and alternative stable equilibria in peatlands. *Am. Nat.* **160**, 553–568 (2002). doi: 10.1086/342814; pmid: 18707507
40. M. J. Donahue, R. A. Desharnais, C. D. Robles, P. Arriola, Mussel bed boundaries as dynamic equilibria: Thresholds, phase shifts, and alternative states. *Am. Nat.* **178**, 612–625 (2011). doi: 10.1086/662177; pmid: 22030731
41. J. Van de Koppel, P. M. J. Herman, P. Thoolen, C. H. R. Heip, Do alternate stable states occur in natural ecosystems? Evidence from a tidal flat. *Ecology* **82**, 3449–3461 (2001). doi: 10.1890/0012-9658(2001)082[3449:DASSOI]2.0.CO;2
42. O. D. Vinent, L. J. Moore, Barrier island bistability induced by biophysical interactions. *Nat. Clim. Chang.* **5**, 158–162 (2015). doi: 10.1038/NCLIMATE2474
43. T. Van der Heide *et al*., Positive feedbacks in seagrass ecosystems: Implications for success in conservation and restoration. *Ecosystems* **10**, 1311–1322 (2007). doi: 10.1007/s10021-007-9099-7
44. T. van der Heide *et al*., Spatial self-organized patterning in seagrasses along a depth gradient of an intertidal ecosystem. *Ecology* **91**, 362–369 (2010) doi: 10.1890/08-1567.1; pmid: 20392001
45. E. D. Lazarus, S. Armstrong, Self-organized pattern formation in coastal barrier washover deposits. *Geology* **43**, 363–366 (2015). doi: 10.1130/G36329.1
46. D. Ruiz-Reynés *et al*., Fairy circle landscapes under the sea. *Sci. Adv.* **3**, e1603262 (2017). doi: 10.1126/sciadv.1603262; pmid: 28782035
47. V. Reijers, "A song of sand and mud: How plant-mediated feedbacks dictate landscape formation and dynamics of barrier islands," thesis, Radboud University, Nijmegen (2019).
48. P. D. L. Ritchie, J. J. Clarke, P. M. Cox, C. Huntingford, Overshooting tipping point thresholds in a changing climate. *Nature* **592**, 517–523 (2021). doi: 10.1038/s41586-021-03263-2; pmid: 33883733
49. H. Hillebrand *et al*., Thresholds for ecological responses to global change do not emerge from empirical data. *Nat. Ecol. Evol.* **4**, 1502–1509 (2020). doi: 10.1038/s41559-020-1256-9; pmid: 32807945
50. G. Bel, A. Hagberg, E. Meron, Gradual regime shifts in spatially extended ecosystem. *Theor. Ecol.* **5**, 591–604 (2012). doi: 10.1007/s12080-011-0149-6
51. Y. R. Zelnik, E. Meron, Regime shifts by front dynamics. *Ecol. Indic.* **94**, 544–552 (2018). doi: 10.1016/j.ecolind.2017.10.068
52. M. Ghil, Climate stability for a Sellers-type model. *J. Atmos. Sci.* **33**, 3–20 (1976). doi: 10.1175/1520-0469(1976)033<0003:CSFAST>2.0.CO;2
53. T. Bódai, V. Lucarini, F. Lunkeit, R. Boschi, Global instability in the Ghil–Sellers model. *Clim. Dyn.* **44**, 3361–3381 (2015). doi: 10.1007/s00382-014-2206-5
54. S. H. R. Rosier *et al*., The tipping points and early-warning indicators for Pine Island Glacier, West Antarctica. *Cryosphere* **15**, 1501–1516 (2021). doi: 10.5194/tc-15-1501-2021
55. C. Fernandez-Oto, O. Tzuk, E. Meron, Front instabilities can reverse desertification. *Phys. Rev. Lett.* **122**, 048101 (2019). doi: 10.1103/PhysRevLett.122.048101; pmid: 30768298
56. M. E. Cates, J. Tailleur, Motility-induced phase separation. *Annu. Rev. Condens. Matter Phys.* **6**, 219–244 (2015). doi: 10.1146/annurev-conmatphys-031214-014710
57. Q.-X. Liu *et al*., Phase separation explains a new class of self-organized spatial patterns in ecological systems. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11905–11910 (2013). doi: 10.1073/pnas.1222339110; pmid: 23818579
58. Q.-X. Liu *et al*., Phase separation driven by density-dependent movement: A novel mechanism for ecological patterns. *Phys. Life Rev.* **19**, 107–121 (2016). doi: 10.1016/j.plrev.2016.07.009; pmid: 27478087
59. E. Sheffer, J. von Hardenberg, H. Yizhaq, M. Shachak, E. Meron, Emerged or imposed: A theory on the role of physical templates and self-organisation for vegetation patchiness. *Ecol. Lett.* **16**, 127–139 (2013). doi: 10.1111/ele.12027; pmid: 23157578
60. R. J. Scholes, S. R. Archer, Tree-grass interactions in savannas. *Annu. Rev. Ecol. Syst.* **28**, 517–544 (1997). doi: 10.1146/annurev.ecolsys.28.1.517
61. J. D. Anadon, O. E. Sala, F. T. Maestre, Climate change will increase savannas at the expense of forests and treeless vegetation in tropical and subtropical Americas. *J. Ecol.* **102**, 1363–1373 (2014). doi: 10.1111/1365-2745.12325
62. M. Garcia Criado, I. H. Meyers-Smith, A. D. Bjorkman, C. E. R. Lehman, N. Stevens, Woody plant encroachment intensifies under climate change across tundra and savanna biomes. *Glob. Ecol. Biogeogr.* **29**, 925–943 (2020). doi: 10.1111/geb.13072
63. A. C. Staver, Prediction and scale in savanna ecosystems. *New Phytol.* **219**, 52–57 (2018). doi: 10.1111/nph.14829; pmid: 29027662
64. F. Accatino, C. De Michele, R. Vezzoli, D. Donzelli, R. J. Scholes, Tree–grass co-existence in savanna: Interactions of rain and fire. *J. Theor. Biol.* **267**, 235–242 (2010). doi: 10.1016/j.jtbi.2010.08.012; pmid: 20708629
65. M. Baudena, F. D'Andrea, A. Provenzale, An idealized model for tree-grass coexistence in savannas: The role of life stage structure and fire disturbances. *J. Ecol.* **98**, 74–80 (2010). doi: 10.1111/j.1365-2745.2009.01588.x
66. V. de L. Dantas, M. Hirota, R. S. Oliveira, J. G. Pausas, Disturbance maintains alternative biome states. *Ecol. Lett.* **19**, 12–19 (2016). doi: 10.1111/ele.12537; pmid: 26493189
67. D. D'Onofrio, J. von Hardenberg, M. Baudena, Not only trees: Grasses determine African tropical biome distributions via water limitation and fire. *Glob. Ecol. Biogeogr.* **27**, 714–725 (2018). doi: 10.1111/geb.12735
68. A. W. Cardoso *et al*., A distinct ecotonal tree community exists at central African forest-savanna transitions. *J. Ecol.* **109**, 1170–1183 (2021). doi: 10.1111/1365-2745.13549
69. T. Charles-Dominique, G. F. Midgley, K. W. Tomlinson, W. J. Bond, Steal the light: Shade vs fire adapted vegetation in forest-savanna mosaics. *New Phytol.* **218**, 1419–1429 (2018). doi: 10.1111/nph.15117; pmid: 29604213
70. T. A. Groen, "Spatial matters: How spatial patterns and processes affect savanna dynamics," thesis, Wageningen University (2007).
71. M. Baudena, M. Rietkerk, Complexity and coexistence in a simple spatial model for arid savanna ecosystems. *Theor. Ecol.* **6**, 131–141 (2013). doi: 10.1007/s12080-012-0165-1
72. R. Martínez-García, J. M. Calabrese, C. López, Spatial patterns in mesic savannas: The local facilitation limit and the role of demographic stochasticity. *J. Theor. Biol.* **333**, 156–165 (2013). doi: 10.1016/j.jtbi.2013.05.024; pmid: 23747988
73. T. A. Groen, C. A. D. M. van de Vijver, F. van Langevelde, Do spatially homogenizing and heterogenizing processes affect transitions between alternative stable states? *Ecol. Modell.* **65**, 119–128 (2017). doi: 10.1016/j.ecolmodel.2017.10.002
74. S. Eby, A. Agrawal, S. Majumder, A. P. Dobson, V. Guttal, Alternative stable states and spatial indicators of critical slowing down along a spatial gradient in a savanna ecosystem. *Glob. Ecol. Biogeogr.* **26**, 638–649 (2017). doi: 10.1111/geb.12570
75. L. Eigentler, J. A. Sherratt, Metastability as a coexistence mechanism in a model for dryland vegetation patterns. *Bull. Math. Biol.* **81**, 2290–2322 (2019). doi: 10.1007/s11538-019-00606-z; pmid: 31012031
76. O. Tzuk, H. Uecker, E. Meron, The role of spatial self-organization in the design of agroforestry systems. *PLOS ONE* **15**, e0236325 (2020). doi: 10.1371/journal.pone.0236325; pmid: 32692773
77. B. Wuyts, A. R. Champneys, N. Verschueren, J. I. House, Tropical tree cover in a heterogeneous environment: A reaction-diffusion model. *PLOS ONE* **14**, e0218151 (2019). doi: 10.1371/journal.pone.0218151; pmid: 31246968
78. V. Yatat, P. Couteron, J. J. Tewa, S. Bowong, Y. Dumont, An impulsive modelling framework of fire occurrence in a size-structured model of tree-grass interactions for savanna ecosystems. *J. Math. Biol.* **74**, 1425–1482 (2017). doi: 10.1007/s00285-016-1060-y; pmid: 27659304
79. V. Yatat, P. Couteron, Y. Dumont, Spatially explicit modelling of tree–grass interactions in fire-prone savannas: A partial differential equations framework. *Ecol. Complex.* **36**, 290–313 (2018). doi: 10.1016/j.ecocom.2017.06.004
80. C. De Michele, F. Accatino, R. Vezzoli, R. J. Scholes, Savanna domain in the herbivores-fire parameter space exploiting a tree–grass–soil water dynamic model. *J. Theor. Biol.* **289**, 74–82 (2011). doi: 10.1016/j.jtbi.2011.08.014; pmid: 21875600
81. Q. Li, A. C. Staver, E. Weinan, S. A. Levin, Spatial feedbacks and the dynamics of savanna and forest. *Theor. Ecol.* **12**, 237–262 (2019). doi: 10.1007/s12080-019-0428-1
82. E. Siero *et al*., Grazing away the resilience of patterned ecosystems. *Am. Nat.* **193**, 472–480 (2019). doi: 10.1086/701669; pmid: 30794443
83. R. M. Holdo, R. D. Holt, J. M. Fryxell, Herbivore-vegetation feedbacks can expand the range of savanna persistence: Insights from a simple theoretical model. *Oikos* **122**, 441–453 (2013). doi: 10.1111/j.1600-0706.2012.20735.x
84. J. Sherratt, Numerical continuation methods for studying periodic travelling wave (wavetrain) solutions of partial differential equations. *Appl. Math. Comput.* **218**, 4684–4694 (2012). doi: 10.1016/j.amc.2011.11.005
85. S. Van der Stelt, A. Doelman, G. M. Hek, J. Rademacher, Rise and fall of periodic patterns for a generalized Klausmeier–Gray–Scott model. *J. Nonlinear Sci.* **23**, 39–95 (2013). doi: 10.1007/s00332-012-9139-0
86. A. Staal, S. C. Dekker, C. Xu, E. Van Nes, Bistability, spatial interactions and the distribution of tropical forests and savannas. *Ecosystems* **19**, 1080–1091 (2016). doi: 10.1007/s10021-016-0011-1
87. J. C. Aleman, A. C. Staver, Spatial patterns in the global distributions of savanna and forest. *Glob. Ecol. Biogeogr.* **27**, 792–803 (2018). doi: 10.1111/geb.12739
88. D. R. Uden *et al*., Spatial imaging and screening for regime shifts. *Front. Ecol. Evol.* **7**, 407 (2019). doi: 10.3389/fevo.2019.00407
89. A. C. Staver, G. P. Asner, I. Rodriguez-Iturbe, S. A. Levin, I. P. J. Smit, Spatial patterning among savanna trees in high-resolution, spatially extensive data. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10681–10685 (2019). doi: 10.1073/pnas.1819391116; pmid: 31085650
90. W. Zhang *et al*., From woody cover to woody canopies: How Sentinel-1 and Sentinel-2 data advance the mapping of woody plants in savannas. *Remote Sens. Environ.* **234**, 111465 (2019). doi: 10.1016/j.rse.2019.111465
91. G. Xanthopoulos, G. Maheras, V. Gouma, M. Gouvas, Is the Keetch–Byram drought index (KBDI) directly related to plant water stress? *For. Ecol. Manage.* **234**, S27 (2006). doi: 10.1016/j.foreco.2006.08.043
92. L. Wang, J. J. Qu, Satellite remote sensing applications for surface soil moisture monitoring: A review. *Front. Earth Sci. China* **3**, 237–247 (2009). doi: 10.1007/s11707-009-0023-7
93. T. Long *et al*., 30 m Resolution Global Annual Burned Area Mapping Based on Landsat Images and Google Earth Engine. *Remote Sens.* **11**, 489 (2019). doi: 10.3390/rs11050489
94. R. B. Hoyle, *Pattern Formation: An Introduction to Methods* (Cambridge Univ. Press, 2006).
95. A. Doelman, in *Complexity Science*, M. Peletier, R. van Santen, E. Steur, Eds. (World Scientific, 2019), pp. 129–192.
96. R. Bastiaansen, A. Doelman, The dynamics of disappearing pulses in a singularly perturbed reaction-diffusion systems with parameters that vary in time and space. *Physica D* **388**, 45–72 (2019). doi: 10.1016/j.physd.2018.09.003
97. R. Bastiaansen, P. Carter, A. Doelman, Stable planar vegetation stripe patterns on sloped terrain in dryland ecosystems. *Nonlinearity* **32**, 2759–2814 (2019). doi: 10.1088/1361-6544/ab1767
98. O. Jaïbi, A. Doelman, M. Chirilus-Bruckner, E. Meron, The existence of localized vegetation patterns in a systematically reduced model for dryland vegetation. *Physica D* **412**, 132637 (2020). doi: 10.1016/j.physd.2020.132637
99. M. Scheffer *et al*., Anticipating critical transitions. *Science* **338**, 344–348 (2012). doi: 10.1126/science.1225244; pmid: 23087241
100. J. N. Kutz, S. L. Brunton, B. W. Brunton, J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (Society for Industrial and Applied Mathematics, 2016).
101. P. Van Heijster, B. Sandstede, Bifurcations to travelling planar spots in a three-component FitzHugh–Nagumo system. *Physica D* **275**, 19–34 (2014). doi: 10.1016/j.physd.2014.02.001
102. B. Wuyts, A. R. Champneys, J. I. House, Correction: Author Correction: Amazonian forest-savanna bistability and human impact. *Nat. Commun.* **9**, 16179 (2018). doi: 10.1038/ncomms16179; pmid: 29465082
103. A. Doelman, P. van Heijster, J. Shen, Pulse dynamics in reaction–diffusion equations with strong spatially localized

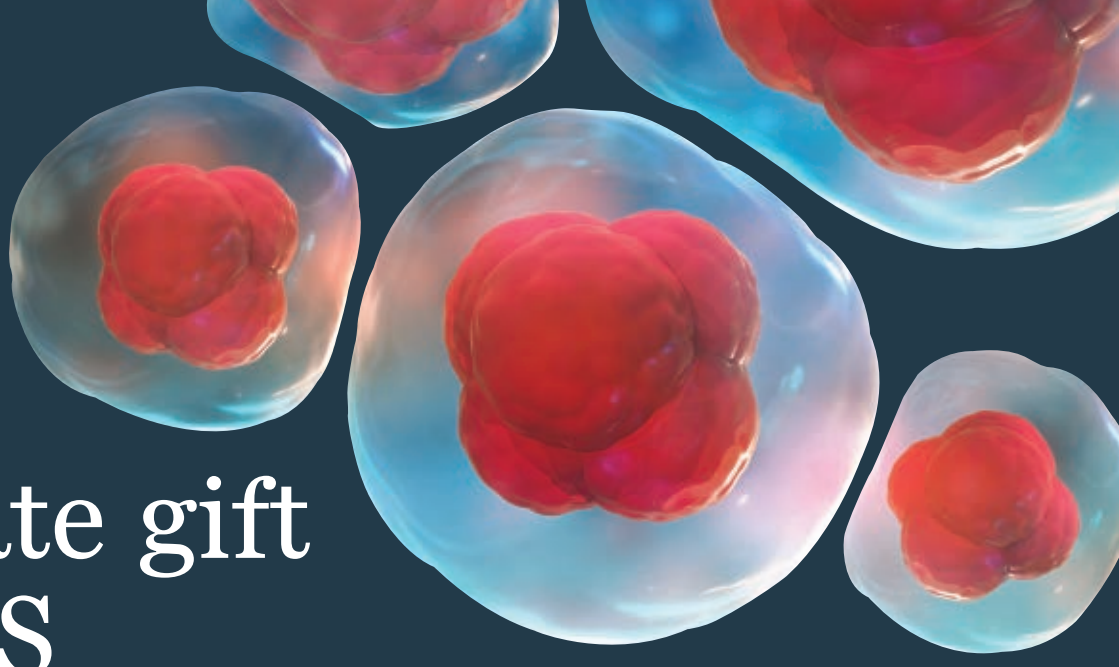
impurities. *Phil. Trans. R. Soc. A* **376**, 20170183 (2018). doi: 10.1098/rsta.2017.0183; pmid: 29507168
104. L. Eigentler, J. A. Sherratt, Spatial self-organisation enables species coexistence in a model for savanna ecosystems. *J. Theor. Biol.* **487**, 110122 (2020). doi: 10.1016/j.jtbi.2019.110122; pmid: 31862451
105. M. I. Budyko, The effect of solar radiation variations on the climate of the Earth. *Tellus* **21**, 611–619 (1969). doi: 10.3402/tellusa.v21i5.10109
106. W. D. Sellers, A global climatic model based on the energy balance of the earth-atmosphere system. *J. Appl. Meteorol.* **8**, 392–400 (1969). doi: 10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2
107. H. Stommel, Thermohaline convection with two stable regimes of flow. *Tellus* **13**, 224–230 (1961). doi: 10.1111/j.2153-3490.1961.tb00079.x
108. H. Gildor, E. Tziperman, Physical mechanisms behind biogeochemical glacial-interglacial CO_2 variations. *Geophys. Res. Lett.* **28**, 2421–2424 (2001). doi: 10.1029/2000GL012571
109. H. M. Alkhayoun, P. Ashwin, L. C. Jackson, C. Quinn, R. A. Wood, Basin bifurcation, oscillatory instability and scale-induced thresholds for Atlantic meridonal overturning circulation in a global box model. *Proc. R. Soc. A* **475**, 20190051 (2019). doi: 10.1098/rspa.2019.0051
110. P. Valdes, Built for stability. *Nat. Geosci.* **4**, 414–416 (2011). doi: 10.1038/ngeo1200

ACKNOWLEDGMENTS

M. te Beest and A. Staal critically reviewed earlier versions of this manuscript. M. Eppinga, J. Rademacher, E. Siero, K. Siteur, and S. van der Stelt contributed to research that led to new insights leading to this paper. We thank T. Markus for designing figures. **Funding:** The research leading to this paper was funded by the Netherlands Organization of Scientific Research, NWO Complexity and NWO Mathematics of Planet Earth programs. This project is TiPES contribution no. 83. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 820970. **Author contributions:** All authors contributed to the conceptualization and research ideas. M.R., A.D., and R.B. wrote the original draft, and all authors contributed to reviewing and editing the paper. M.R., R.B., S.B., J.v.d.K., and M.B. prepared visual materials. **Competing interests:** The authors declare no competing interests.

10.1126/science.abj0359

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

RESEARCH ARTICLE SUMMARY

CIRCADIAN RHYTHMS

Reconstitution of an intact clock reveals mechanisms of circadian timekeeping

Archana G. Chavan†, Jeffrey A. Swan†, Joel Heisler†, Cigdem Sancar, Dustin C. Ernst, Mingxu Fang, Joseph G. Palacios, Rebecca K. Spangler, Clive R. Bagshaw, Sarvind Tripathi, Priya Crosby, Susan S. Golden, Carrie L. Partch*, Andy LiWang*

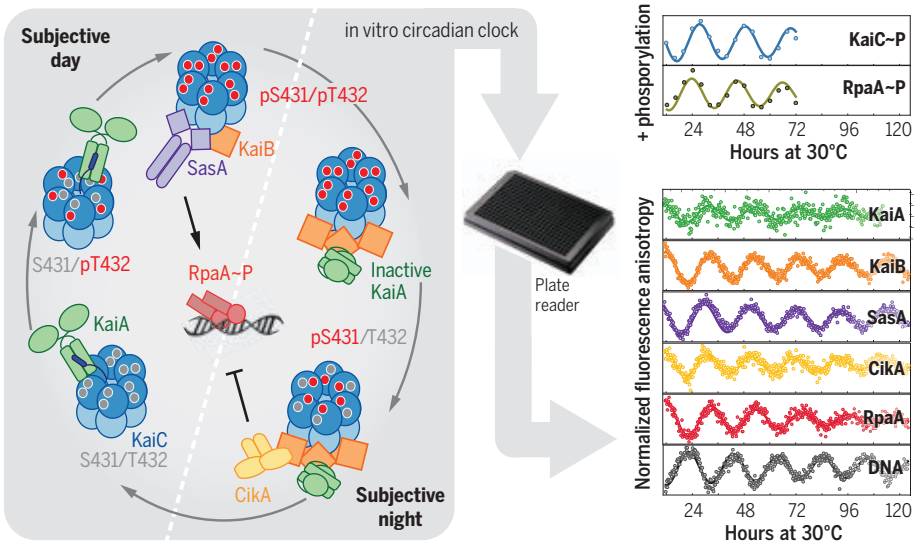
INTRODUCTION: Circadian clocks provide an internal representation of local time inside cells and control the timing of gene expression in anticipation of sunrise and sunset. In cyanobacteria, timekeeping is achieved by way of an oscillator comprising three Kai (meaning "cycle" in Japanese) proteins, KaiA, KaiB, and KaiC, which relay temporal information downstream through two sensor histidine kinases, SasA and CikA, to regulate the transcription factor RpaA. Identifying the specific mechanisms by which circadian clocks exert temporal control over gene expression has proven challenging in the complex milieu of cells. Thus, we reassembled an intact clock, including the core oscillator and signal transduction components, under defined conditions in vitro. Together with structural studies and biochemical analyses of partial clock reactions, we acquired insights into mechanisms by which the cyanobacterial circadian clock functions to control gene expression.

RATIONALE: Although the core oscillator can be reconstituted in vitro with just the Kai proteins, these reactions assayed by SDS–polyacrylamide gel electrophoresis report only on the phosphorylation status of KaiC and have therefore not offered mechanistic insight into clock protein interactions or signal transduction. Thus, we developed a fluorescence polarization–based in vitro whole-clock system that includes the KaiABC oscillator, SasA and/or CikA, RpaA, and a clock promoter–bearing DNA fragment. This in vitro clock (IVC) oscillates autonomously for days and allows monitoring of each component individually in real time, revealing the formation and phase of macromolecular assemblies from the timekeeper to DNA binding by RpaA.

RESULTS: The IVC was used to examine the sufficiency of SasA or CikA to maintain clock output under constant conditions and identify how each contributes to the phase of DNA binding by RpaA; it even allowed the dissection of the breakdown in signaling of an arrhythmic mutant of RpaA. We challenged a long-standing paradigm that timekeeping by the oscillator depends solely on the Kai proteins, with SasA and CikA providing only input-output signaling. The in vitro oscillator is known to function under a relatively narrow set of Kai protein concentrations; we show here that a KaiABC-only mixture that fails to oscillate in a sustained manner owing to limiting levels of KaiB can be rescued by SasA, which acts to recruit KaiB to the KaiC hexamer through heterotropic cooperativity. Cooperativity is based on structural mimicry between SasA and KaiB, and mutations that eliminate heterocooperativity in vitro profoundly affect circadian rhythms in vivo. CikA also rescues period defects under low levels of KaiA. Together, our data help to explain how the clock compensates in vivo for changes in concentrations of oscillator components that occur as part of the transcription-translation feedback loop and protein turnover. The intimate coupling between oscillator and input-output components blurs their distinction, although the extent to which this coupling can be generalized to eukaryotic clocks remains an open question.

CONCLUSION: We developed the reconstituted IVC to establish causal links between clock biochemistry and in vivo phenotypes. Our system provides a powerful platform to explore, for example, how changes in factors such as temperature or adenosine triphosphate levels are compensated to allow generation of reliable circadian rhythms of transcription.

In fact, this IVC is just a starting point, one with many possibilities. For example, the real-time data presented here will allow mathematical models of the cyanobacterial clock to extend beyond the KaiABC oscillator to address the events that effect temporal control of physiology and metabolism. As a next step, transplanting this IVC to artificial cells could enable single-cell microscopy experiments of this streamlined clock under conditions that are much closer to physiological. In an artificial cellular system, transcriptional and translational networks could be included, allowing higher-complexity customized frequencies and outputs to be engineered, opening the door to synthetic biology applications. ■



Reconstituting an intact clock in vitro. Cyanobacterial circadian proteins KaiA, KaiB, KaiC, SasA, CikA, RpaA, and a promoter-bearing DNA fragment were reconstituted into an in vitro clock. Together with structural studies and biochemical assays, this system allowed dissection of a circadian clock system to demonstrate that the concept of the core oscillator should be extended to include output proteins SasA and CikA. The dashed line separates subjective day and subjective night, which refer to the halves of a circadian period under constant conditions that correspond to day and night in a light-dark cycle. S, serine; T, threonine; p, phosphorylated.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: aliwang@ucmerced.edu (A.L.); cpartch@ucsc.edu (C.L.P.)
†These authors contributed equally to this work.
Cite this article as A. G. Chavan *et al.*, *Science* **374**, eabd4453 (2021). DOI: 10.1126/science.abd4453

S READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd4453

RESEARCH ARTICLE

CIRCADIAN RHYTHMS

Reconstitution of an intact clock reveals mechanisms of circadian timekeeping

Archana G. Chavan[1]†, Jeffrey A. Swan[2]†, Joel Heisler[3]†, Cigdem Sancar[4], Dustin C. Ernst[4], Mingxu Fang[4], Joseph G. Palacios[2], Rebecca K. Spangler[2], Clive R. Bagshaw[2], Sarvind Tripathi[2], Priya Crosby[2], Susan S. Golden[4,5], Carrie L. Partch[2,4]*, Andy LiWang[1,3,4,6,7]*

Circadian clocks control gene expression to provide an internal representation of local time. We report reconstitution of a complete cyanobacterial circadian clock in vitro, including the central oscillator, signal transduction pathways, downstream transcription factor, and promoter DNA. The entire system oscillates autonomously and remains phase coherent for many days with a fluorescence-based readout that enables real-time observation of each component simultaneously without user intervention. We identified the molecular basis for loss of cycling in an arrhythmic mutant and explored fundamental mechanisms of timekeeping in the cyanobacterial clock. We find that SasA, a circadian sensor histidine kinase associated with clock output, engages directly with KaiB on the KaiC hexamer to regulate period and amplitude of the central oscillator. SasA uses structural mimicry to cooperatively recruit the rare, fold-switched conformation of KaiB to the KaiC hexamer to form the nighttime repressive complex and enhance rhythmicity of the oscillator, particularly under limiting concentrations of KaiB. Thus, the expanded in vitro clock reveals previously unknown mechanisms by which the circadian system of cyanobacteria maintains the pace and rhythmicity under variable protein concentrations.

Circadian clocks are intracellular systems that provide organisms with an internal representation of local time and have profound consequences to health (*1*). The clock in the cyanobacterium *Synechococcus elongatus* generates circadian rhythms of genetic, physiological, and metabolic activities that enhance fitness (*2*, *3*). The core circadian clock genes of cyanobacteria, *kaiA*, *kaiB*, and *kaiC*, are essential for rhythmic gene expression (*4*), and their proteins generate an autonomous ~24-hour rhythm of KaiC phosphorylation in vivo (*5*), which can be recapitulated in vitro (*6*). Ordered phosphorylation of KaiC at residues S431 and T432 (*7*, *8*), stimulated by KaiA during the day (*9*, *10*) and suppressed by KaiB at night (*11*, *12*), helps set the circadian pace and phase of the oscillator (*13*–*15*). The input-output sensor histidine kinases, *Synechococcus* adaptive sensor A, SasA (*16*), and circadian input kinase A, CikA (*17*), interact with Kai proteins (*14*, *18*) to regulate the master transcription factor, RpaA (*19*), and generate circadian rhythms of gene expression in cyanobacteria (*20*, *21*).

Although the reconstituted KaiABC oscillator has been a powerful tool to study mechanisms of circadian timekeeping in cyanobacteria (*6*, *15*, *22*–*24*), it does not include the signal transduction machinery necessary to provide a readout of timing encoded by the biochemical oscillator. To explore how each component interacts in the network and how output proteins integrate with the core oscillator, we reconstituted clock-controlled DNA binding in a cell-free system that enables real-time reporting from different partners. In addition to the core oscillator components KaiA, KaiB, and KaiC, this in vitro clock (IVC) includes SasA and/or CikA, RpaA, and a DNA duplex that carries a clock-controlled promoter, demonstrating that these components are sufficient to form the fundamental regulatory network for autonomous timekeeping and temporal relay to downstream events. We then use this IVC to identify how a loss-of-function allele in RpaA disrupts circadian rhythms, highlighting its power to provide mechanistic insight for in vivo studies. Moreover, through careful deconstruction of our expanded IVC, we discovered that SasA helps to cooperatively recruit KaiB to the KaiC hexamer, thus defining a previously unknown role for SasA in the core oscillator that it is required for proper circadian rhythms in vivo.

Signal transduction in a reconstituted IVC

To monitor protein interactions in the IVC, we attached fluorescent probes (*25*) to each of the individual clock protein components, KaiA, KaiB, SasA, CikA, and RpaA (see methods in the supplementary materials and tables S1 to S3) using sortase A–mediated ligation (*26*, *27*) and monitored their association into larger complexes using fluorescence anisotropy (FA) (Fig. 1, A to C, and fig. S1).

Fluorescently labeled synthetic double-stranded DNA representing the promoter sequence of the *kaiBC* operon, P*kaiBC*, was used as a representative RpaA target (*20*), and each labeled species was monitored in separate but parallel clock reactions (table S4). Control experiments were performed to verify the specific binding events that cause changes in FA of each labeled species (figs. S2 and S3), assigning their corresponding peaks to known interactions and circadian properties.

We used a multimode microplate reader to monitor in vitro oscillations in parallel, with individual reactions containing 50 to 100 nM of a single fluorescently labeled species for each protein or DNA monitored, as well as physiological concentrations (*20*) of each unlabeled clock protein. This method allowed measurement of FA for each clock component in a single plate and in real time over a duration of several days.

The data from these experiments were fit to a damped cosine function to extract period and phase relative to maximal KaiC phosphorylation, which was referenced to its peak at circadian time (CT) 14 in vivo (*9*), with CT 0 and CT 12 representing subjective dawn and dusk, respectively, under constant lab conditions. Combining FA measurements with traditional time courses of RpaA and KaiC phosphorylation rhythms measured in parallel by gel electrophoresis (Fig. 1 and figs. S4 and S5), we present a comprehensive view of a fully reconstituted IVC.

KaiC adenosine triphosphatase (ATPase) activity was monitored in real time by proton nuclear magnetic resonance (^{1}H-NMR) imaging (Fig. 1D and fig. S6) (*28*) using the core oscillator to avoid conflation with ATP consumption by the output kinases. We found that the ATPase activity of KaiC peaks during subjective midday with the hydrolysis of ~13 to 18 ATPs per KaiC monomer per day but decreases to a nadir of ~7 or 8 ATPs per KaiC monomer per day during subjective night, which is consistent with earlier studies (*29*). Using this information, we next established the phase relationships of all the clock components under each IVC condition (Fig. 1, E to G). We found that the FA of RpaA increased when it was dephosphorylated or interacting with SasA in the absence of KaiC and that it decreased when RpaA was phosphorylated by CikA or a SasA-KaiC complex (fig. S3), which explains the observed phase reversal between peak RpaA anisotropy and RpaA phosphorylation.

[1]School of Natural Sciences, University of California, Merced, CA 95343, USA. [2]Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, USA. [3]Department of Chemistry and Biochemistry, University of California, Merced, CA 95343, USA. [4]Center for Circadian Biology, University of California, San Diego, La Jolla, CA 92093, USA. [5]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA. [6]Center for Cellular and Biomolecular Machines, University of California, Merced, CA 95343, USA. [7]Health Sciences Research Institute, University of California, Merced, CA 95343, USA.
*Corresponding author. Email: aliwang@ucmerced.edu (A.L.); cpartch@ucsc.edu (C.L.P.)
†These authors contributed equally to this work.

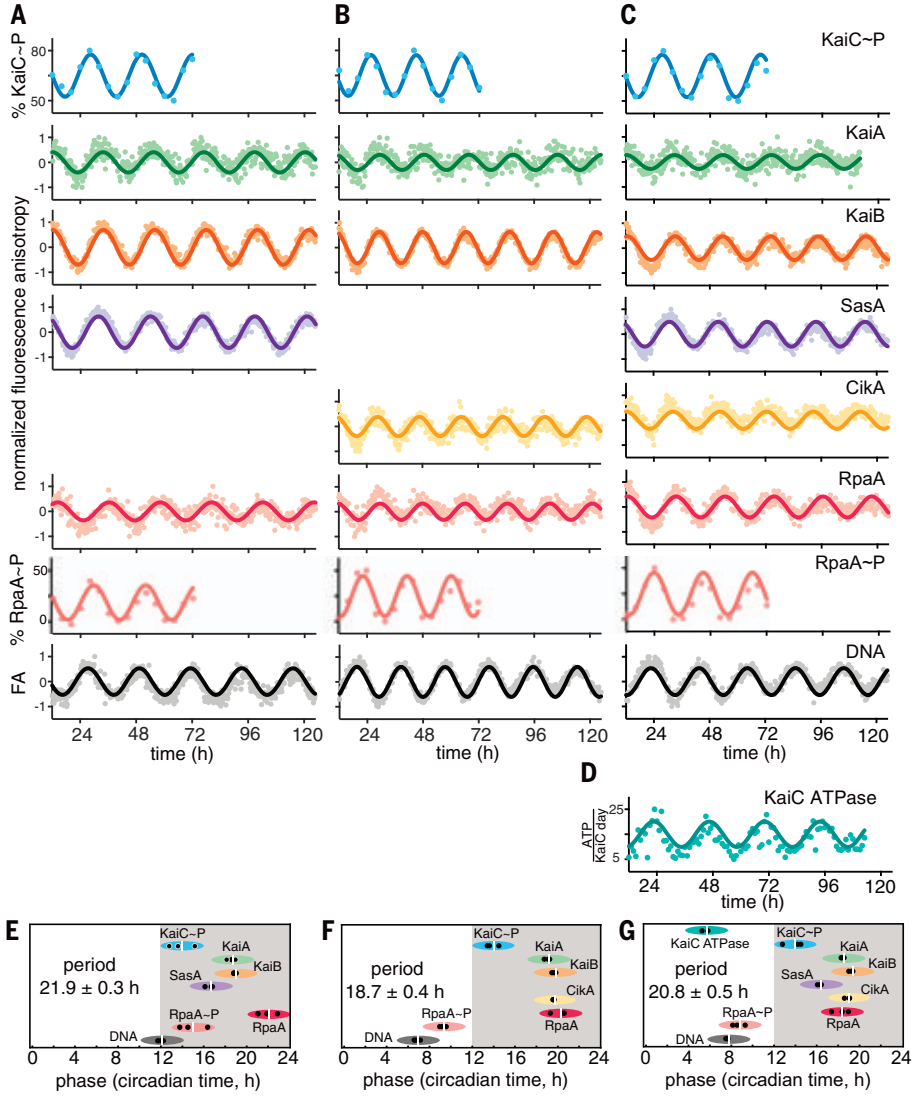


Fig. 1. Reconstitution of an intact clock that controls rhythmic DNA binding in vitro. (**A** to **C**) IVC reactions containing KaiA, KaiB, KaiC, RpaA, P*kaiBC* DNA, 1 mM ATP, 5 mM $MgCl_2$, and SasA and/or CikA. 50 nM fluorescently labeled KaiA (green), KaiB (orange), SasA (purple), CikA (yellow), RpaA (red), or 100 nM P*kaiBC* (gray) were included in separate reactions, and their fluorescence anisotropies (FA) were measured in a microplate reader at 30°C. For phosphorylation experiments, KaiC~P (blue) and RpaA~P (salmon) were quantified by gel electrophoresis on SDS-PAGE and Zn^{2+}-Phos-tag gels respectively. After truncating the first 12 hours of data collection during which samples approached stable limit cycles, data were baseline-corrected and fit to a damped cosine function to calculate periods and phases (see materials and methods in the supplementary materials). FA values were normalized before fitting. (**D**) KaiC-ATPase was measured by real-time ^{1}H-NMR in a KaiABC-only reaction to isolate the KaiC ATPase activity. (**E** to **G**) Relative phases for each component in (A) to (C), determined from three separate experiments, are shown by black dots, with the average denoted by a white vertical line, and the peak of KaiC~P is referenced to CT14 (*9*). The width of each ellipse is ±2 hours centered at the peak. Full dataset is available in data S1.

Exploring clock output by SasA and CikA

Transcriptional output in cyanobacterial circadian rhythms is based on the antagonistic activity of two sensor histidine kinases; SasA is a KaiC-dependent kinase that phosphorylates RpaA to stimulate DNA binding and activate gene expression at dusk, whereas CikA is stimulated to act as a phosphatase that dephosphorylates RpaA when it associates with the Kai complex at night (*16*, *17*, *19*, *21*, *30*). To better understand the activities of each sensor histidine kinase in clock output, we first studied the reconstituted clock under conditions where only one was present at a time. In an IVC that contained KaiA, KaiB, KaiC, and SasA alone with RpaA and DNA (Fig. 1A), the relative phases of the protein interaction rhythms (Fig. 1E) were largely consistent with those reported in vivo (*31*). For example, FA rhythms of KaiA, KaiB, and SasA peaked from CT 15 to 22, indicating the formation of large protein complexes during subjective night owing to interaction with KaiC. These proteins also form large complexes in vivo around Zeitgeber time (ZT) 16 to 22 (*32*), where ZT 0 and ZT 12 correspond to dawn and dusk, respectively, under a light-dark cycle. Although SasA and KaiB bind to overlapping sites on KaiC (*33*), we observed that the rhythms of SasA binding to KaiC peak ~3 hours earlier than those of KaiB (Fig. 1, A and E). In principle, this phase difference could provide a temporal window for the KaiC-dependent activation of SasA autophosphorylation (*20*), allowing the phosphoryl transfer from SasA to RpaA (*19*) that is observed around subjective dusk in the IVC (Fig. 1E). RpaA dephosphorylation in the SasA-only IVC is likely due to a combination of the inherent autophosphatase activity typically seen in response regulators (*34*, *35*) and weak phosphatase activity of SasA (*21*).

When SasA was replaced by CikA in the IVC, we still saw circadian rhythms in the interaction of KaiA and KaiB with KaiC in the same relative phase with KaiC phosphorylation (Fig. 1, B and F), although the period was ~3 hours shorter. The FA of CikA peaked in phase with KaiB, consistent with recruitment of the pseudoreceiver (PsR) domain of CikA to the KaiB-KaiC complex in the subjective night (*36*). We verified that formation of this complex stimulates the phosphatase activity of CikA toward RpaA and that CikA can act as an RpaA kinase when not associated with the KaiB-KaiC complex (figs. S7 and S8), as previously reported (*21*, *37*). The bifunctional switching of CikA between phosphatase and kinase in the IVC likely causes the phase advance in the rhythms of RpaA phosphorylation and DNA binding relative to the SasA-only IVC by ~5 hours (Fig. 1F and fig. S9). CikA levels in vivo are normally low under high-intensity light (*38*) (fig. S10, B and C), suggesting that it may be limited to acting predominantly as a phosphatase at night. However, Δ*sasA* strains generate circadian rhythms in low light (*16*), suggesting that environmental conditions that favor CikA stability can rescue circadian rhythms. We demonstrated robust CikA-dependent circadian rhythms in strains that lack SasA but still express CikA and in which KaiB and KaiC were boosted to wild-type (WT) levels by circumventing autoregulation; notably, these rhythms occurred with a phase advance similar to that of the CikA-only IVC (fig. S10A). Taken with the observation that Δ*sasA*Δ*cikA* strains are arrhythmic (*39*), these data suggest that the circadian rhythms observed in Δ*sasA* strains arise from a CikA-mediated regulatory circuit on RpaA. When CikA is present constitutively in the IVC, it can drive rhythms of RpaA

phosphorylation both in the absence (Fig. 1B) and presence of SasA (Fig. 1C) to push the phase of peak DNA binding to midday instead of dusk (Fig. 1, F and G). Thus, the IVC recapitulates to a reasonable extent the relative in vivo phases of the circadian clock of *S. elongatus* and enables aspects of the clock—input (*40*, *41*), oscillator (*6*), and output (*20*, *21*, *39*, *42*)—to be measured under defined conditions with multiple readouts simultaneously in real time, including the phosphorylation-dependent interactions of RpaA (fig. S11) with a cognate promoter element in DNA.

Leveraging the extended oscillator to study clock mechanisms

Reconstitution of the cyanobacterial clock offers the possibility to ask questions about how downstream signal transduction components are coupled to the KaiABC oscillator to provide deeper mechanistic insight into circadian biology. First, we sought to explore the mechanistic origins of clock disruption in an arrhythmic *S. elongatus* strain that was originally reported as a *crm1* transposon-insertion allele (*43*). However, in addition to disruption of *crm1*, we also noted a single-nucleotide polymorphism in the *rpaA* gene located immediately downstream, resulting in an Arg^{121}→Gln (R121Q) amino acyl substitution in RpaA. When we reconstructed the allele of *rpaA-R121Q* in a WT reporter strain by markerless CRISPR-Cas12a editing (*44*), it displayed the same arrhythmic phenotype (Fig. 2A), demonstrating that mutation of *rpaA*, not *crm1*, disrupted clock function. Despite having a WT RpaA phosphorylation pattern (Fig. 2B and fig. S12), the mutant strain did not produce rhythmic gene expression (Fig. 2A). We used the IVC to test the hypothesis that the RpaA-R121Q variant is deficient in DNA binding, even in its phosphorylated form. Although we observed circadian rhythms of FA (fig. S13) and phosphorylation (Fig. 2B and fig. S14) in the mutant that were similar to WT RpaA, the R121Q mutant bound poorly to the P*kaiBC* promoter, as indicated by the low amplitude of DNA FA (Fig. 2C). Thus, the IVC helped to identify the mechanism of clock disruption by the R121Q amino acyl substitution in RpaA—it does not hinder regulatory interactions with SasA or CikA but instead prevents the phosphorylated sensor domain from regulating its DNA binding domain.

Effect of protein concentrations on rhythmicity and period of the clock

The core oscillator has a well-established, albeit relatively narrow, range of concentrations of KaiA, KaiB, and KaiC under which sustained biochemical rhythms can be observed in vitro (*45*). Using our newly developed fluorescence-based readout, we next measured how rhythms in the core oscillator depend on concentrations of KaiA and KaiB (Fig. 3). We found a notable

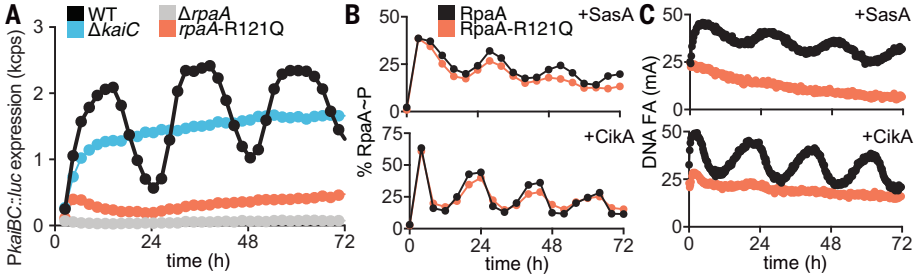


Fig. 2. The IVC reveals a defect in DNA binding by the RpaA-R121Q mutant. (**A**) Rhythms of bioluminescence generated by P*kaiBC::luc* expression were monitored for entrained WT (black), Δ*kaiC* (blue), Δ*rpaA* (gray), and *rpaA*-R121Q (salmon) cyanobacterial strains. Each plot represents an average of six biological replicates. kcps, kilocounts per second. (**B**) Rhythms of RpaA phosphorylation or (**C**) FA of P*kaiBC* DNA in IVC reactions containing either SasA or CikA, as in Fig. 1, A and B.

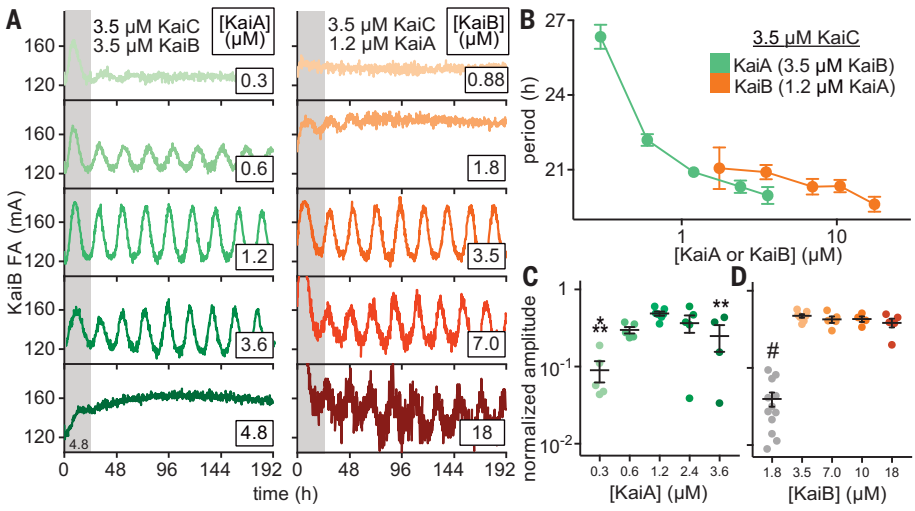


Fig. 3. Rhythmicity and period of the oscillator depend in different ways on the concentrations of KaiA and KaiB. (**A**) In vitro oscillator reactions containing 3.5 μM KaiC, 3.5 μM KaiB (including 50 nM fluorescently labeled KaiB as a probe), and titrations of KaiA from 0.3 to 6.0 μM (green scale, left column). Representative assay from $n = 3$ are shown; the first 24 hours of the reactions are shaded gray. Assays were also performed with 3.5 μM KaiC, 1.2 μM KaiA, and titrations of KaiB from 0.7 to 18 μM (orange scale, right panel). FA data were fit to a cosine function to extract period (**B**) and amplitude (**C** and **D**) as a function of KaiA or KaiB concentration. Amplitudes for KaiB titrations were normalized to account for an increasing ratio of unlabeled KaiB to fluorescently labeled probe KaiB. Reactions that did not oscillate were not included. Data are shown as mean ± SD ($n \geq 3$). One-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test was used for comparison of each suboptimal condition with the 1.2 μM KaiA, 3.5 μM KaiB condition: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $\#P < 0.0001$. Where no symbol is shown, differences were determined not to be statistically significant. Full datasets are available in data S2 and S3.

convergence with previous studies monitoring rhythms of KaiC phosphorylation by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (*45*). Because the fluorescence-based oscillator can be monitored noninvasively in a 384-well format, we could follow rhythms for many days longer than is practical for the KaiC phosphorylation assay and quantify sensitivity of oscillation amplitude and period to concentrations of the oscillator components. In general, the oscillator functions well with substoichiometric concentrations of KaiA, although the period lengthened as KaiA concentration was decreased (Fig. 3, A to C, and table S5). By contrast, oscillator period was not affected by titrating KaiB, although there was a clear requirement for KaiB to be present at concentrations at least equivalent to that of KaiC (Fig. 3D).

At one-half stoichiometry of KaiB relative to KaiC (1.75 μM KaiB and 3.5 μM KaiC), the oscillator functioned at low amplitude for only a few days with a normal circadian period before significantly damping to arrhythmicity (Fig. 3A). As KaiB is found in slight excess over KaiC in vivo (*11*, *46*), it may be essential for the KaiC hexamer to recruit a full stoichiometric complement of six molecules of KaiB to stably

sequester KaiA and facilitate the KaiC dephosphorylation phase in subjective night. Consistent with this model, rhythms abruptly diminished as the concentration of KaiA exceeded that of KaiC (Fig. 3A) (*45*, *47*).

CikA rescues oscillator period under limiting concentrations of KaiA

Although CikA is primarily considered an input-output protein (*17*, *21*, *39*), it is known to affect circadian period in vivo and was recently shown to regulate the period of the core oscillator in vitro (*37*, *48*). We previously determined that the PsR domain of CikA competes for the KaiA binding site on the KaiB-KaiC nighttime complex (*36*). This competition likely underlies period control by CikA; by displacing KaiA from the repressive complex, CikA promotes the activating potential of KaiA to stimulate KaiC phosphorylation (*48*). Using our extended oscillator system, we found that increasing concentrations of CikA led to progressive period shortening, particularly under limiting concentrations of KaiA, bringing it in line with the period found under idealized oscillator conditions (Fig. 4A). This result demonstrates one way that competitive mechanisms can compensate for fluctuating in vivo concentrations of oscillator proteins to influence timekeeping by the cyanobacterial clock.

SasA rescues oscillations under limiting concentrations of KaiB

SasA was originally described as an amplifier of circadian rhythms needed to maintain robust, high-amplitude rhythms, presumably by controlling rhythmic transcription of the *kaiBC* cluster (*16*). In contrast to the effect we observed with CikA, the addition of SasA led to progressive lengthening of the period (Fig. 4A and table S5). The isolated thioredoxin-like domain of SasA was much less effective at lengthening the period than was the full-length SasA dimer, consistent with the avidity observed in binding studies between KaiC and full-length SasA (fig. S15). It should be noted that addition of the isolated PsR domain or full-length CikA caused concentration-dependent changes in period length in vitro that are apparently balanced out by period changes imparted by SasA (Fig. 4 and fig. S16). These data support a model in which the proteins that impart output contribute substantively to the timing of the clock through interactions with the Kai proteins.

In exploring the combined effects of KaiB and SasA concentrations on oscillator period, we unexpectedly observed that amplitude and damping in a weak oscillator with suboptimal concentrations of KaiB could be rescued through the addition of SasA (Figs. 4, B to D). To better understand this phenomenon, we set up assays with increasing concentrations of SasA under typical in vitro oscillator conditions (*6*) as well as two concentrations where KaiB was limiting relative to KaiC. When KaiB concentrations were too low to support oscillations, we observed low FA values for KaiB, consistent with incomplete recruitment of labeled KaiB to KaiC hexamers (Fig. 4C, top left column). Addition of SasA up to 1 μM increased KaiB FA, demonstrating that SasA enhances KaiB binding to KaiC. However, as the concentration of SasA approached or exceeded that of KaiC, FA values for KaiB dropped, demonstrating that SasA eventually outcompetes KaiB for binding to the KaiC hexamer, as previously observed (*36*, *49*).

Using a KaiB concentration equal to half that of KaiC, we saw rapidly damping, low-amplitude rhythms (Fig. 4C, middle column). We also found that addition of substoichiometric amounts of SasA restored full amplitude and eliminated damping of the oscillations. Increasing SasA concentrations from 0.1 to 1 μM under these conditions did not further influence amplitude, and higher concentrations of SasA attenuated oscillations by outcompeting KaiB for binding to KaiC, as reflected by low anisotropy levels of the labeled KaiB. Under standard oscillator conditions with stoichiometric KaiB and KaiC, the amplitude of oscillation was affected by addition of SasA only at 1.0 μM and higher (Fig. 4, B and C, right column). The sharp drop in oscillator amplitude and KaiB FA levels demonstrates that a functional switch from recruitment to competition by SasA occurs at around ~0.65 to 1 μM, depending on the exact oscillator conditions. This setpoint is close to the concentration of SasA in vivo estimated by quantitative Western blotting to be around 0.6 μM (*21*).

We hypothesized that SasA expands the range of permissive KaiB concentrations for biochemical oscillation by directly enhancing the interaction of KaiB and KaiC. As a test, we used a partial oscillator experiment in which unlabeled KaiB was omitted from the reaction, and association of the 50 nM labeled KaiB probe was measured as phosphorylated KaiC accumulated upon stimulation by KaiA (Fig. 4E). We observed a concentration-dependent enhancement in KaiB association as SasA concentration approached the physiological value of ~0.65 μM. This finding suggests that, aside from its role in regulating transcriptional output from the oscillator (*16*), SasA also directly modulates KaiBC association to control formation of the nighttime repressive state and stability of the clock itself. Moreover, we observed a delay in KaiB association with KaiC when SasA was present at higher concentrations, consistent with the period-lengthening effects observed with SasA in the IVC (Fig. 4A and fig. S16).

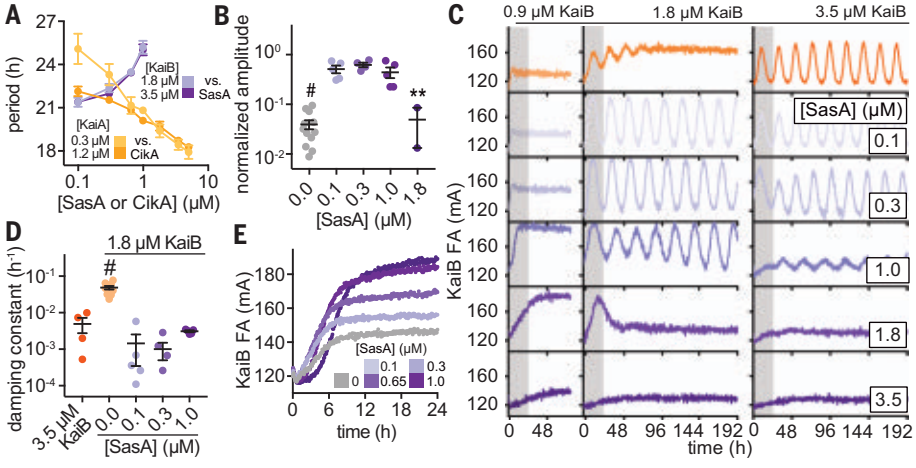


Fig. 4. Output proteins compensate for period and amplitude defects in suboptimal in vitro oscillators. (**A**) Period of clock reactions with varied concentrations of KaiA (light and dark yellow) or KaiB (light and dark purple) supplemented as a function of CikA or SasA concentration, respectively. (**B**) Amplitudes of rhythms with probe KaiB + 1.8 μM unlabeled KaiB in the presence of SasA (gray for KaiB alone; light to dark purple for SasA). Reactions that could not be fit to a sine function because they did not oscillate were not plotted. Statistical analysis was performed exactly as described in Fig. 3, C and D. (**C**) In vitro oscillator assays with 3.5 μM KaiC, 1.2 μM KaiA, and the indicated concentrations of KaiB as well as various amounts of full-length SasA ranging from 0.1 to 3.5 μM. Representative assay from $n \geq 3$ shown; the first 24 hours of the time course are shaded gray. (**D**) Damping constants derived from least-squares fitting of the FA data from (C). Full datasets are available in data S2 and S3. (**E**) "Partial oscillator" reactions containing 50 nM KaiB, 3.5 μM KaiC, and 1.2 uM KaiA with various amounts of SasA.

Structural mimicry in the KaiC-binding region of SasA and KaiB

During subjective day, when KaiB is free, it exists in a distinctive ground-state fold, but the fold-switched monomer that binds KaiC at subjective night is thioredoxin-like, which is the same fold adopted by the KaiC-binding domain of SasA ($SasA_{trx}$) (*37*). A high-resolution crystal structure of the subcomplex comprising a

single KaiC-CI domain and a fold-switch-locked mutant KaiB I88A from the related thermophilic species *T. elongatus* (*36*) (I87A in *S. elongatus*, referred to herein as fsKaiB) illustrates how KaiB docks onto the exposed B-loop of the CI domain (Fig. 5, A and B). To better understand the relationship between SasA and KaiB recruitment to KaiC, we solved a crystal structure of $SasA_{trx}$ bound to the KaiC-CI domain from *T. elongatus* to reveal that SasA binds the B-loop in a similar orientation to KaiB (Fig. 5C and table S6). To investigate the importance of this interface, we probed substitutions in several KaiC residues in the B-loop using affinity assays to KaiC hexamer (Fig. 5D). We performed equilibrium binding assays using the KaiC phosphomimetic mutants KaiC-EE and KaiC-EA, where X and Y in KaiC-XY represent amino acyl residues at positions 431 and 432, respectively. Glutamyl/glutamyl (EE) and glutamyl/alanyl (EA) substitutions at these positions are widely used to approximate the dusk-like phosphoserine/phosphothreonine (pS/pT) and nighttime-like phosphoserine (pS/T) states (*8*). Using these mimetics, we found that KaiB has similar affinity for both states, whereas SasA has a preference for the pS/pT state (fig. S17), consistent with prior results (*50*) and its earlier phase of binding in the IVC (Fig. 1, A and E). Furthermore, although the isolated $SasA_{trx}$ domain is necessary and sufficient for binding to KaiC (*16*), avidity effects in the full-length dimer enhanced affinity for KaiC-EE by at least two orders of magnitude compared with the isolated, monomeric $SasA_{trx}$ domain (fig. S17).

Comparing the KaiB and SasA binding interfaces on KaiC, we found that substitutions at sites conserved between *S. elongatus* and *T. elongatus* (KaiC F122 and D123) decrease affinity for both KaiB and $SasA_{trx}$ (Fig. 5, D to F). We then examined how differences in the structures of KaiB and SasA or their orientations on the B-loop might influence binding to adjacent subunits of the KaiC hexamer (fig. S18). Small changes in the length and orientation of the C-terminal helix between the $SasA_{trx}$ and KaiB could lead to steric clashes of $SasA_{trx}$ with a neighboring subunit. Consistent with this idea, saturation binding experiments showed that $SasA_{trx}$ domains cannot fully occupy all six binding sites on the KaiC hexamer (fig. S18).

SasA enhances KaiBC interaction through heterotropic cooperativity

Six monomers of the active, thioredoxin-like form of KaiB assemble onto the KaiC hexamer of *S. elongatus* (*36*, *51*) to nucleate formation of the nighttime repressive state. Prior native mass spectrometry analyses demonstrated that KaiB binds to the KaiC hexamer with positive homotropic cooperativity (*52*, *53*). Given the similarity between the SasA-KaiC and KaiB-KaiC interactions, we speculated that SasA could facilitate KaiB-KaiC interactions through positive heterotropic cooperativity. We developed an assay to test directly for cooperativity by monitoring 50 nM fluorescently labeled KaiB probe with KaiC-EE using unlabeled SasA or fsKaiB as a secondary titrant (Fig. 6, A and B). Although these assay conditions do not assess the degree of homotropic cooperativity for KaiC-EE with the KaiB probe alone, we found that low concentrations (50 to 100 nM) of SasA or fsKaiB significantly increased binding

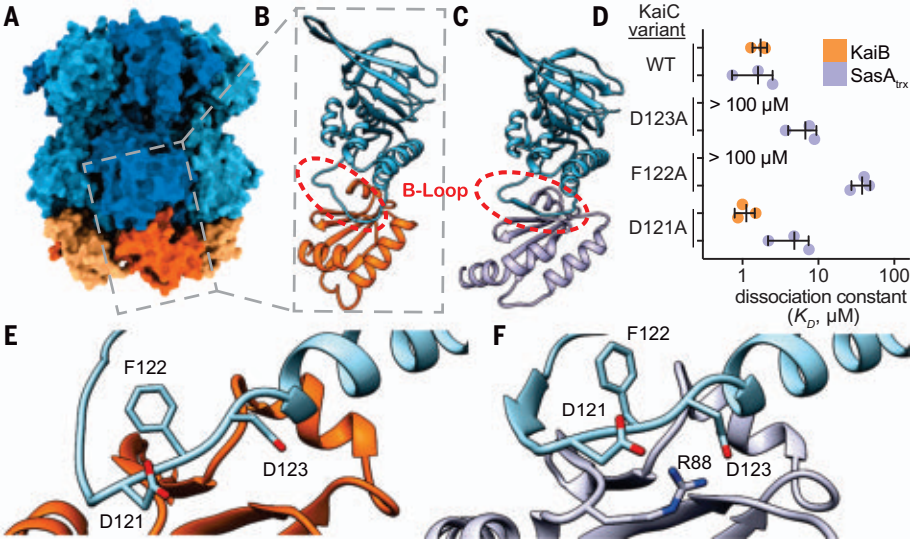


Fig. 5. SasA and KaiB bind to the same site on KaiC. (**A**) The KaiB-KaiC heterododecamer [Protein Data Bank (PDB) ID 5JWQ] with subunits of the KaiC ring depicted in alternating light and dark blue, and fsKaiB subunits in alternating light and dark orange. The gray box indicates the position of (**B**) the isolated KaiC-CI domain–fsKaiB subcomplex (PDB ID 5JWO). (**C**) The KaiC-CI domain–$SasA_{trx}$ subcomplex (PDB ID 6X61), with KaiC (blue) and SasA (purple). (**D**) Equilibrium dissociation constants (K_D) for KaiB (orange) or $SasA_{trx}$ (purple) binding to KaiC-EE mutants (mean ± SD, n = 3). Where indicated, binding was too weak for curve fitting, and K_D is reported as >100 μM. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**E** and **F**) The interface of KaiC-CI with fsKaiB (E) or $SasA_{trx}$ (F) with key residues highlighted.

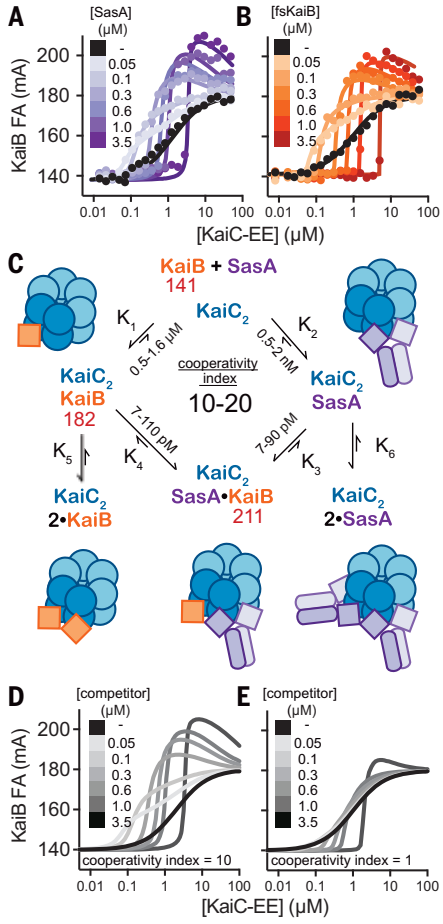


Fig. 6. The thioredoxin-like fold of SasA and fsKaiB cooperatively recruits KaiB to the KaiC hexamer. (**A** and **B**) Equilibrium binding assays of fluorescently labeled KaiB with KaiC-EE in the absence (black) or presence of (A) unlabeled SasA (light to dark purple) or (B) fsKaiB (light to dark orange). Each 2D titration is shown from representative assays with best fit lines from global curve-fitting. (**C**) Summary of a two-site thermodynamic model for SasA heterocooperativity, and mean equilibrium constants derived from least-squares fitting of SasA 2D titration assays (n = 3); values in red are mean peak anisotropy values observed from fitting of all SasA 2D titration datasets using DynaFit software. The cooperativity index is defined as the ratio $K_1/K_3 = K_2/K_4$. This model was used to generate simulated 2D titrations with a cooperativity index of either 10 (**D**) or 1 (**E**). (See data S6 to S8 for DynaFit scripts and full datasets.)

of KaiB probe, demonstrating more efficient recruitment to the KaiC hexamer. Higher concentrations of SasA or fsKaiB resulted in competition with the probe, reducing its binding to the KaiC hexamer; an equivalence point was reached when the concentration of KaiC-EE matched that of the competitor. We observed a considerable overshoot of KaiB probe anisotropy values around this point, consistent with formation of a higher-molecular-weight complex consisting of both KaiB probe and SasA bound to the KaiC hexamer. Association of the KaiB probe to the WT KaiC-CI monomer was not significantly affected by the addition of 300 nM SasA or fsKaiB (fig. S19), nor did SasA interact directly with KaiB in the absence of KaiC, demonstrating that these competitors only enhance KaiB association in the context of the KaiC hexamer, as shown using phosphomimetics, which is consistent with our cooperativity model.

To quantitatively represent the heterotropic cooperativity observed in these two-dimensional (2D) titrations, a simplified two-site thermodynamic model was found by least-squares fitting to be sufficient to account for the data (Fig. 6C and fig. S19). In this model, we defined the heterotropic "cooperativity index" as the fold-increase in KaiB affinity given by the ratio of equilibrium constants K_1/K_3 (= K_2/K_4). Comparison of 2D titrations with SasA or fsKaiB to simulated data representing heterotropic cooperative binding (Fig. 6, A and B), to a model with competition and no cooperativity (Fig. 6, D and E), demonstrates that SasA and fsKaiB similarly influence cooperative recruitment of the KaiB probe on KaiC.

To explore the role of intermolecular KaiB-KaiB interactions in cooperative KaiB-KaiC binding, we characterized KaiB-R22A, a mutant originally identified in *Anabaena* (*54*) that reduces the apparent affinity of KaiB for KaiC although it is not located at the KaiB-KaiC interface (*36*). R22 has been implicated in KaiB cooperativity using native mass spectrometry (*52*) and is situated at the KaiB-KaiB interface in the KaiBC hexamer structure. We tested this mutant in our cooperativity assay and observed a small decrease in the cooperative recruitment of the KaiB probe relative to that of fsKaiB (fig. S19), demonstrating that intersubunit KaiB-KaiB interactions play a role in the cooperative recruitment process and opening the possibility that mutations at the SasA-KaiB interface could also disrupt cooperativity.

SasA-KaiB complementarity is required for proper circadian timing

To identify whether the enhancement of KaiBC association is important for the function of SasA in vivo, we set out to identify point mutations that would separate the cooperativity and output signaling functions of SasA. To this end, we first showed that the catalytically inactive SasA mutant H161A (*19*) is as effective as WT SasA in stimulating cooperative recruitment of KaiB to KaiC in vitro (fig. S19). Next, we constructed a structural model by overlaying crystal structures of $SasA_{trx}$-CI and KaiB-CI onto the lower-resolution structure of the full-length KaiBC hexamer complex (*36*) and used it to predict potential $SasA_{trx}$-KaiB interactions along the clockwise (CW) and counterclockwise (CCW) interfaces of the hybrid complex (Fig. 7A). Within this framework, we investigated a number of SasA substitutions near the CW and CCW interface (fig. S20) and estimated their ability to cooperatively recruit KaiB in simplified 2D titration assays comparing a single SasA concentration to the no-SasA control. The individual mutants H28A (CCW interface) or Q94A (CW interface) showed decreases in their cooperativity indices that were further decreased in the H28A-Q94A double mutant (Fig. 7B). It should be noted that the estimates of cooperativity in wild type were slightly

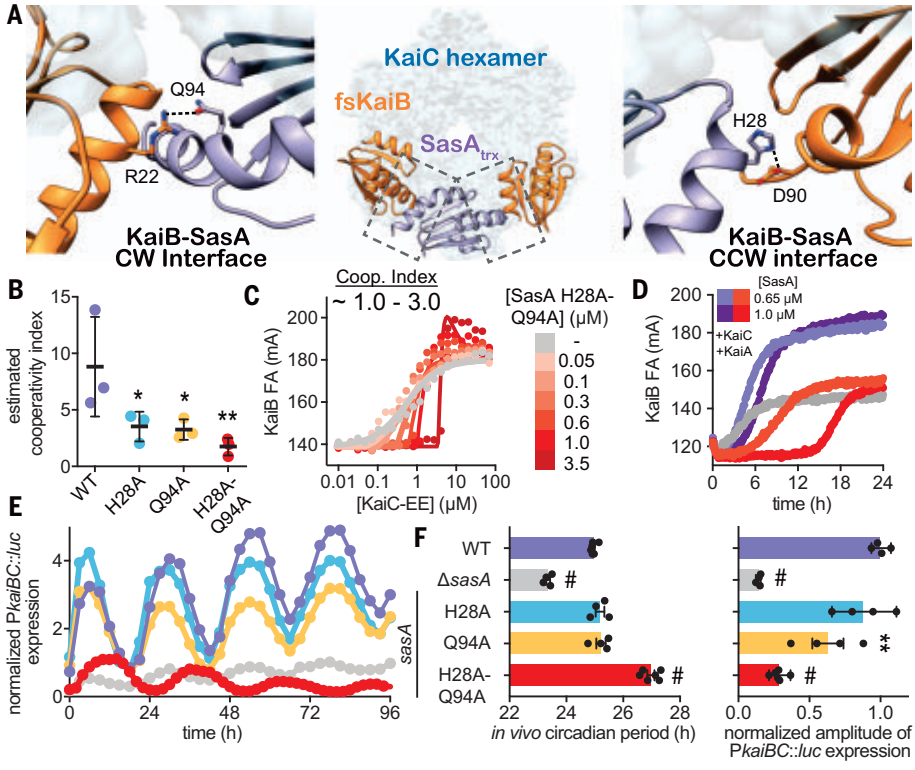


Fig. 7. SasA-KaiB interactions mediate cooperative recruitment of KaiB in vitro and sustain robust circadian rhythms in vivo. (**A**) Structural model of KaiB:SasA interactions on KaiC using KaiC-CI subcomplexes with fsKaiB (orange, PDB ID 5JWO) and $SasA_{trx}$ (purple, PDB ID 6X61) modeled onto adjacent subunits of KaiC (light blue) of the KaiBC heterododecamer (PDB ID 5JWQ). Key residues at clockwise (CW) and counterclockwise (CCW) interfaces are modeled as *S. elongatus* variants on the basis of the alignment in fig. S20. Polar contacts predicted by our hybrid structural model are indicated by dashed lines. (**B**) Cooperativity indices for SasA variants extracted from KaiC titrations done in the presence of 0.3 μM SasA (mean SD, n = 3). Statistical significance of differences between WT and mutant SasA variants was tested using one-way ANOVA with Dunnett's multiple comparisons as described in Fig. 3. (**C**) Representative 2D titration of fluorescently labeled KaiB with KaiC-EE in the absence (gray) or presence (light to dark red) of SasA-H28A-Q94A. (**D**) Time-dependent FA trajectories of 50 nM KaiB in the presence of 3.5 μM KaiC and 1.2 uM KaiA with 0.65 or 1.0 μM SasA WT (purple shades) or H28A-Q94A mutant (red shades). (**E**) Representative time course of bioluminescence driven by P*kaiBC* from *S. elongatus* cultures entrained under 12-hour light-dark cycles for 48 hours and subsequently allowed to free run in LL. (**F**) Raw luminescence curves were fit to a sine function to extract amplitude of P*kaiBC::luc* expression as well as free-running circadian period for each strain. WT control was included for each experiment, and the luminescence of each mutant was normalized to the amplitude of luminescence oscillation for the WT control in that run. Error bars depict SEM among replicate cultures (n = 4 to 6); when error bars are not visible, they were smaller than could be depicted. Symbols indicate significance determined from one-way ANOVA with Dunnett's multiple comparisons between mutant and WT *S. elongatus* with symbols to depict significance as summarized in Fig. 3 legend. Normal distribution was confirmed through quantile-quantile plots showing predicted versus actual residual values in addition to Anderson-Darling, D'Agostino-Pearson omnibus, and Shapiro-Wilk tests, which were all performed in GraphPad PRISM. Equal variance of the data was verified with Brown-Forsythe tests yielding P values above 0.05. Color scheme for data in (E) is consistent with and indicated on left vertical axis of (F).

lower than those determined in the full 2D titrations (Fig. 6, A and C), likely owing to differences in experimental setup. While the H28A-Q94A double mutant did not strongly affect other SasA functions in vitro, such as KaiC association and the KaiC-dependent phosphotransfer to RpaA (fig. S21), it lacked the ability to stimulate KaiB binding in the 2D titration assay (compare Fig. 7C and Fig. 6E) demonstrating that both interfaces contribute to SasA heterotropic cooperativity.

To better understand these mutations in a more native context, we tested the ability of the H28A-Q94A variant to nucleate formation of the repressive complex using a partial oscillator experiment to directly measure the kinetics of KaiBC formation under limiting concentrations of KaiB with WT or H28A-Q94A SasA (Fig. 7D and fig. S21C). Notably, enhancement of KaiB recruitment was essentially nonexistent with the SasA H28A-Q94A mutant, instead showing a lengthened competitive phase that occurs even at physiological concentrations of SasA. When concentration of the double mutant was increased beyond physiological levels, the delay in KaiB association became even more exaggerated, suggesting that complementarity between the SasA-KaiB interfaces is critical for timely replacement of SasA with KaiB on the KaiC hexamer.

With this separation-of-function mutant in hand, we set out to test whether the SasA-induced cooperativity we observed for the expanded oscillator in vitro could influence circadian rhythms in vivo. We introduced the double mutation, as well as several single alanine substitutions, into SasA using markerless CRISPR-Cas12a and monitored bioluminescence from a P*kaiBC* luciferase reporter in constant light (LL) after synchronization in 12-hours light:12-hours dark cycles (Fig. 7E and fig. S22). Small effects on amplitude were seen with the individual H28 or Q94 alanine substitutions on their own, but when these were combined in the H28A-Q94A mutant, the amplitude of circadian rhythms decreased to an extent that was similar to the SasA knockout strain (Fig. 7F, Δ*sasA*), even though the mutant proteins were expressed to a similar level in vivo (fig. S22D), and the double mutant strain exhibited a ~2-hour-longer period (Fig. 7F and fig. S22). Taken together, these results provide a clear link between loss of complementarity at SasA-KaiB interfaces and the long-period phenotype observed in vivo, further supporting our conclusion that cooperative interactions between SasA and KaiB are crucial for proper circadian timing.

Discussion

This extended IVC is an autonomously oscillating system that remains phase coherent for several days without user intervention, allows real-time observation of rhythmic behaviors of several clock components in parallel, and rhythmically activates the interaction of a transcription factor with a promoter element on DNA. Complementing this assay with parallel measurements of phosphorylation and ATPase activities rounds out a movie of an intact cyanobacterial clock. As implemented here, the IVC has revealed a set of minimal components to execute clock-controlled gene expression, helped to dissect the mechanistic origins of a clock-disrupting mutant, and provided insight into mechanisms behind the narrow stoichiometric ratios that are tolerated in the in vitro oscillator (*45*). In particular, our data demonstrate that SasA, previously regarded solely as an output protein, plays a critical role in the enhancing the oscillator amplitude through heterocooperative regulation of KaiB association with KaiC. This newly developed IVC system offers an unparalleled opportunity to explore how the clock maintains consistency in vivo despite rhythmic changes in the concentration of clock components that result from the associated transcription-translation feedback (*11*, *55*) and protein turnover (*56*, *57*) and provides an experimental platform for integrating the oscillator with the upstream and downstream components with which it interacts. By reconstructing, and subsequently deconstructing, this elegant system, we have shown that a dichotomous view of timekeeping and signaling components provides an insufficient representation of the complexity and synergy that has evolved in this ancient circadian clock.

REFERENCES AND NOTES

1. J. C. Dunlap, J. J. Loros, P. J. DeCoursey, Eds., *Chronobiology: Biological Timekeeping* (Sinauer Associates, 2004).
2. M. A. Woelfle, Y. Ouyang, K. Phanvijhitsiri, C. H. Johnson, The adaptive value of circadian clocks: An experimental assessment in cyanobacteria. *Curr. Biol.* **14**, 1481–1486 (2004). doi: 10.1016/j.cub.2004.08.023; pmid: 15324665
3. T. Kondo *et al*., Circadian rhythms in prokaryotes: Luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5672–5676 (1993). doi: 10.1073/pnas.90.12.5672; pmid: 8516317
4. M. Ishiura *et al*., Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* **281**, 1519–1523 (1998). doi: 10.1126/science.281.5382.1519; pmid: 9727980
5. J. Tomita, M. Nakajima, T. Kondo, H. Iwasaki, No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* **307**, 251–254 (2005). doi: 10.1126/science.1102540; pmid: 15550625
6. M. Nakajima *et al*., Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* **308**, 414–415 (2005). doi: 10.1126/science.1108451; pmid: 15831759
7. M. J. Rust, J. S. Markson, W. S. Lane, D. S. Fisher, E. K. O'Shea, Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* **318**, 809–812 (2007). doi: 10.1126/science.1148596; pmid: 17916691
8. T. Nishiwaki *et al*., A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *EMBO J.* **26**, 4029–4037 (2007). doi: 10.1038/sj.emboj.7601832; pmid: 17717528
9. H. Iwasaki, T. Nishiwaki, Y. Kitayama, M. Nakajima, T. Kondo, KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15788–15793 (2002). doi: 10.1073/pnas.222467299; pmid: 12391300
10. S. B. Williams, I. Vakonakis, S. S. Golden, A. C. LiWang, Structure and function from the circadian clock protein KaiA of *Synechococcus elongatus*: A potential clock input mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15357–15362 (2002). doi: 10.1073/pnas.232517099; pmid: 12438647
11. Y. Kitayama, H. Iwasaki, T. Nishiwaki, T. Kondo, KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *EMBO J.* **22**, 2127–2134 (2003). doi: 10.1093/emboj/cdg212; pmid: 12727879
12. Y. Xu, T. Mori, C. H. Johnson, Cyanobacterial circadian clockwork: Roles of KaiA, KaiB and the *kaiBC* promoter in regulating KaiC. *EMBO J.* **22**, 2117–2126 (2003). doi: 10.1093/emboj/cdg168; pmid: 12727878
13. Y.-G. Chang, R. Tseng, N.-W. Kuo, A. LiWang, Rhythmic ring-ring stacking drives the circadian oscillator clockwise. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16847–16851 (2012). doi: 10.1073/pnas.1211508109; pmid: 22967510
14. Y.-G. Chang, N.-W. Kuo, R. Tseng, A. LiWang, Flexibility of the C-terminal, or CII, ring of KaiC governs the rhythm of the circadian clock of cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14431–14436 (2011). doi: 10.1073/pnas.1104221108; pmid: 21788479
15. J. Abe *et al*., Atomic-scale origins of slowness in the cyanobacterial circadian clock. *Science* **349**, 312–316 (2015). doi: 10.1126/science.1261040; pmid: 26113637
16. H. Iwasaki *et al*., A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. *Cell* **101**, 223–233 (2000). doi: 10.1016/S0092-8674(00)80832-6; pmid: 10786837
17. O. Schmitz, M. Katayama, S. B. Williams, T. Kondo, S. S. Golden, CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science* **289**, 765–768 (2000). doi: 10.1126/science.289.5480.765; pmid: 10926536
18. N. B. Ivleva, T. Gao, A. C. LiWang, S. S. Golden, Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17468–17473 (2006). doi: 10.1073/pnas.0606639103; pmid: 17088557
19. N. Takai *et al*., A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12109–12114 (2006). doi: 10.1073/pnas.0602955103; pmid: 16882723
20. J. S. Markson, J. R. Piechura, A. M. Puszynska, E. K. O'Shea, Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* **155**, 1396–1408 (2013). doi: 10.1016/j.cell.2013.11.005; pmid: 24315105
21. A. Gutu, E. K. O'Shea, Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Mol. Cell* **50**, 288–294 (2013). doi: 10.1016/j.molcel.2013.02.022; pmid: 23541768
22. K. Ito-Miwa, Y. Furuike, S. Akiyama, T. Kondo, Tuning the circadian period of cyanobacteria up to 6.6 days by the single amino acid substitutions in KaiC. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20926–20931 (2020). doi: 10.1073/pnas.2005496117; pmid: 32747571
23. E. Leypunskiy *et al*., The cyanobacterial circadian clock follows midday in vivo and in vitro. *eLife* **6**, e23539 (2017). doi: 10.7554/eLife.23539; pmid: 28686160
24. G. K. Chow *et al*., Monitoring protein–protein interactions in the cyanobacterial circadian clock in real time via electron paramagnetic resonance spectroscopy. *Biochemistry* **59**, 2387–2400 (2020). doi: 10.1021/acs.biochem.0c00279; pmid: 32453554
25. J. Heisler, A. Chavan, Y.-G. Chang, A. LiWang, Real-time in vitro fluorescence anisotropy of the cyanobacterial circadian clock. *Methods Protoc.* **2**, 42 (2019). doi: 10.3390/mps2020042; pmid: 31164621
26. C. S. Theile *et al*., Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–1807 (2013). doi: 10.1038/nprot.2013.102; pmid: 23989674
27. C. P. Guimaraes *et al*., Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1787–1799 (2013). doi: 10.1038/nprot.2013.101; pmid: 23989673
28. B. Guo *et al*., Monitoring ATP hydrolysis and ATPase inhibitor screening using ^{1}H NMR. *Chem. Commun.* **50**, 12037–12039 (2014). doi: 10.1039/C4CC04399E; pmid: 25170530
29. K. Terauchi *et al*., ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16377–16381 (2007). doi: 10.1073/pnas.0706292104; pmid: 17901204
30. R. M. Smith, S. B. Williams, Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. *Proc. Natl. Acad.*

Sci. U.S.A. **103**, 8564–8569 (2006). doi: 10.1073/pnas.0508696103; pmid: 16707582
31. S. E. Cohen, S. S. Golden, Circadian rhythms in cyanobacteria. *Microbiol. Mol. Biol. Rev.* **79**, 373–385 (2015). doi: 10.1128/MMBR.00036-15; pmid: 26335718
32. H. Kageyama, T. Kondo, H. Iwasaki, Circadian formation of clock protein complexes by KaiA, KaiB, KaiC, and SasA in cyanobacteria. *J. Biol. Chem.* **278**, 2388–2395 (2003). doi: 10.1074/jbc.M208899200; pmid: 12441347
33. R. Tseng *et al*., Cooperative KaiA-KaiB-KaiC interactions affect KaiB/SasA competition in the circadian clock of cyanobacteria. *J. Mol. Biol.* **426**, 389–402 (2014). doi: 10.1016/j.jmb.2013.09.040; pmid: 24112939
34. P. B. Straughn *et al*., Modulation of response regulator CheY reaction kinetics by two variable residues that affect conformation. *J. Bacteriol.* **202**, e00089-20 (2020). doi: 10.1128/JB.00089-20; pmid: 32424010
35. A. H. West, A. M. Stock, Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.* **26**, 369–376 (2001). doi: 10.1016/S0968-0004(01)01852-7; pmid: 11406410
36. R. Tseng *et al*., Structural basis of the day-night transition in a bacterial circadian clock. *Science* **355**, 1174–1180 (2017). doi: 10.1126/science.aag2516; pmid: 28302851
37. Y.-G. Chang *et al*., A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **349**, 324–328 (2015). doi: 10.1126/science.1260031; pmid: 26113641
38. N. B. Ivleva, M. R. Bramlett, P. A. Lindahl, S. S. Golden, LdpA: A component of the circadian clock senses redox state of the cell. *EMBO J.* **24**, 1202–1210 (2005). doi: 10.1038/sj.emboj.7600606; pmid: 15775978
39. Y. Taniguchi, N. Takai, M. Katayama, T. Kondo, T. Oyama, Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian *kaiBC* expression in cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3263–3268 (2010). doi: 10.1073/pnas.0909924107; pmid: 20133618
40. M. J. Rust, S. S. Golden, E. K. O'Shea, Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331**, 220–223 (2011). doi: 10.1126/science.1197243; pmid: 21233390
41. Y.-I. Kim, D. J. Vinyard, G. M. Ananyev, G. C. Dismukes, S. S. Golden, Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17765–17769 (2012). doi: 10.1073/pnas.1216401109; pmid: 23071342
42. M. L. Paddock, J. S. Boyd, D. M. Adin, S. S. Golden, Active output state of the *Synechococcus* Kai circadian oscillator. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3849–E3857 (2013). doi: 10.1073/pnas.1315170110; pmid: 24043774
43. J. S. Boyd, J. R. Bordowitz, A. C. Bree, S. S. Golden, An allele of the *crm* gene blocks cyanobacterial circadian rhythms. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13950–13955 (2013). doi: 10.1073/pnas.1312793110; pmid: 23918383
44. J. Ungerer, H. B. Pakrasi, Cpf1 is a versatile tool for CRISPR genome editing across diverse species of cyanobacteria. *Sci. Rep.* **6**, 39681 (2016). doi: 10.1038/srep39681; pmid: 28000776
45. M. Nakajima, H. Ito, T. Kondo, *In vitro* regulation of circadian phosphorylation rhythm of cyanobacterial clock protein KaiC by KaiA and KaiB. *FEBS Lett.* **584**, 898–902 (2010). doi: 10.1016/j.febslet.2010.01.016; pmid: 20079736
46. J. Chew, E. Leypunskiy, J. Lin, A. Murugan, M. J. Rust, High protein copy number is required to suppress stochasticity in the cyanobacterial circadian clock. *Nat. Commun.* **9**, 3004 (2018). doi: 10.1038/s41467-018-05109-4; pmid: 30068980
47. J. Lin, J. Chew, U. Chockanathan, M. J. Rust, Mixtures of opposing phosphorylations within hexamers precisely time feedback in the cyanobacterial circadian clock. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3937–E3945 (2014). doi: 10.1073/pnas.1408692111; pmid: 25197081
48. M. Kaur, A. Ng, P. Kim, C. Diekman, Y.-I. Kim, CikA modulates the effect of KaiA on the period of the circadian oscillation in KaiC phosphorylation. *J. Biol. Rhythms* **34**, 218–223 (2019). doi: 10.1177/0748730419828068; pmid: 30755127
49. R. Murakami *et al*., The roles of the dimeric and tetrameric structures of the clock protein KaiB in the generation of circadian oscillations in cyanobacteria. *J. Biol. Chem.* **287**, 29506–29515 (2012). doi: 10.1074/jbc.M112.349092; pmid: 22722936
50. J. Valencia S. *et al*., Phase-dependent generation and transmission of time information by the KaiABC circadian clock oscillator through SasA-KaiC interaction in cyanobacteria. *Genes Cells* **17**, 398–419 (2012). doi: 10.1111/j.1365-2443.2012.01597.x; pmid: 22512339
51. J. Snijder *et al*., Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state. *Science* **355**, 1181–1184 (2017). doi: 10.1126/science.aag3218; pmid: 28302852
52. R. Murakami *et al*., Cooperative binding of KaiB to the KaiC hexamer ensures accurate circadian clock oscillation in cyanobacteria. *Int. J. Mol. Sci.* **20**, 4550 (2019). doi: 10.3390/ijms20184550; pmid: 31540310
53. J. Snijder *et al*., Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interaction. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1379–1384 (2014). doi: 10.1073/pnas.1314326111; pmid: 24474762
54. R. G. Garces, N. Wu, W. Gillon, E. F. Pai, Anabaena circadian clock proteins KaiA and KaiB reveal a potential common binding site to their partner KaiC. *EMBO J.* **23**, 1688–1698 (2004). doi: 10.1038/sj.emboj.7600190; pmid: 15071498
55. Y. Xu, T. Mori, C. H. Johnson, Circadian clock-protein expression in cyanobacteria: Rhythms and phase setting. *EMBO J.* **19**, 3349–3357 (2000). doi: 10.1093/emboj/19.13.3349; pmid: 10880447
56. C. K. Holtman *et al*., High-throughput functional analysis of the *Synechococcus elongatus* PCC 7942 genome. *DNA Res.* **12**, 103–115 (2005). doi: 10.1093/dnares/12.2.103; pmid: 16303742
57. K. Imai, Y. Kitayama, T. Kondo, Elucidation of the role of Clp protease components in circadian rhythm by genetic deletion and overexpression in cyanobacteria. *J. Bacteriol.* **195**, 4517–4526 (2013). doi: 10.1128/JB.00300-13; pmid: 23913328

ACKNOWLEDGMENTS

We thank staff at the 23-ID-D beamline of the Advanced Photon Source, Argonne National Laboratory, for their help with data collection. The Advanced Photon Source (contract DE-AC02-06CH11357) is supported by the US Department of Energy. Molecular graphics and analyses performed with UCSF Chimera and ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (UCSF), with support from National Institutes of Health P41-GM103311(Chimera) and R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (ChimeraX). We thank G. Chow, Y.-G. Chang, M. Rust, L. Hong, S. Sukenik, and M. Zoghbi for helpful discussions. We also thank D. Rice and S. Grimaldi for maintaining the NMR facility and NMR cryoplatform, respectively, at UC Merced. **Funding:** This work was supported by US National Institutes of Health grants R01 GM107521 (to A.L.), R01 GM121507 (to C.L.P.), and R35 GM118290 (to S.S.G.), and NSF-CREST: Center for Cellular and Biomolecular Machines at the University of California, Merced (NSF-HRD-1547848). J.H. was supported by NSF-CREST CCBM HRD-1547848. J.G.P. was supported by NIH IMSD grant R25 GM058903-20. P.C. is supported by EMBO long-term fellowship 57-2019. **Author contributions:** Conceptualization: A.G.C., J.A.S., J.H., S.S.G., C.L.P., and A.L. Methodology: A.G.C., J.H., J.A.S., C.S., D.C.E., M.F., and C.R.B. Investigation: A.G.C., J.A.S., J.H., C.S., D.C.E., M.F., J.G.P., R.K.S., and S.T. Validation: A.G.C., J.A.S., J.H., C.S., D.C.E., M.F., and S.T. Formal analysis: A.G.C., J.H., J.A.S., C.R.B., P.C., C.S., D.C.E., and M.F. Resources: S.S.G., C.L.P., and A.L. Data curation: A.G.C., J.H., J.A.S., and C.S. Writing – original draft: A.G.C., J.H., J.A.S., C.L.P., and A.L. Writing – review and editing: A.G.C., J.A.S., J.H., D.C.E., C.R.B., S.S.G., C.L.P., and A.L. Funding acquisition: J.H., J.G.P., P.C., S.S.G., C.L.P., and A.L. Supervision: S.S.G., C.L.P., and A.L. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd4453
Materials and Methods
Figs. S1 to S22
Tables S1 to S12
References (*58*–*82*)
MDAR Reproducibility Checklist
Data S1 to S8

21 June 2020; accepted 30 August 2021
10.1126/science.abd4453

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Cross-reactive CD4$^+$ T cells enhance SARS-CoV-2 immune responses upon infection and vaccination

Lucie Loyal, Julian Braun†, Larissa Henze†, Beate Kruse†, Manuela Dingeldey†, Ulf Reimer†, Florian Kern†, Tatjana Schwarz†, Maike Mangold, Clara Unger, Friederike Dörfler, Shirin Kadler, Jennifer Rosowski, Kübrah Gürcan, Zehra Uyar-Aydin, Marco Frentsch, Florian Kurth, Karsten Schnatbaum, Maren Eckey, Stefan Hippenstiel, Andreas Hocke, Marcel A. Müller, Birgit Sawitzki, Stefan Miltenyi, Friedemann Paul, Marcus A. Mall, Holger Wenschuh, Sebastian Voigt, Christian Drosten, Roland Lauster‡, Nils Lachman‡, Leif-Erik Sander‡, Victor M. Corman‡, Jobst Röhmel‡, Lil Meyer-Arndt‡, Andreas Thiel*‡, Claudia Giesecke-Thiel*‡

INTRODUCTION: The clinical manifestations of COVID-19 vary from asymptomatic infection to respiratory failure. Severe disease courses are primarily associated with advanced age, immune dysfunctions, and comorbidities. Initially, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was thought to encounter an immunologically unprotected population. However, SARS-CoV-2 has significant homologies with endemic seasonal cold coronaviruses (HCoVs), and recent HCoV infection is associated with a less severe course of COVID-19, suggesting a protective role of cross-reactive immunity.

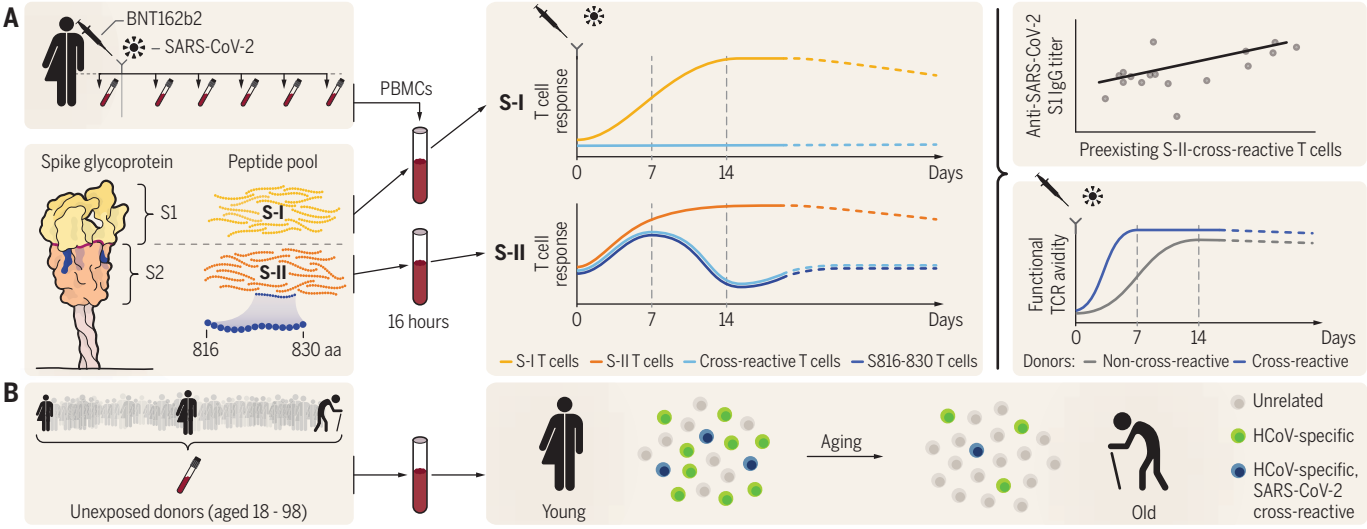
RATIONALE: The preexistence of SARS-CoV-2–cross-reactive CD4$^+$ T cells in unexposed individuals has been repeatedly demonstrated, but their contribution to host responses is an active area of investigation. Here, we investigated the functional role of preexisting SARS-CoV-2–cross-reactive and HCoV-reactive CD4$^+$ T cells with high resolution.

RESULTS: We demonstrate broad CD4$^+$ T cell cross-reactivity in unexposed individuals, with the spike glycoprotein serving as one of the immunodominant targets. Although the N-terminal part of spike (covered by the S-I peptide pool) did not elicit cross-reactive T cell responses, the more HCoV-homologous C-terminal section (covered by the S-II peptide pool) induced T cell responses in unexposed donors. We identified a universal immunodominant coronavirus peptide located within the fusion peptide domain of spike (S816-830) recognized by CD4$^+$ T cells in 20% of unexposed individuals, 50 to 60% of SARS-CoV-2 convalescents, and 97% of BNT162b2-vaccinated individuals. S816-830– and spike–cross-reactive T cells were recruited in immune responses to SARS-CoV-2 infection and BNT162b2 COVID-19 mRNA vaccination. S816-830-reactive T cells initially contributed up to 100% of the S-II–reactive CD4$^+$ T cells but their proportion decreased during the course of the S-II–specific immune response. Upon primary vaccination, cross-reactive cellular and humoral immunity exhibited kinetics typical for secondary immune responses. The frequencies of preexisting cross-reactive T cells correlated positively with functional avidity of the T cell receptor, as well as with the induction and stabilization of anti–SARS-CoV-2-S1-IgG and neutralizing antibodies. Although HCoV-responsive T cells are ubiquitous, their frequencies and the frequencies of SARS-CoV-2–cross-reactive CD4$^+$ T cells decreased with age, consistent with an increased vulnerability of the elderly to severe COVID-19 disease.

CONCLUSION: Preexisting cross-reactive CD4$^+$ T cells enhance immune responses in SARS-CoV-2 infection and BNT162b2 vaccination. Because these cells are greatly diminished in the elderly, our results suggest that their decrease may contribute to the increased susceptibility of this population to severe COVID-19. Preexisting cross-reactive immunity may be responsible for the unexpectedly rapid induction of protective immunity after primary SARS-CoV-2 immunization and the high rate of asymptomatic and mild COVID-19 disease courses. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: andreas.thiel@charite.de (A.T.); giesecke@molgen.mpg.de (C.G.-T.)
†These authors contributed equally to this work.
‡These authors jointly supervised this work.
Cite this article as L. Loyal *et al.*, *Science* **374**, eabh1823 (2021). DOI: 10.1126/science.abh1823
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abh1823



Cross-reactive CD4$^+$ T cells enhance SARS-CoV-2 immune responses upon infection and vaccination. (**A**) Preexisting SARS-CoV-2 S-II– and spike (S)816-830–cross-reactive T cells show booster response characteristics upon SARS-CoV-2 infection and vaccination and their abundance correlates with higher protective anti–SARS-CoV-2 S1-IgG titers and higher functional T cell avidity. (**B**) The frequency of HCoV-reactive and SARS-CoV-2–cross-reactive T cells decreases with age.

RESEARCH ARTICLE

CORONAVIRUS

Cross-reactive CD4+ T cells enhance SARS-CoV-2 immune responses upon infection and vaccination

Lucie Loyal[1,2], Julian Braun[1,2]†, Larissa Henze[1,2]†, Beate Kruse[1,2]†, Manuela Dingeldey[1,2]†, Ulf Reimer[3]†, Florian Kern[3,4]†, Tatjana Schwarz[5]†, Maike Mangold[1,2], Clara Unger[1,2], Friederike Dörfler[1], Shirin Kadler[1,6], Jennifer Rosowski[1,6], Kübrah Gürcan[1,6], Zehra Uyar-Aydin[1,6], Marco Frentsch[7,8], Florian Kurth[9,10], Karsten Schnatbaum[3], Maren Eckey[3], Stefan Hippenstiel[9], Andreas Hocke[9], Marcel A. Müller[2,5,11], Birgit Sawitzki[12], Stefan Miltenyi[13], Friedemann Paul[14,15], Marcus A. Mall[16,17], Holger Wenschuh[3], Sebastian Voigt[18,19], Christian Drosten[2,5,11], Roland Lauster[1,6]‡, Nils Lachman[20]‡, Leif-Erik Sander[9]‡, Victor M. Corman[2,5,11]‡, Jobst Röhmel[16]‡, Lil Meyer-Arndt[1,2,14,15,21]‡, Andreas Thiel[1,2]*‡, Claudia Giesecke-Thiel[22]*‡

The functional relevance of preexisting cross-immunity to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a subject of intense debate. Here, we show that human endemic coronavirus (HCoV)–reactive and SARS-CoV-2–cross-reactive CD4+ T cells are ubiquitous but decrease with age. We identified a universal immunodominant coronavirus-specific spike peptide (S816-830) and demonstrate that preexisting spike- and S816-830–reactive T cells were recruited into immune responses to SARS-CoV-2 infection and their frequency correlated with anti–SARS-CoV-2-S1-IgG antibodies. Spike–cross-reactive T cells were also activated after primary BNT162b2 COVID-19 messenger RNA vaccination and displayed kinetics similar to those of secondary immune responses. Our results highlight the functional contribution of preexisting spike–cross-reactive T cells in SARS-CoV-2 infection and vaccination. Cross-reactive immunity may account for the unexpectedly rapid induction of immunity after primary SARS-CoV-2 immunization and the high rate of asymptomatic or mild COVID-19 disease courses.

Most individuals infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) experience an asymptomatic or mild course of COVID-19. However, severe or fatal disease occurs in ~5% of those infected and is primarily associated with advanced age and comorbidities such as diabetes and chronic cardiovascular, pulmonary, and kidney diseases (*1*). Given that SARS-CoV-2 is a newly emerged human pathogen, it was assumed that SARS-CoV-2 encounters an immunologically naive population. However, SARS-CoV-2 displays considerable homologies with endemic human common cold coronaviruses collectively referred to as "HCoV" (*2*, *3*). There is now strong evidence for cellular and humoral cross-reactivity to SARS-CoV-2 (*3*–*14*), although the role of cross-reactive immunity in SARS-CoV-2 infection is unclear (*2*, *8*, *15*, *16*). Recent HCoV infection is associated with less severe COVID-19, suggesting a protective role (*17*). A better understanding of the extent and impact of cross-immunity in SARS-CoV-2 infection and vaccination is needed because cognate cross-immunity may influence the efficacy of vaccination regimens.

Here, we investigated the functional role of preexisting SARS-CoV-2– and HCoV-reactive CD4+ T cells. The SARS-CoV-2 spike glycoprotein (spike) was the dominant target of broad T cell cross-reactivity in unexposed individuals, which decreased with age. We identified an immunodominant coronavirus peptide located within the fusion peptide domain of spike (S816-830) recognized by CD4+ T cells in 20% of unexposed individuals, 50 to 60% of SARS-CoV-2 convalescents, and 97% of BNT162b2-vaccinated individuals. S816-830– and spike–cross-reactive T cells were recruited into primary SARS-CoV-2 immune responses and also into BNT162b2 COVID-19 mRNA vaccination responses. Finally, upon primary vaccination, cross-reactive immunity exhibited kinetics similar to those in secondary immune responses. Already at an early stage of the immune response, the frequencies of preexisting cross-reactive T cells correlated positively with functional avidity as well as with the induction and stabilization of anti–S1-IgG antibodies. Thus, cross-reactive CD4+ T cells accelerate the immune response in SARS-CoV-2 infection and vaccination. These findings add to the discussion surrounding single-dose vaccination of healthy adults and multiple-dose vaccination of the elderly.

Frequent and broad SARS-CoV-2 cross-reactivity in unexposed healthy donors

To determine the extent of cellular cross-reactivity to SARS-CoV-2 antigens, we stimulated CD4+ T cells of 60 unexposed healthy donors and 59 COVID-19 convalescents as controls (table S1) with peptide pools covering all open reading frames (ORFs) of SARS-CoV-2, referred to here as the "SARS-CoV-2 orfeome" (Fig. 1A). The SARS-CoV-2 orfeome consists of 11 ORFs, five of which [N, spike, E, M, and ORF1a/b (encoding for the nonstructural proteins (NSPs) 1 to 16)] are also found in HCoVs 229E, OC43, NL63, and HKU1. Amino acid sequence alignment revealed discrete areas of high homology in almost all SARS-CoV-2 proteins to the corresponding proteins in HCoVs. Parts of the ORF1a/b including NSP8, NSP10, and NSP12-16 displayed the highest degree of homology and thus potential cross-reactive epitopes to all HCoVs (fig. S1A). Nevertheless, COVID-19 convalescents did not show significantly increased CD4+ T cell reactivity against the NSPs compared with unexposed individuals (Fig. 1A and fig. S1B). Reactivity against the combination of spike N-terminal S-I (amino acid residues 1 to 643), C-terminal S-II (amino acid residues 633 to 1273), N, and M peptide pools clearly distinguished COVID-19 convalescents

[1]Si-M/"Der Simulierte Mensch," a Science Framework of Technische Universität Berlin and Charité – Universitätsmedizin Berlin, Berlin, Germany. [2]Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt – Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany. [3]JPT Peptide Technologies GmbH, Berlin, Germany. [4]Department of Clinical and Experimental Medicine, Brighton and Sussex Medical School, Brighton, UK. [5]Institute of Virology, Charité – Universitätsmedizin Berlin, Berlin, Germany. [6]Medical Biotechnology, Institute for Biotechnology, Technische Universität Berlin, Berlin, Germany. [7]Department of Hematology, Oncology and Tumor Immunology, Charité – Universitätsmedizin Berlin, Berlin, Germany. [8]Therapy-Induced Remodeling in Immuno-Oncology, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany. [9]Department of Infectious Diseases and Respiratory Medicine, Charité – Universitätsmedizin Berlin, Berlin, Germany. [10]Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, and Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [11]German Centre for Infection Research (DZIF), Partner Site Charité, Berlin, Germany. [12]Institute of Medical Immunology, Charité – Universitätsmedizin Berlin, Berlin, Germany. [13]Miltenyi Biotec GmbH, Bergisch-Gladbach, Germany. [14]Experimental and Clinical Research Center, Max Delbrueck Center for Molecular Medicine, and Charité – Universitätsmedizin Berlin, Berlin, Germany. [15]Clinical Neuroimmunology, NeuroCure Clinical Research Center, Charité – Universitätsmedizin Berlin, Berlin, Germany. [16]Department of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine, Charité – Universitätsmedizin Berlin, Berlin, Germany. [17]German Center for Lung Research, Associated Partner, Berlin, Germany. [18]Department of Infectious Diseases, Robert Koch Institute, Berlin, Germany. [19]Institute for Virology, Universitätsklinikum Essen, Essen, Germany. [20]Institute for Transfusion Medicine, Tissue Typing Laboratory, Charité – Universitätsmedizin Berlin, Berlin, Germany. [21]Department of Neurology and Experimental Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany. [22]Max Planck Institute for Molecular Genetics, Berlin, Germany.

***Corresponding author. Email: andreas.thiel@charite.de (A.T.); giesecke@molgen.mpg.de (C.G.-T.)**

†These authors contributed equally to this work.

‡These authors jointly supervised this work.

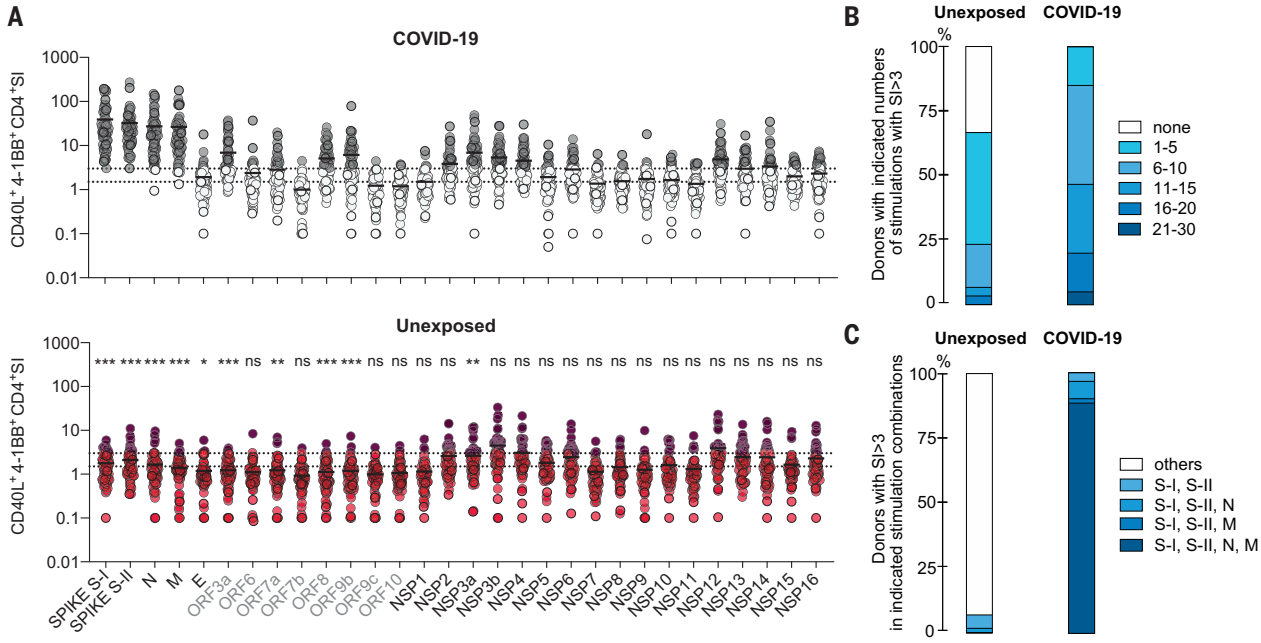


Fig. 1. CD4$^+$ T cell cross-reactivity against the SARS-CoV-2 orfeome. (**A**) Ex vivo stimulation of PBMCs from COVID-19 convalescent patients (top panel, n = 59) and unexposed individuals (bottom panel, n = 60). The percentage of CD40L$^+$4-1BB$^+$ CD4$^+$ T cells among stimulated PBMCs was divided by the percentage of these cells among unstimulated PBMCs to determine the SI shown on the y-axis. The SARS-CoV-2 orfeome peptide pools used for stimulation are shown below the bottom panel. Gray labels highlight proteins exclusive for SARS-CoV-2 (i.e., those not shared with HCoVs). Gray circles (COVID-19) or red circles (unexposed) identify donors with an SI ≥ 3. Dotted lines indicate an SI of 1.5 and 3. Statistically significant differences between COVID-19 convalescents and unexposed groups (with respect to each peptide pool) are indicated above the bottom panel. *P < 0.05, **P < 0.01, and ***P < 0.001; ns, not significant at P > 0.05 by unpaired Student's t test. (**B**) Bars show the proportions of individuals with the indicated number of SARS-CoV-2 orfeome peptide pool stimulations with an SI ≥ 3. (**C**) Proportions of individuals with an SI ≥ 3 for each stimulation in each indicated stimulation combination.

from unexposed individuals irrespective of the disease course (Fig. 1, A to C). In unexposed individuals, we detected variable but low CD4$^+$ T cell reactivity to virtually all SARS-CoV-2 antigens, including those exclusive to SARS-CoV-2 (not shared with HCoVs). However, the degree of amino acid sequence homology between HCoV and SARS-CoV-2 proteins did not correlate with cross-reactivity (fig. S1C). Thus, apart from cognate cross-reactivity (resulting, for example, from previous exposure to similar proteins found in HCoVs), we also identified noncognate cross-reactivity (i.e., cross-reactivity that cannot be explained by the previous exposure to similar proteins in HCoVs). Of all 30 orfeome peptide pools, the spike S-I and S-II pools alone elicited T cell reactivity in all COVID-19 convalescents and in a subset of unexposed individuals. Because antibodies to spike induced by SARS-CoV-2 infection can neutralize the virus (*18*) and most of the recently approved SARS-CoV-2 vaccines are highly effective and include spike as the main vaccine antigen, we examined cellular immunity to spike more closely.

SARS-CoV-2 spike S-II–cross-reactive T cells decrease with age

A distinct feature of SARS-CoV-2 infection is the strong correlation of higher age with disease severity. Immunosenescence is associated with a lack of newly generated T cells and, instead, the expansion of a small number of clones resulting from persistent infections, which limits the breadth and quality of T cell responsiveness (*19*, *20*). To assess the impact of age on SARS-CoV-2–cross-reactive T cell immunity, we examined SARS-CoV-2 spike–specific CD4$^+$ T cell responses in 568 unexposed individuals and 174 COVID-19 convalescents (Fig. 2A and table S1). T cells reacting to a peptide pool representing a mixture of selected T cell epitopes from common pathogens (the CEFX pool) remained relatively stable with age in both cohorts (Fig. 2A). COVID-19 convalescents displayed a significant age-associated increase in spike S-I–reactive T cells that correlated with higher disease severity in the elderly (table S1). However, consistent with our previous findings (*3*) in unexposed individuals, T cell cross-reactivity to spike S-I was rare, close to the limit of detection, and remained stable, albeit at low levels with increasing age. By contrast, reactivity to S-II was more frequent and generally higher in unexposed individuals but significantly decreased with increasing age (Fig. 2A). When total CD4$^+$ T cells were analyzed for activation-induced interferon-γ (IFN-γ) or tumor necrosis factor-α (TNF-α) expression, we did not detect any age-related changes (fig. S2A). However, among bona fide T cell receptor (TCR)–activated antigen-specific CD40L$^+$4-1BB$^+$ CD4$^+$ T cells, IFN-γ$^+$TNF-α$^+$ cells decreased with age (fig. S2B). In contrast to CD40L$^+$4-1BB$^+$ CD4$^+$ T cells, total CD40L$^+$ CD4$^+$ T cells, which can also be induced in part in a TCR-independent manner (*21*), did not decrease with age, consistent with the large compartment of memory T cells in older individuals (fig. S2C). Thus, elderly individuals exhibit decreased cognate cross-reactive immunity to the SARS-CoV-2 spike S-II portion, which exhibits higher homology to HCoV than the S-I portion.

Low CD3 surface expression identifies SARS-CoV-2–reactive T cells with high functional avidity ex vivo

To assess the quality of the spike–cross-reactive T cell response in terms of functional T cell avidity, we examined the level of CD3 surface expression in CD40L$^+$4-1BB$^+$ CD4$^+$ T cells after short-term in vitro stimulation (Fig. 2B). Strong TCR activation, which is characteristic of T cells with high TCR avidity, blocks recycling of the TCR-CD3 complex and can be detected by reduced CD3 surface expression (*22*), a phenomenon known as high functional avidity. Thus, cognate cross-reactivity with higher probability of high functional avidity is distinguishable

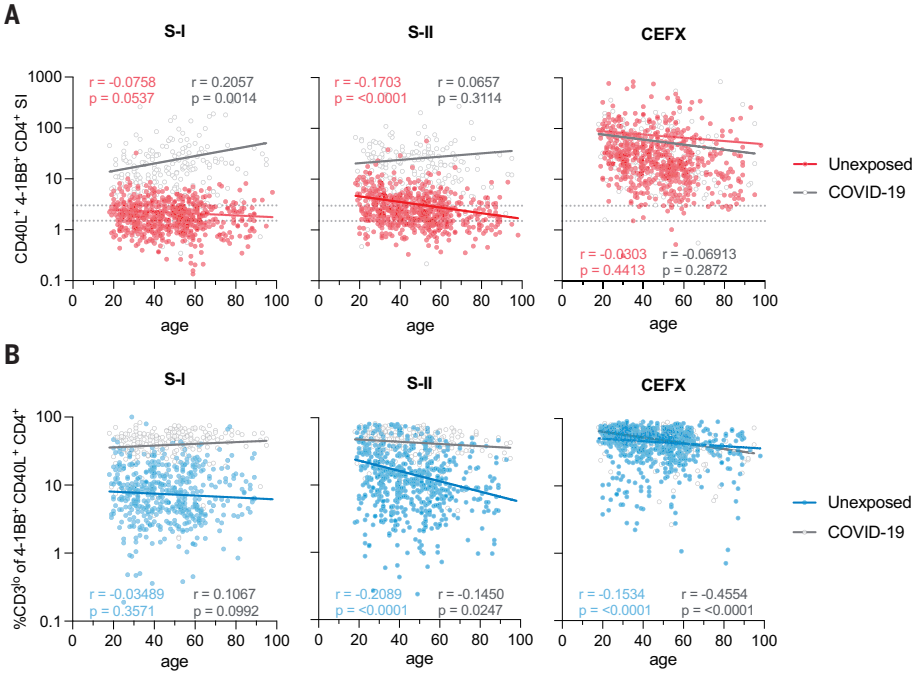


Fig. 2. The magnitude of SARS-CoV-2 cross-reactivity decreases with age. (**A**) Scatter plots showing the SI (CD40L$^+$41BB$^+$ CD4$^+$ T cells) among PBMCs stimulated with SARS-CoV-2 S-I, SARS-CoV-2 S-II, or CEFX (known T cell–stimulating peptides from CMV, EBV, influenza virus, and other common pathogens) plotted against age in n = 568 unexposed donors and n = 174 COVID-19 convalescents. Dotted lines indicate an SI of 1.5 and 3. (**B**) Frequencies of CD3^{lo} cells among S-I–, S-II–, or CEFX-reactive CD40L$^+$4-1BB$^+$ CD4$^+$ T cells over age. CD3^{lo} frequencies are shown for T cell responses with an SI ≥ 1.5. Regression lines denote linear regression on age in each group; the corresponding Pearson correlation coefficients are shown.

from noncognate cross-reactivity with higher probability of low functional avidity by analyzing the frequency of CD3^{lo} T cells among TCR-activated CD4$^+$ T cells (fig. S3, A and B). After stimulation with spike S-I and S-II peptide pools, COVID-19 convalescents showed high frequencies of S-I– and S-II–activated CD4$^+$ T cells that largely lacked CD3 expression characteristic of cognate T cell activation (Fig. 2B). In unexposed individuals, however, the frequency of CD3^{lo} cells among S-I– and S-II–activated CD4$^+$ T cells was markedly lower. Nevertheless, especially in the younger subjects, spike S-II stimulation induced higher frequencies of CD3^{lo} cells than S-I stimulation, indicating that spike S-I–cross-reactive CD4$^+$ T cells have high functional avidity (Fig. 2B). This is consistent with the high degree of homology between the C-terminal S-II portions of SARS-CoV-2 spike and HCoV spike proteins.

Frequency of HCoV spike–reactive high-functional-avidity CD4$^+$ T cells decreases with age

We hypothesized that previous HCoV exposures induce cognate cross-reactive CD4$^+$ T cells. Therefore, we next characterized CD4$^+$ T cell immunity to HCoV spike in unexposed individuals and COVID-19 convalescents. HCoV-S-I– and HCoV-S-II–reactive CD4$^+$ T cells were more readily detectable than SARS-CoV-2 spike–specific T cells and found in 80% (S-I) and 98% (S-II) of SARS-CoV-2–unexposed individuals, respectively (Fig. 3A). Their frequency decreased with age, and SARS-CoV-2 infection did not result in an increase in HCoV-S-I– or HCoV-S-II–reactive T cells. We also examined the functional avidities of HCoV-reactive CD4$^+$ T cells (Fig. 3B). High frequencies of CD3^{lo} T cells were found among both HCoV-S-I– and HCoV-S-II–reactive CD4$^+$ T cells, although they significantly decreased with advancing age. Thus, a high degree of HCoV exposure in the population appears to lead to widespread cross-reactivity to the SARS-CoV-2 spike protein. HCoV-reactive CD4$^+$ T cells frequently contain a subset of cells with high functional avidity but significantly decrease with age.

The immunodominant peptide S816-830 is recognized by SARS-CoV-2 spike S-II–cross-reactive CD4$^+$ T cells

All SARS-CoV-2–cross-reactive unexposed donors showed a response against at least two (S-I) or three (S-II) HCoVs, suggesting that repeated infection with different HCoVs establishes a detectable prominent SARS-CoV-2–cross-reactive T cell pool already early in life and/or that specific T cells are directed against highly homologous sequences shared across multiple HCoVs and SARS-CoV-2 (Fig. 4A). We next investigated whether HCoV spike glycoprotein-specific T cells directly cross-react to SARS-CoV-2 spike glycoprotein. Short-term CD40L$^+$4-1BB$^+$ OC43 S-I– or S-II–reactive CD4$^+$ T cell lines were restimulated with OC43- or SARS-CoV-2 spike pool S-I and S-II, respectively. Six out of 18 OC43 S-II–specific T cell lines displayed cross-reactivity against SARS-CoV-2 S-II, whereas OC43 S-I–specific T cell lines lacked cross-reactivity against SARS-CoV-2 S-I (Fig. 4B). We further identified and validated two overlapping T cell–stimulating peptides, peptides 204 (SKRSFIEDLLFNKVT, amino acids 813 to 827) and 205 (FIEDLLFNKVTLADA, amino acids 817 to 831), derived from the S-II portion of spike, in all five donors analyzed (fig. S4, A to D). Only one donor responded to other identified peptides (peptides 188, 189, and 251) (fig. S4B). Sequence alignment revealed that S-II peptides 204 and 205 together covered the fusion peptide domain of spike, which is characterized by strong homology with HCoV (fig. S4C). By analyzing additional 15-aa peptides along the sequence covered by the peptides 204 and 205, we identified the sequence SFIEDLLFNKVTLAD (amino acids 816 to 830) as an immunodominant coronavirus peptide, hereafter referred to as S816-830 (peptide 204_3; fig. S4D). We next examined direct ex vivo T cell reactivity against S816-830 compared with a control peptide 284 (amino acids 1133 to 1147, hereafter referred to as S1133-1147) and the SARS-CoV-2 spike S-II peptide pool in 48 unexposed individuals and 22 COVID-19 convalescents. S816-830–reactive CD4$^+$ T cells were detected in 50% of convalescents and 20% of unexposed individuals, with significantly higher frequencies in the former (Fig. 4C). Antibodies to the SARS-CoV-2 spike amino acid residues S809 to S826 were previously reported in COVID-19 patients but also in unexposed individuals (*23*, *24*). When we examined the sera of responders and nonresponders to the S816-830 T cell assay, we detected S809-826–binding antibodies in all individuals. However, significantly higher concentrations of these antibodies were found in COVID-19 convalescents with substantially more S816-830–reactive T cells (Fig. 4D). Compared with definite nonresponders [stimulation index (SI) < 1.5], definite S816-830 peptide responders (SI ≥ 3) were more frequently positive for HLA-DPB1*02:01, HLA-DPB1*04:02, and especially homozygous expression of HLA-DPB1*04:01 (Fig. 4E). Because HLA-DPA1*01:03 was found in 100% of the responders and 94.8% of the nonresponders, we investigated whether combinations of HLA-DPA1*01:03 and HLA-DPB1*02:01/DPB1*04:01/DPB1*04:02 were likely to present peptide S816-830 or fragments thereof. HLA-peptide–binding predictions identified excellent potential binders (fig. S4E),

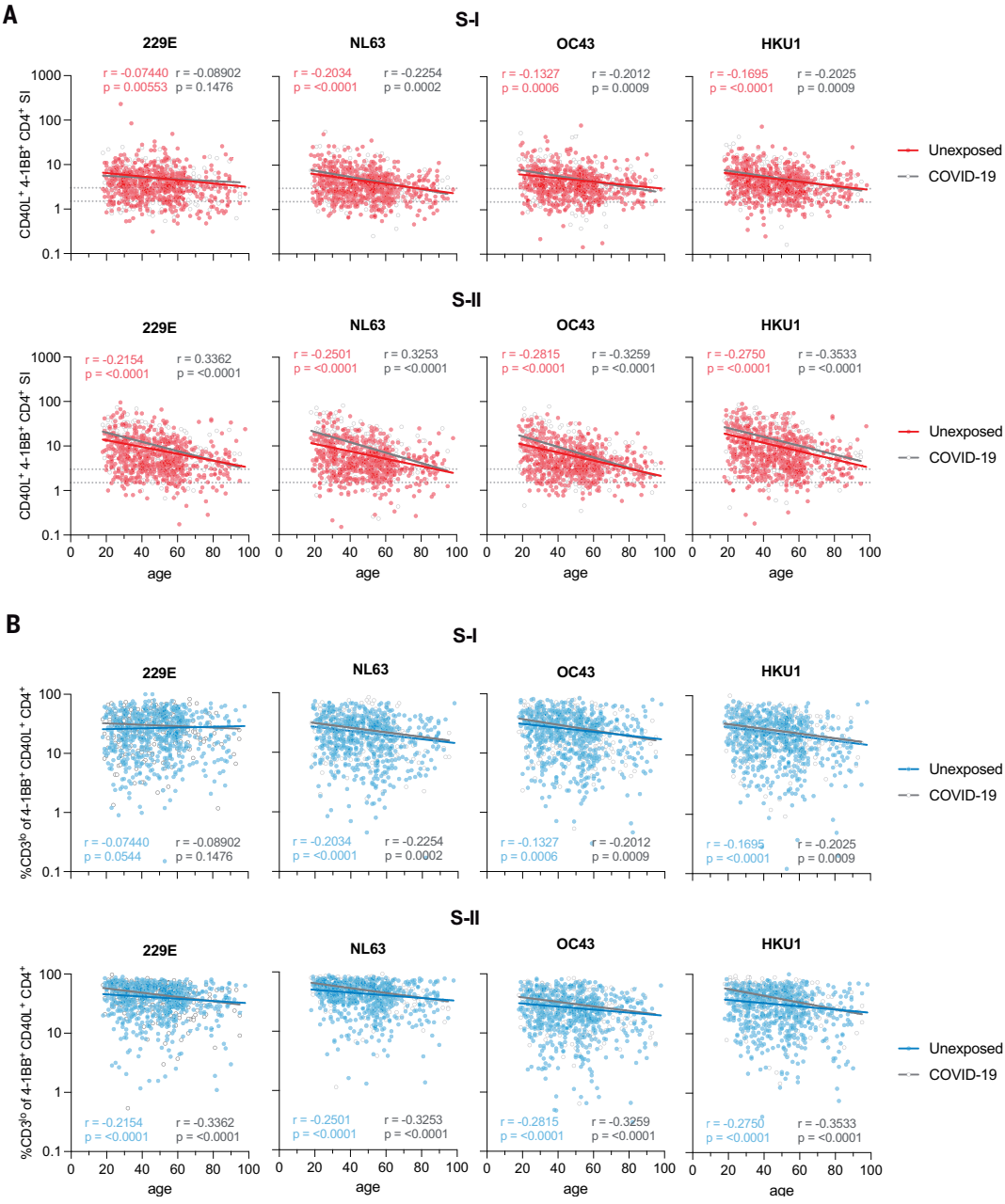


Fig. 3. Frequencies of high-functional-avidity T cells specific for spike S-II from HCoVs decrease with age. (**A**) Scatter plots show the SI of CD40L$^+$4-1BB$^+$ CD4$^+$ T cells in unexposed individuals (n = 568) and COVID-19 convalescents (n = 174) after PBMC stimulation with HCoV (229E, NL63, OC43, and HKU1) spike S-I or S-II peptide pools plotted against age. Dotted lines indicate an SI of 1.5 and 3. (**B**) Frequencies of CD3^{lo} cells in CD40L$^+$4-1BB$^+$ CD4$^+$ T cells from unexposed and COVID-19 convalescents plotted against age. CD3^{lo} frequencies are shown for T cell responses with an SI ≥ 1.5. Regression lines denote linear regression on age in each group; the corresponding Pearson correlation coefficients are shown.

which was also true for the homologous S816-830 peptide in other HCoVs (fig. S4F).

Preexisting SARS-CoV-2 spike S-II–cross-reactive T cells are recruited into primary SARS-CoV-2 immune responses

A still open question is whether and to what extent SARS-CoV-2–cross-reactive T cells influence the disease course of primary SARS-CoV-2 infection. By monitoring the healthy, previously unexposed study participants for primary SARS-CoV-2 infection, we identified 17 cases of acute primary SARS-CoV-2 infection (Fig. 5 and table S2). All 17 patients showed detectable virus titers (fig. S5A) and mild COVID-19 disease course (no hospitalization required) (table S2). Robust CD4$^+$ T cell responses specific to SARS-CoV-2 spike S-I and S-II were detected, and the proportions of HLADR$^+$CD38$^+$ cells among CD40L$^+$4-1BB$^+$ CD4$^+$ T cells significantly increased at follow-up time points 1 and 2 (3 to 16 days), indicating their in vivo activation (Fig. 5, A and B). CD3^{lo} cells substantially increased during acute primary SARS-CoV-2 infection and remained at high levels after the infection resolved (Fig. 5C). Individuals who already had spike S-II–cross-reactive CD4$^+$ T cells with a SI ≥ 3 at baseline showed significantly higher functional avidity throughout the initiation of the T cell response (Fig. 5D). S816-830–reactive T cells increased in both frequency and in functional avidity in 10 of 17 donors after infection (Fig. 5, E and F). Immunoglobulin G (IgG) antibodies against the S809-826 peptide were boosted as early as 3 to 9 days (follow-up time point 1) after the presumed infection (Fig. 5G). Anti–SARS-CoV-2-S1-IgG serum antibodies were detectable at follow-up time point 2 and peaked after day 20 in most individuals, although their kinetics and quantity varied widely (Fig. 5H). Anti–SARS-CoV-2-S1 binding antibody (IgG) units at late time points positively correlated with S-II– but not S-I–cross-reactive T cell levels at day 0, suggesting that preexisting cross-reactive CD4$^+$ T cells enhance SARS-CoV-2–specific humoral immunity (Fig. 5I, left). Moreover, the neutralizing antibody titers also positively correlated

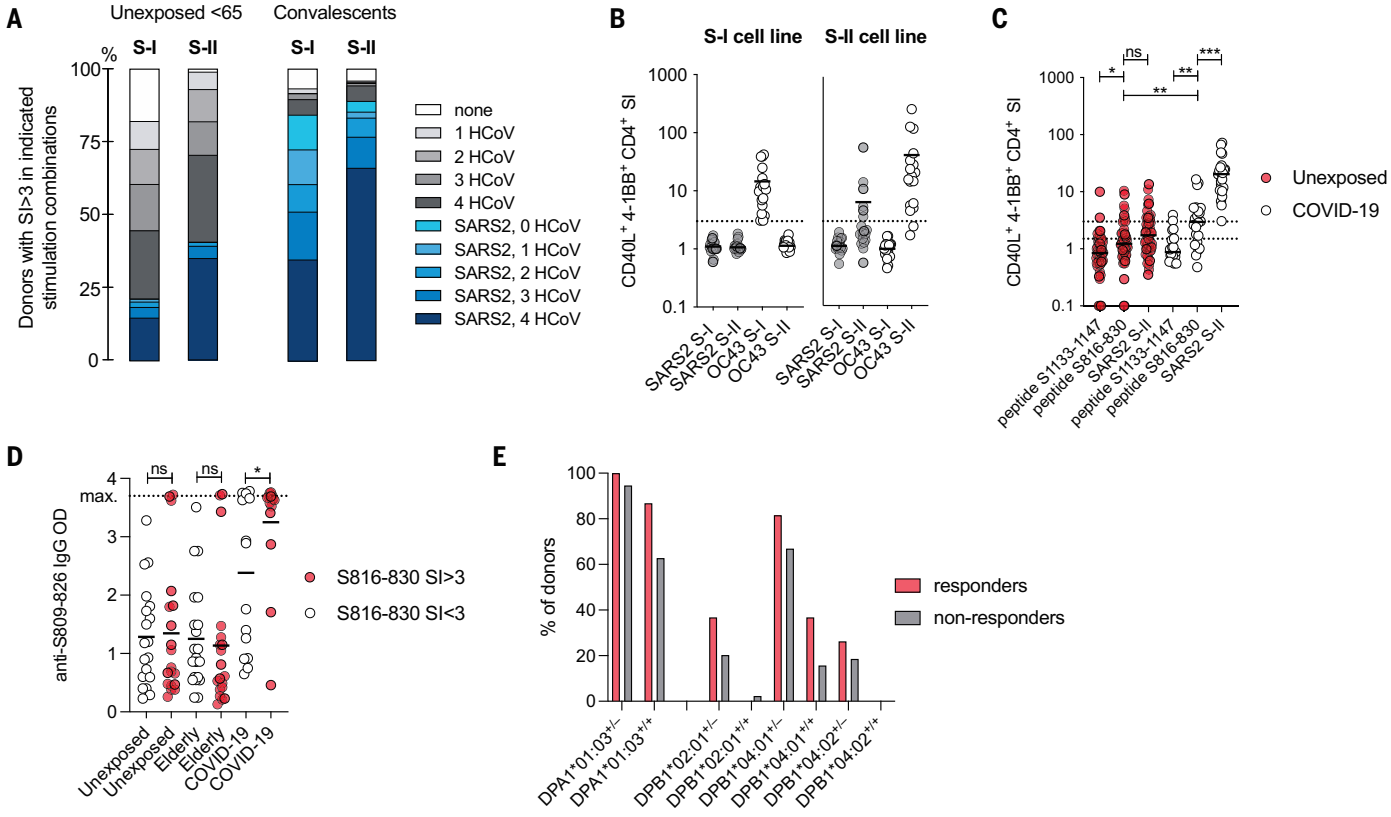


Fig. 4. Peptide S816-830 constitutes an immunodominant epitope of SARS-CoV-2 T cell cross-reactivity. (**A**) Bars show the proportions of unexposed individuals <65 years of age (n = 491) and COVID-19 convalescents (n = 174, 18–79 years) with S-I– or S-II–specific T cell responses to HCoV and/or SARS-CoV-2 with an SI ≥ 3. (**B**) Plots showing the SI ($CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells) of short-term T cell lines derived from OC43 S-I– and S-II–reactive primary T cells after restimulation with autologous antigen-presenting cells in the presence of OC43 or SARS-CoV-2 spike glycoprotein pools S-I and S-II. The dotted line indicates an SI of 3. (**C**) The SIs of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells from unexposed (n = 48) or COVID-19 convalescents (n = 22) after stimulation with the single peptide 204_3 (S816-830), the control single peptide 284 (S1133-1147), or the S-II peptide pool. (**D**) Levels (optical density, OD) of anti–S809-826 peptide IgG (ELISA) in unexposed young (<65 years) and elderly (>65 years) individuals as well as COVID-19 convalescents. ELISA plates were coated with an 18-aa peptide overlapping by 11 aa with S816-30. Serum was diluted 1:100. (**E**) Bars show the frequencies of common class-II HLA alleles in definite S816-830 responders (SI ≥ 3) and definite nonresponders (SI < 1.5) (n = 308). +/+, homozygous; +/–, heterozygous. *P < 0.05, **P < 0.01, ***P < 0.001; ns at P > 0.05 by Student's t test.

with spike S-II– but not spike S-I–cross-reactive $CD4^+$ T cells at baseline, suggesting a protective role of cross-reactive $CD4^+$ T cells (Fig. 5I, middle and right). Finally, the frequency of HCoV-reactive $CD4^+$ T cells also increased in almost all individuals shortly after primary SARS-CoV-2 infection (Fig. 5J). There was a concomitant increase in the frequency of $CD3^{lo}$ cells (fig. S5B) and $HLADR^+CD38^+$ cells (fig. S5C) among HCoV-reactive $CD4^+$ T cells, demonstrating that preexisting HCoV-reactive cellular immunity was activated and transiently expanded during primary SARS-CoV-2 infection. Clearly, preexisting SARS-CoV-2 S-II–cross-reactive $CD4^+$ T cells were recruited into primary SARS-CoV-2 immune responses in healthy, previously unexposed individuals. Thus, the quantity and functional avidity of preexisting cross-reactive cellular immunity corresponds to the quality and magnitude of specific cellular and humoral anti–SARS-CoV-2 responses. It may therefore contribute to a milder course of COVID-19 by limiting viral propagation.

BNT162b2 vaccination reactivates preexisting SARS-CoV-2 spike S-II–cross-reactive T cells

Finally, we investigated how preexisting SARS-CoV-2 S-II–cross-reactive T cells in healthy, unexposed individuals influence the course of BNT162b2 COVID-19 spike mRNA vaccine responses. We monitored baseline and follow-up humoral and T cell responses against SARS-CoV-2 and HCoV spike glycoproteins in 31 healthy adults who underwent primary (day 0) and booster (day 21) vaccination with BNT162b2. At day 21, 30 of 31 donors had detectable anti–SARS-CoV-2 S1 IgG, and all donors had detectable anti–SARS-CoV-2 S1 IgA levels (Fig. 6A). Booster vaccination further increased these antibody levels. Primary vaccination also induced robust S-I– and S-II–reactive $CD4^+$ T cell responses in all individuals that were only slightly enhanced by booster vaccination (Fig. 6B). The kinetics of S-I– and S-II–reactive T cells differed in that S-II–reactive T cells showed a sharp increase from baseline to day 7 but not thereafter, whereas S-I–reactive T cells showed an additional significant increase from day 7 to day 14 (Fig. 6, B and C). This was indicative of the secondary response kinetics of S-II–reactive cells and the primary response kinetics of S-I–reactive cells. High-functional-avidity, $CD3^{lo}$ $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells increased more rapidly in cross-reactive donors (Fig. 6, D and E). Moreover, at day 14, S-I– and S-II–reactive $CD4^+$ T cells included high frequencies of $HLADR^+CD38^+$ cells in all but three donors, indicating their recent in vivo activation (Fig. 6F). Like SARS-CoV-2–specific T cells, HCoV S-II–reactive T cells were significantly increased 7 days after primary vaccination (Fig. 6G). This was associated with an increased frequency of HCoV S-II–reactive $HLADR^+CD38^+$ T cells (Fig. 6H). Thus, cognate cross-reactive T cells were activated early in response to SARS-CoV-2 spike–specific vaccination but did not expand thereafter. All but two of 31 donors (94%) responded with T cells that had high functional avidity to S816-830 at days 7 and 14 (Fig. 6, I and J).

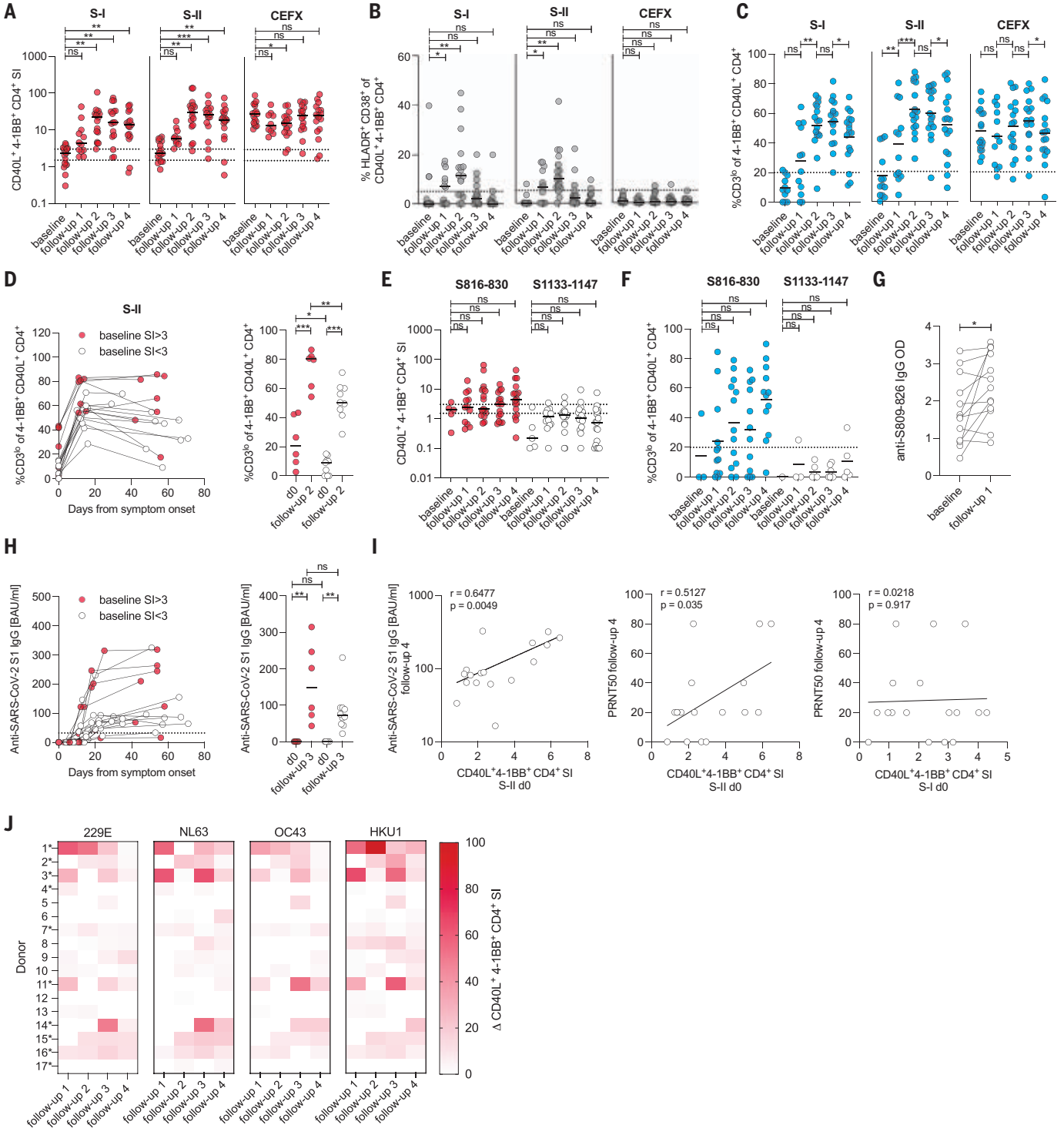


Fig. 5. HCoV-specific SARS-CoV-2–cross-reactive T cells are recruited into the primary SARS-CoV-2 infection response. (**A** to **C**) SI of CD40L$^+$4-1BB$^+$ CD4$^+$ T cells (A), frequencies of HLADR$^+$CD38$^+$ cells (B), and frequencies of CD3^{lo} cells (C) among CD40L$^+$4-1BB$^+$ CD4$^+$ T cells after stimulation with SARS-CoV-2 S-I, S-II, and CEFX peptide pools of donors before infection (baseline) and at four different follow-up time points (table S2) after symptom onset. CD3^{lo} frequencies are shown only for T cell responses with an SI ≥ 1.5. (**D**) Changes to CD3^{lo} frequencies among CD40L$^+$4-1BB$^+$ CD4$^+$ T cells between baseline, follow-up 2 (10 to 16 days after symptom onset), and follow-up 4 (29 to 71 days after symptom onset) (left plot), and statistics (right plot) for baseline and follow-up measurement time point 2 in cross-reactive donors (baseline SI ≥ 3, red circles) and non–cross-reactive donors (baseline SI < 3, white circles). (**E** and **F**) SI (E) and frequency (F) of CD3^{lo} of CD40L$^+$4-1BB$^+$ CD4$^+$ T cells after stimulation with peptide S816-830 or control peptide S1133-1147. CD3^{lo} frequencies are shown for T cell responses with an SI ≥ 1.5. (**G**) Levels (OD) of anti–S809-826 peptide IgG (ELISA) at baseline and follow-up time point 1 (3 to 9 days after symptom onset). ELISA plates were coated with an 18-aa peptide overlapping by 11 aa with S816-830. (**H**) Anti–S1-IgG binding antibody units (BAUs) in cross-reactive (baseline SI ≥ 3, red circles) and non–cross-reactive donors (baseline SI < 3, white circles) were plotted against time (left)

and compared between baseline and follow-up 3 (right). (**I**) Scatter plots showing the relationship between anti–SARS-CoV-2 S1 IgG antibody levels (OD) at follow-up 4 and the SI of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells upon S-II stimulation at baseline (left), the relationship between neutralizing antibody titers (PRNT50) at follow-up 4, and the SI of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells upon S-II stimulation (left) or S-I stimulation (right) at baseline. (**J**) Heatmap showing the change in SI of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells after stimulation with S-II pools of the indicated HCoVs. Δ represents the change of the parameter at the given time point relative to baseline (i.e., white depicts no increase). Asterisks indicate S816-830 peptide responders. For (A), (B), (E), and (F), $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and ns at $P > 0.05$ by repeated-measures one-way-ANOVA with Dunnett's correction. For (C) and (G), $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and ns at $P > 0.05$ by paired Student's t test. For (D) and (H), $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and ns at $P > 0.05$ by Student's t test. For (E), ns at $P > 0.05$ by paired Student's t test. For (I) ns at $P > 0.05$ by Pearson correlation.

S816-830–reactive T cells initially contributed up to 100% of the $CD40L^+4\text{-}1BB^+$ cells in S-II stimulations, but their proportion decreased as other specificities increased during the course of the SARS-CoV-2 S-II–specific immune response (Fig. 6K). Thus, HCoV imprinting does not appear to hamper an immune response tailored to SARS-CoV-2. We observed a correlation between the S816-830–reactive and the S-II–reactive T cell response at day 0 that was even more pronounced at day 7, emphasizing the importance of the S816-830 peptide in the early stages of the anti–SARS-CoV-2 cellular immune response (Fig. 6L). A humoral response to S809-826 (overlapping with S816-830) was detectable as early as 7 days after primary vaccination (Fig. 6M) and this was distinct from the slower anti–SARS-CoV-2-S1-IgG response. This supports the concept that preexisting cross-reactive immunity mediates secondary response kinetics (*26*).

Discussion

The functional relevance of preexisting cognate cross-immunity to SARS-CoV-2 is a subject of intense debate. Noncognate cross-reactivity has been reported but appears to play a minor role compared with HCoV-mediated cognate cross-reactivity (*15*). A recent HCoV infection is associated with a less severe course of COVID-19 (*17*). Unexpectedly, >90% of the population is HCoV-seropositive. Thus, a large proportion of the population might benefit from cross-reactive humoral immunity (*27*, *28*). However, prepandemic serum from PCR-validated HCoV-positive individuals contains neutralizing antibodies against all HCoVs but not SARS-CoV-2 (*27*). In a subsequent study, only low spike-specific cross-reactive antibody activity was detected in just five of 34 donors with recent HCoV infection and in just one of 31 donors without recent HCoV infection, indicating that humoral cross-immunity is weak and decays rapidly (12). Finally, although infection with SARS-CoV-2 increases the prevalence of antibodies against seasonal HCoVs, these antibodies do not provide protection, which highlights the role of cross-reactive cellular immunity (*9*, *27*, *28*).

Recently, T cells cross-reactive to several SARS-CoV-2 antigens were identified in unexposed individuals using predicted peptides individually (*4*, *5*) or as megapools (*8*, *29*). Our work reveals significant cross-reactivity of ORF1a/b-encoded proteins but also shows that most of the anti-SARS-CoV-2 reactivity is directed against the spike, N, and M proteins. We further demonstrate that the magnitude and quality of SARS-CoV-2 cross-reactivity and HCoV reactivity declines with age. The failure of an aging immune system to maintain HCoV-induced SARS-CoV-2–cross-reactive T cells along with a smaller pool of naïve T cells that can be recruited into SARS-CoV-2–specific responses (*20*) may contribute to the increased susceptibility of elderly to severe COVID-19. Our results show that HCoV-specific, SARS-CoV-2–cross-reactive T cells contribute to SARS-CoV-2 immune responses upon infection and vaccination. Additionally, such cognate cross-reactivity correlates with a rapid cellular and enhanced humoral response, both of which may favor mild disease courses. The sequential administration of different haptens sharing the same carrier to mice induced preexisting T cell help for the second hapten, leading to more efficient B cell recruitment in secondary immunization (*30*). Accordingly, B cells recognizing SARS-CoV-2 may benefit from HCoV-reactive T cells cross-reacting with SARS-CoV-2 peptides. Further studies in mice showed that increasing the numbers of antigen-specific T cells at the onset of the immune response also increased B cell activation and proliferation. Moreover, the presence of cognate T cell help during viral infection promotes germinal center formation, which is required for fast and high-affinity antibody generation (*30*–*32*). Because the early induction of SARS-CoV-2 T cell reactivity has been associated with rapid viral clearance and mild disease (*33*), cross-reactive T cells that enhance the immune response to SARS-CoV-2 may well serve as a correlate of immune protection against severe COVID-19 disease courses (*34*, *35*).

Upon BNT162b2 vaccination, we observed immune responses that exceeded the response to actual SARS-CoV-2 infection in terms of spike-specific T cell and antibody levels. Responses to S-II, unlike responses to the non–cross-reactive S-I, however, displayed kinetics reminiscent of a secondary immune response (*25*, *26*). These observations may provide an explanation for the results of large studies showing the high efficacy of SARS-CoV-2 vaccines. Protection levels against SARS-CoV-2 infection have been reported to be >75% as early as 15 to 28 days after primary vaccination with BNT162b2 (*36*). In addition, just one dose of the BNT162b2 or the Astra Zeneca ChAdOx1 vaccine reduced the risk of hospitalization by 85 and 94%, respectively, at days 28 to 34 after primary vaccination, an unusually high efficacy for a primary vaccination (*36*). Similarly, a single-shot vaccination based on AdV26 adenovirus–encoded modified spike protein from Johnson & Johnson has been reported to have a vaccine efficacy of 66% and was recently approved by the U.S. Food and Drug Administration and the European Medicines Agency (*37*, *38*). Our results may provide an immunological explanation for the reported high efficacies. Conversely, in the elderly, who have waning HCoV T cell reactivity and thus reduced SARS-CoV-2 T cell cross-reactivity, additional booster vaccinations may be critical (*39*, *40*).

The immunodominant cross-reactive peptide S816-830 identified here is located within the highly conserved spike fusion peptide domain downstream of the S2′ cleavage site (*41*). We demonstrate that S816-830–reactive T cells are efficiently recruited into the SARS-CoV-2 response in most infected and nearly all vaccinated individuals. Previous reports have also shown that specific antibodies against this region are generated after SARS-CoV-2 infection and vaccination with BNT162b2 (*23*, *24*). In addition, it has been proposed that antibodies specific to the S2 portion of spike have neutralizing activity and may be involved in the early induction of protection before SARS-CoV-2 S1–specific antibodies emerge (*28*, *41*–*43*). In summary, the S816-830 peptide may serve as a conserved universal coronavirus target in the S2 portion of spike for both B cells and T cells. Enhancing the immune response to S816-830 may induce efficient protection and should be a focus of future studies.

Materials and Methods

Study participants

This study was approved by the institutional review board of the Charité (EA/152/20). Written informed consent was obtained from all included participants (*44*), and the study was conducted in agreement with the Declaration of Helsinki. Participants who had tested

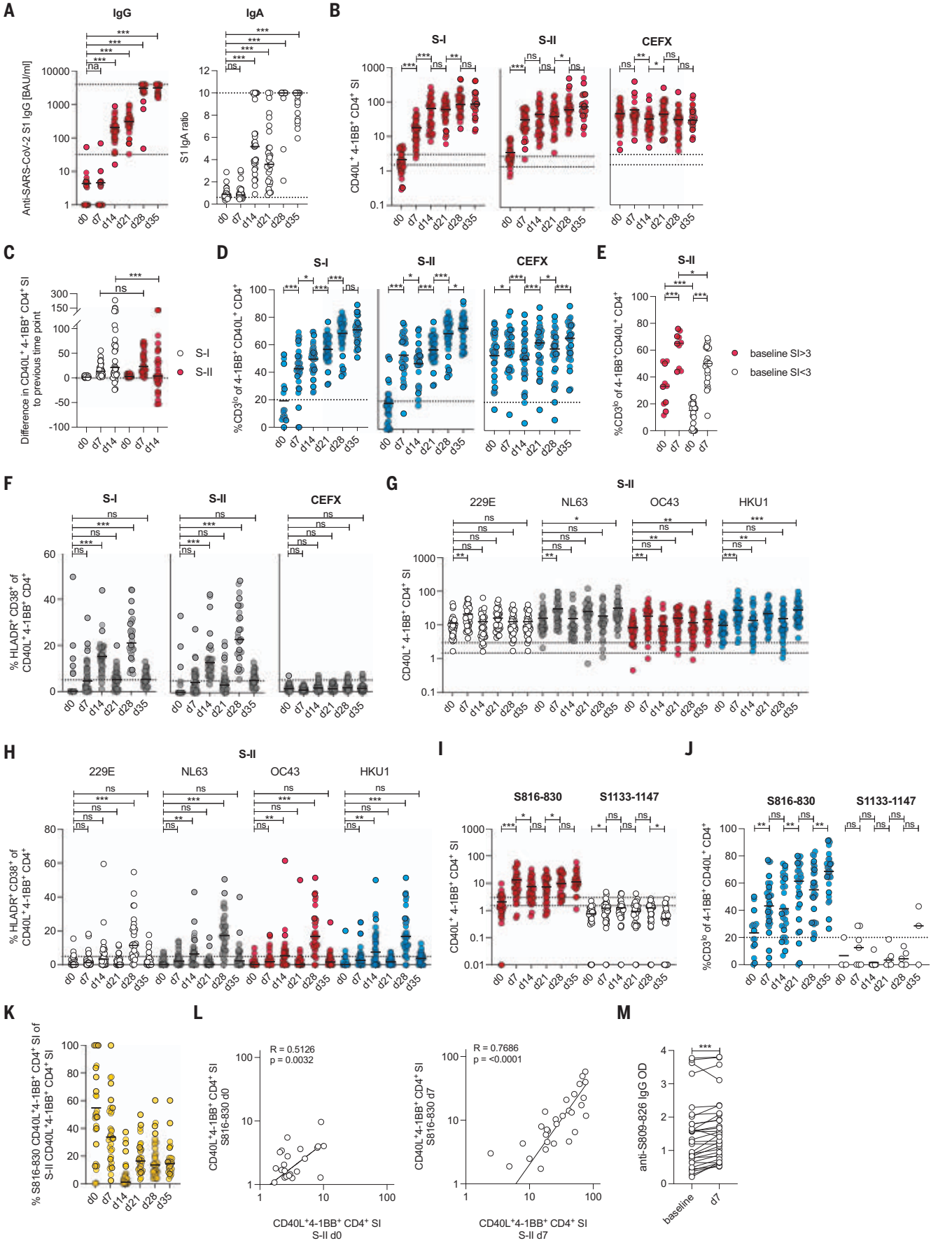

A
IgG
IgA
Anti-SARS-CoV-2 S1 IgG [BAU/ml]
S1 IgA ratio
B
S-I
S-II
CEFX
CD40L+ 4-1BB+ CD4+ SI
C
Difference in CD40L+ 4-1BB+ CD4+ SI to previous time point
S-I
S-II
D
%CD3lo of 4-1BB+CD40L+ CD4+
E
baseline SI>3
baseline SI<3
F
% HLADR+ CD38+ of CD40L+ 4-1BB+ CD4+
G
229E
NL63
OC43
HKU1
H
I
S816-830
S1133-1147
J
K
% S816-830 CD40L+4-1BB+ CD4+ SI of S-II CD40L+4-1BB+ CD4+ SI
L
R = 0.5126
p = 0.0032
CD40L+4-1BB+ CD4+ SI S816-830 d0
CD40L+4-1BB+ CD4+ SI S-II d0
R = 0.7686
p = <0.0001
CD40L+4-1BB+ CD4+ SI S816-830 d7
CD40L+4-1BB+ CD4+ SI S-II d7
M
anti-S809-826 IgG OD
baseline
d7

Fig. 6. HCoV-specific SARS-CoV-2–cross-reactive T cells are recruited into the BNT162b2 vaccine response. (**A**) Serum anti-SARS-CoV-2 S1 IgG BAUs and IgA titer ratio were determined at baseline, day 7, and day 14 after primary vaccination with BNT162b, immediately before secondary vaccination (day 21) as well as 1 (day 28) and 2 weeks (day 35) after secondary vaccination. All values <1 were set to 1. The lower and upper cut-off levels for IgG were set at 32 and 3900, respectively; the corresponding IgA cut-offs were set at 0.6 and 10, respectively, indicated by dotted lines. (**B**) Plots showing the SI of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells after stimulation with S-I, S-II, and CEFX at baseline and at the indicated time points. (**C**) Difference in SI after stimulation with S-I and S-II at each time point relative to the previous time point. (**D**) Plots showing the frequencies of $CD3^{lo}$ of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells after stimulation with S-I, S-II, and CEFX for responses with an SI ≥ 1.5. (**E**) Frequencies of $CD3^{lo}$ of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells at days 0 and 7 in cross-reactive donors (baseline SI ≥ 3, red circles) and non–cross-reactive donors (baseline SI < 3, white circles). (**F**) Frequencies of $HLADR^+CD38^+$ cells among $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells after stimulation with S-I, S-II, and CEFX at the indicated time points. (**G** and **H**) SI of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells (G) and frequencies of $HLADR^+CD38^+$ among these cells (H) after stimulation with HCoV S-II peptide pools at baseline and the indicated time points. (**I** and **J**) SI of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells (I) and frequencies of $CD3^{lo}$ events (SI ≥ 1.5) (J) among these cells after stimulation with peptide S816-830 and control peptide S1133-1147 at baseline and the indicated time points. (**K**) Proportion of S816-830–reactive T cells over SARS-CoV-2 S-II–reactive T cells. (**L**) Relationship between responses to S816-830 and SARS-CoV-2 S-II peptide pool stimulation at day 0 (left) and day 7 (right). (**M**) OD of anti–S809-826-peptide IgG ELISA from sera before and 7 days after primary vaccination. For (A) and (F) to (J), *P < 0.05, **P < 0.01, ***P < 0.001, and ns at P > 0.05 by repeated-measures one-way-ANOVA with Dunnett's correction. For (B) to (D) and (M), *P < 0.05, **P < 0.01, ***P < 0.001, and ns at P > 0.05 by paired Student's t test. For (E), *P < 0.05, **P < 0.01, ***P < 0.001, and ns at P > 0.05 by Student's t test. For (L), ns at P > 0.05 by Pearson correlation.

positive for SARS-CoV-2 RNA [reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) from nasopharyngeal swabs] were classified as convalescent donors. All donors were assessed for age, gender, body-mass index, comorbidities, and medications (table S1). Convalescent donors were subclassified according to their symptoms into World Health Organization severity grades, and information about hospitalization or admission to an intensive care unit is provided in table S1. Day of infection was set as day −3 before reported symptom onset. Measurement day after symptom onset is indicated in the graphs or table S1. Study participants who reported symptoms typical for a SARS-CoV-2 infection were RT-qPCR tested for virus RNA, and positive donors were enrolled for follow-up measurements. Details of the follow-up cohort (age, gender, comorbidities, symptoms, measurement time points after symptom onset) are provided in table S2.

Coronavirus RT-qPCR

RNA was extracted using the MagNA Pure 96 system and the MagNA Pure Viral NA Small Volume Kit (Roche, Germany). RNA extraction was performed from a 200-µl swab dilution (swab suspended in 4.3 ml of Cobas PCR Media, Roche), eluted in 100 µl of elution buffer. Coronavirus detection using 5 µl of the RNA eluate was based on two genomic targets (E- and N gene, TIB Molbiol, Berlin, Germany). An in vitro–transcribed RNA of equine arteritis virus was used as an internal RT and PCR control. SARS-CoV-2 was quantified using the E-gene target and by applying calibration curves and using serial diluted photometrically quantified in vitro–transcribed RNA as described previously (*45*). All RT-qPCRs were performed using a LightCycler 480 II (Roche).

Blood and serum sampling and PBMC isolation

Whole blood was collected in lithium heparin tubes for peripheral blood mononuclear cell (PBMC) isolation and SSTII advance (all Vacutainer, BD Biosciences) tubes for serology. SSTII advance tubes were centrifuged for 10 min at 1000*g* before removing serum. Serum aliquots were frozen at −20°C until further use. PBMCs were isolated by gradient density centrifugation according to the manufacturer's instructions (Leucosep tubes, Greiner; Biocoll, Bio&SELL).

Ex vivo T cell stimulation

Freshly isolated PBMCs were cultivated at a concentration of 5×10^6/ml in AB medium containing RPMI 1640 medium (Invitrogen) supplemented with 10% heat inactivated AB serum (Pan Biotech), 100 U/ml of penicillin (Biochrom), and 0.1 mg/ml of streptomycin (Biochrom). Stimulations were conducted with PepMix overlapping peptide pools (15-aa length with 11-aa overlaps, JPT Peptide Technologies) covering the proteins of interest, including the entire SARS-CoV-2 orfeome: the spike glycoprotein (S), NCAP-1 (N), VEMP-1 (E), VME-1 (M), AP3A (ORF3a), NS6, NS7A, NS7B, NS8, ORF9B, ORF10, Y14 (ORF9c), the ORF1a/b proteins (NSP01, NSP02, NSP03a, NSP03b, NSP04, NSP05, NSP06, NSP07, NSP08, NSP09, NSP10, NSP11, NSP12, NSP13, NSP14, NSP15, and NSP16), as well as the spike glycoproteins of HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 (all JPT Peptide Technologies). Single-peptide stimulations were conducted with the following peptides: 204 (N′-SKRSFIEDLLFNKVT-C′), 204_1 (N′-KRSFIEDLLFNKVTL-C′), 204_2 (N′-RSFIEDLLFNKVTLA-C′), 204_3 (N′-SFIEDLLFNKVTLAD-C′), 205 (N′-FIEDLLFNKVTLADA-C′), and the control peptide 284 (N′-VNNTVYDPLQPELDS-C′) (all JPT Peptide Technologies). All stimulations (peptide pools and single peptides) were performed at final concentrations of 1 µg/ml per peptide. For a negative control, the stimulation peptide solvent dimethyl sulfoxide diluted 1:1 in phosphate-buffered saline was used at the same concentration as in peptide-stimulated tubes. SEB/TSST-1 (1.5 and 1.0 mg/ml, respectively) (Sigma-Aldrich) and/or the CEFX Ultra SuperStim pool (1 µg/ml per peptide) (JPT Peptide Technologies) were used as positive stimulation controls. For optimized costimulation, purified anti-CD28 (clone CD28.2, BD Biosciences) was added to each stimulation at a final concentration of 1 µg/ml. Incubation was performed at 37°C at 5% CO_2 for 16 hours in the presence of 10 µg/ml brefeldin A (Sigma-Aldrich) during the last 14 hours. $CD4^+$ T cell activation was calculated as SI = percentage of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells in the stimulation divided by the percentage of $CD40L^+4\text{-}1BB^+$ $CD4^+$ T cells in the unstimulated control. Dotted lines indicate an SI of 1.5 (positive with uncertainty) and 3 (definite positive).

T cell enrichment and expansion

Activated cells were enriched from stimulated PBMCs by magnetic-activated cell sorting (MACS). Cells were stimulated with indicated PepMixes in the presence of 1 µg/ml of purified anti-CD28 (clone CD28.2, BD Biosciences) and 1 µg/ml of purified anti-CD40 (5C3, BioLegend) for 16 hours, followed by staining with anti-CD40L-APC (5C8, Miltenyi Biotec) and anti–4-1BB-PE (4B4-1, BD Biosciences). The activated cells were enriched using anti-phycoerythrin (anti-PE) MultiSort MicroBeads (Miltenyi Biotec) according to the manufacturer's instructions. After release of anti-PE beads, a second, analogous enrichment step was performed using anti-APC MicroBeads (Miltenyi Biotec). The purity of the enriched population was routinely checked to >80% of live cells. Feeder cells were obtained from the 4-1BB-PE–negative fraction of the initial enrichment step by CD3 MicroBeads (Miltenyi Biotec) depletion and subsequent irradiation at 50 Gy. Enriched $CD40L^+4\text{-}1BB^+$ cells were co-cultured with feeder cells at a ratio of 1:1 in AB medium supplemented with 10 ng/ml of interleukin-7 (IL-7) and 10 ng/ml of IL-15 (both from Miltenyi Biotec) for 10 days, followed by 2 days of cytokine starvation. The cells were

then restimulated in the presence of CD3-depleted autologous feeder cells as described above and as indicated in the figure legends. For spike glycoprotein epitope identification, restimulation was performed with the Epitope Mapping Peptide Set SARS-CoV-2 (JPT) according to the manufacturer's instructions.

Flow cytometry

Stimulations were stopped by incubation in 2 mM EDTA for 5 min. Surface staining was performed for 15 min in the presence of 1 mg/ml of Beriglobin (CSL Behring) with the following fluorochrome-conjugated antibodies titrated to their optimal concentrations as specified in table S3: FITC-conjugated anti-CD3 (Miltenyi Biotec), VioGreen-conjugated anti-CD4 (Miltenyi Biotec), VioBlue-conjugated anti-CD8 (Miltenyi Biotec), APC-conjugated anti-CD38 (Miltenyi Biotec), and PerCP-Vio 700–conjugated anti–HLA-DR (Miltenyi Biotec). During the last 10 min of incubation, Zombie Yellow fixable viability staining (BioLegend) was added. Fixation and permeabilization were performed with eBioscience FoxP3 fixation and PermBuffer (Invitrogen) according to the manufacturer's protocol. Intracellular staining was performed for 30 min in the dark at room temperature with PE-conjugated anti–4-1BB (Miltenyi Biotec), PE-Vio 770–conjugated anti-CD40L (Miltenyi Biotec), Alexa Fluor 700–conjugated anti–IFN-γ (BioLegend), and Brilliant Violet 605–conjugated anti–TNF-α (BioLegend,). All samples were measured on a MACSQuant Analyzer 16 (Miltenyi Biotec). Instrument performance was monitored before every measurement with Rainbow Calibration Particles (BD Biosciences).

Anti–SARS-CoV-2 IgG and IgA ELISA specific for the S subunit 1 (S1) was performed using the commercial kits (QuantiVac for IgG, EUROIMMUN Medizinische Labordiagnostika) according to the manufacturer's instructions and as described previously (*46*). Upper and lower cut-offs were set at 3900 and 32 for IgG, respectively, and at 0.6 and 10 for IgA, respectively.

SARS-CoV-2 neutralization assay

Neutralization activity of SARS-CoV-2–specific antibodies was assessed with a plaque reduction neutralization test (PRNT) as described previously (*45*).

Epitope-specific antibody ELISA

Biotinylated peptide S809-826 (Biotin-Ttds-PSKPSKR*SFIEDLLFNKV*-OH, Ttds linker= N-(3-{2-[2-(3-Amino-propoxy)-ethoxy]-ethoxy}-propyl)-succinamic acid, JPT Peptide Technologies) (400 nM) was immobilized on a 96-well streptavidin plate (Steffens Biotechnische Analysen) for 1 hour at room temperature. After blocking for 1 hour at 30°C, serum samples were diluted 1:100 and incubated for 1 hour at 30°C. Horseradish peroxidase (HRP)–coupled, anti–human IgG secondary antibody (Jackson ImmunoResearch) was diluted 1:5000 (Jackson ImmunoResearch) and added to the serum samples for 1 hour at 30°C, then HRP substrate was added (TMB, Kem-En-Tec). The reaction was stopped by adding sulfuric acid, and absorption was measured at 450 nm using a FlexStation 3.

HLA typing and analysis

HLA typing was performed by LABType CWD assays (One Lambda, West Hills, CA, USA) based on reverse sequence-specific oligonucleotides according to the manufacturer's instructions. Briefly, the HLA genomic region was amplified individually using locus-specific biotinylated primers for HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1, and HLA-DPB1. Amplicons were hybridized to HLA allele- and allele group–specific probes attached to Luminex beads. Complementary binding was detected by addition of R-PE–conjugated streptavidin and acquired using a FLEXMAP 3D flow analyzer (Luminex, Austin, TX, USA). HLA alleles were derived at two-field code resolution (highest probability) as referenced in the catalog of common and well-documented HLA alleles version 2.0.0 *33*. MHC class II binding prediction were performed using the Immune Epitope Database and Analysis Resource (www.IEDB.org) (*47*, *48*), based on the IEDB recommended method version 2.22. For the purposes of this analysis, we refer to an individual as "homozygous" if the two corresponding alleles of the same locus are identical in the first two fields.

Homology score

For the calculation of the homology score, all possible 9-mers were generated for each respective PepMix of SARS-CoV-2. Each of the 9-mers was scored against each unique 9-mer from the proteomes of the corona viruses 229E, NL63, OC43, and HKU1 (isolates N1, N2, and N5) using the PAM30 substitution matrix. The homology score is the percentage of comparisons with a pairwise 9-mer score >30.

Data analysis and statistics

Study data were collected and managed using REDCap electronic data capture tools hosted at Charité (*49*, *50*). Flow cytometry data were analyzed with FlowJo 10.6, and statistical analysis was conducted with GraphPad Prism 9. If not stated otherwise, data are plotted as means. N indicates the number of donors. P values were set as follows: $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

REFERENCES AND NOTES

1. E. J. Williamson *et al*., Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020). doi: 10.1038/s41586-020-2521-4; pmid: 32640463
2. U. Stervbo, S. Rahmann, T. Roch, T. H. Westhoff, N. Babel, Epitope similarity cannot explain the pre-formed T cell immunity towards structural SARS-CoV-2 proteins. *Sci. Rep.* **10**, 18995 (2020). doi: 10.1038/s41598-020-75972-z; pmid: 33149224
3. J. Braun *et al*., SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (2020). doi: 10.1038/s41586-020-2598-9; pmid: 32726801
4. A. Grifoni *et al*., Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501.e15 (2020). doi: 10.1016/j.cell.2020.05.015; pmid: 32473127
5. A. Nelde *et al*., SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat. Immunol.* **22**, 74–85 (2021). doi: 10.1038/s41590-020-00808-x; pmid: 32999467
6. N. Le Bert *et al*., SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020). doi: 10.1038/s41586-020-2550-z; pmid: 32668444
7. J. Mateus *et al*., Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **370**, 89–94 (2020). doi: 10.1126/science.abd3871; pmid: 32753554
8. P. Bacher *et al*., Low-avidity CD4+ T cell responses to SARS-CoV-2 in unexposed individuals and humans with severe COVID-19. *Immunity* **53**, 1258–1271.e5 (2020). doi: 10.1016/j.immuni.2020.11.016; pmid: 33296686
9. E. M. Anderson *et al*., Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *Cell* **184**, 1858–1864.e10 (2021). doi: 10.1016/j.cell.2021.02.010; pmid: 33631096
10. N. Mishra *et al*., Immunoreactive peptide maps of SARS-CoV-2. *Commun. Biol.* **4**, 225 (2021). doi: 10.1038/s42003-021-01743-9
11. A. P. Ferretti *et al*., Unbiased screens show CD8+ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the Spike protein. *Immunity* **53**, 1095–1107.e3 (2020). doi: 10.1016/j.immuni.2020.10.006; pmid: 33128877
12. K. W. Ng *et al*., Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**, 1339–1343 (2020). doi: 10.1126/science.abe1107; pmid: 33159009
13. D. Weiskopf *et al*., Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. *Sci. Immunol.* **5**, eabd2071 (2020). doi: 10.1126/sciimmunol.abd2071; pmid: 32591408
14. T. Sekine *et al*., Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* **183**, 158–168.e14 (2020). doi: 10.1016/j.cell.2020.08.017; pmid: 32979941
15. A. A. Lehmann, G. A. Kirchenbaum, T. Zhang, P. A. Reche, P. V. Lehmann, Deconvoluting the T cell response to SARS-CoV-2: Specificity versus chance and cognate cross-reactivity. *Front. Immunol.* **12**, 635942 (2021). doi: 10.3389/fimmu.2021.635942; pmid: 34127926
16. W. S. Lee, A. K. Wheatley, S. J. Kent, B. J. DeKosky, Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nat. Microbiol.* **5**, 1185–1191 (2020). doi: 10.1038/s41564-020-00789-5; pmid: 32908214
17. M. Sagar *et al*., Recent endemic coronavirus infection is associated with less-severe COVID-19. *J. Clin. Invest.* **131**, e143380 (2021). doi: 10.1172/JCI143380; pmid: 32997649
18. L. Liu *et al*., Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020). doi: 10.1038/s41586-020-2571-7; pmid: 32698192
19. D. L. Farber, N. A. Yudanin, N. P. Restifo, Human memory T cells: Generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.* **14**, 24–35 (2014). doi: 10.1038/nri3567; pmid: 24336101
20. S. Kohler, A. Thiel, Life after the thymus: CD31+ and CD31- human naive CD4+ T-cell subsets. *Blood* **113**, 769–774 (2009). doi: 10.1182/blood-2008-02-139154; pmid: 18583570
21. A. Sattler *et al*., Cytokine-induced human IFN-γ-secreting effector-memory Th cells in chronic autoimmune inflammation. *Blood* **113**, 1948–1956 (2009). doi: 10.1182/blood-2008-02-139147; pmid: 19104082
22. H. Liu, M. Rhodes, D. L. Wiest, D. A. A. Vignali, On the dynamics of TCR:CD3 complex cell surface expression and downmodulation. *Immunity* **13**, 665–675 (2000). doi: 10.1016/S1074-7613(00)00066-2; pmid: 11114379
23. P. Holenya *et al*., Peptide microarray based analysis of antibody responses to SARS-CoV-2 identifies unique epitopes with potential for diagnostic test development. medRxiv 2020.11.24.20216663 [Preprint]. 27 November 2020. doi: 10.1101/2020.11.24.20216663

24. E. Shrock *et al*., Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**, eabd4250 (2020). doi: 10.1126/science.abd4250; pmid: 32994364
25. A. A. Minervina *et al*., Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. *eLife* **9**, e53704 (2020). doi: 10.7554/eLife.53704; pmid: 32081129
26. K. Fink, Origin and function of circulating plasmablasts during acute viral infections. *Front. Immunol.* **3**, 78 (2012). doi: 10.3389/fimmu.2012.00078; pmid: 22566959
27. D. Poston *et al*., Absence of severe acute respiratory syndrome coronavirus 2 neutralizing activity in prepandemic sera from individuals with recent seasonal coronavirus infection. *Clin. Infect. Dis.* ciaa1803 (2020). doi: 10.1093/cid/ciaa1803; pmid: 33270134
28. W. R. Morgenlander *et al*., Antibody responses to endemic coronaviruses modulate COVID-19 convalescent plasma functionality. *J. Clin. Invest.* **131**, e146927 (2021). doi: 10.1172/JCI146927; pmid: 33571169
29. A. Tarke *et al*., Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep. Med.* **2**, 100204 (2021). doi: 10.1016/j.xcrm.2021.100204; pmid: 33521695
30. T. A. Schwickert, B. Alabyev, T. Manser, M. C. Nussenzweig, Germinal center reutilization by newly activated B cells. *J. Exp. Med.* **206**, 2907–2914 (2009). doi: 10.1084/jem.20091225; pmid: 19934021
31. T. A. Schwickert *et al*., A dynamic T cell-limited checkpoint regulates affinity-dependent B cell entry into the germinal center. *J. Exp. Med.* **208**, 1243–1252 (2011). doi: 10.1084/jem.20102477; pmid: 21576382
32. K. Fink *et al*., B cell activation state-governed formation of germinal centers following viral infection. *J. Immunol.* **179**, 5877–5885 (2007). doi: 10.4049/jimmunol.179.9.5877; pmid: 17947661
33. A. T. Tan *et al*., Early induction of functional SARS-CoV-2-specific T cells associates with rapid viral clearance and mild disease in COVID-19 patients. *Cell Rep.* **34**, 108728 (2021). doi: 10.1016/j.celrep.2021.108728; pmid: 33516277
34. M. Lipsitch, Y. H. Grad, A. Sette, S. Crotty, Cross-reactive memory T cells and herd immunity to SARS-CoV-2. *Nat. Rev. Immunol.* **20**, 709–713 (2020). doi: 10.1038/s41577-020-00460-4; pmid: 33024281
35. A. T. Huang *et al*., A systematic review of antibody mediated immunity to coronaviruses: Kinetics, correlates of protection, and association with severity. *Nat. Commun.* **11**, 4704 (2020). doi: 10.1038/s41467-020-18450-4; pmid: 32943637
36. S. Amit, G. Regev-Yochay, A. Afek, Y. Kreiss, E. Leshem, Early rate reductions of SARS-CoV-2 infection and COVID-19 in BNT162b2 vaccine recipients. *Lancet* **397**, 875–877 (2021). doi: 10.1016/S0140-6736(21)00448-7; pmid: 33610193
37. G. Alter *et al*., Immunogenicity of Ad26.COV2.S vaccine against SARS-CoV-2 variants in humans. *Nature* **596**, 268–272 (2021). doi: 10.1038/s41586-021-03681-2; pmid: 34107529
38. J. Sadoff *et al*., Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N. Engl. J. Med.* **384**, 2187–2201 (2021). doi: 10.1056/NEJMoa2101544; pmid: 33882225
39. J. L. Bernal *et al*., Early effectiveness of COVID-19 vaccination with BNT162b2 mRNA vaccine and ChAdOx1 adenovirus vector vaccine on symptomatic disease, hospitalisations and mortality in older adults in England. MedRxiv 2021.03.01.21252652 [Preprint]. 2 March 2021. doi: 10.1101/2021.03.01.21252652
40. C. M. Saad-Roy *et al*., Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science* **372**, 363–370 (2021). doi: 10.1126/science.abg8663; pmid: 33688062
41. A. L. Lai, J. K. Millet, S. Daniel, J. H. Freed, G. R. Whittaker, The SARS-CoV fusion peptide forms an extended bipartite fusion platform that perturbs membrane order in a calcium-dependent manner. *J. Mol. Biol.* **429**, 3875–3892 (2017). doi: 10.1016/j.jmb.2017.10.017; pmid: 29056462
42. C. M. Poh *et al*., Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat. Commun.* **11**, 2806 (2020). doi: 10.1038/s41467-020-16638-2; pmid: 32483236
43. S. Ravichandran *et al*., Longitudinal antibody repertoire in "mild" versus "severe" COVID-19 patients reveals immune markers associated with disease severity and resolution. *Sci. Adv.* **7**, eabf2467 (2021). doi: 10.1126/sciadv.abf2467; pmid: 33674317
44. F. Kurth *et al*., Studying the pathophysiology of coronavirus disease 2019: A protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). *Infection* **48**, 619–626 (2020). doi: 10.1007/s15010-020-01464-x; pmid: 32535877
45. R. Wölfel *et al*., Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020). doi: 10.1038/s41586-020-2196-x; pmid: 32235945
46. N. M. A. Okba *et al*., Severe acute respiratory syndrome coronavirus 2–specific antibody responses in coronavirus disease patients. *Emerg. Infect. Dis.* **26**, 1478–1488 (2020). doi: 10.3201/eid2607.200841; pmid: 32267220
47. P. Wang *et al*., A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLOS Comput. Biol.* **4**, e1000048 (2008). doi: 10.1371/journal.pcbi.1000048; pmid: 18389056
48. P. Wang *et al*., Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* **11**, 568 (2010). doi: 10.1186/1471-2105-11-568; pmid: 21092157
49. P. A. Harris *et al*., The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inform.* **95**, 103208 (2019). doi: 10.1016/j.jbi.2019.103208; pmid: 31078660
50. P. A. Harris *et al*., Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009). doi: 10.1016/j.jbi.2008.08.010; pmid: 18929686

ACKNOWLEDGMENTS

We thank S. Kolling and the Charité Corona Cross (CCC) Study Group associates J. Schlotterbeck, E. Baysal, T. Panne, F. Legler, M. Girod, B. Bohnen, T. Nguyen, J. Schmitz, R. Klemencic, L. Hintze, N. Avinc, P. Resch, A. Maraj, A. Farghaly, P. Schulz, K. Du, N. Matuschewski, K.Gutwenger, I. Katsianas, Z. Cheng, A. Braginets, G. Janach, and A. Bogdanski for their contributions to donor recruitment, sample processing, and measurement; F. Gutsch, T. Bleicker, J. Tesch, P. Tscheak, M. L. Schmidt, and J. Riege for their support; P. Jordan for the illustration of the graphical abstract figure; and U. Klein for the critical reading of the manuscript. **Funding:** This work was supported by the Federal Ministry of Health through a resolution of the German Bundestag (CCC and CCC 2.0). Parts of the work were funded by the German Research Foundation through grant KFO339 to J.B.; SFB-TR84 projects A4, B6, C8, C10 to L.E.S., A.H., and S.H. and B8 to M.A.Ma.; the European Union's Horizon 2020 research and innovation program through project RECOVER (GA101003589) to C.D.; the German Ministry of Research through the projects RAPID (01KI1723A) to C.D., A.H., and S.H., DZIF (301-4-7-01.703) to C.D., DZL (82DZL0098B1) to M.A.Ma., VARIPath (01KI2021) to V.M.C., RECAST to B.S., M.A.Ma., J.Rö., V.M.C., and L.E.S., and NaFoUniMedCovid19 - COVIM, FKZ: 01KX2021 to L.E.S., C.D., V.M.C., and A.T. **Author contributions**: Conceptualization: L.L., C.G.T., A.T.; Data curation: L.L., J.B., L.H., V.M.C., T.S.; Formal analysis: L.L., J.B., L.H., F.Ke.; Funding acquisition: A.T.; Investigation: L.L., L.H., J.B.; Resources: B.K., M.D., U.R., F.Ke., M.M., C.U., T.S., F.D., R.L., S.K., J.Ro., K.G., Z.U.-A., M.F., F.Ku., K.S., M.E., S.H., A.H., B.S., S.M., F.P., M.A.Ma., H.W., S.V., M.A.Mu., C.D., N.L., L.M.-A., L.E.S., V.M.C., J.Rö.; Visualization: L.L., J.B., U.R., F.Ke.; Writing: L.L., F.Ke., A.T., C.G.T. **Competing interests**: L.L., J.B., L.H., and A.T. filed a patent application regarding the usage of flow cytometrically assessed CD3 down-regulation as indicator for functional TCR avidity. L.L. and A.T. filed a patent application regarding S816-830 as dominant cross-reactive T cell epitope. K.S., M.E., U.R., and F.Ke. are employees of JPT. H.W. is chief executive officer of JPT. F.K. is co-inventor and co-owner of a patent describing the use of protein-spanning, overlapping peptide pools for detecting antigen-specific T-cells as applied in this study (WO2001063286A2, now expired). F.K. also holds a part-time position as a research fellow with the Department of Laboratory Medicine, Medizinische Hochschule Brandenburg, Neuruppin/Germany. S.M. is chief executive officer of Miltenyi Biotec. M.A.Mu. and V.M.C. are named together with Euroimmun GmbH on a patent application filed recently regarding the diagnostic of SARS-CoV-2 by antibody testing. F.P. reports research grants and speaker honoraria from Bayer, Teva, Genzyme, Merck, MedImmune, and Novartis unrelated to this work and is member of the steering committee of the OCTIMS study (Novartis). J.R. reports personal fees or grants from Vertex Pharmaceuticals during the conduct of the study. The remaining authors declare no competing financial interests. **Data and materials availability**: All data are available in the main text or in the supplementary materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh1823
Figs. S1 to S6
Tables S1 to S3
MDAR Reproducibility Checklist

14 March 2021; accepted 24 August 2021
Published online 31 August 2021
10.1126/science.abh1823

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

RESEARCH ARTICLES

BATTERIES

Solvation sheath reorganization enables divalent metal batteries with fast interfacial charge transfer kinetics

Singyuk Hou[1]†, Xiao Ji[1]†, Karen Gaskell[2], Peng-fei Wang[1], Luning Wang[2], Jijian Xu[1], Ruimin Sun[1], Oleg Borodin[3]*, Chunsheng Wang[1]*

Rechargeable magnesium and calcium metal batteries (RMBs and RCBs) are promising alternatives to lithium-ion batteries because of the high crustal abundance and capacity of magnesium and calcium. Yet, they are plagued by sluggish kinetics and parasitic reactions. We found a family of methoxyethyl-amine chelants that greatly promote interfacial charge transfer kinetics and suppress side reactions on both the cathode and metal anode through solvation sheath reorganization, thus enabling stable and highly reversible cycling of the RMB and RCB full cells with energy densities of 412 and 471 watt-hours per kilogram, respectively. This work provides a versatile electrolyte design strategy for divalent metal batteries.

Energy upgrades and the implementation of electric transportation demand electrochemical storage systems with higher accessibility, energy density, and safety (*1*). Rechargeable divalent metal batteries, including rechargeable magnesium and calcium metal batteries (RMBs and RCBs), are appealing alternatives to lithium-ion batteries (LIBs) because of the crustal abundances of Mg and Ca, which are more than 1000 times that of Li; larger anode capacity from two-electron transfer (3832 mA·hours cm^{-3} for Mg and 2052.6 mA·hours cm^{-3} for Ca) (*2*); low reduction potential of Mg and Ca metal [−2.38 V versus standard hydrogen electrode (SHE) for Mg and −2.76 V versus SHE for Ca]; and reduced safety concerns from the potentially dendrite-free plating of Mg metal anode (*3*). To fabricate RMBs and RCBs with energy densities comparable to or higher than that of commercial LIBs, the metal anodes need to be paired with cathodes with a potential of >2.5 V versus Li^+/Li (fig. S1 and table S1), which necessitates the use of high-voltage metal oxide cathodes (*4*).

In LIBs, a Li^+ desolvates and transports through the electron-insulating solid-electrolyte interphase (SEI) to the electrode materials (*5*, *6*).

[1]Department of Chemical and Biomolecular Engineering, University of Maryland, College Park, MD 20740, USA. [2]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. [3]Battery Science Branch, Sensors and Electron Devices Directorate, US Army Combat Capabilities Development Command Army Research Laboratory, Adelphi, MD 20783, USA.
*Corresponding authors: cswang@umd.edu (C.W.); oleg.a.borodin.civ@army.mil (O.B.)
†These authors contributed equally to this work.

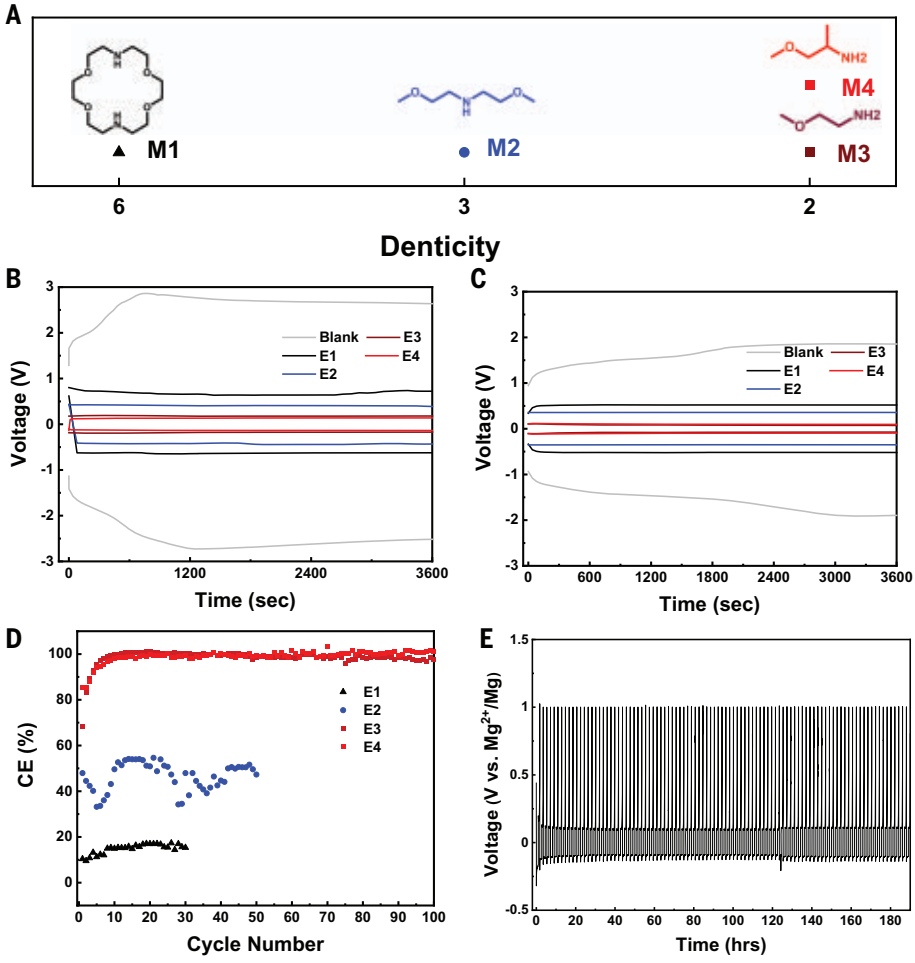


Fig. 1. Overpotentials and CEs of Mg plating and stripping in different electrolytes. (**A**) The molecular structure and denticity of the chelants. (**B** and **C**) The overpotentials at the 10th cycle for plating and stripping in Mg||Mg cells at (B) 1.5 mA cm^{-2} for 1.5 mA·hours cm^{-2} and (C) 0.1 mA cm^{-2} for 0.1 mA·hours cm^{-2} in the blank and **Ex**. (**D**) CEs for Mg plating and stripping in **Ex** in Mg||SS cells cycled at 0.1 mA cm^{-2}. (**E**) Long-term cycling of the Mg||SS cell in **E4** at a current density of 0.1 mA cm^{-2}.

Desolvation and transport of the divalent ions through SEI and active materials, however, are more challenging owing to stronger electrostatic interactions (*7*–*12*), leading to large overpotentials, electrolyte decomposition (*8*, *9*), and irreversible phase transformations or failure to intercalate in cathode materials (*13*–*15*). Reductive and chloride-containing nonaqueous electrolytes avoid SEI formation on Mg anodes but suffer from low anodic stability and incompatibility with current collectors and battery casings (*16*–*18*). Aqueous electrolytes assist Mg^{2+} intercalation in high-voltage cathodes (*14*, *15*, *19*, *20*) but do not support reversible metal anodes (*16*, *17*). To bridge the incompatibility among electrolytes to the cathode, anode, and current collectors, electrolytes consisting of noncorrosive electron-delocalizing anions, such as magnesium bis(trifluoromethanesulfonimide) [$Mg(TFSI)_2$] (*21*, *22*), boron clusters (*23*–*25*), alkoxyborates and alkoxyaluminates (*26*–*29*), and an artificial SEI on a Mg anode, were reported (*30*). However, Mg and Ca anodes still suffer from insufficient coulombic efficiency (CE) while large hysteresis is observed on the cathode side.

We demonstrate that the multidentate methoxyethyl-amine chelants [–$(CH_2OCH_2CH_2N)_n$–] in the first solvation sheath of Mg^{2+} and Ca^{2+} enable both highly reversible Mg and Ca anodes as well as fast (de)intercalation of Mg^{2+} and Ca^{2+} into high-voltage layered oxide cathodes. These chelants demonstrate 6 to 41 times higher affinity for Mg^{2+} than traditional ether solvents, yet the chelant-rich solvation sheaths bypass the energetically unfavorable desolvation process through reorganization, thus reducing the overpotential and eliminating the concomitant parasitic reactions for both the anode and cathode. The reorganization energy of these electrolytes can be tuned by changing the dielectric constants and sizes of the chelants.

Results and discussion

Mg plating and stripping in chelating electrolytes

A 0.5 M $Mg(TFSI)_2$ in 1,2-dimethoxyethane (DME) electrolyte was adopted as the baseline (blank) because of its chemical stability and commercial accessibility. The proposed electrolytes are 0.5 M $Mg(TFSI)_2$ in DME-**M*x*** (x = 1, 2, 3, or 4), in which **M1**, **M2**, **M3**, and **M4** are hexadentate, tridentate, and bidentate chelants for Mg^{2+} (Fig. 1A). The quantities of **M*x*** are varied to keep the molar ratios between donor atoms in **M*x*** to Mg^{2+} at 8:1 (materials and methods). These electrolytes are referred to as **E*x*** (x = 1, 2, 3, or 4).

Mg plating and stripping overpotentials were evaluated in Mg||Mg symmetric cells at a variety of currents (fig. S2). Consistent with previous studies, the overpotentials are around 2.0 V in the blank (*21*) but are substantially reduced in **E*x***, dropping to below 0.1 V in **E4** (Fig. 1, B and C). Because the ionic conductivities of **E*x*** (4.0 to 5.3 mS cm^{-1}) are smaller than that of the blank (6.5 mS cm^{-1}) (fig. S3), the reduced overpotentials in **E*x*** are not attributed to bulk ion transport but rather to the interfacial charge transfer kinetics. As overpotentials reduce, the cycling CEs (cCEs, the average CE after 10 cycles) in Mg||stainless steel (SS) cells increase from 0% in the blank (owing to no Mg stripping at the cutoff potential of 1.0 V versus Mg^{2+}/Mg) to more than 99.5% in **E4** (Fig. 1D) with stable overpotentials (Fig. 1E and fig. S4) and dendrite-free Mg deposits (fig. S5).

The surfaces of the cycled Mg were characterized with high-resolution x-ray photoelectron spectroscopy (XPS). The decomposition products from DME and $TFSI^-$ are summarized in table S2. No signs of **M*x*** decomposition are found in the N 1s spectrum (fig. S6). The species from electrolyte decomposition, in particular –C=O (C 1s and O 1s), crystalline $Mg(OH)_2$ (Mg 2p), $-CF_x$ (F 1s), $-SO_x$ (x < 2, S 2p), and MgS_x (S 2p) are abundant on Mg cycled in the blank but diminish on Mg cycled in **E*x*** (Fig. 2A). The decomposition layer is also much thicker in

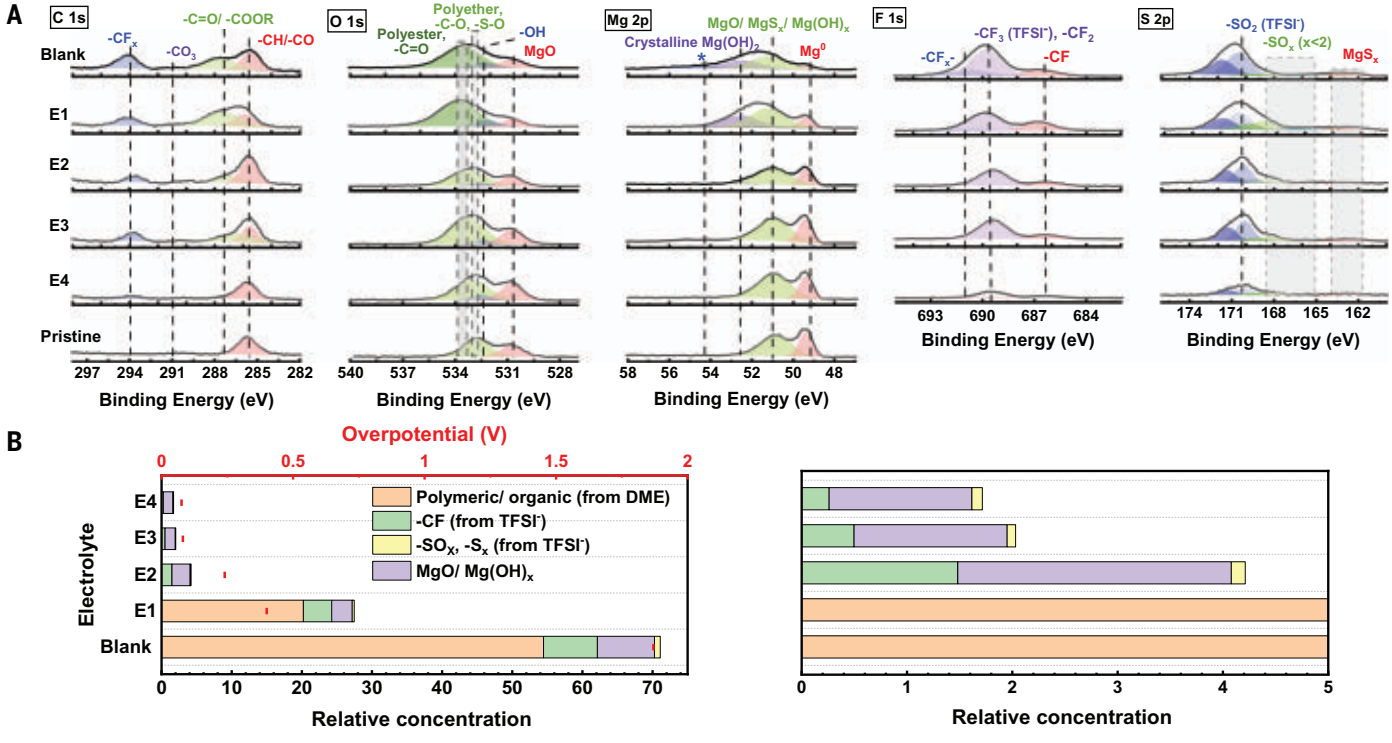


Fig. 2. Tailoring Mg surface composition with the chelants. (**A**) High-resolution XPS spectra of C 1s, O 1s, Mg 2p, F 1s, and S 2p of cycled (100 cycles at 0.1 mA cm^{-2}) and pristine Mg. For S 2p high-resolution XPS, only S 2p 3/2 were labeled; the unlabeled peaks are corresponding S 2p 1/2 spin split peaks. (**B**) The relative concentrations of decomposition species found on the cycled Mg surface. The peaks of –C=O in C 1s and O 1s are used to characterize polymeric and organic decomposition products from DME. All peak areas are normalized to those of Mg^0 to obtain relative concentrations. The red horizontal bars indicate the Mg plating and stripping overpotentials. The low relative concentration region is shown on the right.

the blank, as evidenced by the almost unchanged F and O accumulations after 12-min Ar^+ sputtering (fig. S7). By contrast, the F and O accumulations on Mg cycled in **E3** and **E4** are negligible and can be removed with a much shorter sputtering time (fig. S7). The relative concentrations of these decomposed species show substantial dependence on the overpotentials (Fig. 2B). Thus, by reducing the overpotentials for Mg plating and stripping, electrolyte decomposition is suppressed and cCE is enhanced in **E*x***.

Solvation sheath structure

Without effective SEI, the solvation environment of Mg^{2+} serves as the key to understanding how **M*x*** improves charge transfer kinetics of the Mg anode, which was probed using ^{13}C nuclear magnetic resonance (NMR). Chemical shifts of methylene carbons in DME (δ_{DME}) and **M*x*** (δ_{Mx}) shift up-field in the presence of $Mg(TFSI)_2$ (Fig. 3A). These changes ($\Delta\delta_{DME}$ and $\Delta\delta_{Mx}$) reflect the stoichiometric average between bound and free states (supplementary text). Thus, a smaller $\Delta\delta_{DME}$ in **E*x*** indicates that DME is partially freed from the solvation sheath (Fig. 3A, left). Meanwhile, the substantial association of **M*x*** in the solvation sheath is evidenced by values of $\Delta\delta_{Mx}$ that are >10 times greater than those of $\Delta\delta_{DME}$ (Fig. 3B). Using $\Delta\delta_{Mx}$ and $\Delta\delta_{DME}$, the relative affinities of **M*x***-Mg^{2+} over DME-Mg^{2+} interactions were quantified (table S3), in which **M1**, **M2**, **M3**, and **M4** show affinities that are 41, 12, 10, and 6 times greater with Mg^{2+} than with DME, respectively (Fig. 3B). Separate analysis of N and O stoichiometry through $\Delta\delta_{Mx-N}$ (Fig. 3A, middle) and $\Delta\delta_{Mx-O}$ (Fig. 3A, right) ($\Delta\delta_{Mx}$ of carbon next to N and O, respectively) shows that the association of N in the solvation sheath greatly reduces from **E1** to **E4** (pink circles in Fig. 3B), but the association of O from **M*x*** is, in general, greater than that from DME in the solvation sheath (Fig. 3B). Conversely, bond lengths calculated from molecular dynamics simulations show that Mg^{2+}–N bonds in **M*x*** are 0.2 to 0.3 Å shorter than those in monodentate amines (*31*, *32*), whereas the Mg^{2+}–O bond lengths remain the same for DME and **M*x*** (figs. S8 to S10). Therefore, the higher affinity between **M*x*** and Mg^{2+} is likely to be initiated by the strengthened Mg^{2+}–N bonds that in turn promote the Mg^{2+}–O association. The formation of contact ion pairs (CIPs) between $TFSI^-$ and Mg^{2+} was also characterized with Raman spectroscopy because CIPs potentially facilitate the decomposition of $TFSI^-$ (*33*) and reduce anode reversibility. The S–N stretching vibration in $TFSI^-$ that is sensitive to ionic interactions was deconvoluted to identify the solvent-separated ion pairs (SSIPs) and CIPs (Fig. 3C, fig. S11, and supplementary text). CIP percentages obtained by normalizing CIPs to total peak areas show that only 15 to 20% of the $TFSI^-$ are in contact with the cations in both the blank and **E*x*** (Fig. 3D). Thus, the major role of **M*x*** is to displace DME instead of $TFSI^-$ from the solvation sheath. Owing to the high affinity between **M*x*** and Mg^{2+}, the solvation sheaths containing **M*x*** are more stable than the $Mg^{2+}(DME)_3$ solvates in the blank by 0.3 to 1 eV (30 to 100 kJ mol^{-1}), as determined from density functional theory (DFT) calculations (Fig. 3E and figs. S12 and S13). **M1** has the strongest affinity to Mg^{2+} followed by **M2** and **M4** ≈ **M3**, in agreement with NMR results and molecular dynamics simulations.

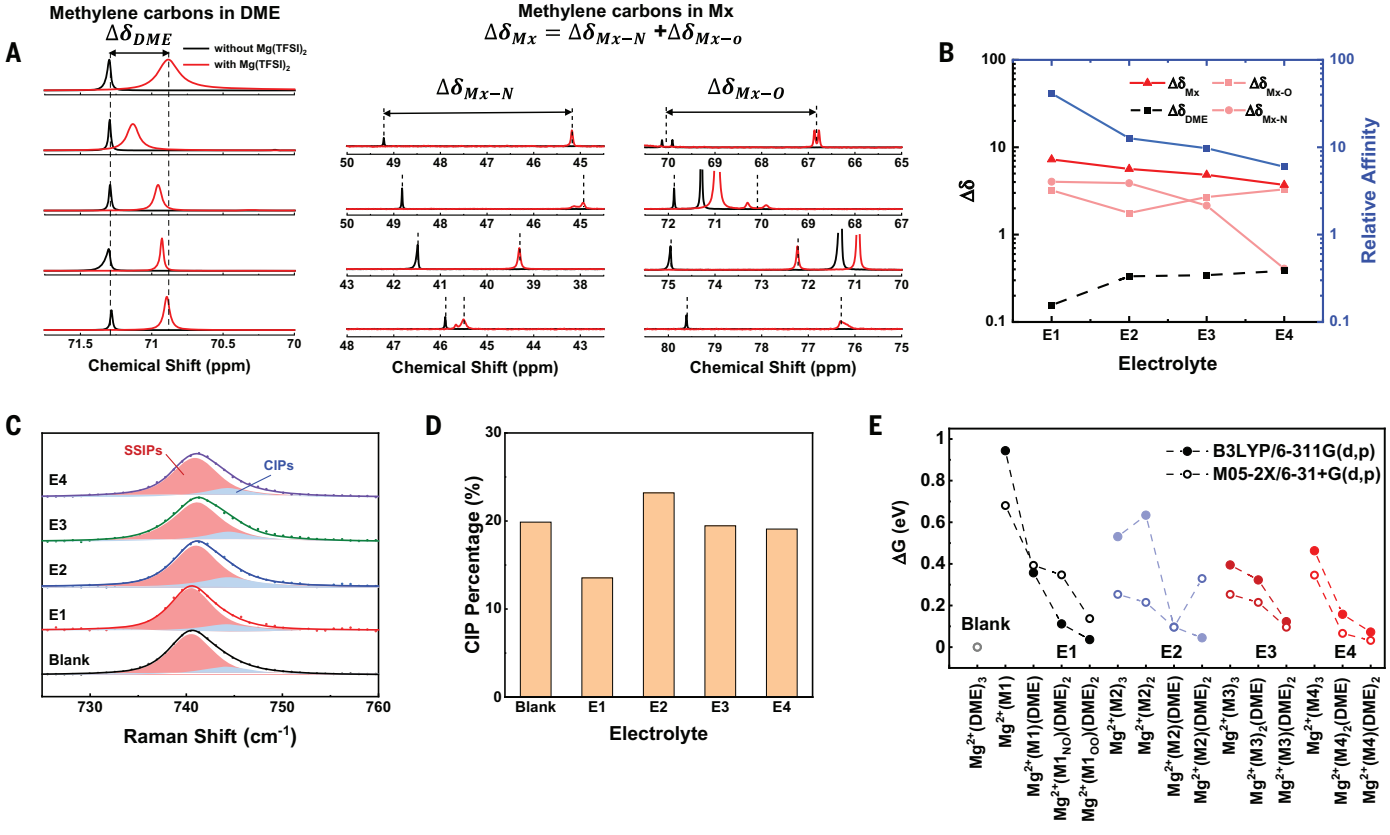


Fig. 3. Characterization of the Mg^{2+} solvation sheath in different electrolytes. (**A**) ^{13}C NMR spectra of methylene carbons in DME (left) and **M*x*** (middle and right) with and without $Mg(TFSI)_2$. ppm, parts per million. (**B**) Summary of Δδ and the relative affinities of **M*x***-Mg^{2+} interactions compared with that of the DME-Mg^{2+} interaction. (**C**) Raman spectra for the S–N stretching vibration of $TFSI^-$. The SSIPs and CIPs are fitted through a pseudo-Voigt function. (**D**) The percentage of $TFSI^-$-forming CIPs with Mg^{2+}in the solvation sheath determined by the ratio between the CIP peak area and the total peak area from the Raman spectra. (**E**) The relative binding free energies (ΔG) for the Mg^{2+} solvates relative to $Mg^{2+}(DME)_3$ from M05-2X/6-31+G(d,p) and B3LYP/6-311G(d,p) DFT calculations with polarized continuum model (ε = 7.19); higher values indicate more stable solvates.

Solvation sheath reorganization mediated charge transfer on the Mg anode

To understand why solvates in **Ex** with much higher binding energies than those in the blank (Fig. 3E) support Mg plating and stripping with smaller overpotentials, the electron transfer from an electrode to solvated Mg^{2+} was analyzed. It is approximated as nonadiabatic (*34*), involving ionic intermediates with the reorganized solvation sheath according to Marcus theory (*35*). For an electrode-bound electrochemical reaction, the overpotential (η) represents the potential required to vary the Fermi level (E_F) of the electrode for each electron transferred (*36*) and the reorganization energy (λ) represents the energy to reorganize the solvation sheath to accept the electron (Fig. 4A). The values of λ were obtained by fitting the Tafel plots for the Mg anode with Marcus-Hush-Chidsey kinetics (*37*, *38*) (equation S2 in fig. S14). To minimize interference from the morphological difference of Mg deposits and the electrolyte decomposition during Mg plating, only the stripping overpotentials of the freshly polished Mg were collected for fitting (fig. S14). The fitted λ showed distinct correlations with the η in three regions, marked as I, II, and III in Fig. 4B. Reactions in **E3** and **E4** fall into region I, in which λ for Mg^{2+} redox is considerably smaller than the onset of decomposition potential (E_D), resulting in little electrolyte decomposition (Fig. 2), and the η agrees well with the λ, reflecting that the electrons are transferred reversibly to Mg^{2+}; reactions in **E2** and **E1** fall into region II, in which λ overlaps with E_D and the η slightly deviates from the postulated λ for Mg^{2+} redox, reflecting that some side reactions take place; and reactions in the blank fall into region III, in which the λ is larger than E_D and the η greatly deviates from the postulated λ for Mg^{2+} redox because the major reaction is now dominated by electrolyte decomposition (Fig. 4B).

The solvation sheath reorganization energy (λ) can be expressed as a dielectric continuum formulation:

$$\lambda = \frac{e^2}{4\pi\varepsilon_0}\left(\frac{1}{a_0}-\frac{1}{R}\right)\left(\frac{1}{\varepsilon_{op}}-\frac{1}{\varepsilon_s}\right)$$

where ε_0 is the vacuum permittivity, e is electron charge, a_0 is the solvated radius of the cations, R is the distance between the solvated ions and electrode surface, ε_{op} is the optical dielectric constant, and ε_s is the static dielectric constant (*39*), in which the λ

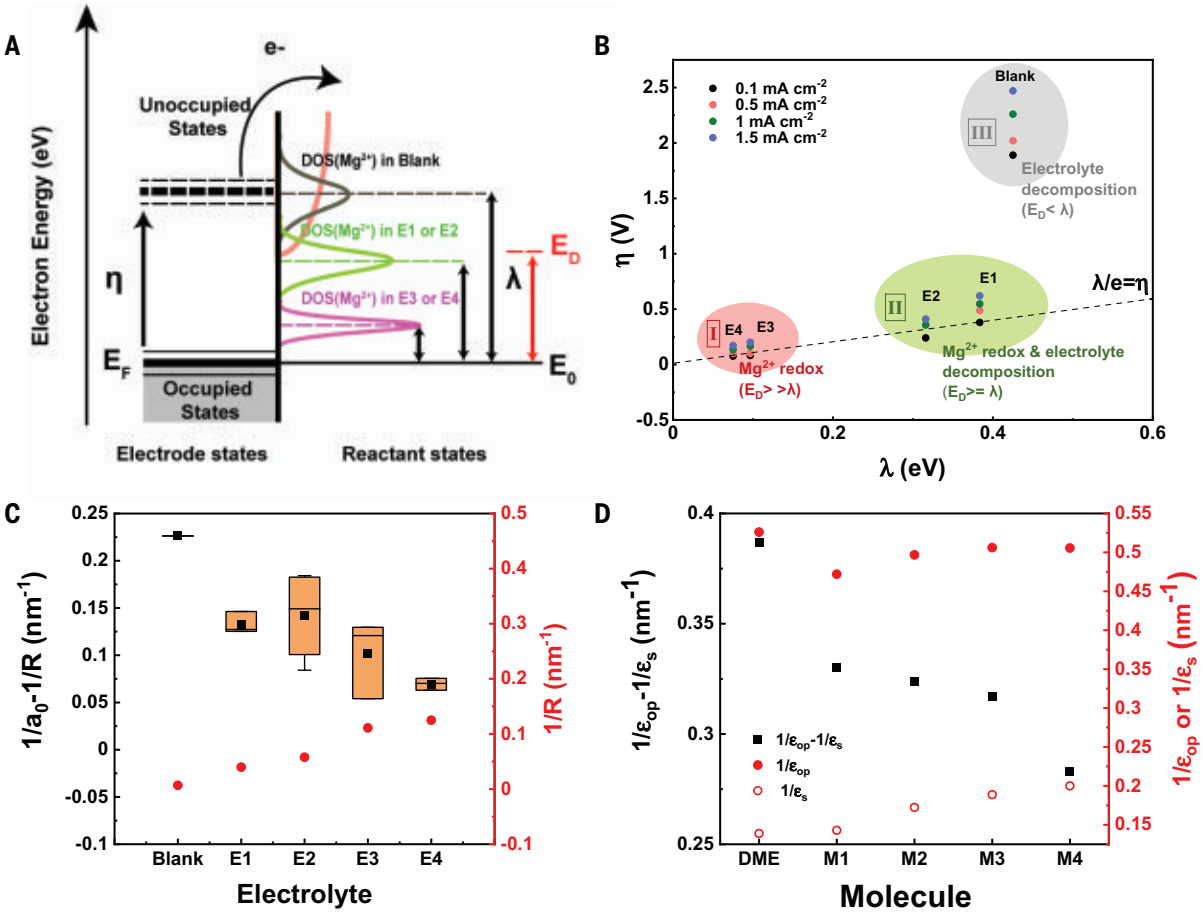


Fig. 4. Reorganization of the solvation sheath. (**A**) The Marcus-Gerischer diagram of electron transfer at the metal electrode–electrolyte interface. E_F refers to the Fermi level of the metal electrode, and E_0 refers to the energy level of the electrolyte and electrode at equilibrium. The red, green, and black Gaussians represent the density of state (DOS) of Mg^{2+} in the electrolyte. The orange curve represents the DOSs of other electrolyte components, and E_D refers to the onset potential of the electrolyte decomposition. (**B**) The correlation between overpotential (η) and reorganization energy (λ) for different reaction modes of the Mg anodes. The dashed line represents when the λ for electron transfer is equal to η. (**C**) The reciprocal of the distance between solvated ion and electrode surface $1/R$ (red dots) and the reciprocal of the distance part $(1/a_0 - 1/R)$ in different electrolytes (box plot). Owing to multiple solvation radii $(1/a_0)$ in each electrolyte, a box plot is used to account for the distribution of $(1/a_0 - 1/R)$ between the first and the third quartile; the mean values of $(1/a_0 - 1/R)$ are marked with black squares. (**D**) The reciprocal of the optical dielectric constant $(1/\varepsilon_{op})$ (red solid circles), static dielectric constant $(1/\varepsilon_s)$ (red open circles), and dielectric part $(1/\varepsilon_{op} - 1/\varepsilon_s)$ (black squares) for DME and **Mx**.

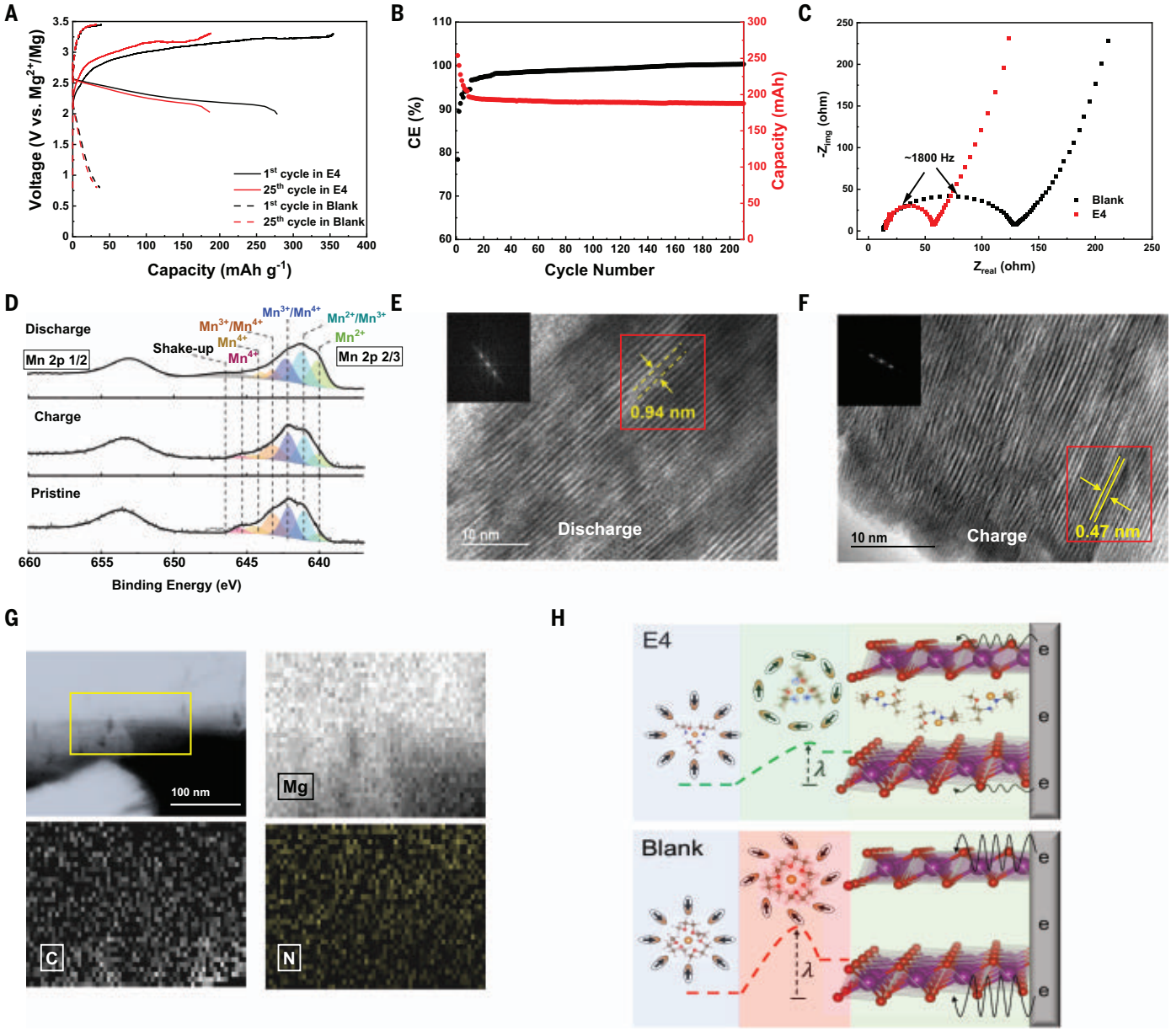


Fig. 5. Intercalation of Mg^{2+} and chelant in layered oxide. (**A**) The charge-discharge curve of the $Mg_{0.15}MnO_2$||Mg cell in **E4** and the blank at 0.5C (1C = 200 mA g^{-1}). (**B**) The capacity and CE of $Mg_{0.15}MnO_2$||Mg cycled in **E4** at 0.5C. (**C**) Electrochemical impedance spectra for $Mg_{0.15}MnO_2$ at a 0% state of discharge in the three-electrode cell after 25 cycles in the blank and **E4**, with a Mg counter electrode and magnesiated Mo_6S_8 reference electrode. EIS was measured between 10^6 and 0.1 Hz, with an ac oscillation of 10 mV amplitude. (**D**) Characterization of the Mn oxidative state with Mn 2p 3/2 high-resolution XPS spectra of discharge, charge, and pristine Mg_xMnO_2. (**E** and **F**) Transmission electron microscopy (TEM) image of $Mg_{0.15}MnO_2$ after (E) discharge (Mg^{2+} intercalation) and (F) charge (Mg^{2+} disintercalation) at the 25th cycle. The insets are the diffraction patterns from fast Fourier transformation of the marked area. (**G**) Scanning TEM (STEM)–EDX spectroscopy mappings of the elemental distributions of Mg, N, C, and O in the discharged Mg_xMnO_2 at the 25th cycle. (**H**) The scheme of concerted ion and electron transfer in the cathode host limited by the solvation sheath reorganization.

is proportional to the distance part $\left(\frac{1}{a_0}-\frac{1}{R}\right)$ and the dielectric part $\left(\frac{1}{\varepsilon_{op}}-\frac{1}{\varepsilon_s}\right)$. The distance part gradually reduces from the blank to **E4** (Fig. 4C) owing to a less compact solvation sheath (smaller $1/a_0$ except for **M3**; table S4). In addition, the formation of decomposition products during cycling increases the distance between the solvated ion and electrode R (Fig. 4C, fig. S15, table S5, and supplementary text). This leads to an increase in λ. The dielectric part also reduces progressively from the blank to **E4** (Fig. 4D and table S6). Considering that $\frac{1}{\varepsilon_{op}}$ and $\frac{1}{\varepsilon_s}$ represent fast and slow solvent motions upon electron transfer (*40*), respectively, the smaller difference demonstrated by **M*x*** suggests that the asymmetric chelants provide more polarizable environments compared with DME. To summarize, the less compact and more polarizable solvation sheath reduces the λ for electron transfer, which consequently reduces overpotential by preventing electrolyte decomposition and promoting stable Mg plating and stripping.

In addition to the λ, substantial differences were observed for the initial Mg^{2+}/Mg^+ reduction potentials for the $Mg^{2+}(\mathbf{M4})_3$ and $Mg^{2+}(DME)_3$ solvates (figs. S16 to S18 and

table S7), with Mg^{2+}/Mg^{+} reduction most readily occurring in Mg^{2+}(**M4**)$_3$, whereas reduction of $Mg^{2+}(DME)_3$ required a much larger overpotential by 0.4 to 0.6 V. An intermediate reduction potential is observed for Mg^{2+}(**M1**)(DME) (fig. S16F) that is consistent with the intermediate performance of the **E1** electrolyte (Fig. 1, B to D). The larger overpotential to initiate electron transfer in the blank likely contributes to the slower interfacial kinetics and electrolyte decomposition. Because MgO is often present on Mg metal, solvent reactivity with MgO was examined in DFT calculations and it was found that the defect-free surfaces are not reactive toward DME or chelants (fig. S19 and table S8).

The solvation sheath mediated charge transfer in the layered oxide cathode

On the anode, the solvation sheath reorganized to the activated state at which the electron transfer occurred. In the oxide cathode, however, the ion accepts an electron indirectly through concerted intercalation near the transitional metal center, yet we found that the **M*x*** in electrolyte can also improve the charge transfer kinetics on the cathode. Mg-pillared layered MnO_2 ($Mg_{0.15}MnO_2$) was used as the cathode material, and **E4** was used to evaluate the electrochemical performance of the $Mg||Mg_{0.15}MnO_2$ cell owing to its higher anodic stability (3.8 V versus Mg^{2+}/Mg) (fig. S20). The $Mg||Mg_{0.15}MnO_2$ cell was charged-discharged between 2 and 3.3 V (versus Mg^{2+}/Mg) at 0.5C (1C corresponds to a 200 mA g^{-1} active material), and a reversible capacity of 190 mA·hours g^{-1} was achieved and maintained for 200 cycles (Fig. 5, A and B). Contrary to the performance in **E4**, the charge-discharge of the $Mg||Mg_{0.15}MnO_2$ cell in the blank showed little capacity at the same C-rate (Fig. 5A). Charge transfer resistances (R_{ct}) of the $Mg_{0.15}MnO_2$ cathode in the blank and **E4** were measured using electrochemical impedance spectroscopy (EIS) in a three-electrode cell with magnesiated Mo_6S_8 reference electrodes (Fig. 5C) to exclude potential variation from the Mg metal reference (fig. S21). The charge-transfer processes of the $Mg_{0.15}MnO_2$ cathode in **E4** and the blank are represented by the semicircles in the EIS spectra with similar characteristic frequencies. The values of R_{ct} determined by the diameters of the semicircles are two to three times smaller in **E4** than in the blank (Fig. 5C), suggesting a much faster charge transfer kinetics.

We further characterized the $Mg_{0.15}MnO_2$ cathode to identify the cause of faster charge transfer kinetics in **E4**. The reversible electron transfer to the $Mg_{0.15}MnO_2$ cathode was demonstrated by manganese (Mn) 2p 3/2 high-resolution XPS spectra, in which the peaks associated with Mn^{4+} reduced and those peaks associated with Mn^{3+} and Mn^{2+} intensified (Fig. 5D). A shake-up satellite characteristic of Mn^{2+} also appeared after discharge. The change of the Mn oxidation states was reversed after charge, and the pristine state was regenerated (Fig. 5D; see peak assignments in table S2). Concerted with the electron transfer is the Mg^{2+} (de)intercalation, demonstrated first by reversible incorporation and extraction of 0.3 Mg^{2+} to the cathode after discharge and charge measured by inductively coupled plasma optical emission spectroscopy (table S9). Because the number of Mg^{2+} incorporated or extracted is in accord with the coulombic capacity (Fig. 5A), Mg^{2+} is the only ion that participated in the redox reaction. Unlike Mo_6S_8 and TiS_2, which demonstrate very little change of lattice indexes upon Mg^{2+} (de)intercalation (*41*, *42*), a substantial enlargement of the interlayer distance was observed after discharging the cathode in **E4** (Fig. 5E), suggesting that larger species were intercalated with Mg^{2+}. The species were then extracted from the oxide cathode together with Mg^{2+} after the charge and the interlayer distance was reduced (Fig. 5F). The elemental distribution of the discharged cathode from energy-dispersive x-ray (EDX) spectroscopy demonstrated the presence of C and N in the materials together with Mg^{2+} (Fig. 5G), in combination with the characteristic peaks for **M4** observed in the attenuated total reflectance Fourier-transform infrared spectrum in the discharge cathode (fig. S22), confirming the presence of **M4** when an electron is transferred to the Mg^{2+} intercalated cathode. The fact that **M4** in the solvation sheath greatly facilitates the charge transfer kinetics of the cathode suggests that solvation sheath reorganization also occurs in the cathode and limits the interfacial charge transfer reaction kinetics (Fig. 5H).

The generality of the electrolyte design is demonstrated in other solvents used in RMBs such as diglyme (fig. S23) and tetrahydrofuran (fig. S24). Extension to RCBs was validated by adding **M4B** to 0.5 M calcium tetrabis(hexafluoroisopropyloxy)borate-DME, in which the Ca^{2+} plating and stripping cCE increases from 80 (*43*, *44*) to 96% (fig. S25). **M4B** is a derivative of **M4** with an extra methylene carbon to accommodate the larger size of Ca^{2+} (fig. S26). This electrolyte was used to pair the Ca metal anode with a $Mg_{0.15}MnO_2$ full cell and demonstrates a 2.6-V average potential with a reversible capacity of 210 mA·hours g^{-1} at 0.5C for 50 cycles (fig. S27).

In this study, we identified several essential factors that are required for the cation solvation sheath to successfully undergo reorganization for a fast Mg^{2+}/Mg redox reaction, including (i) the solvation free energy for Mg^{2+} serving as a descriptor for salt dissociation, (ii) the reduction potential for the $Mg^{2+}(solvent)_n \rightarrow Mg^{+}(solvent)_n$ reaction to ensure that the Mg^{2+}/Mg redox occurs within the electrolyte cathodic stability window, and (iii) the reorganization free energy (λ). To discover additional chelants, several compounds were screened using DFT (fig. S16) and validated with cycling tests (fig. S27). Ethylenediamine was found as a promising chelant or solvent owing to its high free energy of solvation, high reduction potential for Mg^{2+}, and low λ.

Methoxyethyl-amines were demonstrated to circumvent commonly observed challenges such as high interfacial impedance and overpotentials for plating and stripping Mg and Ca by preferential solvation of metal cations and facilitating the initial reduction step. This approach simultaneously solved two critical challenges for divalent metal batteries: the low reversibility for anodes and sluggish kinetics of metal-oxide cathodes, enabling the energy densities of RMBs and RCBs to be comparable to that of LIBs. The design principle is generally applicable for divalent metal batteries. The reorganization energy can be tuned by tailoring the molecular structures through the introduction of heterogeneous donor atoms and less compact structures to further enhance the kinetics and reversibility to match the application standard for commercial LIBs.

REFERENCES AND NOTES

1. M. Armand, J.-M. Tarascon, *Nature* **451**, 652–657 (2008).
2. J. Muldoon, C. B. Bucur, T. Gregory, *Chem. Rev.* **114**, 11683–11720 (2014).
3. M. Li *et al.*, *Nat. Rev. Mater.* **5**, 276–294 (2020).
4. P. Canepa *et al.*, *Chem. Rev.* **117**, 4287–4341 (2017).
5. K. Xu, *Chem. Rev.* **104**, 4303–4417 (2004).
6. Y. Li, Y. Qi, *Energy Environ. Sci.* **12**, 1286–1295 (2019).
7. S. H. Lapidus *et al.*, *Phys. Chem. Chem. Phys.* **16**, 21941–21945 (2014).
8. Z. Lu, A. Schechter, M. Moshkovich, D. Aurbach, *J. Electroanal. Chem.* **466**, 203–217 (1999).
9. D. Aurbach, R. Skaletsky, Y. Gofer, *J. Electrochem. Soc.* **138**, 3536–3545 (1991).
10. E. Levi, Y. Gofer, D. Aurbach, *Chem. Mater.* **22**, 860–868 (2010).
11. G. S. Gautam *et al.*, *Chem. Commun.* **51**, 13619–13622 (2015).
12. Z. Rong *et al.*, *Chem. Mater.* **27**, 6016–6021 (2015).
13. G. Gershinsky, H. D. Yoo, Y. Gofer, D. Aurbach, *Langmuir* **29**, 10964–10972 (2013).
14. X. Sun, V. Duffort, B. L. Mehdi, N. D. Browning, L. F. Nazar, *Chem. Mater.* **28**, 534–542 (2016).
15. L. Yu, X. Zhang, *J. Colloid Interface Sci.* **278**, 160–165 (2004).
16. J. Song, E. Sahadeo, M. Noked, S. B. Lee, *J. Phys. Chem. Lett.* **7**, 1736–1749 (2016).
17. C. B. Bucur, T. Gregory, A. G. Oliver, J. Muldoon, *J. Phys. Chem. Lett.* **6**, 3578–3591 (2015).
18. J. Muldoon *et al.*, *Energy Environ. Sci.* **6**, 482–487 (2013).
19. K. W. Nam *et al.*, *Nano Lett.* **15**, 4071–4079 (2015).
20. X. Ji *et al.*, *Nano Lett.* **18**, 6441–6448 (2018).
21. S.-Y. Ha *et al.*, *ACS Appl. Mater. Interfaces* **6**, 4063–4073 (2014).
22. N. Sa *et al.*, *RSC Advances* **6**, 113663–113670 (2016).
23. T. J. Carter *et al.*, *Angew. Chem. Int. Ed.* **53**, 3173–3177 (2014).
24. O. Tutusaus *et al.*, *Angew. Chem. Int. Ed.* **54**, 7900–7904 (2015).
25. H. Dong *et al.*, *Nat. Energy* **5**, 1043–1050 (2020).
26. Z. Zhao-Karger, M. E. Gil Bardaji, O. Fuhr, M. A. Fichtner, *J. Mater. Chem. A Mater. Energy Sustain.* **5**, 10815–10820 (2017).
27. J. T. Herb, C. A. Nist-Lund, C. B. Arnold, *ACS Energy Lett.* **1**, 1227–1232 (2016).
28. Z. Zhang *et al.*, *Adv. Energy Mater.* **7**, 1602055 (2017).
29. J. Luo, Y. Bi, L. Zhang, X. Zhang, T. L. Liu, *Angew. Chem. Int. Ed.* **58**, 6967–6971 (2019).
30. S.-B. Son *et al.*, *Nat. Chem.* **10**, 532–539 (2018).

31. M. Bremer, H. Noth, M. Warchhold, *Eur. J. Inorg. Chem.* **2003**, 111–119 (2003).
32. A. Xia, J. E. Knox, M. J. Heeg, H. B. Schlegel, C. H. Winter, *Organometallics* **22**, 4060–4069 (2003).
33. N. N. Rajput, X. Qu, N. Sa, A. K. Burrell, K. A. Persson, *J. Am. Chem. Soc.* **137**, 3411–3420 (2015).
34. T. Van Voorhis *et al.*, *Annu. Rev. Phys. Chem.* **61**, 149–170 (2010).
35. R. A. Marcus, *Annu. Rev. Phys. Chem.* **15**, 155–196 (1964).
36. R. A. Marcus, *J. Chem. Phys.* **43**, 679–701 (1965).
37. C. E. Chidsey, *Science* **251**, 919–922 (1991).
38. P. Bai, M. Z. Bazant, *Nat. Commun.* **5**, 3585 (2014).
39. Y.-P. Liu, M. D. Newton, *J. Phys. Chem.* **98**, 7162–7169 (1994).
40. E. R. Barthel, I. B. Martini, B. J. Schwartz, *J. Phys. Chem. B* **105**, 12230–12241 (2001).
41. E. Levi, A. Mitelman, D. Aurbach, M. Brunelli, *Chem. Mater.* **19**, 5131–5142 (2007).
42. X. Sun, P. Bonnick, L. F. Nazar, *ACS Energy Lett.* **1**, 297–301 (2016).
43. A. Shyamsunder, L. E. Blanc, A. Assoud, L. F. Nazar, *ACS Energy Lett.* **4**, 2271–2276 (2019).
44. Z. Li, O. Fuhr, M. Fichtner, Z. Zhao-Karger, *Energy Environ. Sci.* **12**, 3496–3501 (2019).

ACKNOWLEDGMENTS

We acknowledge support from the Maryland NanoCenter and helpful advice on the manuscript from Y. Qi at Brown University. **Funding:** This work was supported by the US Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE) through the Battery 500 Consortium under award number DE-EE0008202. The modeling part of this work that was conducted at the US Army Research Laboratory (ARL) was supported by the Joint Center for Energy Storage Research (JCESR), a US Department of Energy, Energy Innovation Hub, under cooperative agreement number W911NF-19-2-0046. **Author contributions:** S.H. conceived the idea; performed electrochemical, NMR, and Raman experiments; analyzed data; and wrote the manuscript. X.J. and O.B. performed molecular dynamics simulations and DFT calculations. K.G. performed XPS and analysis of results. P.W. performed the TEM and STEM tests. L.W. helped with the NMR experiment and data analysis. J.X. and R.S. helped with the manuscript editing. O.B. and C.W. supervised the study and contributed to the manuscript writing. All authors discussed the results. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions are present in the paper and/or supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg3954
Materials and Methods
Supplementary Text
Figs. S1 to S27
Tables S1 to S9
References (*45*–*85*)

4 January 2021; accepted 3 August 2021
10.1126/science.abg3954

CHEMICAL PHYSICS

Conformer-specific photochemistry imaged in real space and time

E. G. Champenois[1]†, D. M. Sanchez[1,2]†‡, J. Yang[1,3,4], J. P. Figueira Nunes[5], A. Attar[3], M. Centurion[5], R. Forbes[3], M. Gühr[6], K. Hegazy[1,7], F. Ji[3], S. K. Saha[5], Y. Liu[8], M.-F. Lin[3], D. Luo[3], B. Moore[5], X. Shen[3], M. R. Ware[1], X. J. Wang[3]*, T. J. Martínez[1,2]*, T. J. A. Wolf[1]*

Conformational isomers (conformers) of molecules play a decisive role in biology and organic chemistry. However, experimental methods for investigating chemical reaction dynamics are typically not conformer-sensitive. We report on a gas-phase megaelectronvolt ultrafast electron diffraction investigation of α-phellandrene undergoing an electrocyclic ring-opening reaction. We directly imaged the evolution of a specific set of α-phellandrene conformers into the product isomer predicted by the Woodward-Hoffmann rules in real space and time. Our experimental results are in quantitative agreement with nonadiabatic quantum molecular dynamics simulations, which provide considerable detail of how conformation influences the time scale and quantum efficiency of photoinduced ring-opening reactions.

Conformational isomers, or conformers, can interconvert via rotations around single chemical bonds. Interconversion between conformers represents an important step in many bimolecular reactions, where reactant species must encounter one another not only in a particular orientation but also in a specific structural conformation (*1*). Moreover, conformer dynamics arise naturally in self-ordering of macromolecular structures (e.g., protein folding) (*2*). However, investigation of conformer-specific dynamics on their natural time and length scales of femtoseconds and angstroms is hindered by the insensitivity of established experimental methods to conformers.

The influence of conformers on photochemical reactivity is well known, such as from electrocyclic reactions according to the Woodward-Hoffmann (WH) rules (*3*). These rules predict them to be concerted and conformer-specific (i.e., leading to different reaction products depending on the reactant conformer). Moreover, the WH rules predict an inverted conformer specificity for electrocyclic ring opening in the electronically excited state versus the ground state of a molecule. Electrocyclic reactions play an important role in chemical synthesis (*4*) and vitamin D production in human skin (*5*).

An instructive example of conformer specificity is the photochemical ring opening of the monoterpene α-phellandrene (αPH) (*6*), which is produced by plants and used in the fragrance, food, and pharmaceutical industries (*7*). α-Phellandrene consists of a 1,3-cyclohexadiene (CHD)–like ring moiety with two substituents: an isopropyl group at its sp^3 -hybridized C_1 position and a methyl group at the C_8 position (Figs. 1 and 2). The substitution gives rise to two conformers with the isopropyl group being in either quasi-axial (ax) or quasi-equatorial (eq) orientation (i.e., approximately perpendicular or parallel to the ring plane, respectively). The WH rules predict the ring opening to take place in a concerted, conrotatory motion; that is, both ends of the newly formed open-ring molecule rotate away from each other in the same clockwise or counterclockwise direction (Fig. 1A). This motion leads to different photoproducts depending on the reactant conformer: the isomers (3*Z*,5*E*)-3,7-dimethylocta-1,3,5-triene (ZEDOT, from eq-αPH) and (3*Z*,5*Z*)-dimethylocta-1,3,5-triene (ZZDOT, from ax-αPH) (Fig. 1).

Evidence for the conformer specificity of the ring opening in αPH (*6*) and other reactants (*8*, *9*) has been observed in solution-phase measurements of photoproduct ratios. Moreover, resonance Raman investigations revealed evidence for the importance of conrotatory motion in the depopulation of the Franck-Condon region of the reactant excited state (*10*). The photochemical dynamics leading from the Franck-Condon region to the ring-opened photoproducts of αPH have been investigated by several ultrafast spectroscopic studies in both the gas and condensed phase, although largely without sensitivity to the structural dynamics, reactant conformers, or photoproduct isomers (*11*–*15*).

Here, we report on a combined megaelectronvolt ultrafast electron diffraction (UED) and ab initio multiple spawning (AIMS) (*16*–*18*) study of conformer-specific dynamics in the photochemical ring opening of gas-phase αPH with unambiguous conformer sensitivity, confirming but going far beyond the qualitative orbital symmetry–based WH rule predictions. In both our static diffraction measurements and quantum chemical calculations, we found the eq-αPH conformer to dominate in our gas-phase sample (text S1). As opposed to other studies of electrocyclic reactions, our time-resolved measurements directly image concerted and exclusive formation of the ZEDOT photoproduct predicted by the WH rules on the femtosecond time scale with sub-angstrom resolution and in quantitative agreement with AIMS simulations of the photochemical dynamics.

[1]Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [2]Department of Chemistry, Stanford University, Stanford, CA, USA. [3]SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [4]Center of Basic Molecular Science, Department of Chemistry, Tsinghua University, Beijing, China. [5]Department of Physics and Astronomy, University of Nebraska, Lincoln, NE, USA. [6]Institut für Physik und Astronomie, Universität Potsdam, Potsdam, Germany. [7]Department of Physics, Stanford University, Stanford, CA, USA. [8]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA.
***Corresponding author. Email: thomas.wolf@stanford.edu (T.J.A.W.); toddjmartinez@gmail.com (T.J.M.); wangxj@slac.stanford.edu (X.J.W.)**
†These authors contributed equally to this work.
‡Present address: Design Physics Division, Lawrence Livermore National Laboratory, Livermore, CA, USA.

The simulations reveal that after photoexcitation into the lowest excited state (S_1) with $\pi\pi^*$ character, αPH nonadiabatically relaxes to the electronic ground state (S_0) through three different conical intersection (CI) geometries (*19*). Only one of them yields ring-opened photoproducts; the other two form vibrationally "hot" reactants. The presence of the isopropyl group redirects significant (40 ± 4 %) amounts of the S_1 population toward the non-WH CIs, substantially reducing the WH-photoproduct yield relative to CHD (*20*). Additionally, the rotational orientation of the isopropyl group gives rise to three rotational conformers (rotamers) for each of the eq- and ax-αPH structures, labeled as gauche+/– (G+/G–) and trans (T) (Fig. 2), which were simultaneously present in our sample. These rotamers show different ring-opening time scales.

Identification of the reactant conformer

Figure 2A shows the extracted static atomic pair distribution function [PDF(r)] from a gas-phase sample of randomly oriented αPH molecules. For comparison, we plotted simulated PDFs of all six ax-αPH/eq-αPH rotamer geometries. We analyzed the structural information contained in the PDFs in terms of carbon coordination shells, which are additionally plotted as bars for the eq-αPH (blue) and ax-αPH (red) conformers in Fig. 2A: The many C-C bonds (first carbon coordination shell, labeled as α) in the system give rise to a strong maximum in all PDFs at 1.4 Å. A second peak at 2.4 Å corresponds to the second carbon coordination shell (the distances between carbon atoms two atomic sites away from each other; β in Fig. 2A) with some contributions from the third coordination shell (distances of carbons three atomic sites apart from each other). The signatures above 3 Å belong to the third and higher coordination shells.

We found the strongest differences between the PDFs within the third coordination shell (γ in Fig. 2A). All eq-αPH rotamers, the ax-G–αPH rotamer, and the experimental PDF exhibit a minimum around 3.5 Å, in contrast to the two remaining ax-αPH rotamers that produce a shoulder. The minimum results from a splitting of the third coordination shell in the eq-αPH and ax-G– αPH rotamers into distances between carbons in cis configuration (small dihedral angles, 3c in Fig. 2A) (e.g., the C_3-C_9 distance) and in trans configuration (large dihedral angles, 3t in Fig. 2A) (e.g., the C_3-C_{10} distance). The splitting collapses for the ax-G+ and ax-T rotamers because one of the methyl groups of the isopropyl substituent is directly below the carbon ring and thereby moves many of the ring carbon–methyl carbon distances into the γ region (see structures in Fig. 2). The high level of agreement of the experimental PDF with those of eq-αPH and disagreement with some of the ax-αPH PDFs suggest that the sample is dominated by eq-αPH rotamers, in agreement with our quantum chemical calculations (text S1) and previous results (*21*).

Evolution into the Woodward-Hoffmann predicted product isomer

Figure 2B shows an experimental ΔPDF 260 fs after photoexcitation with an ultrashort 266-nm optical pulse (experimental instrumental response function: 150 fs full width at half maximum; see supplementary materials). Atomic distance changes appear in ΔPDFs as a combination of a negative contribution at the initial atomic distance and a positive contribution at the delay-dependent distance. Accordingly, the experimental ΔPDF in Fig. 2B shows signatures from ring opening. The negative signature in the α region directly follows from breaking the C_1-C_4 bond. The negative signature in the β region mainly results from increases in the C_1-C_9, C_2-C_4, and C_3-C_4 distances. Positive signatures in the γ region suggest that at least some of these distances increase to values close to 3.5 Å. The structural implications of these distance changes can be qualitatively understood by comparison to "simple" ΔPDF simulations generated from single photoproduct and reactant geometries. Such simulations for the WH-predicted photoproducts of eq-αPH → ZEDOT (blue line) and ax-αPH → ZZDOT (red line) are plotted in Fig. 2B. Additionally, the WH-forbidden reactant → photoproduct ΔPDF simulations of eq-αPH → ZZDOT

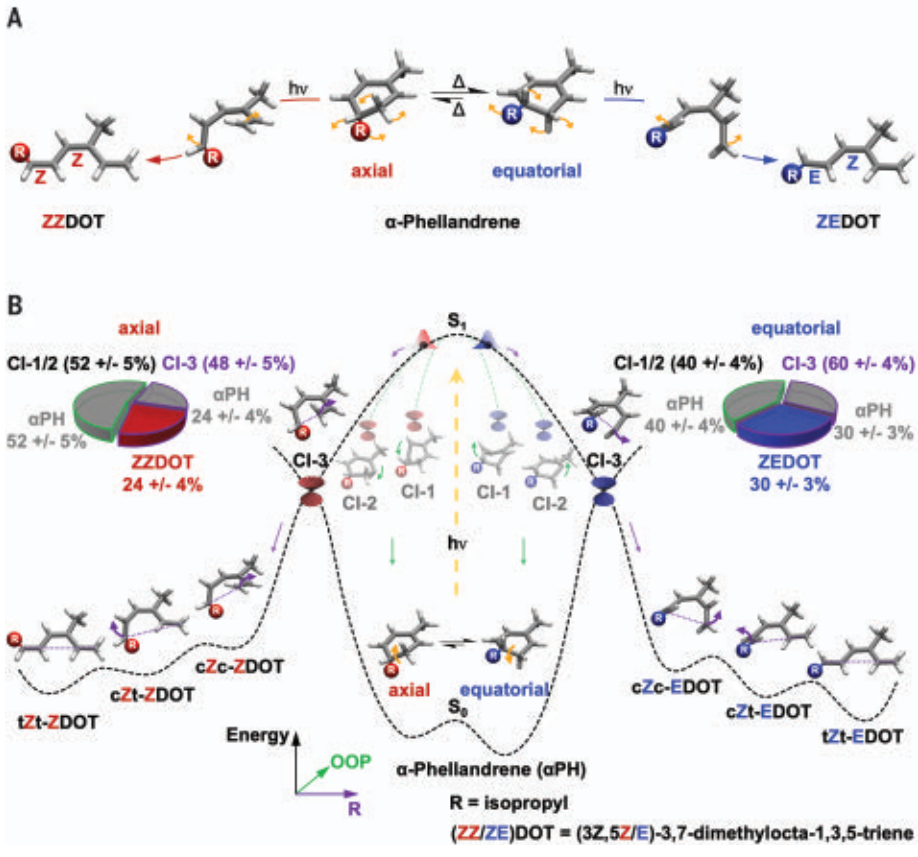


Fig. 1. Conformer-specific photochemistry in α-phellandrene. (**A**) Woodward-Hoffmann predictions for the conformer specificity of photoinduced electrocyclic ring opening in α-phellandrene. Its isopropyl substituent (R) can be in axial or equatorial orientation with respect to the carbon ring. Axial and equatorial conformers are in thermal equilibrium in solution phase (Δ). (6) The Woodward-Hoffmann rules predict a concerted, conrotatory ring-opening motion (orange arrows) yielding isomers with R in different positions depending on the reactant conformer. (**B**) Schematic based on ab initio multiple spawning simulations of the photoinduced ring opening. Equatorial and axial conformers are photoexcited from their respective ground-state (S_0) energy minima to the first excited state (S_1); they evolve along an out-of-plane (OOP, green) coordinate toward conical intersections CI-1 and CI-2 or along the ring-opening coordinate (purple) toward CI-3. CI-1 and CI-2 lead to reformation of α-phellandrene, whereas CI-3 leads to both αPH reformation and ring opening. Several different conformers of the ZZDOT/ZEDOT photoproduct minima (cZc, cZt, and tZt) are accessible in the ground state. The two pie charts visualize the photoproduct distribution for axial and equatorial conformers as well as the distribution among the CI geometries CI-1 to CI-3; errors representing 68% confidence intervals were obtained from bootstrap analysis.

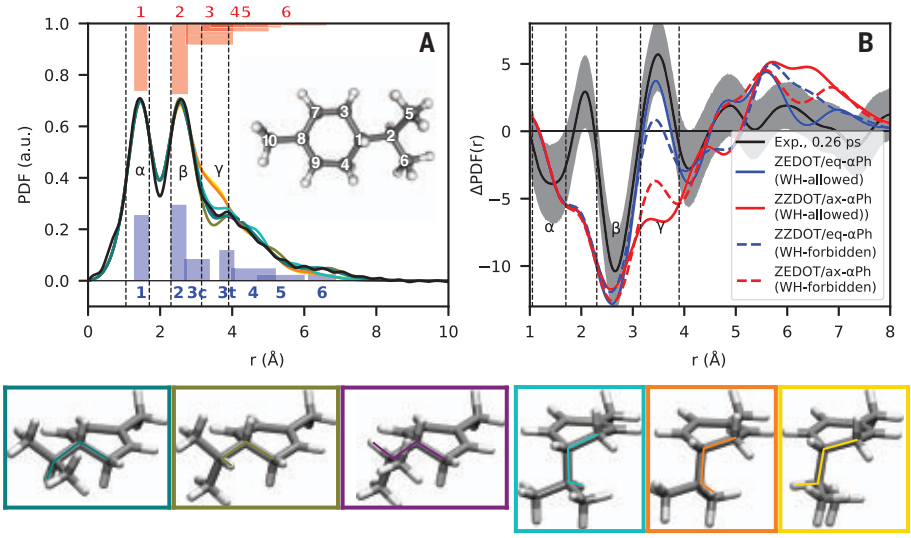


Fig. 2. Comparison of experimental and simulated structural information. (**A**) Experimental (black) and simulated pair distribution functions PDF(r) of six α-phellandrene (αPH) conformers, which are depicted below together with the dihedral angles defining the rotation of the isopropyl group [two gauche orientations (G+/G−) and one trans (T) orientation of the marked isopropyl hydrogen with respect to the marked ring carbon]. Carbon-carbon coordination spheres for axial (red) and equatorial (blue) conformers are shown as bars. Additionally, the α, β, and γ ranges of Fig. 3 are shown. The inset shows the carbon atom numbering used in the text. (**B**) Experimental difference PDF [ΔPDF(r)] at a pump-probe delay of 0.26 ps (black) and simple simulations of the signature of Woodward-Hoffmann (WH)–allowed and WH-forbidden reaction product signatures of the equatorial (eq-αPH) and axial (ax-αPH) reactant conformers and (3Z,5E)-3,7-dimethylocta-1,3,5-triene (ZEDOT) and (3Z,5Z)-dimethylocta-1,3,5-triene (ZZDOT) product isomers. Shaded areas represent a 68% confidence interval obtained from bootstrap analysis.

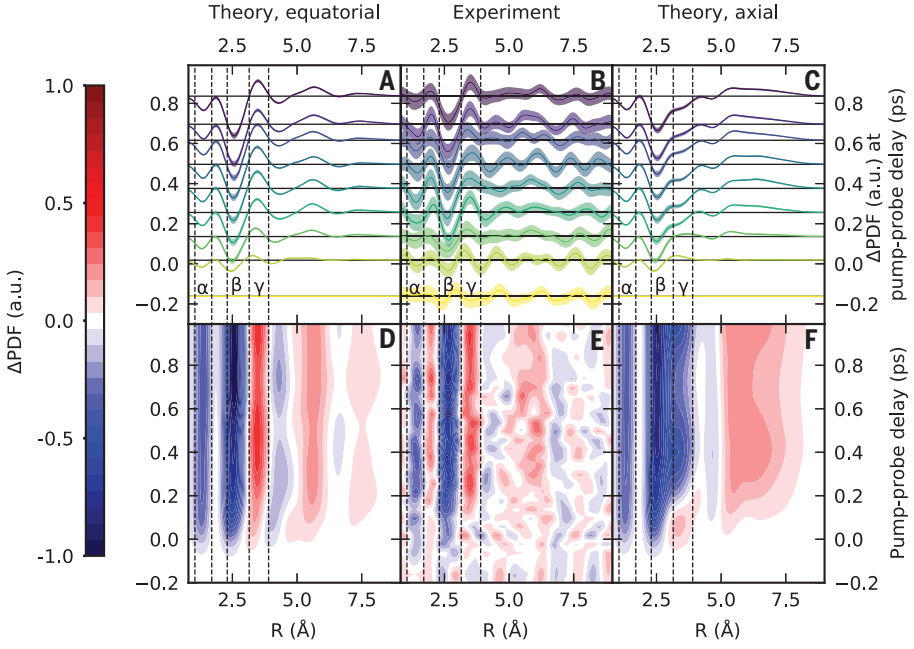


Fig. 3. Comparison of experimental and simulated time-dependent difference pair distribution functions (ΔPDF) of equatorial and axial conformers. (**A** to **C**) ΔPDFs at different pump-probe delays. (**D** to **F**) False-color plots of ΔPDF over the whole investigated time window of 1.2 ps. Simulated ΔPDFs are based on ab initio multiple spawning simulations (see supplementary materials) and represent the averaged dynamics of the G+, G−, and T rotamers. Shaded areas [(A) to (C)] represent a 68% confidence interval obtained from bootstrap analysis. For the simulations, these error bars reflect convergence with respect to initial condition sampling.

(blue dashes) ax-αPH → ZEDOT (red dashes) are shown. Similar to the static PDFs, experimental and simulated ΔPDFs show the most prominent differences in the γ region. Differences between experimental and simulated ΔPDFs in the α and β regions are due to the approximation of the photoproduct by a single geometry. The WH-allowed photoproduct of the eq conformers shows good qualitative agreement with a strong positive signature in the γ-region of the experimental ΔPDF, whereas all other photoproduct ΔPDFs show essentially zero or negative amplitude.

The difference between the simulated photoproduct ΔPDFs can be explained by steric effects from the bulky isopropyl group: In eq-αPH, it points away from the carbon ring. The WH-predicted conrotatory motion (Fig. 1A) rotates it farther away, favoring smaller atomic distance changes into the γ region in the photoproduct (e.g., the cZc-EDOT geometry in Fig. 1B). In ax-αPH, the WH-predicted conrotatory motion rotates the isopropyl group into the carbon ring. This is sterically unfavorable, forcing larger atomic distance changes beyond the γ region in the photoproduct (e.g., the tZt-ZDOT geometry in Fig. 1B). Thus, both effects—the positive signature of eq-αPH → ZEDOT and the negative signature of ax-αPH → ZZDOT—are direct results of the orientation of the isopropyl group in the reactant conformer and therefore an unambiguous signature of conformer-specific ring opening.

Figure 3 shows a comparison of experimental ΔPDFs with those computed from AIMS simulations (*18*) using α state–averaged complete active-space self-consistent field theory (α-CASSCF) (*22*). Experimental ΔPDFs are shown as lineouts in Fig. 3B and as a false-color plot in Fig. 3E. The simulated ΔPDFs of the individual G–, T, and G+ rotamers are averaged assuming roughly equal contributions at room temperature and plotted analogously to the experimental ΔPDFs in Fig. 3, A and D, for eq-αPH and in Fig. 3, C and F, for ax-αPH. Additionally, the delay-dependent integrated experimental ΔPDFs in the α, β, and γ regions are compared in Fig. 4 with the corresponding curves for the averaged and individual eq-αPH rotamer contributions. The averaged AIMS ΔPDFs of the eq conformers show quantitative agreement with the experimental data within their limited signal-to-noise level. The averaged AIMS ΔPDFs of the ax conformers qualitatively disagree with the experimental data in the γ region. The experimental time-dependent signatures are clearly dominated by the eq-αPH rotamers (>80%; see text S1). Thus, we unambiguously observed exclusive formation of the WH-predicted photoproducts of the eq-αPH conformer within a few hundred femtoseconds after optical excitation. Our results constitute direct observation of conformer-specific photochemical reaction dynamics in real space and time.

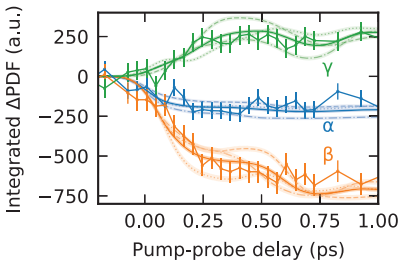


Fig. 4. Comparison of integrated difference pair distribution functions (ΔPDF) from experiment and simulations in the α, β, and γ regions of Fig. 3. Experimental signals (dots with error bars) are compared to corresponding signals of the averaged (lines) and the individual equatorial rotamer simulations with the isopropyl group in gauche+ (G+, dashed lines), gauche− (G−, dotted lines), and trans (T, dash-dotted lines) orientations. The experimental error bars and the shaded areas around the line plots of the simulations represent a 68% confidence interval obtained from bootstrap analysis. For the simulations, these error bars reflect convergence with respect to initial condition sampling.

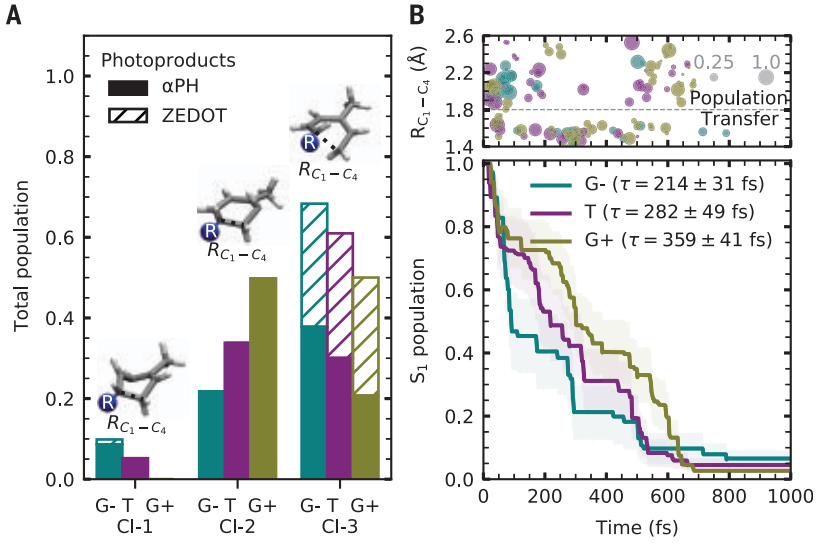


Fig. 5. Characterization of the closed- and open-ring nonradiative relaxation pathways of equatorial (eq) rotamers. (**A**) Histogram of the branching ratio between the closed and open pathways from simulations. Solid and hatched bars represent fractional populations that reform α-phellandrene (αPH) or ZEDOT, respectively. The insets show the minimum geometries of the conical intersections (CI) 1 to 3 of αPH, with the isopropyl substituent depicted as R. Additionally, the C_1-C_4 distance ($R_{C1\text{-}C4}$, Fig. 2A) is marked. (**B**) Top: $R_{C1\text{-}C4}$ versus population transfer time in femtoseconds from the excited state (S_1) to the ground state (S_0) for all 45 equatorial initial conditions (see supplementary materials). The black dashed line corresponds to the threshold used to determine whether the ring is open or closed. The circle radius is proportional to the amount of population involved in the transfer event and separated into the three rotamers. Bottom: The S_1 population decay for the three eq-rotamers with the isopropyl group in gauche (G+/G−) or trans (T) orientation in the first picosecond after photoexcitation. Shaded areas represent a 68% confidence interval obtained from bootstrap analysis.

Conformer-specific dynamics beyond the Woodward-Hoffmann rules

Our AIMS simulations revealed additional aspects of the relaxation process that could not be easily extracted from the experimental data. α-Phellandrene relaxes via the WH-predicted conrotatory ring-opening motion through a CI (CI-3) yielding both ring-opened photoproducts and the closed-ring reactant (Fig. 1B, Fig. 5A, and fig. S2). As a consequence of its differential nature, the ΔPDF observable is preferentially sensitive to the part of the population undergoing ring opening. Additionally, out-of-plane bending in S_1 leads through two CIs (CI-1/CI-2) to the closed-ring reactant (table S1). Furthermore, we observed a hitherto unknown secondary ring-closure reaction in the electronic ground state, yielding 2-ladderane (5-isopropyl-2-methylbicyclo[2.2.0]hex-2-ene). The ring closure was exclusively observed for the fraction of population undergoing internal conversion through CI-1 in the ground state (text S2). Non-WH relaxation pathways have been observed before for the prototypical CHD molecule (*23*) but had marginal contributions to the relaxation process. The presence of the isopropyl substituent redirects 40 ± 4% of the eq-αPH S_1 population toward non-WH pathways and thus substantially reduces the ring-opening quantum yield (Fig. 1B).

The time dependence of the measured and computed ΔPDFs in the α, β, and γ regions are shown in Fig. 4 and fig. S3B for the eq and ax conformers, respectively. The averaged simulation assuming equal distribution of the three eq rotamers (G−, T, and G+) leads to quantitative agreement with experiment, whereas individual eq rotamer contributions to the ΔPDFs are quite distinct (Fig. 4). Thus, the rotational orientation of the isopropyl group must influence the mechanism and time scale of ring opening. This fact becomes obvious in Fig. 5B, top, where simulated nonadiabatic population transfer events are categorized with respect to their C_1-C_4 distances at the time of the transfer event—that is, their ability to undergo ring opening and thus to contribute to the ΔPDF observable (see above). In the case of the eq-G− rotamer, almost all of the population undergoing ring opening relaxes to S_0 within 200 fs after photoexcitation. The other two rotamers show substantially less ring opening in this time frame. Their remaining population undergoes ring opening considerably later, 500 to 700 fs after photoexcitation. The difference in ring-opening time scale is clearly reflected in the time-dependent ΔPDF amplitudes in Fig. 4.

Conclusion

By combining UED with AIMS simulations, we directly observed conformer-specific photochemical reaction dynamics on their natural time and length scales in agreement with the Woodward-Hoffmann rules. Moreover, we observed considerable effects on the efficiency and time scale of the reaction from the presence and orientation of an isopropyl substituent of the molecule. Similar conformer-dependent photochemistry has been observed only in a limited number of studies of photoproduct distributions, mostly after conformer-specific resonance-enhanced photoionization (*1*, *24*–*26*). Because organic structures are prone to a large number of ground-state conformers and rotamers, our findings could have a large impact on our understanding of photochemical reaction mechanisms. We speculate that the paucity of previous observations of conformer and rotamer specificity is due to the absence of suitable experimental observables and challenges in preparing conformer-pure samples. Our study now demonstrates a suitable experimental observable. The investigation of specific conformers can be facilitated by methodological developments in the area of conformer-selected molecular beams (*27*, *28*) and single-molecule methods such as Coulomb explosion imaging (*29*) in combination with high–repetition rate x-ray free electron laser sources. Future investigations of photochemical dynamics using such emerging methods will provide general access to conformer-specific photochemistry.

REFERENCES AND NOTES

1. Y.-P. Chang *et al.*, *Science* **342**, 98–101 (2013).
2. K. A. Dill, J. L. MacCallum, *Science* **338**, 1042–1046 (2012).
3. R. B. Woodward, R. Hoffmann, *Angew. Chem. Int. Ed.* **8**, 781–853 (1969).
4. T. Bach, J. P. Hehn, *Angew. Chem. Int. Ed.* **50**, 1000–1045 (2011).
5. E. Havinga, J. L. M. A. Schlatmann, *Tetrahedron* **16**, 146–152 (1961).
6. J. E. Baldwin, S. M. Krueger, *J. Am. Chem. Soc.* **91**, 6444–6447 (1969).
7. M. M. de Christo Scherer *et al.*, *J. Tissue Viability* **28**, 94–99 (2019).

8. W. G. Dauben, J. Rabinowitz, N. D. Vietmeyer, P. H. Wendschuh, *J. Am. Chem. Soc.* **94**, 4285–4292 (1972).
9. C. W. Spangler, R. P. Hennis, *Chem. Commun.* **1972**, 24–25 (1972).
10. M. K. Lawless, S. D. Wickham, R. A. Mathies, *Acc. Chem. Res.* **28**, 493–502 (1995).
11. P. J. Reid, S. J. Doig, R. A. Mathies, *J. Phys. Chem.* **94**, 8396–8399 (1990).
12. M. Garavelli *et al.*, *J. Phys. Chem. A* **105**, 4458–4469 (2001).
13. B. C. Arruda, B. Smith, K. G. Spears, R. J. Sension, *Faraday Discuss.* **163**, 159–171 (2013).
14. O. Njoya, S. Matsika, T. Weinacht, *ChemPhysChem* **14**, 1451–1455 (2013).
15. B. C. Arruda, R. J. Sension, *Phys. Chem. Chem. Phys.* **16**, 4439–4455 (2014).
16. M. Ben-Nun, T. J. Martínez, *Adv. Chem. Phys.* **121**, 439–512 (2002).
17. M. Ben-Nun, T. J. Martínez, *J. Chem. Phys.* **108**, 7244–7257 (1998).
18. M. Ben-Nun, J. Quenneville, T. J. Martínez, *J. Phys. Chem. A* **104**, 5161–5175 (2000).
19. S. Matsika, P. Krause, *Annu. Rev. Phys. Chem.* **62**, 621–643 (2011).
20. A. Hofmann, R. de Vivie-Riedle, *Chem. Phys. Lett.* **346**, 299–304 (2001).
21. K. M. Marzec, I. Reva, R. Fausto, L. M. Proniewicz, *J. Phys. Chem. A* **115**, 4342–4353 (2011).
22. J. W. Snyder Jr., R. M. Parrish, T. J. Martínez, *J. Phys. Chem. Lett.* **8**, 2432–2437 (2017).
23. T. J. A. Wolf *et al.*, *Nat. Chem.* **11**, 504–509 (2019).
24. M. H. Kim, L. Shen, H. Tao, T. J. Martinez, A. G. Suits, *Science* **315**, 1561–1565 (2007).
25. S. T. Park, S. K. Kim, M. S. Kim, *Nature* **415**, 306–308 (2002).
26. K.-W. Choi, D.-S. Ahn, J.-H. Lee, S. K. Kim, *Chem. Commun.* **2007**, 1041–1043 (2007).
27. F. Filsinger, G. Meijer, H. Stapelfeldt, H. N. Chapman, J. Küpper, *Phys. Chem. Chem. Phys.* **13**, 2076–2087 (2011).
28. Y.-P. Chang, D. A. Horke, S. Trippel, J. Küpper, *Int. Rev. Phys. Chem.* **34**, 557–590 (2015).
29. T. Endo *et al.*, *Science* **370**, 1072–1077 (2020).
30. T. J. A. Wolf, Figures for "Conformer-specific photochemistry imaged in real space and time". DOI:10.5281/zenodo.5214798.

ACKNOWLEDGMENTS

We thank J. P. Cryan, M. C. Hoffmann, R. K. Li, M. Niebuhr, S. Weathersby, and T. Weinacht for their help and fruitful discussions. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the US Department of Energy (DOE), National Nuclear Security Administration, under contract DE-AC52-07NA27344. **Funding:** This work was supported by the AMOS program within the DOE Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division. The experimental part of this research was performed at the SLAC MeV UED facility, which is supported in part by the DOE BES SUF Division Accelerator & Detector R&D program, the Linac Coherent Light Source (LCLS) Facility, and SLAC under contracts DE-AC02-05-CH11231 and DE-AC02-76SF00515. Also supported by a Volkswagen Foundation Lichtenberg Professorship (M.G.); a NSF graduate fellowship (D.M.S.); and DOE Office of Science, Basic Energy Sciences awards DE-SC0014170 (J.P.F.N. and M.C.) and DE-FG02-08ER15984 (Y.L.). **Author contributions:** Conceptualization: D.M.S., T.J.M., and T.J.A.W. Methodology: J.Y., J.P.F.N., M.C., M.G., B.M., X.S., X.J.W., and T.J.A.W. Investigation: E.G.C., D.M.S., J.Y., J.P.F.N., A.A., M.C., R.F., M.G., K.H., F.J., S.K.S., Y.L., M.-F.L., D.L., B.M., X.S., M.R.W., X.J.W., T.J.M., and T.J.A.W. Visualization: E.G.C., D.M.S., and T.J.A.W. Funding acquisition: M.C., M.G., X.J.W., T.J.M., and T.J.A.W. Project administration: T.J.A.W. Supervision: T.J.M. and T.J.A.W. Writing–original draft: E.G.C., D.M.S., T.J.M., and T.J.A.W. Writing–review and editing: E.G.C., D.M.S., J.Y., J.P.F.N., A.A., M.C., R.F., M.G., K.H., F.J., S.K.S., Y.L., M.-F.L., D.L., B.M., X.S., M.R.W., X.J.W., T.J.M., and T.J.A.W. **Competing interests:** None declared. **Data and materials availability:** All data underlying the figures are deposited in the Zenodo repository (*30*). The raw experimental data are archived at the SLAC MeV-UED facility, and the raw simulated data are stored at the Martinez laboratory at Stanford University and SLAC. All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk3132
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S3
References (*31*–*41*)
Movies S1 to S26

6 July 2021; accepted 24 August 2021
10.1126/science.abk3132

PALEOGENOMICS

Ten millennia of hepatitis B virus evolution

Arthur Kocher[1,2,3]*, Luka Papac[2,3], Rodrigo Barquera[2,3], Felix M. Key[2,4], Maria A. Spyrou[2,3,5], Ron Hübler[2], Adam B. Rohrlach[2,3,6], Franziska Aron[2], Raphaela Stahl[2], Antje Wissgott[2], Florian van Bömmel[7], Maria Pfefferkorn[7], Alissa Mittnik[2,8,9], Vanessa Villalba-Mouco[2,10], Gunnar U. Neumann[2,3], Maïté Rivollat[2,11], Marieke S. van de Loosdrecht[2], Kerttu Majander[2,12], Rezeda I. Tukhbatova[2,13], Lyazzat Musralina[2,3,14,15], Ayshin Ghalichi[2,3], Sandra Penske[2,3], Susanna Sabin[2], Megan Michel[2,3,9], Joscha Gretzinger[2,3], Elizabeth A. Nelson[2], Tiago Ferraz[2,16], Kathrin Nägele[2,3], Cody Parker[2,17], Marcel Keller[2,18], Evelyn K. Guevara[2,19], Michal Feldman[2,5], Stefanie Eisenmann[2,3], Eirini Skourtanioti[2,3], Karen Giffin[2,3], Guido Alberto Gnecchi-Ruscone[2,3], Susanne Friederich[20], Vittoria Schimmenti[21], Valery Khartanovich[22], Marina K. Karapetian[23], Mikhail S. Chaplygin[24], Vladimir V. Kufterin[25], Aleksandr A. Khokhlov[26], Andrey A. Chizhevsky[27], Dmitry A. Stashenkov[28], Anna F. Kochkina[28], Cristina Tejedor-Rodríguez[29], Íñigo García-Martínez de Lagrán[30], Héctor Arcusa-Magallón[31], Rafael Garrido-Pena[32], José Ignacio Royo-Guillén[33], Jan Nováček[34,35], Stéphane Rottier[11], Sacha Kacki[11,36], Sylvie Saintot[37], Elena Kaverzneva[38], Andrej B. Belinskiy[39], Petr Velemínský[40], Petr Limburský[41], Michal Kostka[42], Louise Loe[43], Elizabeth Popescu[44], Rachel Clarke[44], Alice Lyons[44], Richard Mortimer[44], Antti Sajantila[19,45], Yadira Chinique de Armas[46], Silvia Teresita Hernandez Godoy[47,48], Diana I. Hernández-Zaragoza[49,50], Jessica Pearson[51], Didier Binder[52], Philippe Lefranc[53], Anatoly R. Kantorovich[54], Vladimir E. Maslov[55], Luca Lai[56,57], Magdalena Zoledziewska[58], Jessica F. Beckett[59], Michaela Langová[41], Alžběta Danielisová[41], Tara Ingman[60], Gabriel García Atiénzar[61], Maria Paz de Miguel Ibáñez[61], Alejandro Romero[61,62], Alessandra Sperduti[63,64], Sophie Beckett[65,66,67], Susannah J. Salter[65,68], Emma D. Zilivinskaya[25], Dmitry V. Vasil'ev[69], Kristin von Heyking[70], Richard L. Burger[71], Lucy C. Salazar[71], Luc Amkreutz[72], Masnav Navruzbekov[73], Eva Rosenstock[74], Carmen Alonso-Fernández[75], Vladimir Slavchev[76], Alexey A. Kalmykov[39], Biaslan Ch. Atabiev[77], Elena Batieva[78], Micaela Alvarez Calmet[79], Bastien Llamas[80,81,82], Michael Schultz[83,84], Raiko Krauß[85], Javier Jiménez-Echevarría[75], Michael Francken[86], Svetlana Shnaider[87], Peter de Knijff[88], Eveline Altena[88], Katrien Van de Vijver[89,90,91], Lars Fehren-Schmitz[92,93], Tiffiny A. Tung[94], Sandra Lösch[95], Maria Dobrovolskaya[55], Nikolaj Makarov[55], Chris Read[96], Melanie Van Twest[65], Claudia Sagona[97], Peter C. Ramsl[98], Murat Akar[99], K. Aslihan Yener[100], Eduardo Carmona Ballestero[101,102], Francesco Cucca[58,103], Vittorio Mazzarello[103], Pilar Utrilla[104], Kurt Rademaker[105], Eva Fernández-Domínguez[36], Douglas Baird[51], Patrick Semal[89], Lourdes Márquez-Morfín[106], Mirjana Roksandic[46,107,108], Hubert Steiner[109], Domingo Carlos Salazar-García[110,111,112], Natalia Shishlina[22,38], Yilmaz Selim Erdal[113], Fredrik Hallgren[114], Yavor Boyadzhiev[115], Kamen Boyadzhiev[115], Mario Küßner[34], Duncan Sayer[116], Päivi Onkamo[117,118], Robin Skeates[36], Manuel Rojo-Guerra[29], Alexandra Buzhilova[23], Elmira Khussainova[15], Leyla B. Djansugurova[15], Arman Z. Beisenov[119], Zainolla Samashev[120,121], Ken Massy[122], Marcello Mannino[123,124], Vyacheslav Moiseyev[22], Kristiina Mannermaa[125], Oleg Balanovsky[126,127,128], Marie-France Deguilloux[11], Sabine Reinhold[129], Svend Hansen[129], Egor P. Kitov[25,119], Miroslav Dobeš[41], Michal Ernée[41], Harald Meller[20], Kurt W. Alt[130,131,132], Kay Prüfer[2,3], Christina Warinner[2,3,133], Stephan Schiffels[2,3], Philipp W. Stockhammer[2,3,122], Kirsten Bos[2,3], Cosimo Posth[2,5], Alexander Herbig[2,3], Wolfgang Haak[2,3,134], Johannes Krause[2,3]*, Denise Kühnert[1,2,3,135]*

Hepatitis B virus (HBV) has been infecting humans for millennia and remains a global health problem, but its past diversity and dispersal routes are largely unknown. We generated HBV genomic data from 137 Eurasians and Native Americans dated between ~10,500 and ~400 years ago. We date the most recent common ancestor of all HBV lineages to between ~20,000 and 12,000 years ago, with the virus present in European and South American hunter-gatherers during the early Holocene. After the European Neolithic transition, Mesolithic HBV strains were replaced by a lineage likely disseminated by early farmers that prevailed throughout western Eurasia for ~4000 years, declining around the end of the 2nd millennium BCE. The only remnant of this prehistoric HBV diversity is the rare genotype G, which appears to have reemerged during the HIV pandemic.

The World Health Organization (WHO) estimates that in 2015, 257 million people were living with chronic hepatitis B virus (HBV) infection, which causes close to 1 million deaths each year (*1*). HBV is transmitted through contact with bodily fluids, mainly in sexual and perinatal contexts (*2*), and has no known environmental or animal reservoir. Its spread is therefore tightly linked to the dispersal of humans, whose past population dynamics and migrations have likely shaped the genetic diversity of this partially

double-stranded DNA virus, which is currently classified into nine genotypes associated with characteristic ethno-geographic ranges (Fig. 1) (*3*, *4*). However, the temporal and geographic context of HBV origins in humans, as well as its major routes of dissemination in the past, remain widely debated (*5*–*10*). Recent studies have retrieved HBV DNA from archaeological human remains (*11*–*16*), providing new avenues to address questions about HBV evolution and phylogeographic history. In particular, these studies revealed the presence of HBV in Europe as early as the Neolithic and ancient HBV lineages that are now seemingly extinct. Ancient DNA data permits molecular clock calibration, and the time to the most recent common ancestor (tMRCA) of all known HBV lineages has been dated to between ~21 thousand years ago (ka) and ~9 ka (*14*). However, the extent of the past diversity of this virus remains generally unknown because only 19 ancient HBV genomes with a limited temporal and geographic distribution have been reconstructed to date.

The MRCA of all known HBV lineages

Here, we report genomic evidence of HBV in the skeletal remains of 137 individuals from Eurasia and the Americas dated to between ~10,500 and ~400 years ago (Fig. 1, fig. S1 and data S1). Despite advances in molecular virology and numerous sequences from present-day HBV genomes, assessing the phylogenetic relationships among HBV genotypes has proven challenging (*7*, *17*–*20*), and doubts have been cast about its evolutionary rate and molecular clock–like behavior (*9*, *16*, *21*). Nevertheless,

[1]Transmission, Infection, Diversification and Evolution Group, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [3]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [4]Max Planck Institute for Infection Biology, 10117 Berlin, Germany. [5]Archaeo- and Palaeogenetics group, Institute for Archaeological Sciences, Eberhard Karls University Tübingen, 72070 Tübingen, Germany. [6]ARC Centre of Excellence for Mathematical and Statistical Frontiers, School of Mathematical Sciences, University of Adelaide, Adelaide, SA 5005, Australia. [7]Division of Hepatology, Department of Medicine II, Leipzig University Medical Center, Leipzig, Germany. [8]Department of Genetics, Harvard Medical School, Boston, MA, USA. [9]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [10]Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona, Spain. [11]Université de Bordeaux, CNRS, PACEA UMR 5199, Pessac, France. [12]Institute of Evolutionary Medicine (IEM), University of Zürich, 8057 Zürich, Switzerland. [13]Laboratory of Structural Biology, Kazan Federal University, Kazan, Russia. [14]Al-Farabi Kazakh National University, Almaty, Kazakhstan. [15]Institute of Genetics and Physiology, 050060 Almaty, Kazakhstan. [16]Departmento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil. [17]Arizona State University School of Human Evolution and Social Change, Tempe Arizona, USA. [18]Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia. [19]Department of Forensic Medicine, University of Helsinki, Helsinki, Finland. [20]State Office for Heritage Management and Archaeology Saxony-Anhalt and State Museum of Prehistory, D-06114 Halle, Germany. [21]Museo Archeologico Regionale "Antonino Salinas", 90133 Palermo, Italy. [22]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, 199034 St. Petersburg, Russia. [23]Anuchin Research Institute and Museum of Anthropology, Lomonosov Moscow State University, Moscow, Russia. [24]Sterlitamak Museum of Local History, Sterlitamak, Russia. [25]Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow, Russia. [26]Samara State University of Social Sciences and Education, Samara, Russia. [27]Institute of Archaeology named after A. Kh. Khalikov, Tatarstan Academy of Sciences, Kazan, Russia. [28]Samara Museum for Historical and Regional Studies named after P. V. Alabin, Samara, Russia. [29]Department of Prehistory and Archaeology, Faculty of Philosophy and Letters, University of Valladolid, Spain. [30]Department of Prehistory and Archaeology, National Distance Education University (UNED), Madrid, Spain. [31]Private Technical Archaeologist, 50196 La Muela, Spain. [32]Department of Prehistory and Archaeology, Faculty of Philosophy and Letters, Autonomous University of Madrid, Spain. [33]Technical Archaeologist, Government of Aragón, Spain. [34]Thuringian State Office for Heritage Management and Archaeology, 99423 Weimar, Germany. [35]University Medical School Göttingen, Institute of Anatomy and Cell Biology, 37075 Göttingen, Germany. [36]Department of Archaeology, Durham University, South Road, Durham. DH1 3LE. UK. [37]INRAP, ARAR UMR 5138, Maison de l'Orient et de la Méditerranée, Lyon, France. [38]State Historical Museum, Moscow, Russia. [39]Nasledie Cultural Heritage Unit, 355006 Stavropol, Russia. [40]Department of Anthropology, The National Museum, Prague, Czech Republic. [41]Institute of Archaeology of the Czech Academy of Sciences, Prague, Czech Republic. [42]The City of Prague Museum, Prague, Czech Republic. [43]Oxford Archaeology South, Janus House, Osney Mead, Oxford, OX2 0ES, UK. [44]Oxford Archaeology East, Bar Hill, Cambridge, CB23 8SQ, UK. [45]Forensic Medicine Unit, Finnish Institute of Health and Welfare, Helsinki, Finland. [46]Department of Anthropology, University of Winnipeg, Winnipeg, MB, Canada. [47]Grupo de Investigación y Desarrollo, Dirección Provincial de Cultura, Matanzas, Cuba. [48]Universidad de Matanzas, Matanzas, Cuba. [49]Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia (ENAH), Mexico City, Mexico. [50]Immunogenetics Unit, Técnicas Genéticas Aplicadas a la Clínica (TGAC), Mexico City, Mexico. [51]Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool L69 7WZ, UK. [52]Université Côte d'Azur, CNRS, CEPAM UMR 7264, Nice, France. [53]Université de Strasbourg, CNRS, Archimède UMR 7044, Strasbourg, France. [54]Department of Archaeology, Faculty of History, Lomonosov Moscow State University, 119192 Moscow, Russia. [55]Institute of Archaeology, Russian Academy of Sciences, , Moscow 117292, Russia. [56]Department of Anthropology, University of South Florida, Tampa, FL, USA. [57]Department of Anthropology, University of North Carolina at Charlotte, Charlotte, NC, USA. [58]Istituto di Ricerca Genetica e Biomedica–CNR, Monserrato, Italy. [59]Private contractor, Cagliari, Sardinia, Italy. [60]Koç University, Research Center for Anatolian Civilizations, Istanbul 34433, Turkey. [61]Institute for Research in Archaeology and Historical Heritage (INAPH), University of Alicante, 03690, Alicante, Spain. [62]Departamento de Biotecnología, Facultad de Ciencias, Universidad de Alicante, 03690, Alicante, Spain. [63]Bioarchaeology Service, Museum of Civilizations, Rome, Italy. [64]Dipartimento Asia Africa e Mediterraneo, Università di Napoli L'Orientale, Napoli, Italy. [65]Sedgeford Historical and Archaeological Research Project, Old Village Hall, Sedgeford, Hunstanton PE36 5LS, UK. [66]Melbourne Dental School, University of Melbourne, Victoria 3010 Australia. [67]Cranfield Forensic Institute, Cranfield Defence and Security, Cranfield University, College Road, Cranfield, MK43 0AL, UK. [68]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK. [69]Astrakhan State University, Astrakhan, Russia. [70]SNSB, State Collection for Anthropology and Palaeoanatomy, 80333 Munich, Germany. [71]Department of Anthropology, Yale University, New Haven, CT 06511, USA. [72]National Museum of Antiquities, 2301 EC Leiden, Netherlands. [73]Institute of History, Archaeology and Ethnography, Dushanbe, Tajikistan. [74]Freie Universität Berlin, Einstein Center Chronoi, 14195 Berlin, Germany. [75]Cronos S.C. Arqueología y Patrimonio, Burgos, Spain. [76]Regional Historical Museum Varna, Varna, BG-9000, Bulgaria. [77]Institute for Caucasus Archaeology, 361401 Nalchik, Republic Kabardino-Balkaria, Russia. [78]Azov History, Archaeology and Palaeontology Museum-Reserve, Azov 346780, Russia. [79]Museo de Sitio Pucllana, s/n Miraflores, 18 Lima, Peru. [80]Australian Centre for Ancient DNA, School of Biological Sciences and The Environment Institute, Adelaide University, Adelaide, SA 5005, Australia. [81]Centre of Excellence for Australian Biodiversity and Heritage (CABAH), University of Adelaide, Adelaide, SA 5005, Australia. [82]National Centre for Indigenous Genomics, Australian National University, Canberra, ACT 0200, Australia. [83]University Medical School Göttingen, Institute of Anatomy and Embryology, 37075 Göttingen, Germany. [84]Institute of Biology, University of Hildesheim, Germany. [85]Institute for Prehistory, Early History and Medieval Archaeology, University of Tübingen, 72070 Tübingen, Germany. [86]State Office for Cultural Heritage Baden-Württemberg, 78467 Konstanz, Germany. [87]ArchaeoZoology in Siberia and Central Asia–ZooSCAn, CNRS–IAET SB RAS International Research Laboratory, IRL 2013, Novosibirsk, Russia. [88]Department of Human Genetics, Leiden University Medical Center, Leiden, 2333 ZC, Netherlands. [89]Royal Belgian Institute of Natural Sciences, Brussels, Belgium. [90]Center for Archaeological Sciences, University of Leuven, Belgium. [91]Dienst Archeologie–Stad Mechelen, Belgium. [92]UCSC Paleogenomics Laboratory, Department of Anthropology, University of California at Santa Cruz, Santa Cruz, CA 95064, USA. [93]UCSC Genomics Institute, University of California at Santa Cruz, Santa Cruz, CA 95064, USA. [94]Department of Anthropology, Vanderbilt University, Nashville, TN 37235, USA. [95]Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Bern, Switzerland. [96]Applied Archaeology School of Science, Institute of Technology Sligo, Ireland. [97]School of Historical and Philosophical Studies, University of Melbourne, Victoria 3010, Australia. [98]Institute of Prehistoric and Historical Archaeology, University of Vienna, Austria. [99]Department of Archaeology, Hatay Mustafa Kemal University, Alahan-Antakya, Hatay 31060, Turkey. [100]Institute for the Study of the Ancient World (ISAW), New York University, New York, NY 10028, USA. [101]Territorial Service of Culture and Tourism from Valladolid, Castilla y León Regional Government, C/ San Lorenzo, 5, 47001, Valladolid, Spain. [102]Department of History, Geography and Comunication, University of Burgos, Paseo de Comendadores, s/n 09001 Burgos (Burgos), Spain. [103]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy. [104]Área de Prehistoria, P3A DGA Research Group, IPH, University of Zaragoza, C/ Pedro Cerbuna 12, 50009, Zaragoza, Spain. [105]Department of Anthropology, Michigan State University, East Lansing, MI 48824, USA. [106]Osteology Laboratory, Post Graduate Studies Division, Escuela Nacional de Antropología e Historia (ENAH), Mexico City, Mexico. [107]Caribbean Research Institute, Univeristy of Winnipeg, Winnipeg, MB, Canada. [108]DFG Center for Advanced Studies "Words, Bones, Genes, Tools," University of Tübingen, Tübingen, Germany. [109]South Tyrol Provincial Heritage Service, South Tyrol, Italy. [110]Grupo de Investigación en Prehistoria IT-1223-19 (UPV-EHU)/IKERBASQUE-Basque Foundation for Science, Vitoria, Spain. [111]Departament de Prehistòria, Arqueologia i Història Antiga, Universitat de València, València, Spain. [112]Department of Geological Sciences, University of Cape Town, Cape Town, South Africa. [113]Human_G Laboratory, Department of Anthropology, Hacettepe University, Ankara 06800, Turkey. [114]The Cultural Heritage Foundation, 722 12 Västerås, Sweden. [115]National Archaeological Institute with Museum at the Bulgarian Academy of Sciences, Sofia 1000, Bulgaria. [116]School of Natural Sciences, University of Central Lancashire, Preston, UK. [117]Department of Biosciences, University of Helsinki, 00014 Helsinki, Finland. [118]Department of Biology, University of Turku, 20500 Turku, Finland. [119]Institute of archaeology named after A. Kh. Margulan, 44 Almaty, Kazakhstan. [120]Branch of Institute of Archaeology named after A.Kh. Margulan, 24 of 511 Nur-Sultan, Kazakhstan. [121]State Historical and Cultural Museum-Reserve "Berel," Katon-Karagay district, East Kazakhstan region, Kazakhstan. [122]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig-Maximilians-Universität München, 80539 Munich, Germany. [123]Department of Archeology and Heritage Studies, Aarhus University, 8270 Højbjerg, Denmark. [124]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig Germany. [125]Department of Cultures, University of Helsinki, 00014 Helsinki, Finland. [126]Research Centre for Medical Genetics, Moscow, Russia. [127]Biobank of North Eurasia, Moscow, Russia. [128]Vavilov Institute of General Genetics, Moscow, Russia. [129]Eurasia Department, German Archaeological Institute, Berlin, Germany. [130]Danube Private University, Center of Natural and Cultural Human History, A - 3500 Krems-Stein, Austria. [131]Integrative Prehistory and Archaeological Science, Spalenring 145, CH-4055 Basel, Switzerland. [132]Department of Biomedical Engineering (DBE), Universitätsspital Basel (HFZ), CH-4123 Allschwil, Switzerland. [133]Department of Anthropology, Harvard University, Cambridge, MA 02138, USA. [134]School of Biological Sciences, University of Adelaide, Adelaide, SA, 5005, Australia. [135]European Virus Bioinformatics Center (EVBC), Jena, Germany.
***Corresponding author. Email: kocher@shh.mpg.de (A.K.); krause@shh.mpg.de (J.K.); kuehnert@shh.mpg.de (D.K.)**

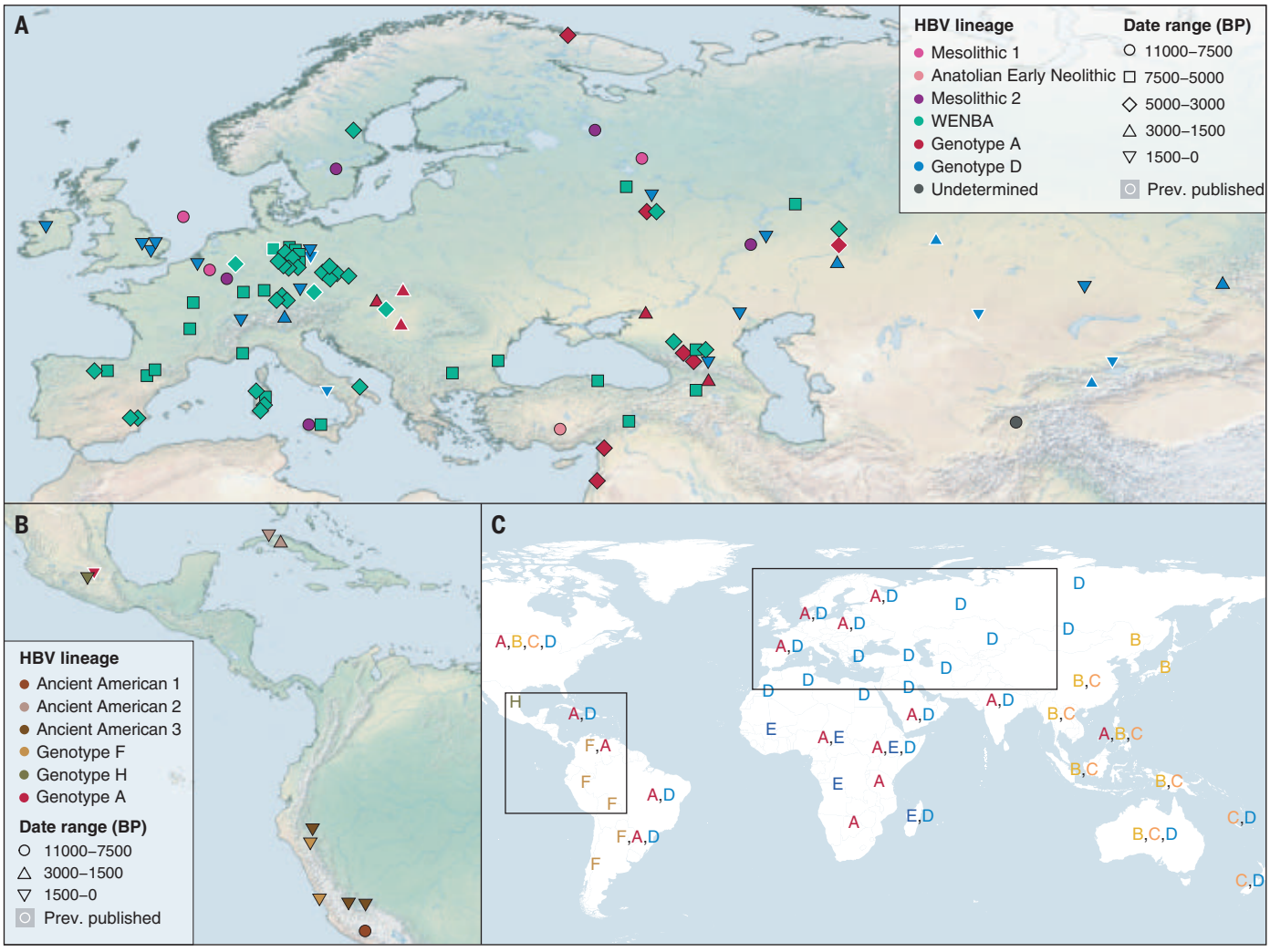


Fig. 1. Geographic location, time period, and lineage of ancient HBV genomes. (**A** and **B**) Lineages from (A) Eurasia and (B) the Americas. (**C**) Main distribution of present-day HBV genotypes [adapted from (*4*, *14*)].

most HBV phylogenetic reconstructions have recovered a topology in which HBV genotypes typically found in Native Americans (genotypes F and H) represent a sister clade to the rest of worldwide HBV diversity (which we refer to as the Eurasian branch) (*18*). This topology was supported by a study incorporating 12 ancient HBV genomes (*14*) and was also reconstructed in this work (Fig. 2 and figs. S2 and S3). In particular, the monophyly of the American HBV branch, comprising all ancient genomes from the Americas dating back to as early as ~9 ka from the Cuncaicha rock shelter in the Andean highlands (CUN002), was highly supported. However, deep nodes within the Eurasian branch were not well resolved, pointing to plausible alternative topologies in which some of the earliest Eurasian lineages would have diverged before the American branch (figs. S4 and S5) (*22*). Our results confirm that HBV genomic data do exhibit a clear temporal structure when incorporating samples spanning several thousand years (fig. S3). Using the best-fitting uncorrelated relaxed clock model, we estimate the tMRCA of HBV, corresponding to the divergence of American and Eurasian HBV branches, to be between ~16 and ~12 ka [95% highest posterior density (HPD)] (table S1), which is within the range of previous findings (*14*). This suggests that contacts between ancestral Eurasians and First Americans occurred until at least shortly before the Bølling-Allerød interstadial (~15 to 13 ka), a period of warming corresponding to widespread human expansion in North America (*23*, *24*). However, studies of ancient human genomes indicate that the ancestors of the First Americans likely began diverging from their closest Eurasian relatives between ~25 and 18 ka, possibly reflecting an extended isolation in a Beringian refugium during the Last Glacial Maximum, before dispersing into and across the Americas (*25*–*27*). The use of a time-dependent rate (TDR) model yielded an estimate of ~20 to 17 ka for the HBV tMRCA (95% HPD), which was more consistent in this regard. This suggests that not accounting for the time dependency of the evolutionary rate may have led to an underestimation of deep divergence times. However, model selection favored the use of a relaxed clock over a TDR model (log BF: 405) (*22*). Taken together, these results point to a scenario in which the MRCA of all HBV strains examined to date existed around the end of the Pleistocene and gave rise to one or several lineages that spread across Eurasia and eventually reached Africa and Oceania, and to another lineage that spread into the Americas with early settlers of this continent.

Our findings challenge the view that current HBV diversity reflects early human dispersals out of Africa. This model is supported in particular by the exclusive association of HBV

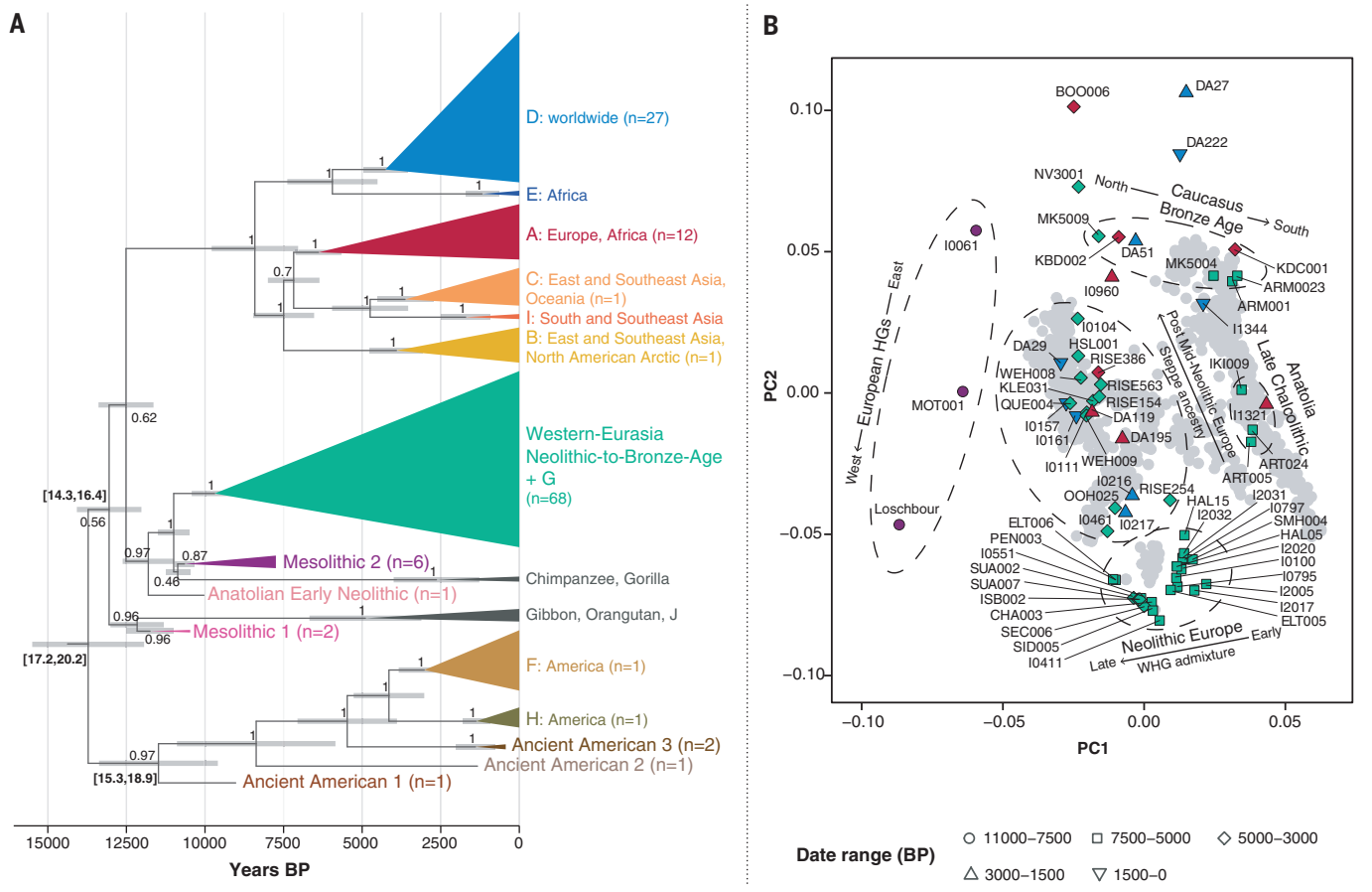


Fig. 2. HBV phylogeny and genetic profile of infected individuals. (**A**) Time-calibrated phylogenetic tree of HBV obtained by using a skyline coalescent tree prior and a lognormal relaxed clock. Main clades were collapsed and annotated with their typical present-day geographic location and the number of ancient genomes they contain. Posterior node supports and date estimates (gray bars, indicating 95% HPD) are reported. Additional time intervals written on deep nodes are 95% HPD estimates obtained with a time-dependent rate model. (**B**) Principal components analysis plot of modern and ancient western Eurasians summarizing the genetic variation of a subset of individuals for which human genetic data was available. Individuals are colored according to the lineage of the HBV strain they carried, as in the tree.

subgenotype C4 with the Aboriginal people of Australia, suggesting that this subgenotype may have been carried by the first settlers of Australia at least ~50 ka (*5*, *20*). Instead, in accordance with previous findings (*14*), our results indicate that all known modern and ancient HBV strains descend from a lineage that began to diversify at a more recent stage of human history and that subgenotype C4 was introduced in the Australian continent after ~4.5 ka (Fig. 2). Nevertheless, the age of the observed MRCA only represents a lower limit for the earliest presence of HBV in humans. Whether the latter has been preceded by long coevolution, a recent spillover from another animal species, or any intermediate scenario remains an open question. Other viruses from the Hepadnaviridae family have been recovered from a wide range of vertebrates, but none of them appear to represent an ancestral zoonotic source for the human HBV (*8*).

HBV circulated widely in western Eurasia as early as 10 ka

The retrieval of HBV genomes from around 10 ka in different parts of Europe and Anatolia indicates that the virus was widespread in western Eurasia at that time (Fig. 1 and fig. S1). The oldest HBV strains recovered in Europe form two distinct clades (Fig. 2, fig. S2, and table S2): one that was found in three hunter-gatherers (HGs) from northwestern Russia, Belgium, and Doggerland (Mesolithic 1) and another that was found in an HG from western Russia (Mesolithic 2). These two lineages are placed within the Eurasian branch as sister groups to the modern strains found in nonhuman primates (NHPs) from Southeast Asia and Africa, respectively. The position of NHP HBV lineages within human HBV diversity has been observed in most previous phylogenetic reconstructions and is thought to reflect spillover events from humans to NHPs (*7*, *22*, *28*). The HBV genome reconstructed from an early Anatolian farmer forms a separate lineage recovered at a phylogenetic position intermediate to the two European Mesolithic clades. Between ~9 and 7.5 ka, HBV strains found in HGs from Karelia (northwestern Russia), Sweden, Luxembourg, and Sicily all belonged to the Mesolithic 2 clade. Thus, although our data do not allow detailed phylogeographic inference, they suggest that during the early Holocene, HBV strains could spread over large parts of western Eurasia within a few thousand years. This is consistent with evidence of genetic connections between Europe and the Near East that predate the Neolithic transition (*29*, *30*) and with the observed genetic cline from Western to Eastern HGs (*31*). Our results further highlight that Mesolithic populations likely formed a network through which pathogens could spread.

It has been suggested that most human-adapted pathogens emerged after the Neolithic

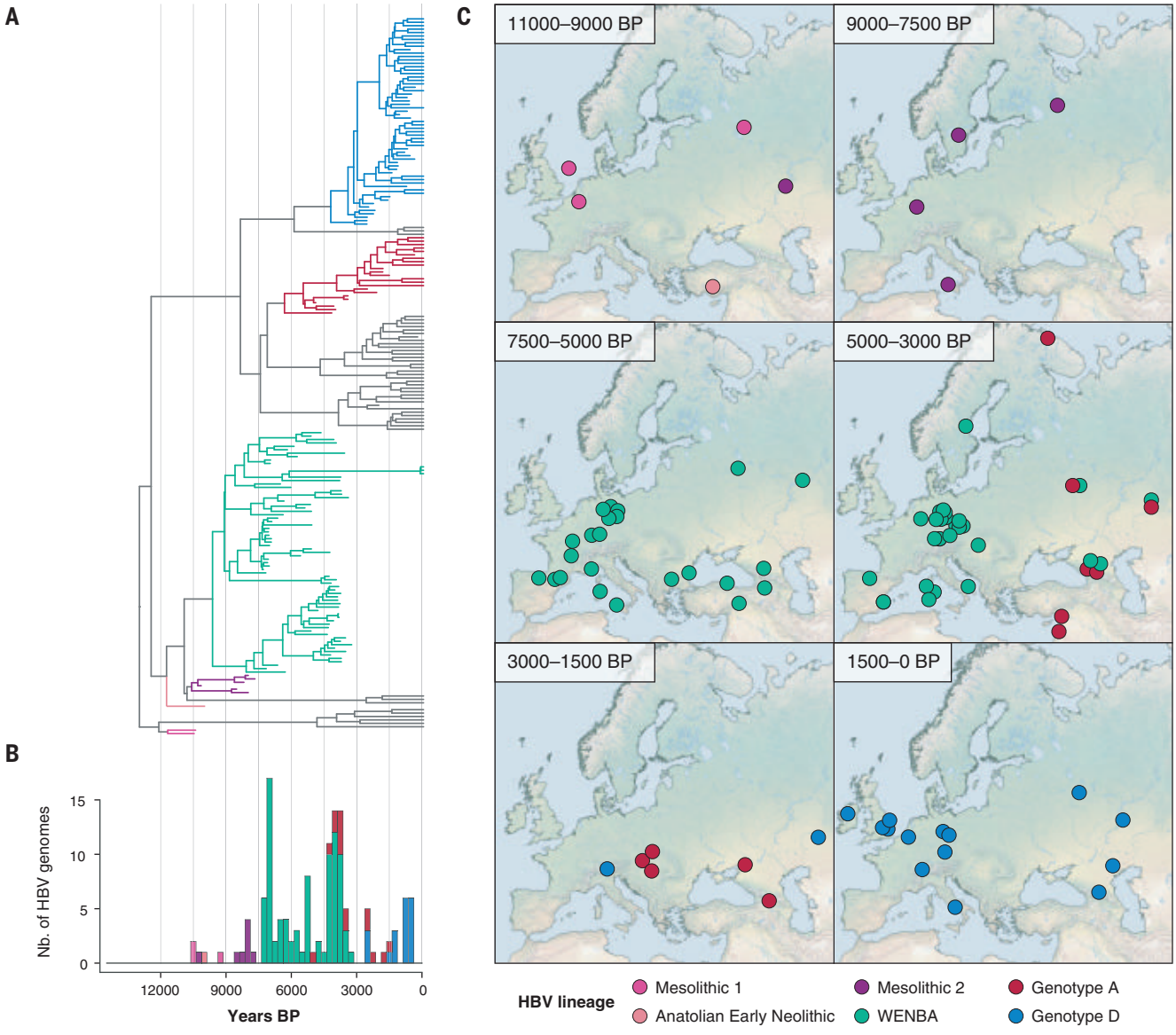


Fig. 3. Spatiotemporal distribution of ancient western Eurasian HBV strains. (**A**) Time-calibrated phylogenetic tree (Eurasian branch). Lineages containing ancient HBV genomes are colored. (**B**) Histogram showing the number of recovered ancient HBV genomes belonging to each lineage through time. (**C**) Geographic distribution of ancient HBV genomes within different time-periods, colored by lineage.

transition in association with sedentary lifestyles, increased contact with domesticated animals, and higher population densities, a phenomenon sometimes referred to as the "first epidemiological transition" (*32*–*34*). Our finding of widespread HBV in HG populations indicates that HBV was present before the advent of agriculture and animal husbandry in different parts of the world. Today, HBV rarely causes lethal fulminant hepatitis but rather asymptomatic infections that may evolve into chronic forms, sometimes developing into liver complications and possible liver failure after decades of infection (*1*, *2*). Although it is difficult to extrapolate from present-day medical studies what the clinical impact of a pathogen would have been in the past—given different diets, disease burdens, and life expectancies—the virus has likely exhibited similar pathophysiological features. Consequently, our findings are consistent with the view that although small HG communities could not sustain highly epidemic "crowd" diseases, they could maintain chronic infectious agents (*35*, *36*).

A replacement of HBV diversity occurred with the Neolithic transition in Europe

Our data show that HBV remained widespread in Europe after the Neolithic transition (8 to 7 ka), with numerous strains recovered from early European farmers (EEF) across the continent (Fig. 3, fig. S1, and data S1). All of these strains belong to a single HBV lineage that does not descend from previously observed Mesolithic strains (Figs. 2 and 3 and fig. S2). We refer to this HBV lineage as the Western-Eurasian Neolithic–to–Bronze Age (WENBA) lineage. This transition is also observable at a microscale in Grotta dell'Uzzo (Sicily), where HBV strains recovered from Neolithic individuals are unrelated to a Late Mesolithic strain identified at the same site (figs. S1 and S2). This suggests that the HBV strains observed in EEFs were not acquired

from local HGs in different areas but were rather disseminated by EEFs themselves. Although EEFs ultimately derived from early agricultural populations in the Near East (*37*, *38*), the strain we retrieved from an Anatolian farmer dated to ~10 ka was not ancestral to the WENBA lineage (Fig. 2). Therefore, even if EEFs were indeed key in disseminating WENBA strains, whether this lineage originated in Near Eastern centers of early agriculture or in another location along EEF's expansion routes remains to be determined. Furthermore, given the current sample availability for this period, a scenario in which the WENBA lineage would have originated and disseminated among European HGs shortly before the Neolithic transition cannot be completely excluded.

We also found WENBA HBV strains in two HGs from transitional Neolithic contexts in western Russia dated to ~7.2 and ~6.4 ka (JAZ001 and MUR007), as well as on both sides of the Greater Caucasus Mountain range and in Anatolia as early as ~5.6 ka (fig. S1). In general, phylogenetic relationships among HBV sublineages within the WENBA clade do not exhibit a strong geographical structure (fig. S2), nor do they seem to reflect the material culture or genetic profile of the individuals in which they were found (fig. S6). Furthermore, our phylodynamic reconstruction indicates that after an initial growth phase, the transmission of WENBA HBV reached an equilibrium from ~7.5 to ~3.5 ka (fig. S7). Overall, this suggests that HBV strains disseminated by EEFs quickly spread throughout much of western Eurasia beyond the limits of the European agricultural expansion, where they became endemic and continued to circulate widely across different populations, for several thousand years. In particular, we do not observe substantial changes in the HBV genetic landscape associated with the expansion of steppe-related ancestry that dramatically altered the genetic profile of Europeans from ~5 ka onward (Fig. 2, fig. S2, and data S1) (*37*). Sexual and perinatal transmission have likely always been the major mechanisms of HBV infection in humans, but cultural practices involving contact with blood [such as tattooing (*39*)] or nonsexual violent interactions (*40*) could also have played a role in the spread of the virus in the past. In general, our findings attest to a degree of interconnectivity among prehistoric populations of different origins, subsistence modes, and cultures that allowed for the dissemination of directly transmitted pathogens.

The collapse of WENBA HBV during the 2nd millennium BCE

After the Early Neolithic (8 to 7 ka), the WENBA HBV lineage prevailed in most parts of western Eurasia for more than 4000 years (Fig. 3). However, the latest occurrence of a WENBA strain in our dataset is dated to ~3.3 ka, after which this lineage is no longer observed (figs. S1 and S2). By contrast, genotype A, which we first observed at the eastern edge of Europe and in the Near East between ~5 and ~3.5 ka, still appears after ~2.5 ka, by which time it had reached the Carpathian Basin in central Europe. Around the same date, we first observed genotype D in two individuals from the Italian Alps, as well as in various locations in the western steppe, before prevailing in large parts of Europe during the Medieval period. Thus, it seems that as most WENBA HBV lineages disappeared by the end of the 2nd millennium BCE, genotypes A and D subsequently spread from eastern reservoirs to eventually reach western regions that had previously only harbored WENBA strains (*22*).

The second half of the 2nd millennium BCE bears witness to major cultural shifts in the archaeological record in western Eurasia, including the sudden disappearance of tell settlements in the Carpathian Basin (*41*), the expansion of the Urnfield culture and the increase of military conflicts in large parts of Europe (*42*–*45*), the breakdown of the Terramare culture in northern Italy (*46*), and the so-called Late Bronze Age collapse of most state societies in the eastern Mediterranean region and Near East (*47*, *48*). Some of these societal transformations could have been triggered by underlying phenomena such as climatic events (*49*) or the spread of epidemic diseases (*50*) and were likely associated with substantial shifts in population densities, transregional networks, and modes and scales of human mobility. The observed decline of WENBA HBV diversity, as well as our phylodynamic reconstruction (fig. S7), further point to major changes in epidemiological dynamics over large parts of western Eurasia during this period. However, although our data suggests that new lineages disseminated across Europe only later on, the lack of observations around 3 ka (Fig. 3) could reflect sampling biases related to the widespread adoption of cremation practices around that time (*42*–*44*) rather than a decrease of HBV prevalence. Searching for the virus in a large number of systematically dated samples across this period could help to better characterize the process that ultimately led to the renewal of western Eurasian HBV diversity after the end of the 2nd millennium BCE.

Recent reemergence of the WENBA HBV lineage

The majority of HBV strains circulating in western Eurasia today belong to genotypes A and D (*3*, *4*), thus only reflecting a relatively recent part of the phylogeographic history of this virus. However, our results show that despite the seemingly complete disappearance of WENBA HBV strains around the end of the 2nd millennium BCE, one lineage descending from this clade has persisted to the present. The latter gave rise to a group of modern strains classified as genotype G (Fig. 2 and fig. S2), a rare, recently described genotype for which the biology is poorly understood (*51*). First discovered in patients from France and the United States, genotype G was later found in other parts of Europe, the Americas, and in Asia, making its geographic origin unclear (*52*). Despite its wide distribution, genotype G exhibits remarkably low genetic diversity (*53*), suggesting a recent reemergence after thousands of years of low-level persistence. Furthermore, genotype G has mostly been found in HIV-positive patients, and phylodynamic patterns have pointed to a sharp increase of its dissemination co-occurring with the HIV pandemic, possibly associated with highly sexually active groups and injection-drug users (*52*).

Genotype G has sometimes been referred to as "aberrant" because of its distinctive genomic features: a 36-nucelotide insertion near the 5′ end of the core gene and two nonsense mutations in the precore region (*51*, *54*). These changes inhibit production of the immunotolerogen e antigen (HBeAg), which appears to be essential for the establishment of a persistent HBV infection, and alter the structure of the HBV core protein, which may impair packaging and replication of the viral genetic material (*54*, *55*). This likely explains why in the vast majority of cases genotype G occurs in co-infections with other HBV genotypes, which can provide the HBeAg and core protein production functions lacking in genotype G (*54*–*56*). We identified similar insertions and stop codons in 14 ancient HBV genomes ranging in age between ~7 and 3.5 ka, which form the WENBA subclade from which genotype G descends (fig. S8). Additionally, most of these ancient genomes were found in individuals showing signs of infections with several HBV variants (fig. S8 and data S2) (*22*). Cases of mixed infection were exclusively found in individuals carrying WENBA HBV strains, among which they were very frequent (22 of 83 individuals, likely underestimating the true frequency). In all cases, both major and minor strains appeared to belong to the WENBA lineage, and sequencing data were partially supporting a ~40–base pair insertion at the 5′ end of the core gene (table S3 and data S1).

Therefore, although genotype G is considered rare today, it seems that the cotransmission of its ancestral form together with another $HBeAg^+$ WENBA strain was a common epidemiological feature of HBV between ~7.5 and 3.5 ka. Notably, this functionally limited variant persisted until today, whereas the rest of the WENBA HBV diversity

seemingly went extinct. Virologic studies indicate that genotype G tends to outcompete HBeAg-producing strains during late HBV infection stages after anti-HBeAg seroconversion (*56*–*58*). Although these short-term selection patterns parallel the survival of this lineage over thousands of years, the latter may rather be related to less deterministic factors. One of the closest Bronze Age ancestors of genotype G was recovered at the archaeological site of Shagara in the eastern European forest zone (SGR003) (figs. S1 and S2), a location where the present-day widespread genotype A was already circulating (SGR004). Genotype A is the most common genotype found in mixed infections with genotype G today (*55*, *57*). The discovery of ancestral forms of both genotypes at the same archaeological site, albeit from different individuals and time periods, may indicate that this viral association had already formed during prehistory in eastern Europe.

Conclusions

This study demonstrates the value of large-scale paleogenomic analyses for studying the phylogeographic history of HBV. DNA enrichment allowed us to reconstruct large proportions of more than 100 ancient HBV genomes from a variety of skeletal tissues, opening possibilities for future paleovirologic studies. We show that HBV was already widely present in humans during the early Holocene and that its phylogeographic history reflects several well-known human migrations and demographic events, including the expansion of First American populations in the Americas and the Neolithic transition in Europe, but not others, such as later Bronze Age steppe ancestry expansions. Furthermore, our results reveal patterns that were not expected on the basis of human genetic and archaeological data alone, such as the near complete renewal of western Eurasian HBV diversity around the end of the 2nd millennium BCE. These findings highlight that the reconstruction of ancient viral diversity has great potential to contribute to our understanding of human history.

REFERENCES AND NOTES

1. World Health Organization (WHO), *Global Hepatitis Report* (WHO, 2017).
2. D. Lavanchy, M. Kane, in *Hepatitis B Virus in Human Diseases*, Y.-F. Liaw, F. Zoulim, Eds. (Springer International Publishing, Cham, 2016), *Molecular and Translational Medicine*, pp. 187–203.
3. A. Kramvis, *Intervirology* **57**, 141–150 (2014).
4. S. Velkov, J. J. Ott, U. Protzer, T. Michler, *Genes (Basel)* **9**, 495 (2018).
5. M. Littlejohn, S. Locarnini, L. Yuen, *Cold Spring Harb. Perspect. Med.* **6**, a021360 (2016).
6. S. Locarnini, M. Littlejohn, M. N. Aziz, L. Yuen, *Semin. Cancer Biol.* **23** (6 Pt B), 561–575 (2013).
7. D. Paraskevis *et al*., *Hepatology* **57**, 908–916 (2013).
8. A. Rasche, A.-L. Sander, V. M. Corman, J. F. Drexler, *J. Hepatol.* **70**, 501–520 (2019).
9. P. Simmonds, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 1013–1026 (2001).
10. G. Zehender *et al*., *World J. Gastroenterol.* **20**, 7622–7634 (2014).
11. G. Kahila Bar-Gal *et al*., *Hepatology* **56**, 1671–1680 (2012).
12. R. Barquera *et al*., *Curr. Biol.* **30**, 2078–2091.e11 (2020).
13. B. Krause-Kyora *et al*., *eLife* **7**, e36666 (2018).
14. B. Mühlemann *et al*., *Nature* **557**, 418–423 (2018).
15. J. Neukamm *et al*., *BMC Biol.* **18**, 108 (2020).
16. Z. Patterson Ross *et al*., *PLOS Pathog.* **14**, e1006750 (2018).
17. P. L. Bollyky, E. C. Holmes, *J. Mol. Evol.* **49**, 130–141 (1999).
18. B. A. Godoy, J. R. R. Pinho, N. J. R. Fagundes, *Virus Res.* **276**, 197776 (2020).
19. D. Paraskevis *et al*., *Mol. Phylogenet. Evol.* **93**, 44–54 (2015).
20. L. K. W. Yuen *et al*., *Mol. Biol. Evol.* **36**, 942–954 (2019).
21. R. Bouckaert, M. V. Alvarado-Mora, J. R. Pinho, *Antivir. Ther.* **18** (3 Pt B), 497–503 (2013).
22. Materials and methods are available as supplementary materials.
23. D. Palacios *et al*., *Earth Sci. Rev.* **203**, 103113 (2020).
24. M. R. Waters, *Science* **365**, eaat5447 (2019).
25. B. Llamas *et al*., *Sci. Adv.* **2**, e1501385 (2016).
26. J. V. Moreno-Mayar *et al*., *Nature* **553**, 203–207 (2018).
27. M. Raghavan *et al*., *Science* **349**, aab3884 (2015).
28. B. F. de Carvalho Dominguez Souza *et al*., *J. Hepatol.* **68**, 1114–1122 (2018).
29. M. Feldman *et al*., *Nat. Commun.* **10**, 1218 (2019).
30. Q. Fu *et al*., *Nature* **534**, 200–205 (2016).
31. I. Mathieson *et al*., *Nature* **555**, 197–203 (2018).
32. R. Barrett, C. W. Kuzawa, T. McDade, G. J. Armelagos, *Annu. Rev. Anthropol.* **27**, 247–271 (1998).
33. J. Diamond, *Nature* **418**, 700–707 (2002).
34. F. M. Key *et al*., *Nat. Ecol. Evol.* **4**, 324–333 (2020).
35. M. C. Inhorn, P. J. Brown, *Annu. Rev. Anthropol.* **19**, 89–117 (1990).
36. N. D. Wolfe, C. P. Dunavan, J. Diamond, *Nature* **447**, 279–283 (2007).
37. W. Haak *et al*., *Nature* **522**, 207–211 (2015).
38. I. Lazaridis *et al*., *Nature* **513**, 409–413 (2014).
39. A. Deter-Wolf, B. Robitaille, L. Krutak, S. Galliot, *J. Archaeol. Sci. Rep.* **5**, 19–24 (2016).
40. K. W. Alt *et al*., *Sci. Rep.* **10**, 2131 (2020).
41. F. Gogâltan, *Studia Hercynia* **23**, 198–214 (2019).
42. G. Capuzzo, J. A. Barceló, *World Archaeol.* **47**, 622–641 (2015).
43. F. Falkenstein, in *Ancestral Landscape. Burial mounds in the Copper and Bronze Ages (Central and Eastern Europe–Balkans–Adriatic–Aegean, 4th–2nd millennium B.C.)* Proceedings of the International Conference held in Udine, 15 to 18 May 2008 (Maison de l'Orient et de la Méditerranée Jean Pouilloux, 2012), Travaux de la Maison de l'Orient et de la Méditerranée. Série recherches archéologiques, pp. 329–340.
44. H. Fokkens, *Antiquity* **71**, 360–373 (1997).
45. S. Hansen, in *Hillforts and Weaponry in the Early and Middle Bronze Age*, S. Hansen, R. Krause, Eds. (Habelt, 2019).
46. A. Cardarelli, in *Scienze dell'Antichità* (Edizioni Quasar, 2009), vol. 15, pp. 449–520.
47. E. H. Cline, *1177 B.C.: The Year Civilization Collapsed* (Princeton Univ. Press, 2015).
48. J. Driessen *et al*., *An Archaeology of Forced Migration. Crisis-Induced Mobility and the Collapse of the 13th c. BCE Eastern Mediterranean* (PLU, 2018).
49. D. Kaniewski, E. V. Campo, in *3.2 ka BP Megadrought and the Late Bronze Age Collapse*, H. Weiss, Ed. (Oxford Univ. Press, 2017).
50. P. Norrie, in *A History of Disease in Ancient Times: More Lethal than War*, P. Norrie, Ed. (Springer International Publishing, 2016), pp. 61–101.
51. L. Stuyver *et al*., *J. Gen. Virol.* **81**, 67–74 (2000).
52. J. M. Wolf, S. De Carli, V. R. Z. B. Pereira, D. Simon, V. R. Lunge, *J. Viral Hepat.* **28**, 393–399 (2021).
53. M. Cornelissen *et al*., *BMC Infect. Dis.* **16**, 268 (2016).
54. K. Li *et al*., *J. Virol.* **81**, 9202–9215 (2007).
55. T. Sakamoto *et al*., *J. Viral Hepat.* **20**, e27–e36 (2013).
56. Y. Tanaka *et al*., *Virology* **376**, 408–415 (2008).
57. H. Kato *et al*., *Hepatology* **35**, 922–929 (2002).
58. M. Sugiyama *et al*., *Hepatology* **45**, 929–937 (2007).

ACKNOWLEDGMENTS

We thank M. Harbeck, L. Bondioli, R. Risch, and H. Duday for their contribution to the archaeological interpretation; R. Heyne and H. Heyne for comments concerning clinical manifestation of hepatitis B infections; B. Mühleman and T. C. Jones for helpful discussion on genome assemblies and mixed infections; and M. O'Reilly for her help on the design of the figures. **Funding:** The research was funded by the Max Planck Society, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (771234–PALEoRIDER, to W.H.; 805268–CoDisEASe to K. Bos; 834616–ARCHCAUCASUS to S.H.), the Slovak Academy of Sciences and the European Union's Seventh Framework Programme and Marie Curie Actions under the Programme SASPRO (1340/03/03 to P.C.R.), the ERA.NET RUS Plus–S&T programm of the European Union's Seventh Framework Programme (277–BIOARCCAUCASUS to S.Re. and S.H.), the Werner Siemens Stiftung ("Paleobiochemistry", to CW), the Award Praemium Academiae of the Czech Academy of Sciences (to M.E.), the Institute of Archaeology of the Czech Academy of Sciences (RVO 67985912, to M.Dobe.), the Russian Foundation for Basic Research (19-09-00354a, to M.K.K. and V.V.K.; 19-78-10053 to SSh), the German Research Foundation (DFG-HA-5407/4-1–INTERACT to W.H. and RE2688/2 to S.Re.), the French National Research Agency (ANR-17-FRAL-0010–INTERACT, to M.F.D., M.Ri., S.Ro., S.Sai., D.Bi., and P.Le.), the Wenner-Gren Dissertation Fieldwork Grant (9558 to S.Sab.), and the Ministry of Education and Science of the Republic of Kazakhstan (AP08856654 to L.B.D., L.M., and E.Kh. and AP08857177 to A.Z.B.). **Author contributions:** D.K. and J.K. initiated and supervised the study. S.F., V.Sc., V.K., M.K.K., M.S.C., V.V.K., A.A.Kh., A.A.C., D.A.S., A.F.K., C.T.R., Í.G.-M.d.L., H.A.M., R.G.P., J.I.R.G., J.N., S.Ro., S.K., S.Sai., E.Ka., A.B.B., P.V., P.Li., M.Ko., L.Lo., E.P., R.C., A.L., R.M., Y.C.d.A., S.T.H.G., D.I.H.Z., J.P., D.Bi., P.Le., A.R.K., V.E.M., L.La., M.Z., J.F.B., M.L., A.D., T.I., G.G.A., M.P.d.M.I., A.R., A.Sp., S.B., S.Sal., E.D.Z., D.V.V., K.v.H., R.L.B., L.C.S., L.A., M.N., E.R., C.A.F., V.Sl., A.A.Ka., B.C.A., E.B., M.A.C., M.S., R.K., J.J.E., M.Fr., S.Sh., P.d.K., E.A., K.V.d.V., L.F.S., T.A.T., S.L., M.Dobr., N.M., C.R., M.V.T., C.S., P.C.R., M.A., K.A.Y., E.C.B., F.C., V.Ma., P.U., K.R., D.Ba., P.S., L.M.M., M.Ro., H.S., D.C.S.G., N.S., Y.S.E., F.H., Y.B., K.Boy., M.Kü., D.S., P.O., R.Sk., M.R.G., A.B., L.B.D., A.Z.B., Z.S., K.Mas., M.Ma., V.Mo., K.Man., S.Re., S.H., E.P.K., M.Dobe., M.E., H.M., and K.W.A. conducted the archaeological/anthropological work or the curation of the archeological material. J.K., W.H., C.Po., K.Bos., P.W.S., S.Sc., C.W., M.F.D., O.B., E.Kh., E.F.D., B.L., A.Sa., E.B., E.K.G., L.M., and M.Ri. organized the sample collection for genetic analyses. R.B., L.P., M.A.S., F.A., R.St., A.W., A.M., V.V.M., G.U.N., M.Ri., M.S.v.d.L., K.Maj., R.I.T., L.M., A.G., S.P., S.Sab., M.Mi., J.G., E.A.N., T.F., K.N., C.Pa., M.Ke., E.K.G., M.Fe., S.E., E.S., K.G., G.A.G.R., D.I.H.Z., B.L., E.F.D., O.B., M.F.D., and W.H. participated in the laboratory work, sample management, and gathering of contextual information. A.K., F.M.K., M.A.S., and R.H. performed HBV screening of the sequencing data, which was supervised by A.H.; A.K. and K.P. performed the bioinformatic analyses. A.K. and D.K. performed the phylogenetic analyses. L.P. performed the human population genetics analyses. A.K. and A.B.R. performed the statistical analyses. All authors contributed to the interpretation of the data. A.K. wrote the first version of the manuscript, which was edited by D.K., J.K., W.H., A.H., C.Po., K.Bos., P.W.S., S.Sc., C.W., F.v.B., M.P., L.P., R.B., F.M.K., and M.A.S., and all authors contributed to its improvement. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The sequencing data generated in this study has been deposited at the European Nucleotide Archive (www.ebi.ac.uk/ena/data/view/PRJEB45699). Complete HBV sequence alignments are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi5658
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S5
References (*59*–*254*)
Data S1 to S4
MDAR Reproducibility Checklist

18 March 2021; accepted 17 August 2021
10.1126/science.abi5658

REPORTS

METALLURGY

Grain boundary velocity and curvature are not correlated in Ni polycrystals

Aditi Bhattacharya[1], Yu-Feng Shen[2]†, Christopher M. Hefferan[2]‡, Shiu Fai Li[2]§, Jonathan Lind[2]¶, Robert M. Suter[2], Carl E. Krill III[3], Gregory S. Rohrer[1]*

Grain boundary velocity has been believed to be correlated to curvature, and this is an important relationship for modeling how polycrystalline materials coarsen during annealing. We determined the velocities and curvatures of approximately 52,000 grain boundaries in a nickel polycrystal using three-dimensional orientation maps measured by high-energy diffraction microscopy before and after annealing at 800°C. Unexpectedly, the grain boundary velocities and curvatures were uncorrelated. Instead, we found strong correlations between the boundary velocity and the five macroscopic parameters that specify grain boundary crystallography. The sensitivity of the velocity to grain boundary crystallography might be the result of defect-mediated grain boundary migration or the anisotropy of the grain boundary energy. The absence of a correlation between velocity and curvature likely results from the constraints imposed by the grain boundary network and implies the need for a new model for grain boundary migration.

During the high-temperature processing of polycrystalline materials, grain boundary migration is responsible for creating the final microstructure and therefore influences many of a material's macroscopic properties. This process occurs in a wide range of technologically important materials. For example, the internal configuration of grain boundaries influences the strength and electronic conductivity of polycrystalline Cu within integrated circuits (*1*) and the toughness of high-strength automotive steels (*2*). The arrangement of grain boundaries also has a profound effect on the performance of functional materials used in, for example, dielectric (*3*), thermoelectric (*4*), and photovoltaic (*5*) devices. In the field of geology, models for grain boundary migration are important for understanding the properties of rocks in the Earth's mantle (*6*) and for deriving information about the paleoclimate from measurements of the microstructure of large ice sheets (*7*).

In cases where the energy stored in intragranular defects is minimal and boundary curvature is the most substantial driving force for grain growth, it is assumed that a grain boundary migrates toward its center of curvature at a velocity (v) that is directly proportional to the product of the grain boundary mean curvature (κ) and the grain boundary energy (γ), such that

$$v = M\kappa\gamma \quad (1)$$

where M is a constant of proportionality referred to as the mobility (*8*). The linear relationship between grain boundary velocity and curvature has been verified by measurements (*9*) and simulations (*10*) of grain boundary migration in bicrystals. The simple linear relation also describes the average behavior of polycrystals if one accepts the average grain growth rate as a proxy for grain boundary velocity and the inverse mean grain size as a proxy for curvature (*11*). Although the simple expression in Eq. 1 has been widely used to interpret simulations and experiments for nearly 70 years, mainly based on its validation in bicrystal studies, there is a range of evidence that neither the mobility nor the driving force ($\kappa\gamma$) are constants; instead, both are sensitive to the boundary structure. For example, Gottstein *et al.* (*9*) showed experimentally, and Olmsted *et al.* (*12*) showed using atomistic simulations, that M depends on boundary crystallography. In addition, Frolov (*13*) showed that M can vary with grain boundary complexion (structure), even when the macroscopic grain boundary degrees of freedom are constant. Similarly, γ varies with the grain boundary crystallography (*14*) and, because of this, the driving force is more complex. The actual driving force is proportional to the grain boundary stiffness, which is the sum of the boundary energy and the second derivative of that energy with respect to orientation in the directions of the principal curvatures (*15*, *16*).

Despite its widespread acceptance, it remains an open question whether grain boundaries in polycrystals migrate at rates proportional to their curvatures. Boundaries in polycrystals differ from those in a bicrystal in an important way: Instead of being able to move freely, they are connected (on average) through five triple lines to 10 other grain boundaries (*17*). Therefore, a grain boundary is unable to reduce its area without concomitant changes in the positions and areas of the connected boundaries, but these changes will not necessarily reduce the total energy. These mutual dependencies might explain how a critical population of slow-moving boundaries is able to bring about

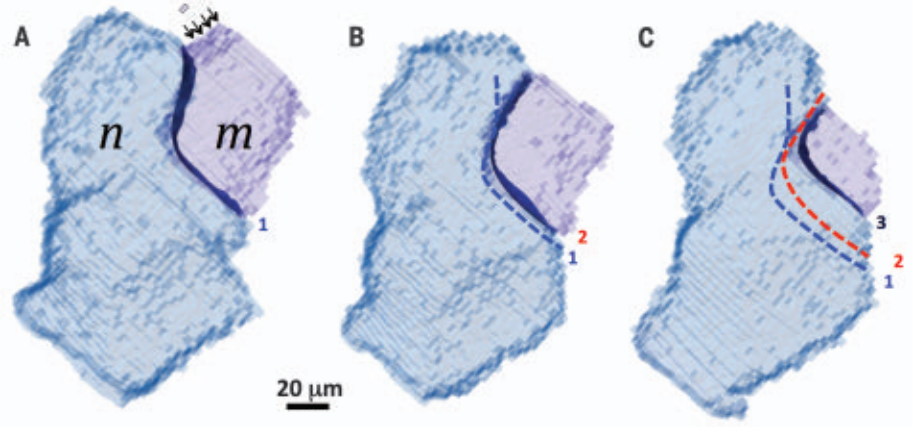


Fig. 1. Grain boundary migration. Two grains (n and m), showing individual voxels, at three different time steps. The boundary separating the grains is denoted by the dark blue shading and labeled "1", "2", and "3" in (A), (B), and (C), respectively. As the boundary between the two grains moves, some voxels that belong to grain m in (**A**) become part of grain n in (**B**) and (**C**). The four arrows at the top of (A) indicate four layers of voxels, and the small rectangle shows the projected area of a single 2.3 × 2.3 × 4.0 μm voxel. The interface position in (A) (labeled "1") is superposed on (B) and (C). The interface position in (B) (labeled "2") is superposed on (C).

[1]Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [2]Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [3]Institute of Functional Nanosystems, Ulm University, 89081 Ulm, Germany.
*Corresponding author. Email: gr20@andrew.cmu.edu
†Present address: Morgan Stanley, New York, NY 10036, USA.
‡Present address: RJ Lee Group, Monroeville, PA 15146, USA.
§Present address: Blue River Technology, Sunnyvale, CA 94086, USA.
¶Present address: Materials Engineering Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA.

grain growth stagnation in the presence of curvature driving forces (*18*) or why boundary mobility measurements based on the assumed validity of Eq. 1 have led to inconsistent results for crystallographically identical boundaries (*19*).

The technical reason that grain boundary velocities have yet to be measured within polycrystalline materials is because there was no way until recently to measure the time dependence of the shapes and volumes of crystals embedded in an opaque material. The development of near-field high-energy diffraction microscopy (HEDM) during the past decade has made this possible (*20*). HEDM and related techniques have been used to study changes that occur during the annealing of polycrystals, including twin formation (*21*), coarsening (*22*), and grain growth (*19*, *23*–*25*). We introduce a simple method to determine grain boundary velocities by measuring the microstructure at different time intervals. Using the three-dimensional microstructure of Ni imaged six times during annealing at 800°C, we measured the grain boundary curvatures and velocities of more than 5×10^4 grain faces of known crystallography (*26*). From these data, we assessed the validity of Eq. 1 for boundaries in polycrystals and determined how the velocities of grain boundaries depend on their crystallographic characteristics. The absence of a correlation between velocity and curvature has implications for modeling microstructure evolution.

The Ni specimen was annealed at 950°C for 6 hours before the grain growth experiment at 800°C to achieve a minimal and constant dislocation density that would not affect the driving force for grain growth in a measurable way (*26*). We measured the grain boundary velocity using a method that we explain with reference to Fig. 1 (and show schematically in fig. S1). The boundary in Fig. 1 is changing, not only its position, but also its shape and area, which made it impossible to track unit areas with different orientations. As a result, we determined the velocities from the volume exchanged between the grains, a method that provides an average velocity for the entire boundary and ameliorates the problems caused by the changing boundary shape (*26*). Each grain in the microstructure is assigned an identification (ID) number. For the boundary between the grain with ID = n and ID = m, any voxel that was part of grain $n(m)$ in the first time step and changed to grain $m(n)$ in the second was counted. This count provides a measure of the volume exchanged from n to m ($V_{n\to m}$) and from m to n ($V_{m\to n}$) in a certain time period. When volume is divided by the average area of the boundary between n and m in the two time steps ($\bar{A}_{n|m}$), and the time interval (Δt), the result has the dimensions of velocity (v)

$$v = \left| \frac{V_{n\to m} + V_{m\to n}}{\bar{A}_{n|m} \times \Delta t} \right| \qquad (2)$$

The values used in Eq. 2 for the case illustrated by Fig. 1 can be found in table S1. Like all grain boundary properties, v is unsigned because it must obey grain exchange symmetry (*26*). Once completed, our analysis encompassed 51,794 grain faces. The distribution of velocities (fig. S1A) had a mean of 0.1 μm/min and an SD of the same order. The distribution of curvatures (fig. S1B) had a mean of 0.05 μm^{-1}, and the SD was 0.08 μm^{-1}. The magnitude of the mean corresponds to grains having a mean radius of 20 μm, consistent with the average grain size in this microstructure. All boundaries are represented by a set of triangular mesh elements, and we determined the orientation, area, curvature, and velocity for each element.

We illustrate the relationship between the grain face–averaged grain boundary velocity and curvature with scatter plots (Fig. 2, A and B). When the data are viewed throughout the entire domain (Fig. 2A), the majority falls near the origin, with no apparent correlation between the quantities. A closer look at the domain near the origin, containing 84% of the data, confirms the lack of correlation (Fig. 2B). To examine average velocities of boundaries with similar curvature, we binned the boundaries into curvature intervals of 0.003 μm^{-1} (Fig. 2C). We again found no obvious correlation. A linear fit to the mean values yields a slope of −0.025 and a correlation coefficient of only 0.17. These data provide no evidence for a positive correlation between the curvature and velocity, as required by Eq. 1.

We also considered variations of the velocity with crystallographic parameters. We found the most substantial influence of the grain boundary crystallographic parameters on the velocity when we specified both the misorientation and the grain boundary plane orientation. For example, the mean velocities of [111] twist boundaries roughly doubles for boundaries with an ~40° twist angle (Fig. 3A). Similarly, the velocities of all symmetric tilt grain boundaries (STGBs) with rotations about the [100] axis and with $\Sigma \leq 49$, where Σ equals the inverse lattice coincidence (*27*), are relatively constant until tilt angles between 65 and 75°, where the boundary velocities are two to three times greater (Fig. 3B).

We can examine variations in the velocity with grain boundary plane orientation at fixed misorientation, as illustrated for the Σ3 misorientation (60° around [111]) (Fig. 4A). The coherent twin boundary at the (111) orientation has the minimum velocity, and tilt boundaries 90° from [111] have the maximum velocities. This observation is consistent with our expectations for Σ3 boundaries in face-centered cubic metals (*12*, *28*). Our data also show that the coherent twin has the minimum curvature (Fig. 4B), consistent with the tendency of

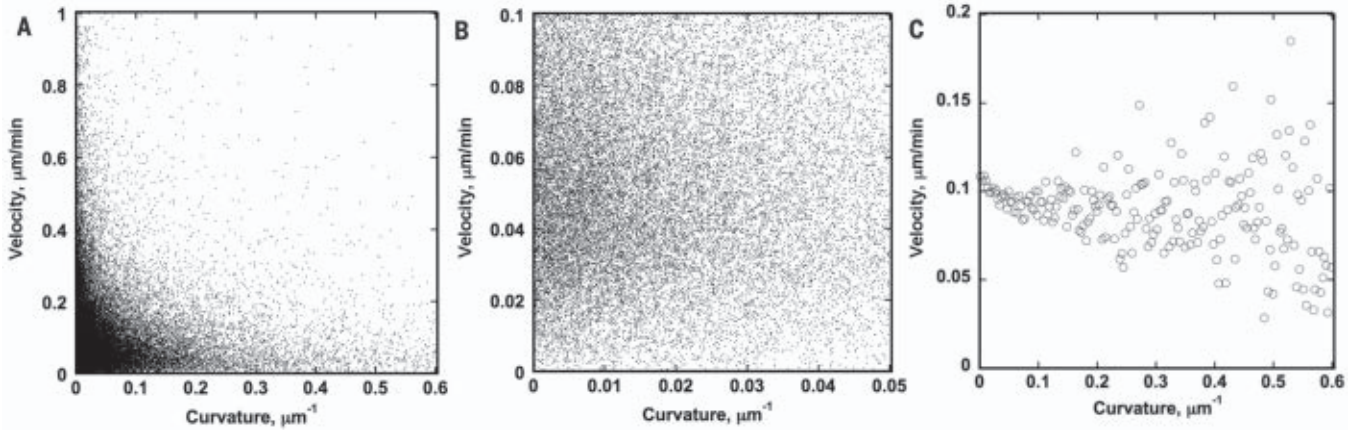


Fig. 2. Grain boundary velocity as a function of curvature. (**A**) Scatter plot showing the entire domain of the data. (**B**) Magnified view illustrating 84% of the data. (**C**) Mean velocities of boundaries classified into curvature groups with widths of 0.003 μm^{-1}.

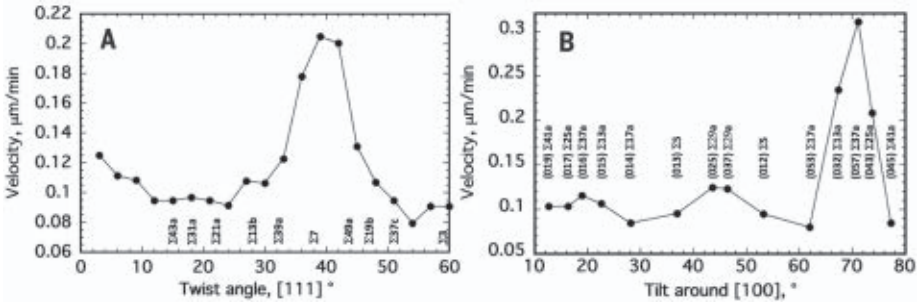


Fig. 3. Mean velocities as a function of crystallographic parameters. (**A**) Twist grain boundaries for rotations about the [111] axis in 3° intervals. These are boundaries with planes (111)||$(\bar{1}\bar{1}\bar{1})$. For reference, the positions of coincident site lattice (*27*) boundaries with $\Sigma \leq 49$ are indicated. (**B**) STGBs for all coincident site lattice boundaries with $\Sigma \leq 49$ and rotations about the [100] axis. The symmetric plane is listed for each boundary.

twin boundaries to be flat, as documented, for example, by curvature measurements of twins in an austenitic steel (*29*). The coherent twin is the most common boundary in Ni, which is reflected in its large relative area of 600 multiples of a random distribution (Fig. 4C). Finally, when our measured velocities are compared with previously computed grain boundary mobilities (*12*), we found that the twin boundaries had both low mobility and low velocity, with the higher mobility boundaries 90° away. However, the $\{\bar{2}11\}$ type boundaries with the maximum velocity (Fig. 4A) are predicted (Fig. 4D) to have low mobilities. Therefore, the velocities that we observed are not well correlated with the computed mobilities even if we divide the velocity by the curvature, which is relatively constant along the zone of tilt boundaries.

We also compared the velocity, curvature, population, and calculated mobility of the Σ7 boundaries (38.2° around [111]) (Fig. 4, E to H) because they were among the highest-velocity boundaries. The maximum velocities are in a broad peak near the orientation of the (111) twist grain boundary, where there is also a maximum in population (Fig. 4G) that can be attributed to a low grain boundary energy (*14*). All other orientations have smaller velocities. The curvature distribution (Fig. 4F) is nearly opposite to the velocity, with a local minimum in curvature at the twist position and maximum curvature along the positions of the tilt boundaries. The computed mobility distribution (Fig. 4H) contradicts the velocity distribution, and the differences cannot be explained by the observed curvature, which reinforces the differences. Although the Σ3 boundary had the greatest area among all grain boundary types, the area of the Σ7 is more typical of all the other boundaries. We determined the Σ7 distribution from the evolution of 252 individual grain faces. Examples of some additional boundaries, consistent with the data for the Σ7 boundary, can be found in figs. S6 and S7.

Finding literature data that can be used to make an exact comparison with our observed velocities of grain boundaries in a network is challenging. The closest existing data are those produced by molecular dynamics simulations of Ni at 1200 K using an embedded atom potential (*12*). The calculations do not compare well to our observations (Figs. 4, D and H). The largest difference between them is that the simulations are of bicrystals, so the boundary could move without the cooperative motion of attached boundaries. The high velocity that we observed for boundaries in the vicinity of 40° for [111] misorientations (Figs. 3A and 4E) is reminiscent of previous results for grain boundary migration in Al bicrystals (*30, 31*). However, in this previous work, the high-velocity boundaries were cited to be of tilt character, whereas our high-velocity boundaries have twist character.

Overall, we found little overlap between our observations of boundary migration in the polycrystal compared with experiments and simulations of bicrystals. This might not be too surprising considering the connectivity of the grain boundary network. Although a boundary in a bicrystal clearly moves toward its center of curvature to reduce the interface free energy, a network of grain boundaries must evolve to lower the global energy of the network, which might not be compatible with all boundaries moving toward their centers of curvature. When a grain boundary migrates, it is also potentially changing the areas and orientations of all of the boundaries to which it is connected. These boundaries will have different energies, and those energies are affected not only by the change in area, but also by changes in the grain boundary orientation (*14*, *32*). The success of classical grain growth theories can be attributed to these details being averaged out, with the evolution of an ensemble of grains being subject only to the constraint of reducing the total grain boundary energy rather than requiring each boundary to move toward its center of curvature. An example of this behavior is the formation and growth of corner twins in Ni, which increases the total grain boundary area while simultaneously reducing the grain boundary energy (*21*). Similarly, it has also been shown that the migration of a boundary in recrystallizing Al is more dependent on the boundary plane inclination than on the stored energy driving force (*33*).

One clear result from our observations is the dependence of the velocity on the five crystallographic grain boundary parameters. This finding suggests a strong dependence on grain boundary structure and would be a natural consequence of grain boundary migration mediated by disconnections (*34*–*36*). According to this theory, the rate of boundary migration remains proportional to curvature in a bicrystal but is mechanistically determined by the rate of disconnection nucleation and glide along the boundary. Each boundary has been shown to have a spectrum of disconnection nucleation energies (*36*). When the energy for disconnection nucleation is small and there are numerous modes that, acting together, can advance the boundary without accumulating large strains at the triple lines, boundaries are expected to have higher velocities than in cases where more energy for nucleation is required and there are fewer available modes. A detailed comparison between our results and disconnection mode spectra might lead to a better understanding of the relationship between grain boundary structure and velocity.

The influence of grain boundary crystallography on the observed velocities indicates that the driving force for grain boundary migration is more complex than depicted by Eq. 1. Because the grain boundary energies of Ni are anisotropic (*14*, *32*), the microstructure is expected to contain a mixture of flat singular boundaries (orientations at a deep cusp in the grain boundary energy) and smoothly curved, nonsingular boundaries. For singular boundaries, the driving force for migration is proportional to the weighted mean curvature, a quantity that depends more on the energies of the connected interfaces than on those of the moving interface (*37*). In this case, we do not expect a positive correlation between a grain boundary's velocity and its curvature, because Eq. 1 does not apply. Even for nonsingular boundaries, differences between the grain boundary stiffness and the energy will degrade the predicted correlation between curvature and velocity (*15*).

Considering that the boundary properties of Ni, Cu, Al, and Au are known to be highly correlated (*38*), our observations may well also apply to other technologically important metals with the face-centered cubic structure. The discovery that grain boundary velocities

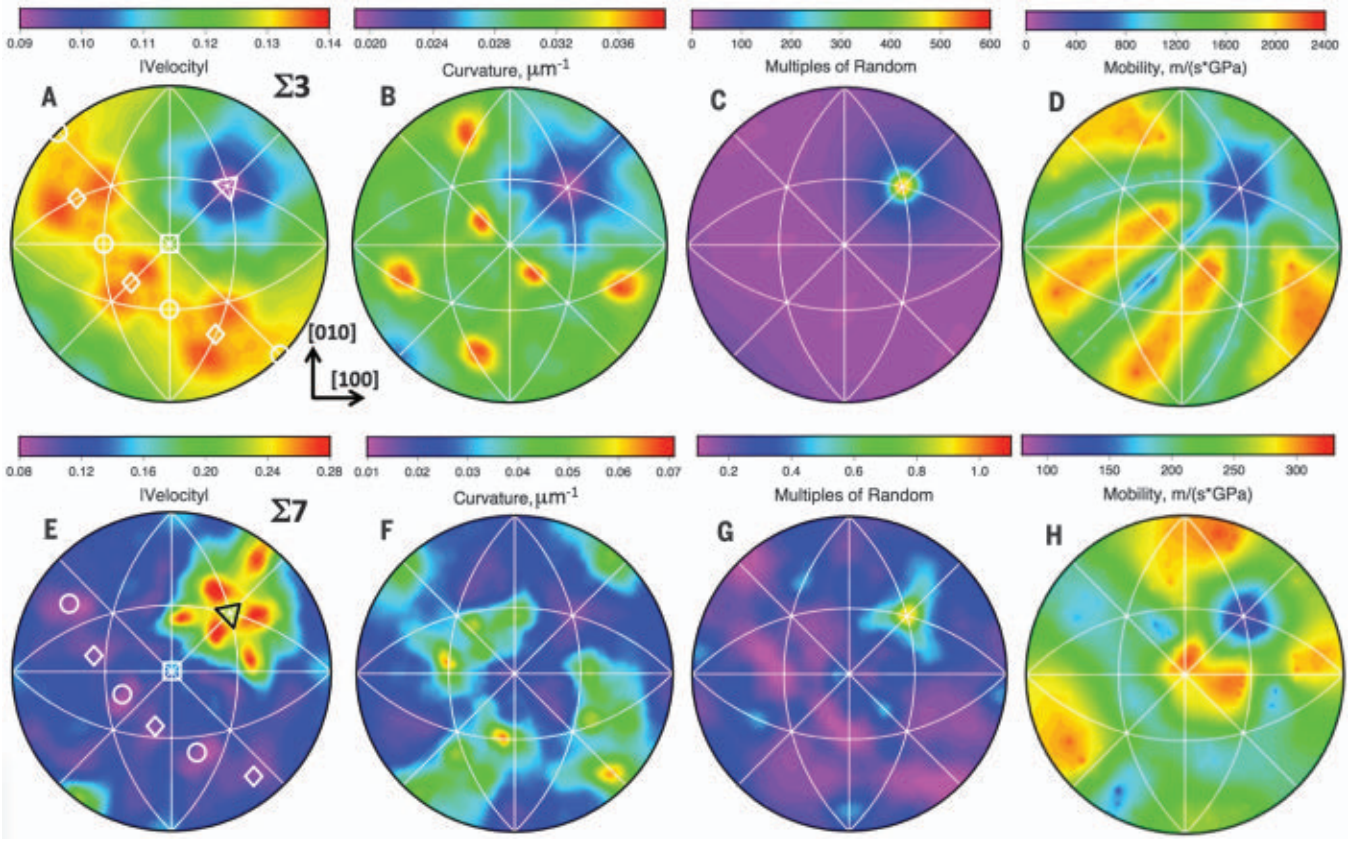


Fig. 4. Selected properties of Σ3 and Σ7 grain boundaries. The properties of Σ3 (**A** to **D**) and Σ7 (**E** to **H**) grain boundaries plotted on stereographic projections along [001]. The velocity [(A) and (E)] in micrometers per minute, curvature [(B) and (F)], and relative areas [(C) and (G)] are from the current data. The grain boundary mobilities [(D) and (H)] are from Olmsted *et al.* (*12*) and were determined from molecular dynamics simulations at 1200 K and a driving force of 25 meV. In (A) and (E), the [001] direction is at the position of the white square and the [111] misorientation axis is at the position of the triangle. In (A), the three white circles mark the orientations of the $\{\bar{1}10\}$ improper quasisymmetric tilt boundaries and the three white diamonds mark the orientations of the $\{\bar{2}11\}$ STGBs. In (E), the three white circles mark the orientations of the $\{\bar{3}21\}$ STGBs and the three white diamonds mark the orientations of the $\{\bar{5}14\}$ improper quasisymmetric tilt boundaries.

are not correlated to mean curvatures suggests that, for polycrystals, the traditional fundamental assumption about grain boundary migration is incorrect. Models based on this assumption are therefore potentially flawed. This might explain the failure of such models to predict microstructure-level details in the few cases in which grain-by-grain comparisons have been performed between observations and simulations (*24*, *39*). As an example of the limits of such models, curvature-driven grain growth simulations continue to evolve until all interfaces have been eliminated, a phenomenon that almost never occurs during real grain growth. More realistic models, which incorporate the influence of very slow-moving grain boundaries (*18*) or the effects of shear-coupled motion (*40*), are able to reproduce the stagnation of grain size that occurs almost universally in isothermal growth experiments. Our findings suggest that a new model for grain boundary migration is needed to predict microstructure evolution. A necessary condition for the new model will be that it includes grain boundary crystallography, which is shown here to influence grain boundary migration kinetics. The disconnection-mediated grain boundary migration theory meets this condition (*36*).

REFERENCES AND NOTES

1. K. Lu, L. Lu, S. Suresh, *Science* **324**, 349–352 (2009).
2. L. Liu *et al.*, *Science* **368**, 1347–1352 (2020).
3. C. A. Randall, N. Kim, J. P. Kucera, W. W. Cao, T. R. Shrout, *J. Am. Ceram. Soc.* **81**, 677–688 (1998).
4. S. I. Kim *et al.*, *Science* **348**, 109–114 (2015).
5. D. W. de Quilettes *et al.*, *Science* **348**, 683–686 (2015).
6. V. S. Solomatov, R. El-Khozondar, V. Tikare, *Phys. Earth Planet. Inter.* **129**, 265–282 (2002).
7. G. Durand *et al.*, *J. Geophys. Res. Earth Surf.* **111**, F01015 (2006).
8. P. A. Beck, in *Metal Interfaces* (American Society for Testing Materials, 1952), pp. 208–247.
9. G. Gottstein, D. A. Molodov, L. S. Shvindlerman, D. J. Srolovitz, M. Winning, *Curr. Opin. Solid State Mater. Sci.* **5**, 9–14 (2001).
10. M. Upmanyu, R. W. Smith, D. J. Srolovitz, *Inter. Sci.* **6**, 41–58 (1998).
11. J. D. Powers, A. M. Glaeser, *Interface Sci.* **6**, 23–39 (1998).
12. D. L. Olmsted, E. A. Holm, S. M. Foiles, *Acta Mater.* **57**, 3704–3713 (2009).
13. T. Frolov, *Appl. Phys. Lett.* **104**, 211905 (2014).
14. J. Li, S. J. Dillon, G. S. Rohrer, *Acta Mater.* **57**, 4304–4311 (2009).
15. F. Abdeljawad *et al.*, *Acta Mater.* **158**, 440–453 (2018).
16. C. Herring, in *Structure and Properties of Solid Surfaces: Proceedings of a Conference Arranged by the National Research Council and Held in September, 1952, in Lake Geneva, Wisconsin, USA.*, R. Gomer, C. S. Smith, Eds. (University of Chicago Press, 1953), pp. 5–81.
17. W. M. Williams, C. S. Smith, *J. Met.* **4**, 755–765 (1952).
18. E. A. Holm, S. M. Foiles, *Science* **328**, 1138–1141 (2010).
19. J. Zhang *et al.*, *Acta Mater.* **191**, 211–220 (2020).
20. J. V. Bernier, R. M. Suter, A. D. Rollett, J. D. Almer, *Annu. Rev. Mater. Res.* **50**, 395–436 (2020).
21. B. Lin *et al.*, *Acta Mater.* **99**, 63–68 (2015).
22. J. M. Dake *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5998–E6006 (2016).
23. A. Bhattacharya *et al.*, *Acta Mater.* **167**, 40–50 (2019).
24. I. M. McKenna, S. O. Poulsen, E. M. Lauridsen, W. Ludwig, P. W. Voorhees, *Acta Mater.* **78**, 125–134 (2014).
25. Y. F. Shen *et al.*, *Phys. Rev. Mater.* **3**, 063611 (2019).
26. Materials and methods are available as supplementary materials.
27. B. L. Adams, J. Zhao, H. Grimmer, *Acta Crystallogr. A* **46**, 620–622 (1990).
28. K. G. F. Janssens *et al.*, *Nat. Mater.* **5**, 124–127 (2006).
29. X. Zhong, D. J. Rowenhorst, H. Beladi, G. S. Rohrer, *Acta Mater.* **123**, 136–145 (2017).
30. Y. Huang, F. J. Humphreys, *Acta Mater.* **47**, 2259–2268 (1999).
31. D. A. Molodov, U. Czubayko, G. Gottstein, L. S. Shvindlerman, *Scr. Metall. Mater.* **32**, 529–534 (1995).

32. D. L. Olmsted, S. M. Foiles, E. A. Holm, *Acta Mater.* **57**, 3694–3703 (2009).
33. C. Hong *et al.*, *Scr. Mater.* **205**, 114187 (2021).
34. K. Chen, J. Han, D. J. Srolovitz, *Acta Mater.* **194**, 412–421 (2020).
35. K. Chen, D. J. Srolovitz, J. Han, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33077–33083 (2020).
36. J. Han, S. L. Thomas, D. J. Srolovitz, *Prog. Mater. Sci.* **98**, 386–476 (2018).
37. J. E. Taylor, *Acta Metall. Mater.* **40**, 1475–1485 (1992).
38. E. A. Holm, D. L. Olmsted, S. M. Foiles, *Scr. Mater.* **63**, 905–908 (2010).
39. M. C. Demirel, A. P. Kuprat, D. C. George, A. D. Rollett, *Phys. Rev. Lett.* **90**, 016106 (2003).
40. E. Rabkin, D. J. Srolovitz, *Scr. Mater.* **180**, 83–87 (2020).
41. Data and code for: A. Bhattacharya, Y.-F. Shen, C. M. Hefferan, S. F. Li, J. Lind, R. M. Suter, C. E. Krill III, G. S. Rohrer, Grain boundary velocity and curvature are not correlated in Ni polycrystals, Grain Boundary Data Archive (2021); http://mimp.materials.cmu.edu/~gr20/Grain_Boundary_Data_Archive.

ACKNOWLEDGMENTS

Funding: This work was supported by the National Science Foundation under grant DMR 1628994 to G.R. and R.M.S. The Advanced Photon Source is a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357 (to R.M.S.).
Author contributions: Conceptualization: G.S.R., R.M.S., C.E.K.; Funding acquisition: G.S.R., R.M.S.; Investigation: A.B., Y.-F.S., C.M.H., S.F.L., J.L.; Methodology: A.B., Y.-F.S., C.M.H., S.F.L., J.L., R.M.S., C.E.K., G.S.R.; Supervision: G.S.R., R.M.S.; Project administration: G.S.R., R.M.S.; Visualization: A.B., G.S.R.; Writing – original draft: A.B., G.S.R.; Writing – review & editing: A.B., R.M.S., C.E.K., G.S.R. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The data and code used for this study will be made available upon publication at the Grain Boundary Data Archive website (*41*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj3210
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*42*–*45*)

5 May 2021; accepted 3 September 2021
10.1126/science.abj3210

POLYMER SCIENCE

Toughening hydrogels through force-triggered chemical reactions that lengthen polymer strands

Zi Wang[1,2], Xujun Zheng[1], Tetsu Ouchi[1,2], Tatiana B. Kouznetsova[1,2], Haley K. Beech[2,3], Sarah Av-Ron[3], Takahiro Matsuda[4], Brandon H. Bowser[1,2], Shu Wang[1,2], Jeremiah A. Johnson[2,5]*, Julia A. Kalow[2,6]*, Bradley D. Olsen[2,3]*, Jian Ping Gong[2,4,7,8]*, Michael Rubinstein[1,2,7,8,9,10,11]*, Stephen L. Craig[1,2,7]*

The utility and lifetime of materials made from polymer networks, including hydrogels, depend on their capacity to stretch and resist tearing. In gels and elastomers, those mechanical properties are often limited by the covalent chemical structure of the polymer strands between cross-links, which is typically fixed during the material synthesis. We report polymer networks in which the constituent strands lengthen through force-coupled reactions that are triggered as the strands reach their nominal breaking point. In comparison with networks made from analogous control strands, reactive strand extensions of up to 40% lead to hydrogels that stretch 40 to 50% further and exhibit tear energies that are twice as large. The enhancements are synergistic with those provided by double-network architectures and complement other existing toughening strategies.

The physical breakdown of polymeric materials such as elastomers and hydrogels occurs during periods of mechanical strain (*1*, *2*) and limits their useful scope and lifetime. This has resulted in strategies that improve the durability of polymer networks (*3*), including the use of multinetwork architectures (*3*, *4*), reversible cross-linking, noncovalent domain unfolding (*5*), and the addition of fillers (*6*). All of these approaches have in common that the primary, covalent chemical structure of the network remains fixed up until the point that the polymer strand breaks.

Polymer networks that actively remodel their primary, covalent structure during use therefore have the potential to provide property enhancements that complement existing mechanisms. Recent advances in covalent polymer mechanochemistry now enable the synthesis of polymer strands that extend the covalent bonding along their backbone when stretched (*7*), as a result of force-coupled chemical reactions that are triggered at or just before the strands reach their nominal breaking point. Because these reactions are typically associated with high forces (approximately nanonewtons), the amount of energy absorbed at the single-molecular level can be quite large relative to the energy absorbed by a covalent strand in the absence of the reaction-enabled extension (*8*). Existing models of network fracture draw a direct relationship between the energy required to propagate a crack through a rubbery material (the tear energy, T) and the energy required to stretch an individual polymer strand to the point at which it breaks (the chain fracture energy, U) (*9*). We therefore hypothesized that by turning a polymer network into a literal network of responsive chemical reactions, reactive strand extension (RSE) might improve the properties of an existing bulk material.

The RSE concept is shown in Fig. 1. Strand extension is provided by the sodium salts of bicyclo[6.2.0]decane (BCD) mechanophores (fig. S1), which react by means of a force-coupled [2+2] cycloreversion to release stored length (*10*). BCD copolymers are formed through the radical addition copolymerization of the corresponding bicyclic cyclobutene carboxylate monomer **1a**. Mechanophore reactivity was quantified by using single-molecule force spectroscopy (SMFS) implemented through an atomic force microscope (AFM). Copolymerization of **1a** with 2-acrylamido-2-methylpropane sulfonate (AMPS) **2** and acetoacetate methacrylate ***3*** (for strong attachments) provided linear copolymers **P1** with **1a:2:*3*** = 0.22:0.62:0.16 (Fig. 2). At forces of up to ~1500 pN, the SMFS curves are typical of a flexible polymer strand overcoming conformational entropy (~1 to 10 pN) to reach end-to-end distances that approach its fully extended contour length (~10^2 pN). Subsequent stretching is dominated by the deformation of bond angles and bond lengths, as the polymer nears its physical stretching limits before breaking (~10^3 pN). With the BCD mechanophore, however, the force-separation curve shows a reproducible release of stored length of up to ~40% of the initial polymer contour length at 1500 to 2000 pN. This transition is consistent with that expected from the BCD cycloreversion, and the transition force is similar to that reported in other cyclobutane mechanophores (*11*). Preextension curves of **P1** overlap well with those of control strands **P2** made from **1b** (no fused ring) (multiple pulls of **P2** are available in fig. S2), and simulated pulling calculations (fig. S3) confirm that the details of scission for the fused and unfused cyclobutanes are effectively indistinguishable. As the polymer strands reach their nominal stretching limit, the presence of selective mechanophores converts strand scission events into force-coupled covalent rearrangements that allow the strands to survive forces and extensions that would otherwise be catastrophic.

[1]Department of Chemistry, Duke University, Durham, NC, USA. [2]NSF Center for the Chemistry of Molecularly Optimized Networks, Duke University, Durham, NC, USA. [3]Department of Chemical Engineering, Massachussetts Institute of Technology (MIT), Boston, MA, USA. [4]Faculty of Advanced Life Science, Hokkaido University, N21W11, Kita-ku, Sapporo 001-0021, Japan. [5]Department of Chemistry, MIT, Boston, MA, USA. [6]Department of Chemistry, Northwestern University, Evanston, IL, USA. [7]Soft Matter GI-CoRE, Hokkaido University, N21W11, Kita-ku, Sapporo 001-0021, Japan. [8]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, N21W11, Kita-ku, Sapporo 001-0021, Japan. [9]Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA. [10]Department of Biomedical Engineering, Duke University, Durham, NC, USA. [11]Department of Physics, Duke University, Durham, NC, USA.
***Corresponding author. Email: jkalow@northwestern.edu (J.A.K.); jaj2109@mit.edu (J.A.J.); bdolsen@mit.edu (B.D.O.); gong@sci.hokudai.ac.jp (J.P.G.); michael.rubinstein@duke.edu (M.R.); stephen.craig@duke.edu (S.L.C.)**

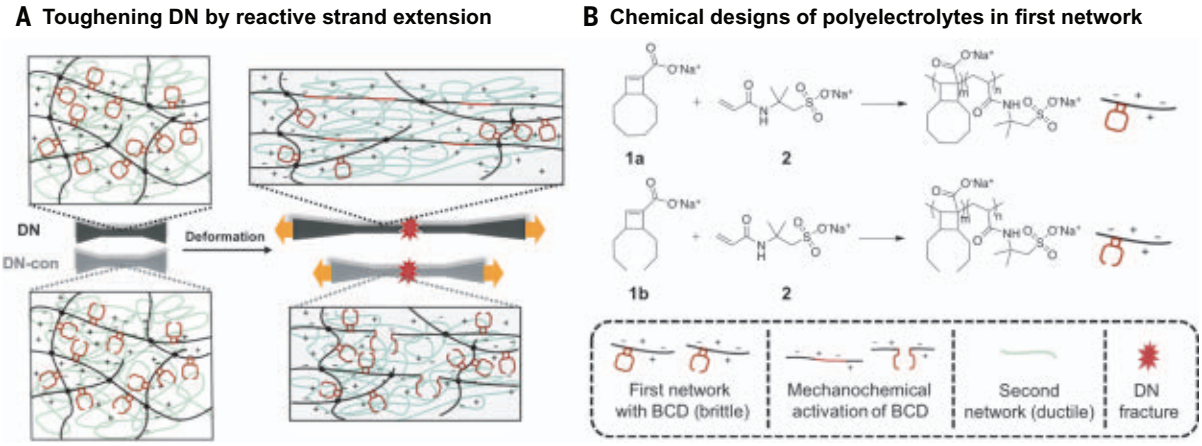


Fig. 1. The RSE concept in DN hydrogels. Strand scission events are not shown. (**A**) Schematic of DNs with RSE and analogous controls. (**B**) Chemical structures of the first network copolymer strands used in this work.

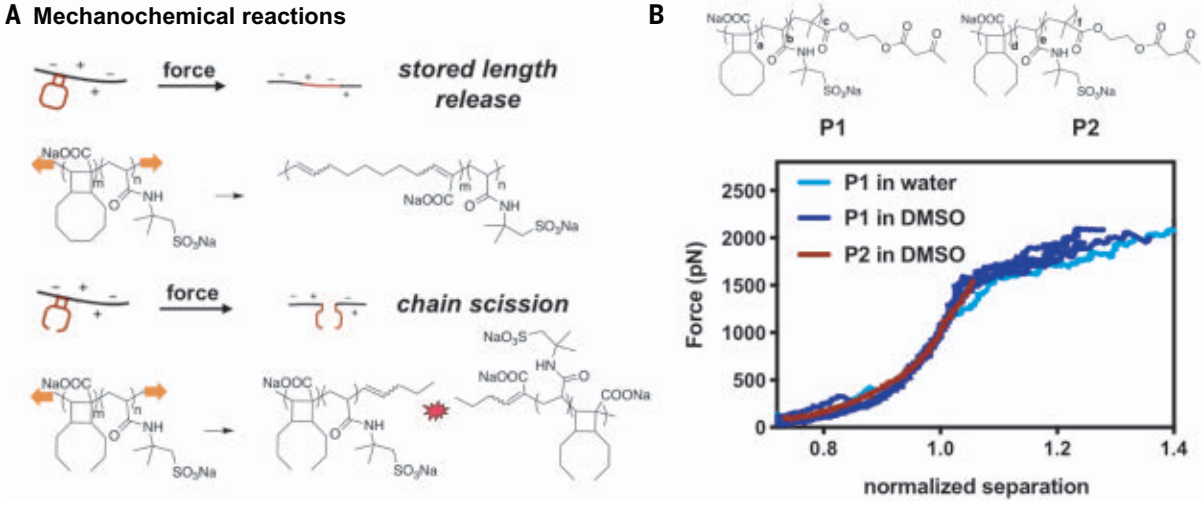


Fig. 2. Mechanochemical reactions driving RSE. (**A**) Mechanically coupled [2+2] cycloreversion leads to chain extension in BCD mechanophores and to chain scission in analogous acyclic cyclobutane mechanophores. (**B**) Representative single-molecule force-extension curves of linear RSE copolymer **P1** in water and dimethyl sulfoxide (DMSO), and control copolymer **P2** in DMSO.

The mechanochemically active copolymers were incorporated into double-network (DN) hydrogels, whose robust mechanical properties facilitate the characterization of mechanical response. DNs consist of two interpenetrating networks with contrasting structure: The first network comprises strands in a highly stretched state and is rigid and brittle. The second network comprises strands in a coiled state and as a result is soft and stretchable (*12*). This DN structure allows multiple strands within the first brittle network to break independently without causing catastrophic breakdown of the entire brittle network because the local strain amplification created by the breaking of the brittle strand is suppressed by transferring the load to the soft network strands nearby rather than propagating through the first network (*12*). Hydrophilic polymer networks were made through the radical addition copolymerization of **1a**, NaAMPS **2** (feed concentrations of 0.3 and 0.7 M, respectively), and *N*,*N*′-methylenebisacrylamide (MBAA) cross-linker. In the DNs, this BCD-AMPS first network was swollen in a precursor solution of acrylamide (AAm, 4 M; MBAA, 0.01 mol %) that was polymerized to form an interwoven, sparsely cross-linked, and highly stretchable polyacrylamide (PAAm) second network (fig. S4). Two first networks, **N1** and **N2**, of the same BCD-AMPS content [BCD:AMPS = 3:17 mol/mol (single network and DN synthesis in supplementary materials); ~40% contour length increase expected from calculations (fig. S3)] but different cross-linking density (8 and 4 mol % of MBAA for **N1** and **N2**, respectively) were prepared (table S1 and fig. S4).

Control networks **N1-con** and **N2-con** were synthesized by using acyclic cyclobutene carboxylate **1b**. The structural similarity of **1a** and **1b** leads to similar reactivity in the polymerization (fig. S5), and the similar network structure in **N1** and **N1-con**, or in **N2** and **N2-con**, was verified through equilibrium swelling ratios (V/V_0) of **N1** (12.4 ± 0.1) and **N2** (32.7 ± 0.7) in water that are indistinguishable from those of **N1-con** (12.1 ± 0.9) and **N2-con** (32.7 ± 1.8), respectively, and a gel permeation chromatograph (GPC) comparison of the disassembly products of the first networks (fig. S6). These single networks were too brittle for reliable mechanical characterization, but the moduli of **DN1** (288 ± 4 kPa) and **DN1-con** (290 ± 8 kPa) fall within experimental uncertainty, as do those of **DN2** (243 ± 5 kPa) and **DN2-con** (249 ± 13 kPa) (Table 1). Thus, any differences in ultimate stretch and tear resistance between the active and control networks are attributed to the conversion of scission to molecular extension

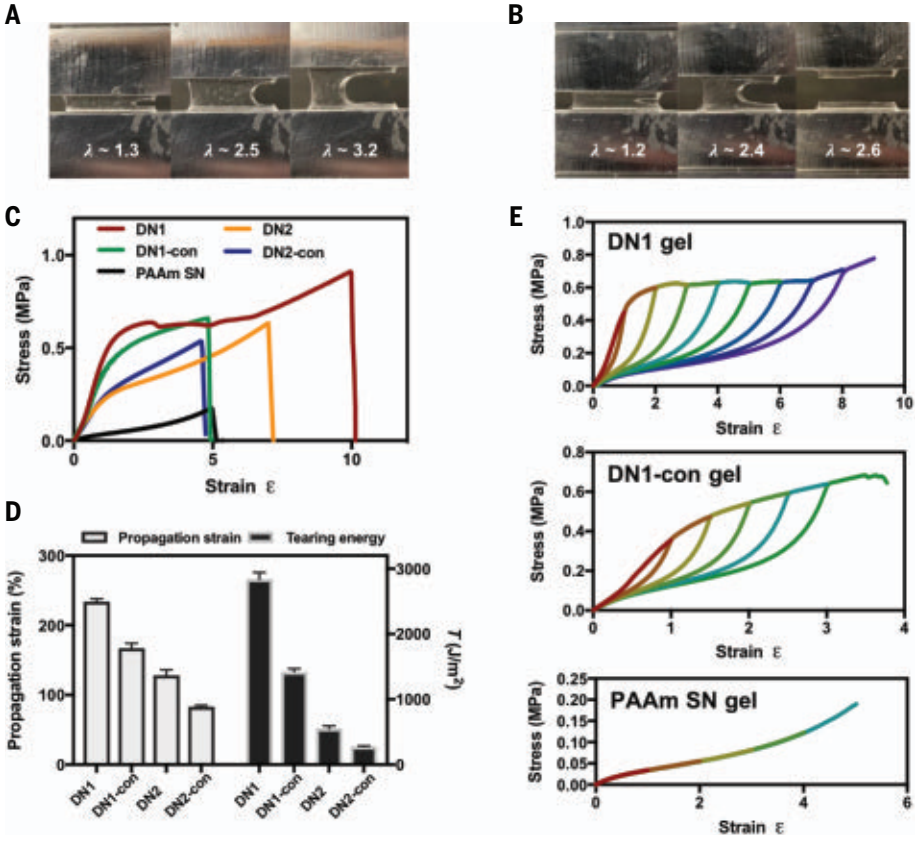


Fig. 3. Response of DNs to uniaxial tension. (**A**) Single-edge notched **DN1** at stretch ratios λ (millimeter/millimeter) of 1.3, 2.5, and 3.2. (**B**) Single-edge notched **DN1-con** at stretch ratios (millimeter/millimeter) of 1.2, 2.4, and 2.6. (**C**) Representative stress-strain curves (pure shear, 2-mm-thick films) of unnotched DNs and a single-PAAm network prepared identically to the second network of the DNs. (**D**) Propagation strain and tear energy of DNs. (**E**) Hysteresis in cyclic loading-unloading of **DN1**, **DN1-con**, and a single-PAAm network prepared identically to the second network of the DNs. The scales of the strain axes differ for the three loading-unloading curves.

Table 1. Young's modulus, nominal strain, nominal stress, and input energy density at break of unnotched films in unconstrained uniaxial tension, and crack propagation strain and tearing energy of films in a single-edge notched geometry. *E*, Young's modulus; ε_b, nominal strain; σ_b, nominal stress; W_b, input energy density; ε_p crack propagation strain; *T*, tearing energy. Tearing energy of PAAm gel is listed as reported (*23*). Dash indicates value not measured.

	E (kPa)	ε_b	σ_b (MPa)	W_b (MJ m^{-3})	ε_p	*T* (J m^{-2})
DN1	288 ± 4	10.0	0.92	6.4	2.3	2840 ± 103
DN1-con	290 ± 8	4.9	0.61	2.3	1.7	1419 ± 52
DN2	243 ± 5	7.0	0.64	2.6	1.3	554 ± 40
DN2-con	249 ± 13	4.7	0.54	1.6	0.8	276 ± 10
PAAm	57 ± 6	5.0	0.18	0.4	—	~100

and higher breaking force that is allowed in the BCD copolymers by virtue of the designed release of length stored in the bicyclic ring structure.

The incorporation of the BCD mechanophore leads to substantial changes in the mechanical limits of its double networks. The resistance to tearing was evaluated by introducing a notch into DN films and measuring the strain necessary for the notch to propagate under pure shear loading. Deformations of notched **DN1** and **DN1-con** under uniaxial extension are shown in Fig. 3, A and B (stress-strain curves are provided in fig. S7). The critical stretch for crack propagation in **DN1** is 234 ± 4%, which is 40% greater than the 167 ± 7% observed in **DN1-con**. The gain in stretchability is found as well in **DN2**, for which the critical stretch for crack propagation is 128 ± 8%, versus 83 ± 2% in **DN2-con**. The resistance to tearing can be further quantified in terms of the tear energy *T*, or the energy required per area of crack growth, which was determined with the method of Rivlin and Thomas (*13*). Both **DN1** and **DN2** have roughly double the *T* of their respective controls: 2840 ± 103 J/m^2 versus 1419 ± 52 J/m^2 for **DN1** versus **DN1-con**, and 554 ± 40 J/m^2 versus 276 ± 10 J/m^2 for **DN2** versus **DN2-con** (Fig. 3D and Table 1).

The improved mechanical properties due to RSE are also observed in unnotched **DN1** relative to **DN1-con** (ultimate stretch of 890 ± 120% and 479 ± 99% before breaking, respectively) (Fig. 3C and fig. S8). **DN1** shows a clear yielding, suggesting that the first network ruptures into a global discontinuous phase that enables the load to be transferred to the second network in the yielded region (*12*, *14*). However, **DN1-con** shows a gradual increase in stress without a clear plateau stress region, suggesting that the internal rupture of the first network is relatively random and does not achieve a global discontinuous structure up to sample failure. We attribute this to fracture of **N1-con** within **DN1-con** beginning at a lower strain than that of **N1**. When a small defect forms, that stress largely remains in **N1-con** because the second network is still too soft for stress transfer (the local strain at the defect is not great enough). Thus, the fracture propagates further across **N1-con** within the DN and only begins to transfer to the second network after the defect becomes larger. In **DN1**, fracture of **N1** initiates at higher strain, at which point a greater fraction of the stress is transferred to the second network, and so the fracture event is blunted and the next fracture occurs elsewhere, leading to the desired fractal-domain patterns for greater toughness (*12*).

The presence of RSE should alter the energy dissipation behavior from that observed in conventional DN hydrogels. The two systems (for example, **DN1** versus **DN1-con**) start with strands that, as seen in the SMFS experiments (Fig. 2B), are effectively identical elastically. In **DN1-con**, a given strand breaks upon reaching a critical strain. In **DN1**, the analogous RSE strand requires the same energy up to that point but, instead of breaking, starts to grow longer. This molecular picture leads to two expectations. First, each strand's contribution to the elastic modulus is a function of the ratio of its initial (relaxed) end-to-end distance to final contour length (*15*). Activation of RSE preserves the initial end-to-end distance but increases the final contour length, and RSE therefore reduces a strand's contribution to the tensile modulus, but that reduction is not

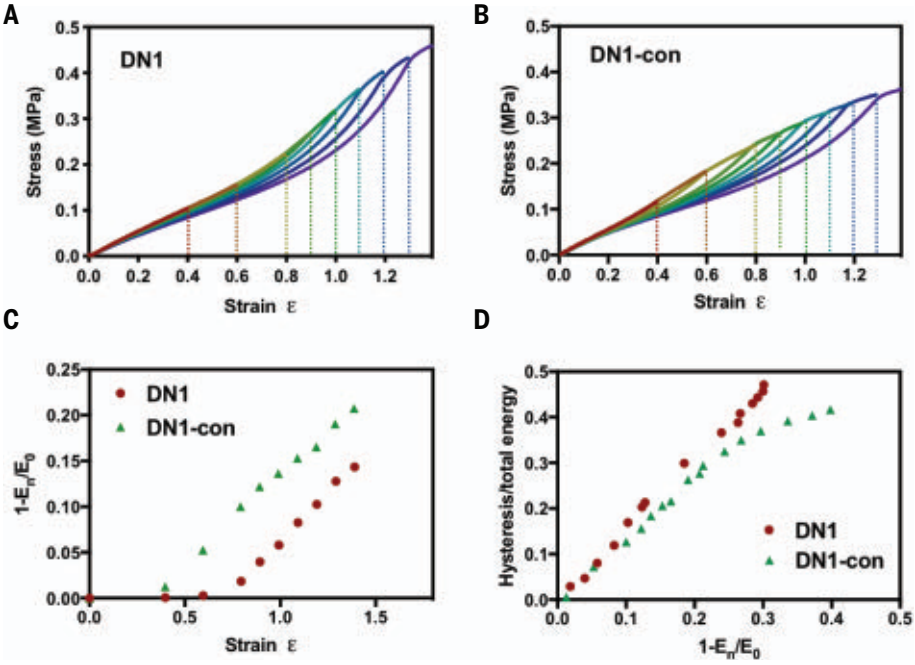


Fig. 4. Hysteresis analysis of DN1 and DN1-con. (**A** and **B**) Cyclic loading-unloading of (A) **DN1**, and (B) **DN1-con** in small strain range. (**C**) Fractional loss of Young's modulus of **DN1** and **DN1-con** as a function of strain. (**D**) Fractional hysteresis as a function of fractional loss in modulus for **DN1** and **DN1-con**.

as great as in the comparison strand that has broken. Therefore, we expect a smaller decrease in modulus with strain under cyclic loading in **DN1** than in **DN1-con**. Second, by the time the RSE strand fully breaks, it will have absorbed and dissipated more energy than the analogous strand that broke without RSE; we also expect that the RSE strands (**DN1**) will dissipate greater energy per loss of modulus than that of their non-RSE controls (**DN1-con**).

The expected effects are observed in sequential tensile loading-unloading cycles, in which the maximum strain increases from one cycle to the next. When a DN gel is stretched to a given strain for the first time and then unloaded, there is hysteresis (Fig. 3E; low strain behavior is expanded in Fig. 4, A and B) (*16*). Typically, hysteresis in DN gels is attributed almost entirely to chain scission in the first network (*12*) and is correlated with a decrease in modulus because of the loss of stress-bearing network strands (figs. S9 and S10) (*17*). Longer elastically effective strands in the first network result in less modulus loss at a given strain (*16*, *18*). The initial strand lengths in **DN1** and **DN1-con** are indistinguishable, but RSE in **DN1** suppresses the loss in modulus as expected; greater stretch is required to trigger a measurable change in modulus in **DN1** than **DN1-con**, and the loss in modulus in **DN1** consistently lags that observed in **DN1-con** as a function of strain (Fig. 4C). Further, **DN1** exhibits greater hysteresis per loss in modulus than **DN1-con** (Fig. 4D), which is consistent with the ability of RSE to dissipate energy without chain scission.

The concept of reactive chain extension demonstrated here has numerous implications for network design. First, reactive chain extension occurs only at large strains and only at the specific molecular locations that are about to break without it. Unlike the incorporation of weak, reversible cross-linking or noncovalent interstrand interactions, the enhancement does not come with a concomitant increase in the viscous modulus of the network [as evidenced by overlap of repeated unloading-loading cycles at the same maximum strain (fig. S11)]. That is, the energy dissipation is triggered only when needed to resist fracture and need not be overcome in routine functions of the material when fracture is not imminent. Second, the property enhancements are achieved within networks that otherwise have identical structure and modulus to their nonmechanophore controls, meaning that reactive chain extension might be introduced into at least some materials whose properties are already optimized for a given function, including optimizing the ratio of first- and second-network swelling to achieve maximum dissipation (*12*, *14*). Relative to control DN gels, RSE provides fracture resistance while also suppressing the loss in modulus that accompanies energy dissipation. Third, prior examples of mechanochemically induced changes in macroscopic physical properties involve either activation in solution (*8*, *19*), irreversible plastic deformation (*20*), and/or the presence of reagents beyond the initial network (*19*, *21*). Mechanophore designs offer an increasing range of responses (*22*), and the chemical changes that accompany reactive chain extension might be designed to produce additional, desirable responses.

REFERENCES AND NOTES

1. M. M. Caruso *et al.*, *Chem. Rev.* **109**, 5755–5798 (2009).
2. S. Lin *et al.*, *Sci. Adv.* **5**, eaau8528 (2019).
3. X. Zhao, *Soft Matter* **10**, 672–687 (2014).
4. E. Ducrot, Y. Chen, M. Bulters, R. P. Sijbesma, C. Creton, *Science* **344**, 186–189 (2014).
5. H. Lei *et al.*, *Nat. Commun.* **11**, 4032 (2020).
6. Z. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5967–5972 (2019).
7. S. Garcia-Manyes, A. E. M. Beedle, *Nat. Rev. Chem.* **1**, 0083 (2017).
8. Z. Chen *et al.*, *Science* **357**, 475–479 (2017).
9. S. Wang, S. Panyukov, M. Rubinstein, S. L. Craig, *Macromolecules* **52**, 2772–2777 (2019).
10. B. H. Bowser, C.-H. Ho, S. L. Craig, *Macromolecules* **52**, 9032–9038 (2019).
11. Y. Tian *et al.*, *J. Am. Chem. Soc.* **142**, 18687–18697 (2020).
12. J. P. Gong, *Soft Matter* **6**, 2583–2590 (2010).
13. R. S. Rivlin, A. G. Thomas, *J. Polym. Sci.* **10**, 291–318 (1953).
14. T. Matsuda *et al.*, *Macromolecules* **49**, 1865–1872 (2016).
15. P. J. Flory, J. Rehner Jr., *J. Chem. Phys.* **11**, 521–526 (1943).
16. T. Nakajima, T. Kurokawa, S. Ahmed, W. Wu, J. P. Gong, *Soft Matter* **9**, 1955–1966 (2013).
17. T.-T. Mai, T. Matsuda, T. Nakajima, J. P. Gong, K. Urayama, *Macromolecules* **51**, 5245–5257 (2018).
18. X. Wang, W. Hong, *Soft Matter* **7**, 8576–8581 (2011).
19. Z. Wang, J. Ayarza, A. P. Esser-Kahn, *Angew. Chem. Int. Ed.* **58**, 12023–12026 (2019).
20. A. L. B. Ramirez *et al.*, *Nat. Chem.* **5**, 757–761 (2013).
21. T. Matsuda, R. Kawakami, R. Namba, T. Nakajima, J. P. Gong, *Science* **363**, 504–508 (2019).
22. M. A. Ghanem *et al.*, *Nat. Rev. Mater.* **6**, 84–98 (2020).
23. J.-Y. Sun *et al.*, *Nature* **489**, 133–136 (2012).

ACKNOWLEDGMENTS

We thank Y. J. Lin and P. Zhang for the assistance with density functional theory calculations, and M. Walters for the preparation of the experimental devices. **Funding:** This work was supported by the NSF Center for the Chemistry of Molecularly Optimized Networks (MONET), CHE-2116298, to J.A.J., J.A.K., B.D.O., M.R., and S.L.C. This work was partially supported by JSPS KAKENHI grant (JP17H06144) to J.P.G. **Author contributions:** Z.W., B.H.B., and S.L.C. conceived the concept. Z.W., T.O., H.K.B., T.M., S.W., J.A.J., J.A.K., B.D.O., J.P.G., M.R., and S.L.C designed the experiments. Z.W., X.Z., T.B.K., and S.A.-R. performed experiments and analyzed the data. Z.W., J.A.J., J.A.K., B.D.O., J.P.G., M.R., and S.L.C. wrote the manuscript. **Competing interests:** T.M. is also affiliated with DSM Engineering Materials, Royal DSM, 6160 BB Geleen, Netherlands. Z.W. is also affiliated with Lawrence Berkeley National Laboratory, Molecular Foundry, 67 Cyclotron Road, Berkeley, CA 94720, USA. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg2689
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S4
References (*24*–*32*)
Movies S1 and S2

22 December 2020; accepted 27 August 2021
10.1126/science.abg2689

CANCER

The G protein signaling regulator RGS3 enhances the GTPase activity of KRAS

Chuanchuan Li[1], Alberto Vides[1], Dongsung Kim[1], Jenny Y. Xue[1,2], Yulei Zhao[1], Piro Lito[1,2,3,4]*

Recently reported to be effective in patients with lung cancer, KRAS^{G12C} inhibitors bind to the inactive, or guanosine diphosphate (GDP)–bound, state of the oncoprotein and require guanosine triphosphate (GTP) hydrolysis for inhibition. However, KRAS mutations prevent the catalytic arginine of GTPase-activating proteins (GAPs) from enhancing an otherwise slow hydrolysis rate. If KRAS mutants are indeed insensitive to GAPs, it is unclear how KRAS^{G12C} hydrolyzes sufficient GTP to allow inactive state–selective inhibition. Here, we show that RGS3, a GAP previously known for regulating G protein–coupled receptors, can also enhance the GTPase activity of mutant and wild-type KRAS proteins. Our study reveals an unexpected mechanism that inactivates KRAS and explains the vulnerability to emerging clinically effective therapeutics.

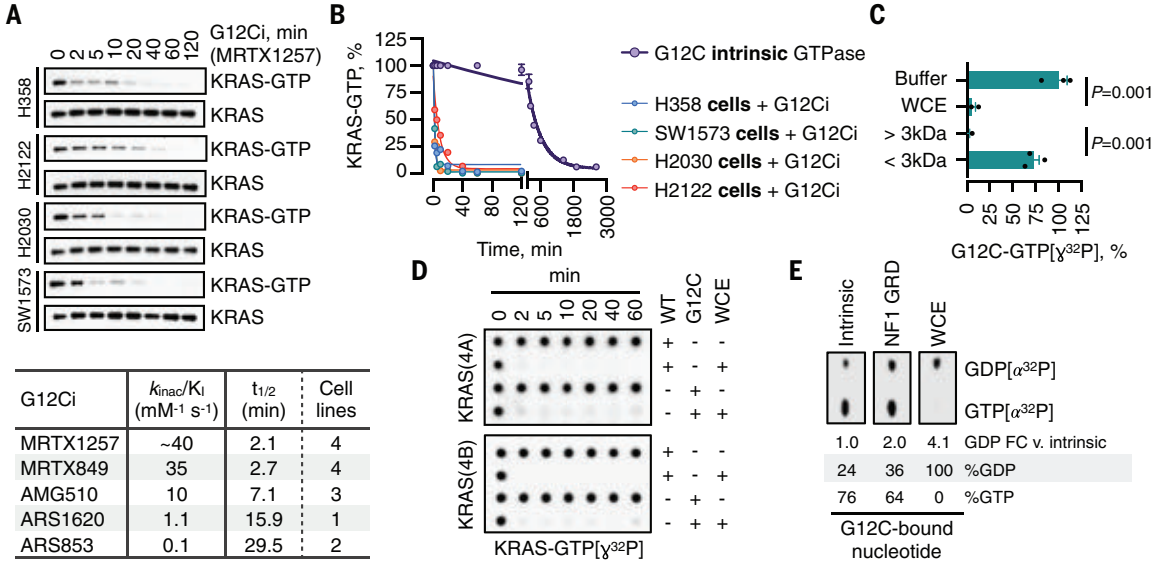
KRAS hydrolyzes guanosine triphosphate (GTP) to guanosine diphosphate (GDP) and controls diverse cellular functions by cycling between an active (GTP-bound) and an inactive (GDP-bound) conformation (*1*–*4*). Its weak intrinsic GTPase activity is enhanced by GTPase-activating proteins (GAPs), and the exchange of GDP for GTP is enhanced by guanine-nucleotide exchange factors (*5*–*7*). KRAS mutations are among the most common activating alterations found in cancer (*4*, *8*). Most of these mutations (~80%) result in single–amino acid substitutions of glycine 12 (e.g., G12C/D/V, etc.) and prevent the catalytic arginine residue in GAPs from enhancing GTP hydrolysis (*3*, *9*, *10*). Insensitivity to GAPs is thought to render KRAS oncoproteins constitutively active and independent of upstream input for activation.

Emerging evidence suggests that some KRAS oncoproteins require nucleotide exchange for activation. Perhaps the strongest evidence is provided by KRAS^{G12C} inhibitors (*11*–*17*), which selectively target the inactive state of the oncoprotein to prevent its activation by nucleotide exchange. We have previously shown that these drugs induce responses in about one-third of patients with lung cancer (*18*), and that inhibition requires an intact GTPase activity by KRAS^{G12C} (*12*, *17*). If KRAS mutants are insensitive to GAPs, it is unclear how KRAS^{G12C} hydrolyzes sufficient GTP to allow inactive state–selective inhibition.

Intrinsic hydrolysis alone may be sufficient to enable inactive state–selective inhibition. To investigate this, we compared the rate of inhibition in cancer cells with the rate of intrinsic GTP hydrolysis by KRAS^{G12C}. The time required for maximal inhibition of KRAS in cells treated with a potent and selective G12C inhibitor (G12Ci) (MRTX1257; fig. S1A) ranged from ~20 min in serum-deprived medium, a condition that suppresses nucleotide exchange (half-life ~2 min; Fig. 1A), to ~60 min in medium with 10% serum (half-life ~10 min; fig. S1B). Similar cellular inhibition kinetics were observed with other potent inhibitors ($k_{inac}/K_I \geq 10$ mM^{-1} s^{-1}) such as MRTX849 and AMG510 (Fig. 1A).

By comparison, purified KRAS^{G12C} hydrolyzed GTP[γ^{32}P] at a much slower rate (half-life 120 to 300 min; Fig. 1B and fig. S1C). The

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer, New York, NY 10065, USA. [2]Tri-Institutional MD-PhD Program, Weill Cornell Medical College and Rockefeller University and Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [3]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [4]Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA.
*Corresponding author. Email: litop@mskcc.org



G12Ci	k_{inac}/K_I (mM^{-1} s^{-1})	$t_{1/2}$ (min)	Cell lines
MRTX1257	~40	2.1	4
MRTX849	35	2.7	4
AMG510	10	7.1	3
ARS1620	1.1	15.9	1
ARS853	0.1	29.5	2

Fig. 1. Cancer cell extracts enhance the GTPase activity of KRAS^{G12C}. (**A**) KRAS^{G12C}-mutant cell lines were serum deprived for 16 hours and treated as shown. The level of active KRAS (GTP-bound) was determined by RAS-binding domain pulldown and quantified by densitometry. Bottom: Potency of drug binding to G12C (k_{inac}/K_I) and the half-life for cellular inhibition (mean from the indicated number of cell lines). (**B**) KRAS^{G12C} was loaded with GTP[γ^{32}P] and its intrinsic GTPase activity was assayed over time, ensuring sufficient time for the reaction to reach completion (mean ± SEM, n = 2). Also shown is the effect of MRTX1257 treatment on the ratio of active to total cellular KRAS. (**C**) GTP[γ^{32}P] hydrolysis by KRAS^{G12C} was assayed in the absence or presence of WCEs from H358 lung cancer cells (mean ± SEM, n = 3, two-tailed P value). WCEs were either added directly or after fractionation with a 3-kDa molecular cut-off column. (**D**) KRAS splice variants 4A or 4B, either WT or G12C mutant, were loaded with GTP[γ^{32}P] and assayed for GTP hydrolysis over time. (**E**) Thin layer chromatography was used to separate α^{32}P-labeled nucleotides that were eluted from KRAS^{G12C} incubated with the NF1 GAP-related domain (GRD) or H358 cellular extracts. Shown are the fold increase in the absolute GDP[α^{32}P] signal and the percentage α^{32}P-labeled nucleotide relative to the total signal per sample. FC, fold change. A representative of at least two experimental repeats is shown in (D) and (E).

intrinsic hydrolysis half-life found for KRAS proteins varies across studies (range: 25 to 200 min; fig. S1D). Regardless of this variation, the rate of cellular $KRAS^{G12C}$ inhibition with potent inhibitors was quicker than even the fastest intrinsic hydrolysis estimate (*19*) reported in the literature (fig. S1E).

The discrepancy between the rate of cellular inactivation and the rate of the intrinsic GTPase reaction suggests the presence of cellular factors that enhance GTP hydrolysis by $KRAS^{G12C}$. To test this possibility, purified KRAS was loaded with GTP[γ^{32}P] and incubated with whole-cell extracts (WCEs) from lung cancer models. Because cytoplasmic RAS-GAPs such as NF1 enhance the GTPase activity of wild-type (WT) KRAS, the latter was used as a positive control. The KRAS-nucleotide complex was immobilized, and the remaining GTP[γ^{32}P] was quantified by autoradiography (hereafter referred to as the γ-phosphate assay; fig. S1C). As shown in Fig. 1C and fig. S2A, little intrinsic GTP[γ^{32}P] hydrolysis was detected for $KRAS^{G12C}$ at 1 hour of incubation. In the presence of cellular extracts, however, there was a near complete hydrolysis of GTP[γ^{32}P] (Fig. 1C and fig. S2A). Under these conditions, the reaction reached a steady-state maximum within 2 min of incubation (Fig. 1D and fig. S2B). Cellular extracts did not affect KRAS protein stability in these experiments (fig. S2, C and D). The activity was greatly reduced in the flow-through from a 3-kDa molecular cut-off fractionation column (Fig. 1C and fig. S2A), suggesting that a cellular protein was responsible.

It has been reported that cellular extracts enhance hydrolysis by WT but not by mutant (G12V/D) NRAS (*5*). Whether the same holds true for $KRAS^{G12C}$ has not been investigated previously. In agreement with previous work, our data show that the GTPase activity of $KRAS^{WT}$ was enhanced in a concentration-dependent manner and that the effect on $KRAS^{WT}$ was greater than that on $KRAS^{G12C}$ (fig. S2D). However, rather than being completely insensitive to cellular extracts, $KRAS^{G12C}$ was also responsive, with activity detected at higher lysate concentrations (fig. S2, D and E).

It is possible that the $KRAS^{G12C}$ effect reflects an enhanced rate of nucleotide dissociation or the exchange of GTP for GDP. To address these possibilities, the GTPase activity of $KRAS^{G12C}$ was assayed by loading it with GTP[α^{32}P] and then incubating it with cellular extracts (hereafter referred to as the α-phosphate assay; fig. S2F). After immobilization of KRAS, the bound nucleotides were eluted and subjected to separation by thin-layer chromatography. As expected, after incubation with the GAP-related domain of NF1, the predominant [α^{32}P]-labeled nucleotide bound to $KRAS^{G12C}$ was GTP (Fig. 1E). By comparison, the predominant [α^{32}P]-labeled nucleotide bound to $KRAS^{G12C}$ incubated with cell extracts was GDP. If nucleotide exchange (or nucleotide dissociation) were responsible, then GTP[α^{32}P] would be replaced with cold nucleotide and no GDP[α^{32}P] signal would be detected. Because an increase in GDP[α^{32}P] was indeed detected (Fig. 1E), the data support the presence of a cellular protein that accelerates GTP hydrolysis by $KRAS^{G12C}$. Moreover, the GTPase-enhancing effect of cellular extracts was attenuated when A59G, a transition state mutation that blocks GTPase activity (*20*), was introduced alongside $KRAS^{G12C}$ (fig. S2G).

To isolate the protein responsible for enhancing the GTPase activity of $KRAS^{G12C}$, we devised a purification scheme in which $KRAS^{G12C}$-mutant cancer cell extracts (see the materials and methods) were subjected to two rounds of size-exclusion chromatography (SEC, steps 1 and 2), desalting (step 3), anion-exchange chromatography (AEC, step 4), and mass spectrometry (MS, step 5) to identify proteins classified as GAPs (Fig. 2A). The eluted fractions from each chromatographic step were assayed for their optical density and their ability to enhance $KRAS^{G12C}$ hydrolysis. Fractions that retained activity were pooled and carried forward to the next step. SEC step 1 elution volumes ranging from 52 to 72 ml had $KRAS^{G12C}$-directed GAP activity (fig. S3A). Four of these were independently subjected to SEC step 2 to confirm the activity and to identify which step 2 fractions to carry forward (Fig. 2B). SEC step 2 elution volumes ranging from 12 to 15 ml were pooled, desalted, and analyzed by AEC with a linear salt gradient (fig. S3B). The latter identified five peaks with $KRAS^{G12C}$-directed GAP activity (Fig. 2C).

The active fractions eluted in step 4 were pooled, acid precipitated, and subjected to MS to identify proteins annotated as GAP (fig. S3C). One of these, RGS3, is a GAP for the $G\alpha_{i/q}$ subunit of heterotrimeric G protein–coupled receptors (*21*–*23*). Many of the RGS3 peptides identified by MS localized in the RGS domain of the protein (RGSD, the GAP domain; fig. S3, D and E). RGS3 lacks the catalytic arginine (R) residue present in canonical RAS-GAPs (*24*, *25*). We thus reasoned that RGS3 might not be impeded by G12 mutations and decided to study it further. The broader RGS family contains 20 members (fig. S3F) that often have alternatively spliced variants. At least nine variants of RGS3 have been reported. As expected, the 75-kDa (p75) and/or the 25-kDa (p25) RGS3 isoforms (both of which contain the RGS GAP domain) eluted in one or more fractions that enhanced GTP[γ^{32}P] hydrolysis by $KRAS^{G12C}$ (Fig. 2D).

In beginning to validate the effect of RGS3, we questioned whether it interacts with $KRAS^{G12C}$ in cells. To investigate this, we relied on the ability of the G12Ci to selectively displace effector proteins from $KRAS^{G12C}$ while sparing those bound to WT KRAS, NRAS, and HRAS (*12*, *13*). Coimmunoprecipitation studies in three $KRAS^{G12C}$-mutant cell lines showed an interaction between endogenous $KRAS^{G12C}$ and RGS3, an interaction that was diminished upon G12Ci treatment (fig. S4A). GST-pulldown experiments confirmed that RGS3

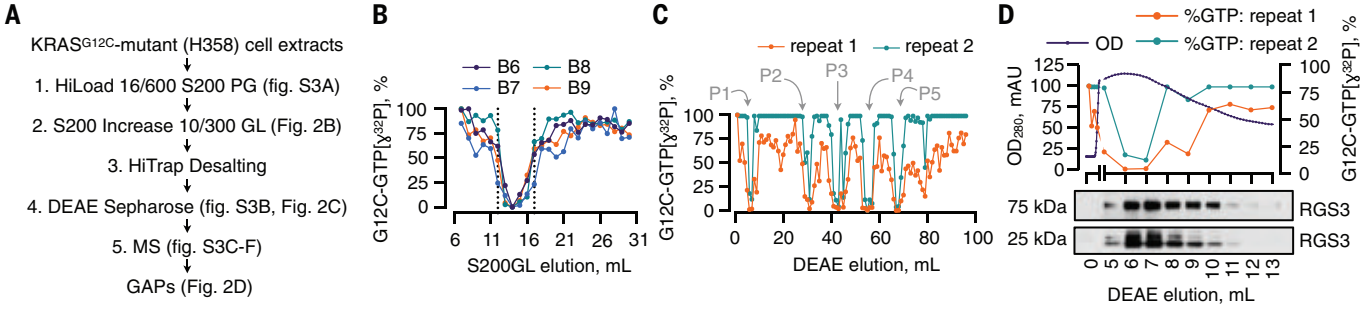


Fig. 2. Identification of a protein that enhances the GTPase activity of $KRAS^{G12C}$. (**A**) The purification schema included SEC (steps 1 and 2), desalting (step 3), AEC (step 4), and MS (step 5). (**B**) H358 cell extracts (2 g) were subjected to SEC, and eluted fractions (1 ml each) were incubated with GTP[γ^{32}P]-loaded $KRAS^{G12C}$ for 1 hour and subjected to the γ-phosphate assay. The G12C-GTP[γ^{32}P] signal in each fraction from step 2 is shown. See fig. S3A for step 1 analysis. (**C**) Active fractions from step 2 (12 to 17 ml) were pooled, desalted, and subjected to AEC using a linear salt gradient. The effect of each fraction on GTPase activity is shown. See also fig. S3B. (**D**) Elution of RGS3 p75 and p25 variants in fractions from peak 1 (P1) and the effect on $KRAS^{G12C}$ GTPase activity.

interacts with either KRAS^{G12C} 4A or 4B splice variants (fig. S4B), an interaction that was less pronounced with KRAS^{WT} (fig. S4C). Direct binding assays with purified proteins suggested a preference for the active (GTPγS-loaded) conformation of KRAS^{G12C} compared with the inactive (GDP-loaded) state (fig. S4D).

To test directly whether RGS3 acts as a GAP for KRAS^{G12C}, we performed GTP hydrolysis assays with purified KRAS in the presence of RGS3 variants. In addition to the radioactive γ- and α-phosphate assays (figs. S1C and S2F), we also used a nonradioactive hydrolysis assay that enables continuous measurement of phosphate release over time (fig. S5A). As expected, RGS3 enhanced GTP hydrolysis by KRAS^{G12C} (Fig. 3, A and B, and fig. S5B), leading to a concentration-dependent increase in the rate constant (fig. S5, C and D). RGS3 was more effective at enhancing KRAS^{G12C} hydrolysis than canonical RAS-GAPs such as NF1 and RASA1 (fig. S5, E and F). Whereas no measurable effect was detected for RASA1, some activity was observed for NF1 (Fig. 3, A and B, and fig. S5, E and F). The latter, however, was not markedly attenuated when the catalytic arginine of NF1 was mutated to alanine (R1276A), supporting our suspicion that KRAS^{G12C}-directed GAP activity occurs independently of the arginine residue, which is also known as the "R-finger."

Rather than having an idiosyncratic effect on KRAS^{G12C}, RGS3 also enhanced GTP hydrolysis by G12D/V and G13C/D mutant KRAS (fig. S6, A to C), which together comprise ~95% of KRAS mutations found in cancer. RGS3 also enhanced the GTPase activity of KRAS^{WT}, but in this setting RGS3 was much less effective than NF1 or RASA1 (Fig. 3, A and B, and fig. S5G).

RGS3 had little effect when the transition state mutation A59G was engineered alongside G12C (Fig. 3C). This suggests that RGS3 enhances hydrolysis by helping the reaction progress past the transition state. To better understand the mechanism, we constructed a molecular model using the crystal structures of KRAS-GMPPNP [*6OB2* (*26*)], RGS3 GAP domain [*2OJ4* (*27*)], and Ga_{ia1}-GDP·AlF_4:RGS4 [*1AGR* (*24*)]. The model identified an asparagine (N) residue in the GAP domain of RGS3 as potentially being involved in the hydrolysis reaction (Fig. 3D). This residue was predicted to orient away from the P-loop and therefore less likely to be impeded by G12 substitutions. Also, asparagine residues serve a catalytic function in other (non-RAS-specific) GAPs, including those enhancing GTP hydrolysis by RAP1 and RHEB GTPases (*28*, *29*).

We thus tested whether the asparagine residue in RGS3 was necessary for the GTPase-activating effect on KRAS^{G12C}. A histidine (H) substitution on N460 in p75 (RGS3) or on N147 in p25 (RGSD) attenuated the effect of RGS3 proteins on GTP hydrolysis by KRAS^{G12C}, as evidenced by the γ-phosphate (Fig. 3E) and α-phosphate (Fig. 3F) hydrolysis assays. Because RGS3 NH mutants had some residual activity against KRAS^{G12C}, it is likely that other residues also contribute to the GAP activity. This is in agreement with reports indicating that multiple RGS residues (in addition to asparagine) participate in enhancing GTP hydrolysis by Gα (*24*). The asparagine residue is conserved between RGS isoforms, suggesting that other members of the family also enhance the KRAS GTPase activity. Indeed, this was the case for RGS4 (Fig. 3E). Together, the biochemical data suggest that RGS3 enhances the GTPase activity of KRAS in a mutant-inclusive manner (Fig. 3G).

If RGS3 enhances KRAS^{G12C} hydrolysis, then it may inactivate this oncoprotein in cells. As expected, single-guide RNA–mediated deletion of RGS3 in KRAS^{G12C}-mutant lung cancer cells led to an increase in KRAS activation (Fig. 4A and fig. S7A). Loss of RGS3 also enhanced proliferation in culture and accelerated tumor growth in xenograft studies (Fig. 4B and fig. S7B). The proliferative advantage of RGS3$^{-/-}$ cells was largely dependent on KRAS^{G12C}, as evidenced by a previously validated (*17*, *30*) KRAS^{G12C}-specific small interfering RNA (siRNA) (fig. S7C). In agreement, siRNA-mediated knockdown of RGS3 increased proliferation in cells expressing KRAS^{G12C} but

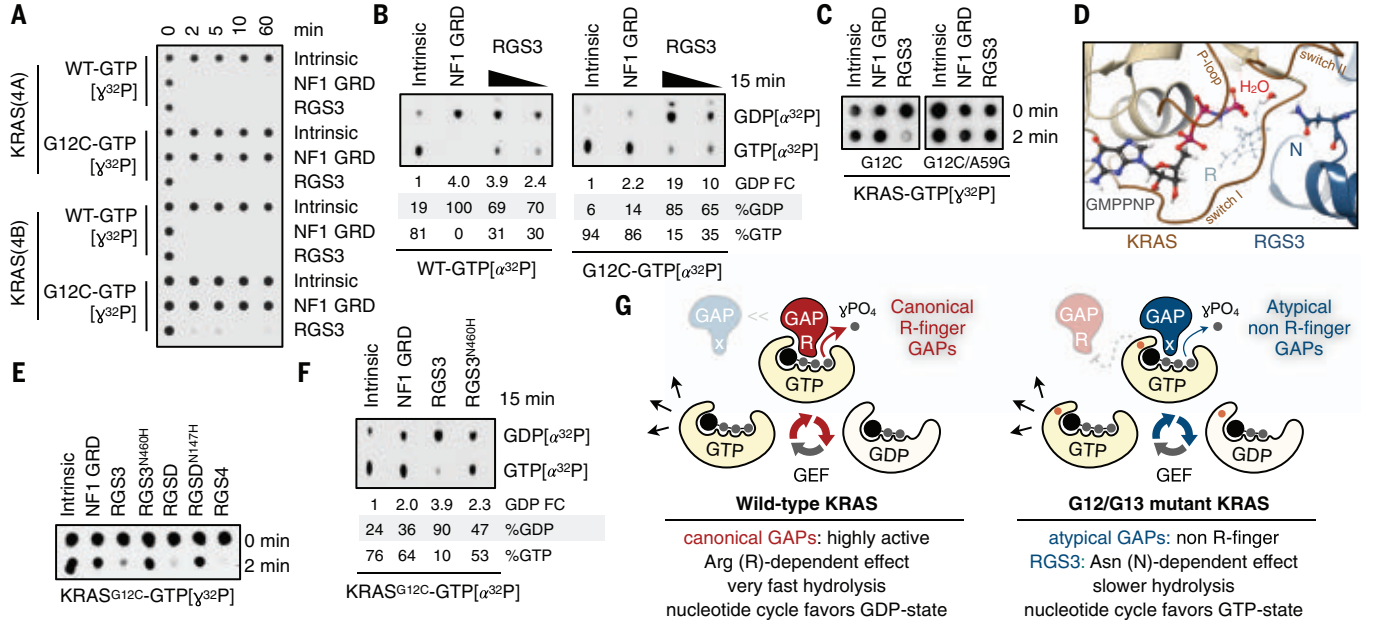


Fig. 3. Mechanistic basis for RGS3-assisted GTP-hydrolysis by KRAS^{G12C}. (**A** and **B**) GTP[γ^{32}P]-loaded KRAS variants were incubated with buffer, NF1 GRD, or RGS3 (p75) for the indicated times, followed by determination of GTPase activity by using the γ-phosphate assay (A) or the α-phosphate assay (B). (**C**) As in (A) except that the hydrolysis transition-state mutation A59G was engineered alongside G12C. (**D**) Molecular docking model of the interaction of KRAS-GMPPNP with the GAP domain of RGS3 (RGSD). The catalytic arginine residue of RASA1 (R) is superimposed as a reference point (not part of the modeling). (**E** and **F**) KRAS^{G12C} was loaded with GTP[γ^{32}P] (E) or GTP[α^{32}P] (F) and then incubated with WT or asparagine-to-histidine (NH) mutant RGS3, followed by determination of GTPase activity using the γ-phosphate (E) or the α-phosphate (F) assay. A representative of at least two experimental repeats is shown in (A) to (C), (E), and (F). (**G**) Schematic of the mechanism that enables mutant KRAS-inclusive GAP activity.

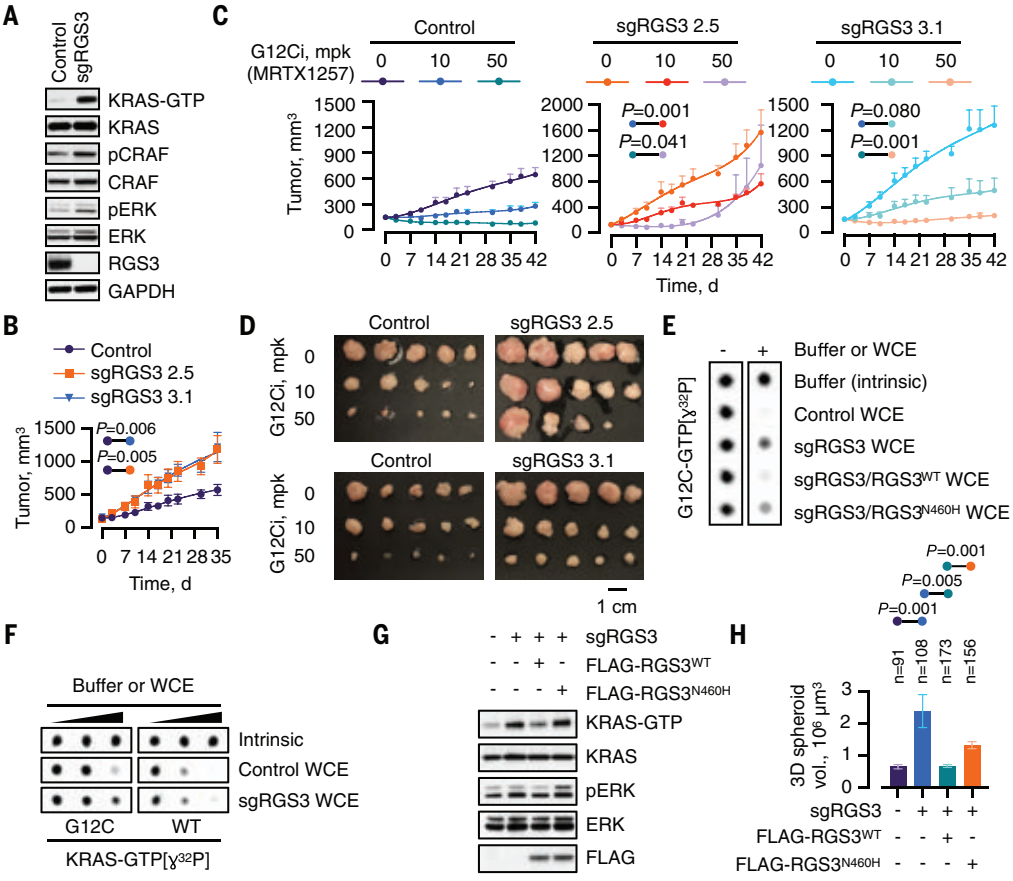


Fig. 4. RGS3 diminishes KRAS^{G12C} activation in cancer cells. (**A**) WCEs from control or RGS3-null H358 cells were immunoblotted to determine the effect on the indicated signaling intermediates. A representative of at least three experimental repeats is shown. (**B**) Control or RGS3$^{-/-}$ clones were implanted in athymic mice to determine the effect on tumor growth (n = 10 for control and n = 5 for each sgRGS3 clone). Mean ± SEM and two-tailed P values are shown. (**C** and **D**) Mice bearing control or RGS3$^{-/-}$ xenografts were treated as shown to determine the effect on tumor growth (n = 10 for the control and n = 5 for each sgRGS3 clone). Two-tailed P values (C) and end point tumor volumes (D) are shown. (**E**) Purified KRAS^{G12C} was loaded with GTP[γ^{32}P] and assayed for GTP hydrolysis in the absence or presence of extracts (0.4 µg/µl) from control and RGS3$^{-/-}$ cells, as well as extracts from RGS3$^{-/-}$ cells engineered to reexpress WT or N460H mutant RGS3. (**F**) Purified KRAS^{WT} or KRAS^{G12C} proteins were loaded with GTP[γ^{32}P] and assayed for hydrolysis in the absence (intrinsic) or presence of WCEs from control or RGS3$^{-/-}$ cells (0, 0.1, and 0.2 µg/µl). (**G**) WCEs from the indicated cell lines were evaluated by RAS-binding assay pulldown to determine the amount of active KRAS. (**H**) The cells were cultured as three-dimensional tumor spheroids in collagen I–containing medium for 17 days. The spheroid volumes were determined from representative images.

not in cells expressing KRAS$^{G12C/A59G}$ (fig. S7D), which is insensitive to the GTPase-enhancing effect of WCEs (fig. S2G) or RGS3 (Fig. 3C).

To determine whether RGS3 expression inversely correlated with KRAS activation in patients with lung cancer, we established a mutant KRAS-dependent transcriptional output score (fig. S8, A and B) and experimentally validated its ability to detect changes in KRAS^{G12C}-driven signaling after G12Ci treatment (fig. S8C). As shown in fig. S8, D and E, a higher RGS3 expression was associated with lower mutant KRAS output in lung cancers harboring G12C or any KRAS mutation. The negative correlation was observed with distinct KRAS up-regulated or down-regulated gene expression signatures (fig. S8, F and G). Little to no correlation was noted in KRAS^{WT} lung cancers.

RGS3$^{-/-}$ cells had an attenuated response to G12Ci treatment compared with their isogenic RGS3 WT cells. This was evidenced by: (i) a diminished inhibition of KRAS-GTP levels and downstream signaling (fig. S9, A and B), (ii) a diminished antiproliferative effect (fig. S9, C and D), and (iii) a less potent antitumor effect in vivo (Fig. 4, C and D, and fig. S9, E and F). Moreover, RGS3 expression correlated with susceptibility to G12Ci treatment in a panel of nine lung cancer patient–derived xenograft (PDX) models. The PDX were established from seven patients, and their treatment response profile (fig. S10A) was similar to that reported in a recent clinical study (*18*). RNA sequencing (fig. S10B) and immunohistochemistry with an RGS3-specific antibody (fig. S10, C and D) revealed that tumors with higher RGS3 expression had a larger magnitude of inhibition.

To provide additional evidence that RGS3 functions as a mutant KRAS-inclusive GAP, we tested the ability of RGS3-depleted extracts to directly enhance GTP[γ^{32}P] hydrolysis by KRAS. As shown in Fig. 4E, almost complete KRAS^{G12C} hydrolysis was observed in the presence of control extracts (row 2 versus row 1), an activity that was attenuated in RGS3$^{-/-}$ extracts (row 3). By comparison, the activity was restored in extracts from RGS3$^{-/-}$ cells expressing RGS3^{WT} (row 4) and diminished again in those expressing RGS3^{N460H} (row 5). RGS3$^{-/-}$ extracts did not significantly affect GTP[γ^{32}P] hydrolysis by KRAS^{WT} (Fig. 4F). The latter agrees with the notion that RGS3$^{-/-}$ cells still express conventional RAS-GAPs (such as NF1), which greatly enhance hydrolysis by KRAS^{WT} (Fig. 3B and fig. S5G: NF1>>RGS3), but only modestly that by KRAS^{G12C} (Fig. 3B and fig. S5E: RGS3>NF1). Last, the increase in cellular KRAS activation (Fig. 4G) and proliferation (Fig. 4H) conferred by the deletion of RGS3 were reversed by reexpression of RGS3^{WT} but much less so by RGS3^{N460H}.

Here, we report that the GTPase activity of KRAS is enhanced by non–arginine-finger dependent GAPs such as RGS3 (see also supplementary text). As evidenced above, RGS3 enhanced GTP hydrolysis in mutant KRAS-inclusive manner, affecting both WT and several oncogenic mutants (including G12C, G12D, G12V, G13C, and G13D). This effect was dependent on a key asparagine residue in the GAP domain of RGS3. KRAS mutations impede the catalytic arginine (or R-finger) of canonical RAS-GAPs (such as NF1 or RASA1) from enhancing an otherwise slow GTP hydrolysis rate. If, however, KRAS mutants were completely insensitive to GAPs, then inactive state–selective KRAS^{G12C} inhibitors would require a long time to take effect, considering that the time to inhibition is limited by the rate of hydrolysis. Although inhibitor potency might have hindered initial observations (*13*), highly potent G12Cis inactivate the oncoprotein with minutes, quicker than even the fastest estimate of intrinsic hydrolysis rate in the literature (*19*). RGS3-assisted

hydrolysis by $KRAS^{G12C}$ was approximately one to two orders of magnitude slower than NF1- or RASA1-stimulated hydrolysis by $KRAS^{WT}$, suggesting that RGS3 is unlikely to inactivate the entire cellular pool of $KRAS^{G12C}$; that is, not to the same extent as canonical R-finger GAPs are able to do for $KRAS^{WT}$. As a result, $KRAS^{G12C}$ has a longer residency time in its active (GTP-bound) conformation than $KRAS^{WT}$, allowing the oncoprotein to drive proliferation despite its susceptibility to "atypical-for-RAS" GAPs (Fig. 3G). Collectively, our experiments uncover an unexpected regulatory mechanism with important implications for understanding and therapeutically targeting KRAS oncoprotein–driven cancers.

REFERENCES AND NOTES

1. M. Malumbres, M. Barbacid, *Nat. Rev. Cancer* **3**, 459–465 (2003).
2. B. Papke, C. J. Der, *Science* **355**, 1158–1163 (2017).
3. D. K. Simanshu, D. V. Nissley, F. McCormick, *Cell* **170**, 17–33 (2017).
4. S. Li, A. Balmain, C. M. Counter, *Nat. Rev. Cancer* **18**, 767–777 (2018).
5. M. Trahey, F. McCormick, *Science* **238**, 542–545 (1987).
6. S. M. Margarit *et al.*, *Cell* **112**, 685–695 (2003).
7. J. L. Bos, H. Rehmann, A. Wittinghofer, *Cell* **129**, 865–877 (2007).
8. Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, *Nat. Rev. Cancer* **11**, 761–774 (2011).
9. G. Bollag, F. McCormick, *Nature* **351**, 576–579 (1991).
10. K. Scheffzek *et al.*, *Science* **277**, 333–338 (1997).
11. J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells, K. M. Shokat, *Nature* **503**, 548–551 (2013).
12. P. Lito, M. Solomon, L. S. Li, R. Hansen, N. Rosen, *Science* **351**, 604–608 (2016).
13. M. P. Patricelli *et al.*, *Cancer Discov.* **6**, 316–329 (2016).
14. M. R. Janes *et al.*, *Cell* **172**, 578–589.e17 (2018).
15. J. Canon *et al.*, *Nature* **575**, 217–223 (2019).
16. J. Hallin *et al.*, *Cancer Discov.* **10**, 54–71 (2020).
17. J. Y. Xue *et al.*, *Nature* **577**, 421–425 (2020).
18. D. S. Hong *et al.*, *N. Engl. J. Med.* **383**, 1207–1217 (2020).
19. J. C. Hunter *et al.*, *Mol. Cancer Res.* **13**, 1325–1335 (2015).
20. B. E. Hall, D. Bar-Sagi, N. Nassar, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12138–12142 (2002).
21. N. Watson, M. E. Linder, K. M. Druey, J. H. Kehrl, K. J. Blumer, *Nature* **383**, 172–175 (1996).
22. E. M. Ross, T. M. Wilkie, *Annu. Rev. Biochem.* **69**, 795–827 (2000).
23. R. R. Neubig, D. P. Siderovski, *Nat. Rev. Drug Discov.* **1**, 187–197 (2002).
24. J. J. Tesmer, D. M. Berman, A. G. Gilman, S. R. Sprang, *Cell* **89**, 251–261 (1997).
25. S. P. Srinivasa, N. Watson, M. C. Overton, K. J. Blumer, *J. Biol. Chem.* **273**, 1529–1533 (1998).
26. D. Rabara *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22122–22131 (2019).
27. L. Rezabkova *et al.*, *J. Struct. Biol.* **170**, 451–461 (2010).
28. O. Daumke, M. Weyand, P. P. Chakrabarti, I. R. Vetter, A. Wittinghofer, *Nature* **429**, 197–201 (2004).
29. Y. Li, K. Inoki, K. L. Guan, *Mol. Cell. Biol.* **24**, 7965–7975 (2004).
30. N. Sunaga *et al.*, *Mol. Cancer Ther.* **10**, 336–346 (2011).

ACKNOWLEDGMENTS

We thank M. Mroczkowski for discussing this work throughout its stages and for reviewing the manuscript; R. Soni from the Columbia University mass spectrometry core facility for helping with protein identification; J. Christensen from Mirati Therapeutics for graciously providing MRTX1257 and MRTX849; and R. Lipford from Amgen for graciously providing AMG510. **Funding:** This work was supported in part by the National Cancer Institute of the National Institutes of Health (grant 1R01CA230745-01 to P.L., grant 1R01CA230267-01A1 to P.L., grant K08CA191082-01A1 to P.L., and grant F30CA232549-01 to J.X.), the Memorial Sloan Kettering Cancer Center (MSKCC) Josie Robertson Investigator Program, and the MSKCC Support Grant-Core Grant program (grant P30 CA008748). P.L. is also supported in part by The Pew Charitable Trusts and the Damon Runyon Cancer Research Foundation.

Author contributions: P.L. conceived and supervised the study. C.L. performed most of the experiments in the study. A.V. assisted with GTPase assays. D.S. and Y.Z. helped to validate the protein-protein interactions. J.Y.X. helped with PDX studies and stable cell line generation. C.L. and P.L. were the main writers of the manuscript. All other authors reviewed and edited the manuscript. **Competing interests:** P.L. reports research grants to his institution from Mirati, Revolution Medicines, and Strategia. P.L. is an inventor on patents filed by MSKCC regarding treatment of KRAS or BRAF mutant cancers. P.L. has received scientific advisory fees from Revolution Medicines and Black Diamond Therapeutics. The remaining authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf1730
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*31*–*37*)
MDAR Reproducibility Checklist

15 October 2020; resubmitted 6 April 2021
Accepted 17 August 2021
10.1126/science.abf1730

NEUROSCIENCE

Population-specific neuromodulation prolongs therapeutic benefits of deep brain stimulation

Teresa A. Spix[1,2], Shruti Nanivadekar[2,3,4], Noelle Toong[2], Irene M. Kaplow[5,2], Brian R. Isett[1,2], Yazel Goksen[1], Andreas R. Pfenning[1,5,2], Aryn H. Gittis[1,2]*

Symptoms of neurological diseases emerge through the dysfunction of neural circuits whose diffuse and intertwined architectures pose serious challenges for delivering therapies. Deep brain stimulation (DBS) improves Parkinson's disease symptoms acutely but does not differentiate between neuronal circuits, and its effects decay rapidly if stimulation is discontinued. Recent findings suggest that optogenetic manipulation of distinct neuronal subpopulations in the external globus pallidus (GPe) provides long-lasting therapeutic effects in dopamine-depleted (DD) mice. We used synaptic differences to excite parvalbumin-expressing GPe neurons and inhibit lim-homeobox-6–expressing GPe neurons simultaneously using brief bursts of electrical stimulation. In DD mice, circuit-inspired DBS provided long-lasting therapeutic benefits that far exceeded those induced by conventional DBS, extending several hours after stimulation. These results establish the feasibility of transforming knowledge of circuit architecture into translatable therapeutic approaches.

Identification of distinct neuronal subpopulations has been essential for understanding brain function. Interventions that target specific neuronal subpopulations (e.g., optogenetics) can alleviate disease symptoms with few side effects in animal models. However, clinical approaches in humans are rarely tailored to achieve population-specific neuromodulation. Electrical deep brain stimulation (DBS) is increasingly used to treat an expanding list of conditions in humans. Although theoretical work has long hypothesized that patterned electrical stimulation can improve its precision in targeting specific neuronal subpopulations (*1*–*5*), these approaches have not yet achieved the desired cell type specificity. Motor symptoms of Parkinson's disease (PD) are acutely responsive to DBS, but symptoms return quickly when stimulation is discontinued (*6*). The need for constant stimulation increases the risk of side effects and drains battery power. Using optogenetics in mice, we identified cellular nodes in the external globus pallidus (GPe) of the basal ganglia, where targeted interventions ameliorate circuit dysfunction and restore movement several hours after stimulation (*7*). This optogenetic approach requires genetically targeted tools to drive opposite firing responses of spatially intermixed GPe subpopulations: Parvalbumin-expressing (PV-GPe) neurons must be excited, whereas lim-homeobox-6–expressing (Lhx6-GPe) neurons must be suppressed. Refining DBS protocols to achieve the same degree of population-specific neuromodulation would provide a treatment approach translatable to humans that could extend the therapeutic duration of stimulation, representing a major therapeutic advance.

To gain insight into the cellular principles that dictate the responses of PV-GPe and Lhx6-GPe neurons to electrical stimulation, we recorded extracellular spiking activity from these populations in acute brain slices from $Lhx6\text{-}GFP^{+/-}$ and $PV\text{-}tdTmt^{+/-}$ mice (Fig. 1A and fig. S1A). Both populations exhibited similar excitement during continuous electrical stimulation (100 Hz, 1 s) (Fig. 1, B and C). However,

[1]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA. [3]Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA, USA. [4]School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [5]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA.
***Corresponding author. Email: agittis@cmu.edu**

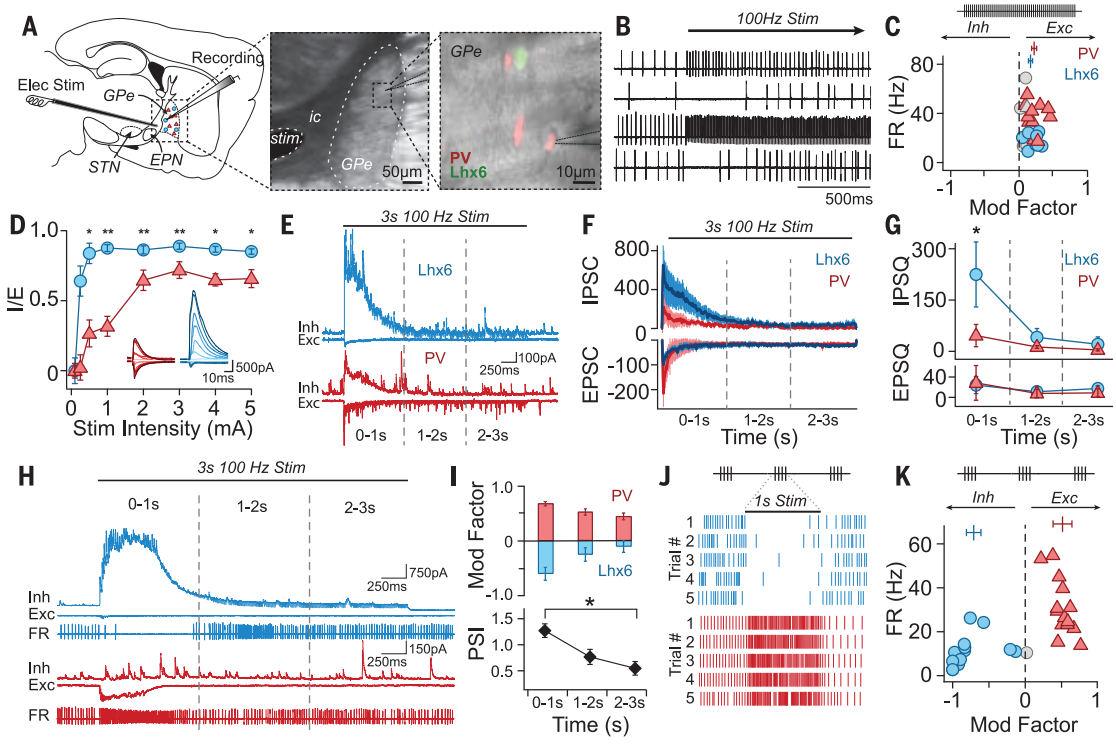


Fig. 1. Bursts of stimulation segregate the firing responses of PV-GPe and Lhx6-GPe neurons. (**A**) Experimental configuration and fluorescent identification of GPe subpopulations (ic, internal capsule). (**B**) Extracellular traces during stimulation (100 Hz for 30 s) after artifact removal. (**C**) Stimulation at 100 Hz for 30 s modestly excited both populations; this was quantified as "modulation factor" (MF) of the firing rate (FR) for each neuron: $(FR_{stim} - FR_{baseline})/(FR_{stim} + FR_{baseline})$ (bars: $MF_{Lhx6} = 0.16 \pm 0.11$; $MF_{PV} = 0.22 \pm 0.12$; Mann-Whitney (MWU) $P = 0.29$; 12 pairs of neurons; six mice, error bars: SEM). Colored symbols indicate neurons significantly modulated during stimulation (paired t test, $P < 0.05$), and gray symbols indicate neurons not significantly modified. (**D**) Ratios of postsynaptic currents for inhibition (IPSC) and excitation (EPSC) were calculated with peak current amplitudes as [(IPSC)/(IPSC + EPSC)], showing that Lhx6-GPe neurons receive more inhibition than PV-GPe neurons [Kruskal-Wallis test (KW), $\chi^2(7) = 57.4$, $P < 0.00001$, post hoc pairwise MWU, $^*P < 0.05$, $^{**}P < 0.001$; 15 pairs; nine mice)]. Insets show currents from representative neurons. (**E** and **F**) Rapid depression of synaptic currents during 100-Hz stimulation (1 mA), from representative neurons (E) and across the population (F) (10 pairs; four mice; medians, error bars: SEM). (**G**) Synaptic charges [inhibitory postsynaptic charge (IPSQ) and excitatory postsynaptic charge (EPSQ)] transferred during the first, second, and third seconds of stimulation. IPSQ was larger in Lhx6-GPe neurons than in PV-GPe neurons during the initial second of stimulation ($IPSQ_{Lhx6} = 284 \pm 300$ pC, $IPSQ_{PV} = 98 \pm 103$ pC, MWU, $^*P = 0.023$) (10 pairs; four mice). (**H**) Three-part recordings with 3 s of 100-Hz stimulation. (Bottom) Extracellular recording traces. (Middle and top) Whole-cell recordings of EPSCs (Exc, middle, $V_{hold} = -70$ mV), and IPSCs (Inh, top, $V_{hold} = 0$ mV) (**I**). (Top) MF by cell type in 1 s time bins. (Bottom) Modulation was most segregated in the initial second of stimulation (PSI = $MF_{PV} - MF_{Lhx6}$, see supplementary methods) [$PSI_{0\text{-}1s} = 1.27 \pm 0.49$; $PSI_{1\text{-}2s} = 0.77 \pm 0.54$; $PSI_{2\text{-}3s} = 0.55 \pm 0.49$; KW, $\chi^2(2) = 12.0$, $^*P = 0.003$; 14 pairs; three mice]. (**J**) Raster plots of extracellular modulation by 1 s bursts of stimulation (100 Hz, 1 mA, 30 s intertrial interval). (**K**) PV-GPe neurons are excited whereas Lhx6-GPe neurons are inhibited ($MF_{PV} = 0.5 \pm 0.1$; $MF_{Lhx6} = -0.7 \pm 0.3$; 14 pairs; three mice).

synaptic recordings showed that Lhx6-GPe neurons received proportionately more inhibition than PV-GPe neurons (Fig. 1D and fig. S1, B and C). This synaptic asymmetry was even more pronounced at high stimulation frequencies, where rapid synaptic depression was observed (Fig. 1, E and F). Notably, Lhx6-GPe neurons experienced significantly greater inhibitory charge than PV-GPe neurons during the initial second of stimulation (Fig. 1G).

To determine whether these synaptic asymmetries could be leveraged to drive population-specific neuromodulation, we took sequential extracellular and intracellular recordings. Firing responses of PV-GPe and Lhx6-GPe neurons were heavily influenced by the dynamics of their underlying synaptic inputs (Fig. 1H). Large inhibitory currents in Lhx6-GPe neurons were well correlated with transient suppressions in extracellular firing rates, which diminished as inhibitory synapses depressed. By contrast, PV-GPe neurons—which lacked large inhibitory currents—were instantly excited by stimulation.

Extracellular recordings revealed additional differences between Lhx6-GPe and PV-GPe neurons after stimulation was initiated. Lhx6-GPe neurons were routinely inhibited during the initial second of stimulation, but less so in subsequent time bins. Notably, some neurons switched from inhibited to excited within 3 s of continuous stimulation (Fig. 1I and fig. S1D). By contrast, PV-GPe neurons were typically excited throughout (Fig. 1I and fig. S1D). The degree to which firing responses of PV-GPe and Lhx6-GPe neurons could be separated during stimulation was calculated as a population separation index (PSI) (Fig. 1I and supplementary methods). The PSI was maximal during the initial second of stimulation and decayed over time, largely a result of decreased suppression of Lhx6-GPe neurons (Fig. 1I).

These results led to the hypothesis that delivering stimulation in short bursts, rather than continuously, could achieve population-specific neuromodulation in the GPe. We therefore delivered repeated bursts (100 Hz, 1 s) to pairs of Lhx6-GPe and PV-GPe neurons. This stimulus pattern drove robust and reproducible suppression of Lhx6-GPe neurons and simultaneous excitation of PV-GPe neurons (Fig. 1, J and K).

To identify the range of parameter combinations that separate the responses of PV-GPe and Lhx6-GPe neurons, we used an iterative machine learning approach (see supplementary methods). Experimental data (Fig. 2A) were used to train models to predict potential nonlinear relationships between parameter combinations and firing responses of PV-GPe and Lhx6-GPe populations. In the first iteration, a dataset of 208 distinctive burst combinations (table S1), collected across 19 pairs of Lhx6-GPe and PV-GPe neurons, was used to train Gaussian process regressions with quadratic polynomial kernels for each population. Compared with other approaches, this approach yielded improved leave-one-out cross-validation performance (table S2). Parameters were then sampled (table S3) and evaluated to identify ranges of parameters predicted to yield

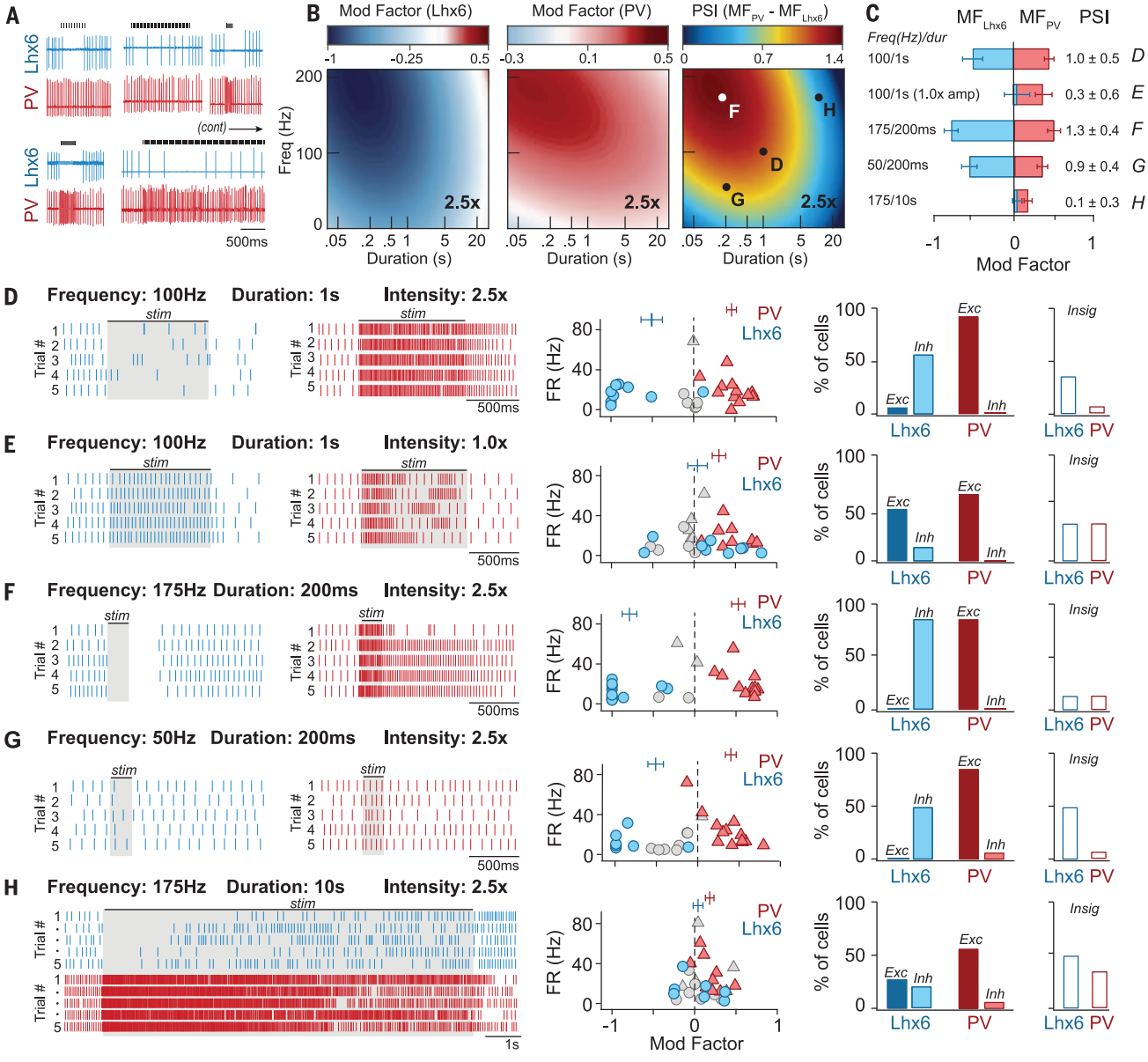


Fig. 2. Stimulus optimization with a machine learning approach accurately predicts burst designs that achieve population-specific neuromodulation in the GPe using electrical stimulation. (**A**) Samples of neural response data used to train models (supplementary methods). (**B**) Surface plots (2.5× intensity) showing predicted FR responses of Lhx6-GPe (MF_{Lhx6}, left) and PV-GPe (MF_{PV}, middle) populations across a range of stimulus parameters. Predicted increases in FR are shown in red, and predicted decreases in FR are shown in blue. (Right) Surface plot showing the degree to which stimulus parameters are predicted to segregate the responses of PV-GPe and Lhx6-GPe populations (PSI: red, more segregation; blue, less segregation). Lettered points indicate burst combinations selected for experimental validation [(D) to (G)]. (**C**) Experimentally measured responses of PV-GPe and Lhx6-GPe populations to burst protocols developed using our model (B). The degree to which burst designs segregated the responses of PV-GPe and Lhx6-GPe neurons (PSI) is shown to the right, with letters on right referencing matching panel letters. Error bars, SEM. (**D** to **H**) Experimental validation of model predictions. (Left) Rasters of extracellularly recorded FR from Lhx6-GPe (blue) and PV-GPe (red) neurons responding to burst designs drawn from regions of parameter space indicated in (B). Shaded boxes denote the time bin used to calculate modulation factors. (Middle) Modulation factors for each neuron (averaged across five trials), plotted against its baseline firing rate (14 pairs of neurons; four mice). Gray symbols, neurons in which firing rate was not significantly modified from baseline ($P > 0.05$, paired t test, $FR_{baseline}$ versus FR_{stim}, trials 1 to 5). Vertical markers show average MF_{Lhx6} and MF_{PV}, error bars: SEM. (Right) Bar graphs indicating the percentage of Lhx6-GPe and PV-GPe neurons that were excited (Exc), inhibited (Inh), or had no significant change in firing rate (Insig) during stimulation.

a large PSI. This range was then used to collect a second iteration of data (table S4), which were used to further refine the models (fig. S3, A and B). Because there was strong convergence of predictions between the first and second iterations of the models, no additional iterations were run. Surface plots with modulation factor predictions for PV-GPe and Lhx6-GPe populations across many thousands of artificially generated parameters, as well as the predicted PSI across the parameter space, are shown in Fig. 2B and fig. S2A.

As predicted by our model, 100 Hz, 1-s bursts, delivered in brain slices, were highly effective at segregating the responses of PV-GPe and Lhx6-GPe neurons at a stimulus intensity of 2.5×, but less effective at weaker stimulus intensities (Fig. 2, C to E). Furthermore, we improved

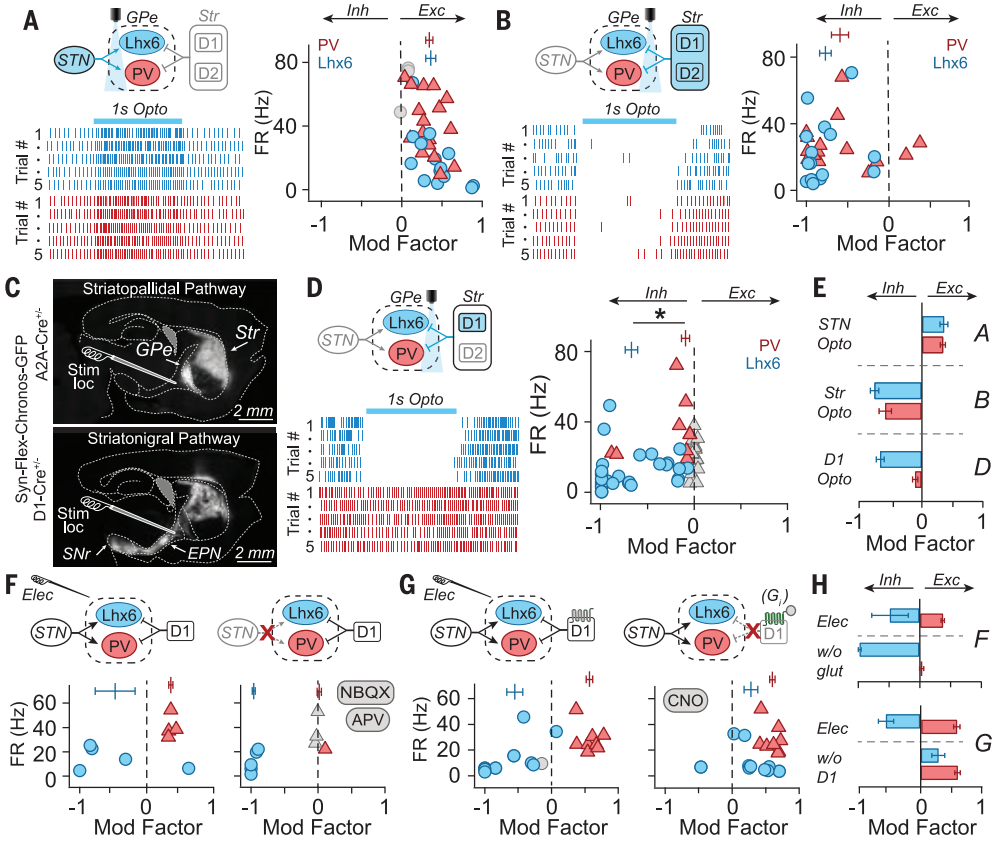


Fig. 3. Population-specific neuromodulation in the GPe is driven by convergent excitation from the STN and inhibition from D1-SPNs. (**A**) (Left) Schematic and rasters from two representative GPe neurons during optogenetic stimulation of STN fibers (100 Hz for 1 s). (Right) Modulation factors (MFs) for individual neurons (symbols) and populations (vertical bars), showing that PV-GPe and Lhx6-GPe neurons are similarly excited (MF_{Lhx6} = 0.36 ± 0.27, MF_{PV} = 0.34 ± 0.17; MWU P = 0.976; 17 pairs of neurons; three mice). (**B**) (Left) Schematic and rasters from two neurons during striatal fiber stimulation (100 Hz for 1 s). (Right) MFs show that both PV-GPe and Lhx6-GPe neurons are similarly inhibited by striatal fiber stimulation (MF_{Lhx6} = −0.77 ± 0.29, MF_{PV}: −0.59 ± 0.41; MWU P = 0.25; 16 pairs; two mice). (**C**) (Top) Fluorescent image of striatopallidal pathway. (Bottom) Fluorescent image of striatonigral pathway. Typical placement of the stimulating electrode is shown for reference. (**D**) (Left) Schematic and rasters from two neurons during D1-SPN striatal fiber stimulation. (Right) MFs show that Lhx6-GPe neurons are preferentially inhibited (MF_{Lhx6} = −0.68 ± 0.34; MF_{PV} = −0.1 ± 0.23; MWU *P < 0.00001; 27 pairs; four mice). (**E**) Summary of MF_{Lhx6} (blue) and MF_{PV} (red) for experiments A to D; error bars: SEM. (**F**) MFs in response to electrical stimulation before (left) and after (right) application of 10 μM NBQX/50 μM APV. Excitation of PV-GPe neurons was blocked (MF_{Ctrl}: 0.36 ± 0.05, $MF_{NBQX/APV}$: 0.01 ± 0.07, paired t test, P = 0.0001), but not Lhx6-GPe inhibition (MF_{Ctrl}: −0.47 ± 0.66, $MF_{NBQX/APV}$: −0.96 ± 0.05, paired t test, P = 0.18) (5 Lhx6-GPe neurons, 4 PV-GPe neurons; two mice). (**G**) MFs in response to electrical stimulation before (left) and after (right) chemogenetic inhibition of D1-SPN fibers [AAV2-hsyn-DIO-hM4D(G_i)-mCherry + CNO, see methods]. Inhibition of Lhx6-GPe neurons was blocked (MF_{pre} = −0.54 ± 0.39, MF_{CNO} = 0.28 ± 0.33, paired t test, P = 0.003), but excitation of PV-GPe neurons was not (MF_{pre} = 0.57 ± 0.15, MF_{CNO} = 0.59 ± 0.13, paired t test, P = 0.7) (n = 10 Lhx6-GPe neurons, n= 8 PV-GPe neurons; n = five mice). (**H**) Summary of MF_{Lhx6} (blue) and MF_{PV} (red) for experiments F to G. Error bars: SEM.

our experimentally measured PSI by using shorter, higher-frequency bursts (175 Hz, 200 ms; Fig. 2, B, C, and F). The model also accurately predicted the boundaries of parameter combinations that could dissociate the responses of PV-GPe and Lhx6-GPe neurons. For example, reducing intraburst frequency to 50 Hz while keeping a 200-ms duration (50 Hz, 200 ms) weakened the PSI but did not eliminate it entirely (Fig. 2, B, C, and G), whereas increasing the burst duration (175 Hz, 10 s) prevented stimulation from driving population-specific responses (Fig. 2, B, C, and H). Deviations from model predictions for experimental outcomes were quantified with root mean squared error (fig. S3, C and D, and table S5).

Given the critical role of inhibitory synaptic transmissions in differentiating the responses of PV-GPe and Lhx6-GPe neurons, we sought to identify the mechanism underlying its cell type specificity. To test the hypothesis that Lhx6-GPe inhibition was driven by the local GPe collateral network, engaged by excitation from the subthalamic nucleus (STN) (*8*, *9*), we used optogenetics to selectively stimulate STN afferents. The fast channelrhodopsin variant Chronos enabled stimulation of STN fibers at high frequencies (*10*). However, 1-s bursts of STN stimulation (100 Hz) excited both populations of GPe neurons to a similar degree (Fig. 3, A and E).

Next, we tested the hypothesis that inhibition from the striatum (*11*) was weighted more heavily toward Lhx6-GPe neurons than PV-GPe neurons. Chronos was globally expressed in the striatum (Fig. 3, B and C). Optogenetic stimulation (100 Hz, 1 s) of striatal afferents in the GPe similarly inhibited both PV-GPe and Lhx6-GPe neurons (*12*–*14*) (Fig. 3, B and E).

These results seem to suggest that striatal inputs lack the cell type specificity needed for population-specific neuromodulation. However, considering the placement of our stimulating electrode (Fig. 3C), electrical stimulation was unlikely to engage all striatal cell types equally. Instead, electrical stimulation was likely biased toward antidromic activation of D1-type dopamine receptor SPNs (D1-SPNs), which are not a canonical source of inhibitory input to the GPe, although they do make some synaptic contacts (*12*, *15*–*18*). To test the hypothesis that striatal afferents from D1-SPNs preferentially inhibit Lhx6-GPe neurons, Chronos expression was restricted to D1-SPNs (Fig. 3D) in the striatum. This time, optogenetic stimulation of D1-SPN afferents differentiated the responses of GPe neurons, driving robust inhibition of Lhx6-GPe but not PV-GPe neurons (Fig. 3, D and E).

These results suggest that the circuit mechanism by which electrical stimulation drives population-specific neuromodulation was through the converging effects of excitatory inputs from the STN and inhibitory inputs from D1-SPNs. We therefore assessed the effects of systematically blocking each input during electrical stimulation. Blocking excitatory inputs eliminated the excitation of PV-GPe neurons during stimulation and unmasked even greater inhibition of Lhx6-GPe neurons (Fig. 3, F and H). Conversely, chemogenetic suppression of D1-SPN fibers during electrical stimulation blocked the inhibitory response of

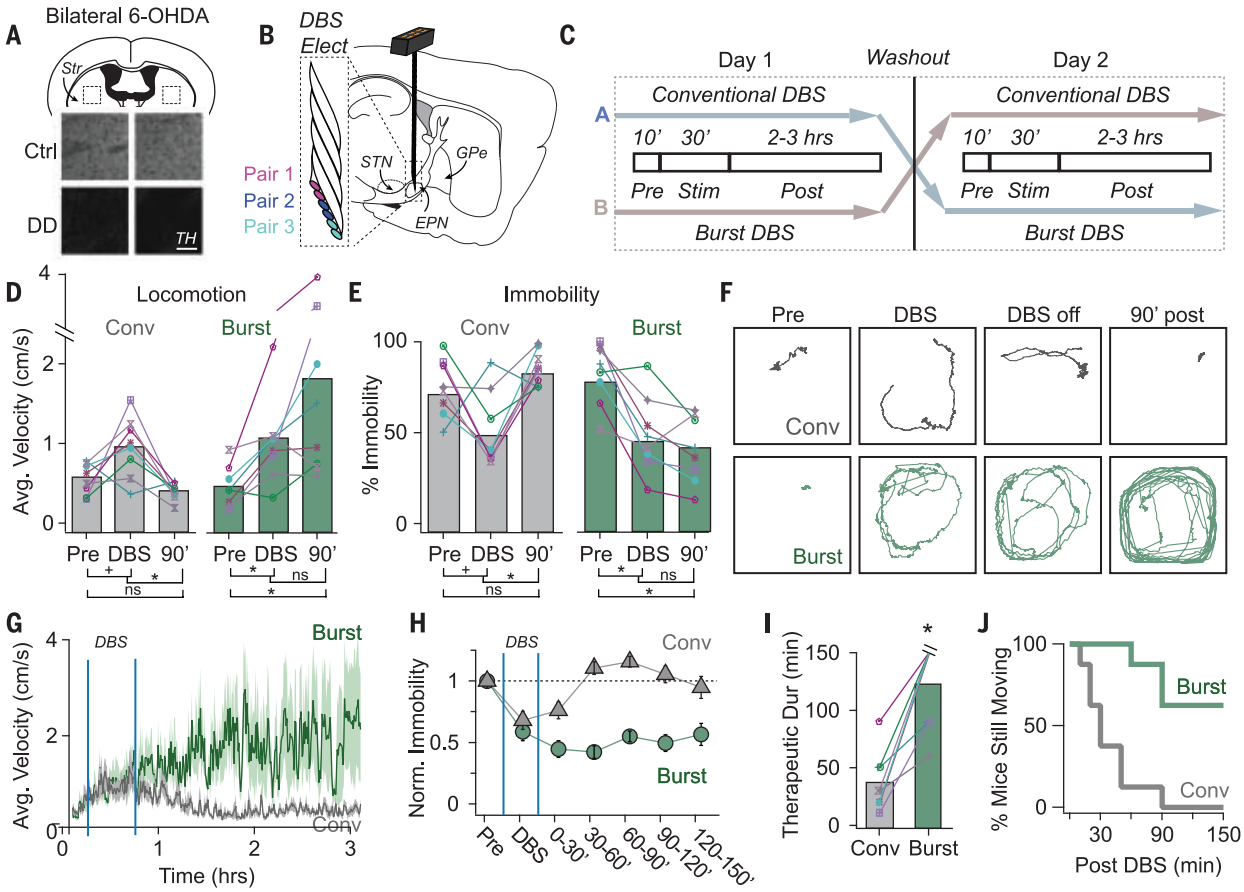


Fig. 4. Burst stimulation restores movement persistently in parkinsonian mice. (**A**) Schematic showing bilateral dopamine depletion (6.43 ± 0.09% tyrosine hydroxylase (TH) remaining compared with littermate controls). (**B**) Schematic showing DBS electrode placement near the EPN. (**C**) Cross-over design illustrating a pseudorandom experimental protocol in which DD mice receive conventional or burst DBS on alternate days. (**D**) Average velocity of DD mice increased during conventional DBS (gray) but did not persist after stimulation (Pre: 0.54 ± 0.07, Stim: 0.94 ± 0.13, Post: 0.40 ± 0.06 cm/s) [Wilcoxon signed rank (WSR): Pre versus Stim, ^{+}P = 0.055; Stim versus 90 min (90') $^{*}P$ = 0.008; Pre versus 90'. P = 0.195, n = eight mice]. Average velocity increased during burst DBS (green) and was still elevated 90 min after stimulation (Pre: 0.47 ± 0.09, Stim: 1.03 ± 0.19, Post: 1.77 ± 0.54 cm/s) (WSR: Pre versus Stim $^{*}P$ = 0.016; Stim versus 90' P = 0.195; Pre versus. 90' $^{*}P$ = 0.008). Data from individual mice are shown, and colors and symbols are consistent throughout (ns, not significant). (**E**) Immobility of DD mice decreased during conventional DBS (gray) but did not persist (Pre: 75 ± 6 Stim: 51 ± 7, Post: 87 ± 4% time immobile) (WSR: Pre versus Stim, ^{+}P = 0.056 Stim versus 90' $^{*}P$ = 0.008, Pre versus 90' = 0.195). Immobility decreased during burst DBS (green) and was still low 90 min after stimulation (Pre: 82 ± 6, Stim: 48 ± 7, Post: 45 ± 9% time immobile) (WSR: Pre versus Stim, $^{*}P$ = 0.016; Stim versus 90'; P = 0.641; Pre versus. 90' $^{*}P$ = 0.016). (**F**) Movement paths over 8-min intervals throughout one trial. Data are from the same mouse treated with conventional DBS (gray) or burst DBS (green). (**G** to **H**) Average movement velocities (G) and immobility (H) are plotted for the duration of behavioral trials. Movement velocities were averaged over 30-s time bins and percent immobility was averaged over 30-min time bins. Gray, conventional DBS; green, burst DBS (n = eight mice). (**I** to **J**) Amount of time after DBS before mice returned to pre-stim levels of immobility ("therapeutic duration"). Therapeutic duration was significantly longer after burst DBS (123.8 ± 37.4 min, green) than after conventional DBS (37.5 ± 25.5 min, gray) (WSR, $^{*}P$ = 0.012, n = eight mice), shown for individual mice (I) and cumulatively across the population (J).

Lhx6-GPe neurons, unmasking an excitatory effect (Fig. 3, G and H).

Finally, to test whether an electrical DBS protocol tailored to drive population-specific neuromodulation would mimic the persistent therapeutic effects induced by cell-type–specific optogenetic manipulations (*7*), we compared the therapeutic efficacy of conventional versus burst DBS (supplementary methods) in mice rendered parkinsonian by bilateral injections of the toxin 6-hydroxydopamine (6-OHDA) into the medial forebrain bundle (MFB) (Fig. 4A, supplementary methods). Burst DBS consisted of short bursts delivered once a second with a duration of 200 ms and an intraburst frequency of 175 Hz. Recordings in brain slices confirmed that the population specificity of this protocol was retained when delivered with a biphasic stimulating electrode, as required in vivo, both for naïve tissue and in tissue from dopamine-depleted (DD) mice (fig. S4, A and B).

The therapeutic effects of burst versus conventional DBS were then compared using a cross-over study design with a 1-day washout period between treatments in DD mice (Fig. 4C), using an approach shown to recapitulate a number of features of DBS in humans (*19*). Six-lead DBS electrodes were implanted bilaterally near the entopeduncular nucleus (EPN) (Fig. 4B and fig. S5), an electrode placement often used in human patients receiving DBS in the internal globus pallidus (GPi) (*20*, *21*) and an area where fibers from the STN and D1-SPNs can be coactivated.

Before stimulation, DD mice were highly immobile. Both conventional and burst DBS reduced this immobility, enabling mice to move around the arena during stimulation (Pre and DBS in Fig. 4, D to F, and movies S1 and S2). After receiving conventional DBS, mice quickly returned to the immobile state (Fig. 4, D to H). Conversely, after receiving burst DBS, mice continued to move around the arena (Fig. 4, D to H), with five out of the eight mice still moving at the end of the trial,

~2.5 hours after stimulation (Fig. 4, I and J, 90'+, movies S1 and S2). On average, the therapeutic effects of burst DBS persisted >4.5 times as long as those of conventional DBS (Fig. 4I), a conservative estimate as only 37% of mice had returned to the immobile state at the end of the trial (Fig. 4J). DBS protocols predicted to be less effective at dissociating the responses of PV-GPe, and Lhx6-GPe neurons were less effective at inducing persistent rescue (fig. S5).

These results demonstrate how fundamental knowledge about the organization and function of basal ganglia circuitry can be used to refine the population specificity of electrical stimulation, ultimately prolonging the therapeutic benefits of DBS beyond those achieved with conventional methods. Other alternative forms of DBS, including adaptive and coordinated reset DBS (*22*–*25*), can induce persistent therapeutic effects with the use of specialized implants, and it is plausible that the long-lasting effect observed in our study shares a common underlying mechanism with these approaches. However, our burst DBS protocol can be delivered through commonly used DBS implants and falls within US Food and Drug Administration–approved stimulus frequencies, enabling immediate testing in PD models across species, including human patients. This work joins the growing field of opto-inspired DBS (*26*–*28*), in which fundamental discoveries about the organizing principles of neural circuits gained through the use of optogenetics guide the development of more robust electrical DBS approaches that can be rapidly translated to humans.

REFERENCES AND NOTES

1. T. Guo *et al*., *Front. Neurosci.* **13**, 413 (2019).
2. C. A. Anastassiou, S. M. Montgomery, M. Barahona, G. Buzsáki, C. Koch, *J. Neurosci.* **30**, 1925–1936 (2010).
3. C. C. McIntyre, W. M. Grill, *Ann. Biomed. Eng.* **28**, 219–233 (2000).
4. T. Radman, R. L. Ramos, J. C. Brumberg, M. Bikson, *Brain Stimul.* **2**, 215–228.e3 (2009).
5. C. C. McIntyre, W. M. Grill, *J. Neurophysiol.* **88**, 1592–1604 (2002).
6. P. Temperli *et al*., *Neurology* **60**, 78–81 (2003).
7. K. J. Mastro *et al*., *Nat. Neurosci.* **20**, 815–823 (2017).
8. C. Miguelez *et al*., *J. Physiol.* **590**, 5861–5875 (2012).
9. H. Kita, Y. Tachibana, A. Nambu, S. Chiken, *J. Neurosci.* **25**, 8611–8619 (2005).
10. N. C. Klapoetke *et al*., *Nat. Methods* **11**, 338–346 (2014).
11. D. R. Smith *et al*., *Neuroscience* **86**, 135–146 (1998).
12. M. Cazorla *et al*., *Neuron* **81**, 153–164 (2014).
13. X.-S. Yuan *et al*., *eLife* **6**, e29055 (2017).
14. R. F. Kovaleski *et al*., *J. Physiol.* **598**, 1897–1927 (2020).
15. Y. Kawaguchi, C. J. Wilson, P. C. Emson, *J. Neurosci.* **10**, 3421–3438 (1990).
16. M. Ketzef, G. Silberberg, *Neuron* **109**, 516–529.e4 (2021).
17. Q. Cui *et al*., *J. Neurosci.* **41**, 3966–3987 (2021).
18. Y. Wu, S. Richard, A. Parent, *Neurosci. Res.* **38**, 49–62 (2000).
19. J. S. Schor, A. B. Nelson, *J. Clin. Invest.* **129**, 3833–3838 (2019).
20. K. L. K. Au *et al*., *Neurol. Ther.* **10**, 7–30 (2021).
21. J. L. Vitek, T. Hashimoto, J. Peoples, M. R. DeLong, R. A. Bakay, *Mov. Disord.* **19**, 907–915 (2004).
22. I. Adamchic *et al*., *Mov. Disord.* **29**, 1679–1684 (2014).
23. P. A. Tass *et al*., *Ann. Neurol.* **72**, 816–820 (2012).
24. J. Wang *et al*., *Brain Stimul.* **9**, 609–617 (2016).
25. B. Rosin *et al*., *Neuron* **72**, 370–384 (2011).
26. M. Creed, V. J. Pascoli, C. Lüscher, *Science* **347**, 659–664 (2015).
27. A. H. Gittis, E. A. Yttri, *Curr. Opin. Biomed. Eng.* **8**, 14–19 (2018).
28. S. Valverde *et al*., *Nat. Commun.* **11**, 2388 (2020).
29. T. Spix *et al*., *Zenodo* (2021); doi:10.5281/zenodo.5211556.
30. T. Spix *et al*., *Zenodo* (2021); doi:10.5281/zenodo.5218517.

ACKNOWLEDGMENTS

We thank J. Schwenk and C. Snyder for expert animal care and technical assistance, J. Schorr for DBS training and guidance, and R. Gerkin for specialized IGOR analysis and acquisition routines. **Funding:** This work was supported by the Richard King Mellon Foundation Presidential Fellowship in the Life Sciences (T.A.S.), Lane Fellowship in Computational Biology (I.M.K.), Michael J. Fox Foundation (A.H.G.), National Institutes of Health grant T32GM008208 (S.N.), and National Institutes of Health grants R01NS101016, R01NS104835, and R01NS117058 (A.H.G.) **Author contributions:** Conceptualization and Methodology: A.H.G., T.A.S., B.R.I., and A.R.P. Investigation: A.H.G., T.A.S., S.N., A.R.P., N.T., I.M.K., and Y.G. Visualization: A.H.G., T.A.S., B.R.I., and S.N. Funding acquisition: A.H.G. and T.A.S. Project administration: A.H.G. Writing: A.H.G., T.A.S., S.N., I.M.K., and N.T. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The processed data used to generate the figure panels (*29*) and the code used for Gaussian regression modeling (*30*) can be found at Zenodo.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi7852
Materials and Methods
Figs. S1 to S6
Tables S1 to S6
References (*31*–*35*)
MDAR Reproducibility Checklist
Movies S1 and S2

30 March 2021; accepted 18 August 2021
10.1126/science.abi7852

METALS IN BIOLOGY

Microfluidic-like fabrication of metal ion–cured bioadhesives by mussels

Tobias Priemel[1], Gurveer Palia[1], Frank Förste[2], Franziska Jehle[1,3], Sanja Sviben[3], Ioanna Mantouvalou[2]†, Paul Zaslansky[4], Luca Bertinetti[3]‡, Matthew J. Harrington[1]*

To anchor in seashore habitats, mussels fabricate adhesive byssus fibers that are mechanically reinforced by protein-metal coordination mediated by 3,4-dihydroxyphenylalanine (DOPA). The mechanism by which metal ions are integrated during byssus formation remains unknown. In this study, we investigated the byssus formation process in the blue mussel, *Mytilus edulis*, combining traditional and advanced methods to identify how and when metals are incorporated. Mussels store iron and vanadium ions in intracellular metal storage particles (MSPs) complexed with previously unknown catechol-based biomolecules. During adhesive formation, stockpiled secretory vesicles containing concentrated fluid proteins are mixed with MSPs within a microfluidic-like network of interconnected channels where they coalesce, forming protein-metal bonds within the nascent byssus. These findings advance our understanding of metal use in biological materials with implications for next-generation metallopolymers and adhesives.

The mussel byssus is a well-established role model for next-generation polymeric materials, including self-healing materials, advanced coatings, and, especially, bio-inspired adhesives (*1*, *2*). In particular, the discovery and characterization of protein-metal coordination complexes as dynamic load-bearing cross-links in the byssus (Fig. 1A) (*3*–*5*) triggered a surge in development of mussel-inspired metallopolymeric materials (*1*, *2*, *6*). The byssus consists of protein-based fibers produced individually through the secretion of around 10 different protein precursors into a groove in a specialized secretory organ called the mussel foot (Fig. 1B) (*7*–*9*). At the distal end of the byssal thread, a flattened disk known as the plaque adheres strongly to almost any surface under seawater conditions (*1*, *8*). Biochemical and functional investigations revealed that most plaque proteins contain high concentrations of 3,4-dihydroxyphenylalanine (DOPA), which contains a catechol group produced by posttranslational hydroxylation of tyrosine (*1*, *8*). DOPA interacts with different surface chemistries through various physical and chemical interactions, playing a crucial role in adhesion. Additionally, the plaque is cured and mechanically stabilized through the formation of DOPA-metal coordination bonds (Fig. 1A) that provide strong, yet reversible, cohesive interactions between the plaque proteins (*10*). Although it was initially believed that mussels preferentially use iron as a coordinating metal ion, recent evidence suggests that mussels are opportunistic, possibly using other metal ions such as vanadium and aluminum (*11*–*13*). Despite the functional importance of these metal ions and the relevance to ongoing efforts to produce mussel-inspired materials, the pathway by which metals are introduced into the byssus remains unknown.

[1]Department of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, Quebec H3A 0B8, Canada. [2]Institute of Optics and Atomic Physics, Technische Universität Berlin, Hardenbergstrasse 36, 10623 Berlin, Germany. [3]Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Potsdam, Germany. [4]Department for Restorative and Preventive Dentistry, Charité-Universitätsmedizin Berlin, 14197 Berlin, Germany.
***Corresponding author. Email: matt.harrington@mcgill.ca**
†Present address: Helmholtz Zentrum Berlin, Albert-Einstein-Str. 15, 12489 Berlin, Germany.
‡Present address: Technische Universität Dresden, 01069 Dresden, Germany.

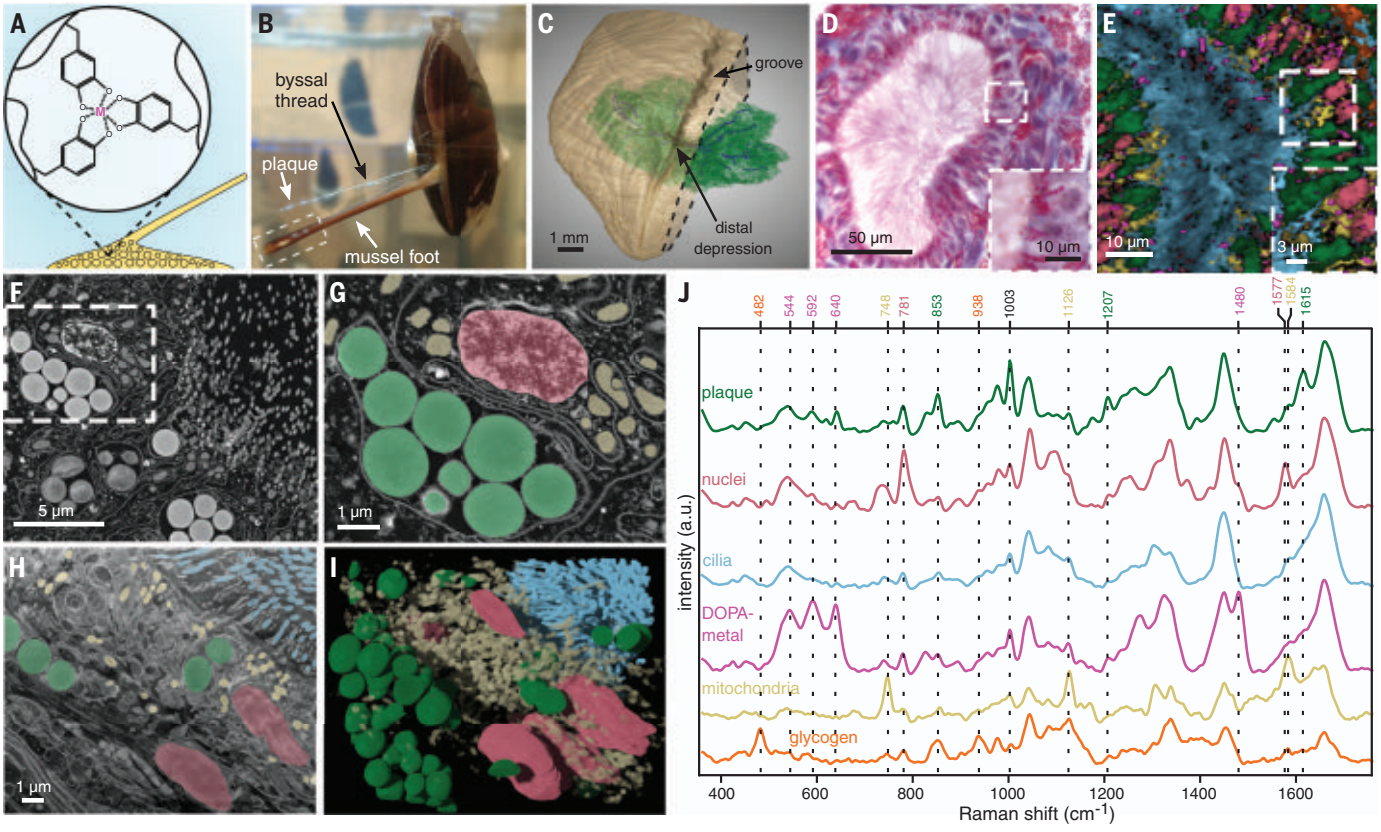


Fig. 1. Characterization of the LDs found in the mussel foot. (**A**) Schematic of a byssal thread showing the DOPA–metal (M) coordination cross-links that are crucial for the mechanical properties of the porous plaque adhesive. (**B**) Photo of a mussel in the process of producing a thread. The mussel is attached with several byssal threads and has its foot fully extended. (**C**) Micro-CT reconstruction of the distal mussel foot from a region similar to that in the dashed box in (B). The right-hand side is cut away, revealing the plaque gland in green and the LDs in blue; the black dashed line indicates the cut-away line. (**D**) Cross section of a LD stained with Masson's trichrome. The inset shows a magnified image of the area enclosed by the dashed box. (**E**) Confocal Raman map of an unstained LD similar to (D) showing the localization of plaque proteins (green), nuclei (red), cilia (blue), metal storage vesicles (pink), mitochondria (yellow), and glycogen (orange). The colors correspond to the spectra in (J). The inset shows a magnified image of the area enclosed by the dashed box. (**F**) STEM image of part of the LD showing cell morphology and cell components. (**G**) Magnified STEM image from the dashed box in (F) highlighting plaque vesicles in green, cell nucleus in red, and mitochondria in yellow. (**H**) FIB-SEM image of a LD highlighting cell components in similar color to (G). (**I**) 3D reconstruction of the FIB-SEM dataset from (H). (**J**) Raman spectra corresponding to cell components shown in (E). a.u., arbitrary units.

The byssus plaque is formed in just a few minutes from an acidic secretion of fluid protein precursors into a small divot at the tip of the foot on the ventral side that is known as the distal depression (Fig. 1, B and C) (*8*, *9*). Before formation, plaque proteins are stockpiled within micron-scale secretory vesicles inside specialized foot gland cells surrounding the distal depression (*9*, *14*). Plaque vesicle contents are highly concentrated and fluid in nature, but during fabrication, they rapidly cure and cross-link, forming a solid, open-cell, porous network (*8*, *9*, *15*). Mature plaques are mechanically stabilized through DOPA-metal coordination cross-links (*10*); yet, the proteins stored in the plaque vesicles do not exhibit metal coordination, suggesting that metals are introduced during formation (*9*, *12*). Initial work published nearly 50 years ago suggested that the plaque vesicles are transported to the distal depression via ciliated channels known as the longitudinal ducts (LDs) that extend from the plaque gland (*16*). However, there has been almost no advancement since then in understanding the cellular secretory pathway and the dynamic role of the LDs in the assembly process—in particular, how the fluid plaque proteins mature into a metal cross-linked porous adhesive during their journey along the LDs. This information will illuminate the mechanisms by which living organisms use metal coordination bonds to fabricate robust materials, which will find direct relevance in the highly active field of mussel-inspired metallopolymers, as well as bio-inspired soft matter in general (*1*, *2*, *6*). Here, we used a combination of methods, including micro–computed tomography (micro-CT), light microscopy, electron microscopy, x-ray fluorescence microscopy, and confocal Raman spectroscopic imaging to examine the tissue-based machinery responsible for this notable feat of adhesive metallopolymer biomanufacturing in the blue mussel, *Mytilus edulis*.

Micro-CT tomographic reconstruction of the mussel foot reveals a complex network of LDs running through the plaque gland that converges at the distal depression where the plaque is formed (Fig. 1C and movie S1), consistent with the previously proposed role of the LDs as a transport network for the secreted plaque proteins (*16*). Light microscopy and scanning transmission electron microscopy (STEM) of foot sections reveal that LD microchannels possess a circular cross section between 10 and 100 µm in diameter formed by a distinctive columnar epithelial tissue with many cilia and microvilli (Fig. 1, D and F). The cilia and microvilli fill almost the entire lumen of the duct, with neighboring cilia lining up in the same

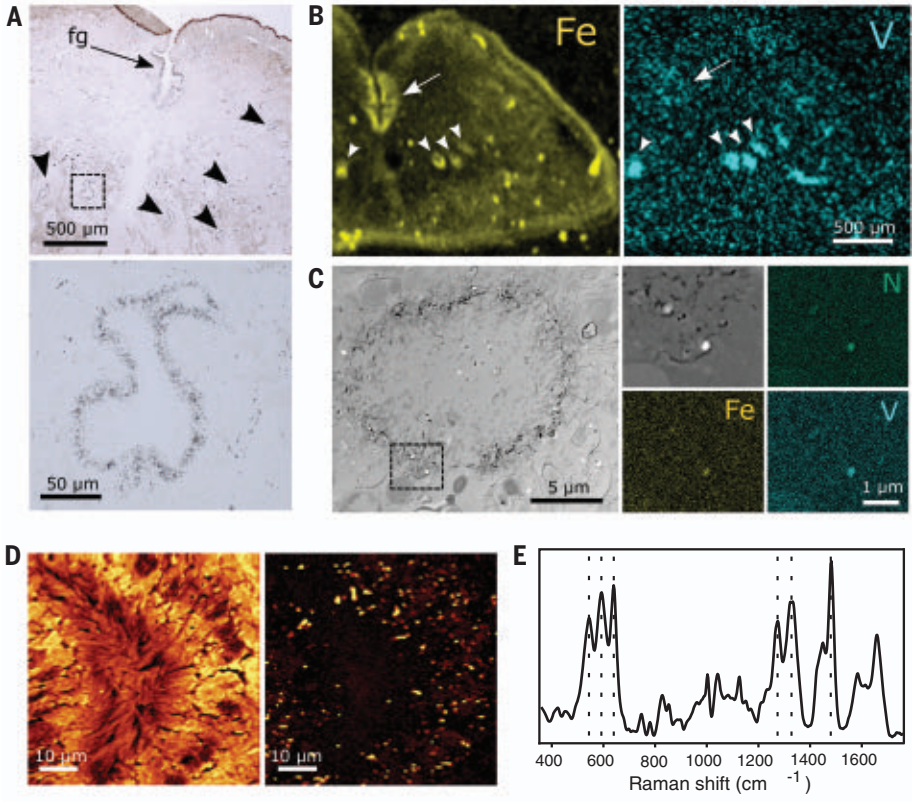


Fig. 2. Metal storage around the LDs. (**A**) Autometallography staining of a mussel foot cross section with positive staining around the LDs (arrowheads) and around the mussel foot groove (fg) (top), with a magnified image of the area enclosed by the dashed box shown directly beneath. (**B**) μXRF measurements on 200-μm-thick mussel foot cross sections showing the distribution of iron (yellow) and vanadium (turquoise), with arrowheads indicating the LDs and an arrow indicating the groove. (**C**) STEM-EDS analysis of a metal storage vesicle from the region enclosed by the dashed box at the periphery of a LD showing distribution of nitrogen in green, iron in yellow. and vanadium in turquoise. (**D**) Raman maps of a LD corresponding to the CH-stretch (2900 to 3000 cm^{-1}) on the left side and DOPA-metal coordination (500 to 700 cm^{-1}) on the right side. (**E**) Raman spectrum of the metal storage vesicles showing DOPA-metal coordination resonance peaks highlighted by dashed lines.

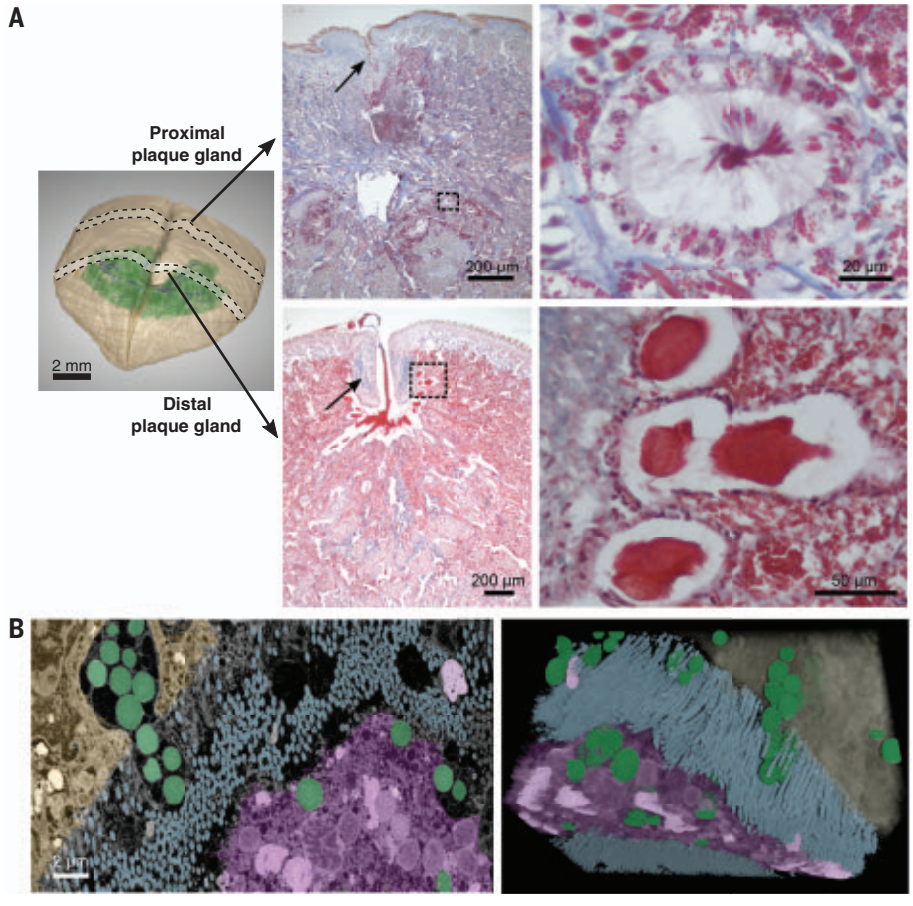


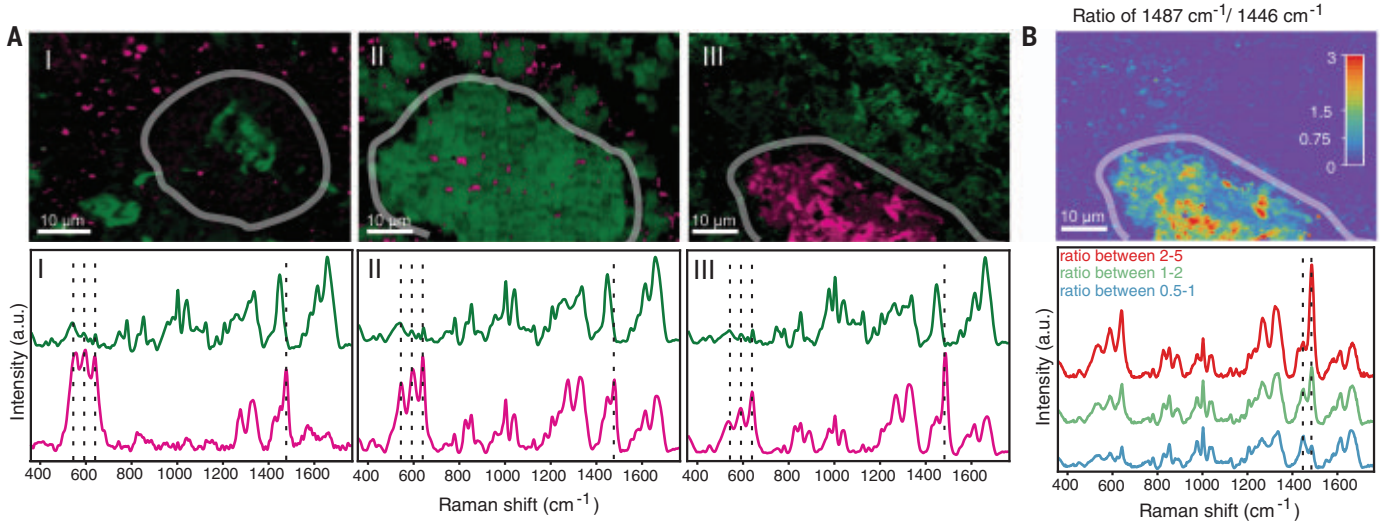
Fig. 3. Secretion of the plaque proteins in the LDs. (**A**) Masson's trichrome staining of mussel foot cross sections after induction of secretion acquired from the proximal (top) and distal (bottom) regions of the plaque gland. The magnified images on the right correspond to the dashed boxes in the left panels and indicate the secretion of plaque proteins into the ducts. (**B**) FIB-SEM analysis of an induced LD showing the secretion of plaque vesicles (green) into the lumen of the LD where they coalesce to form the nascent plaque (purple). Cilia (blue) all line up in the direction of secretion. The 2D image is provided on the left, and the 3D reconstruction is on the right.

direction (Fig. 1I), making them practical for transporting materials similar to other ciliated tissue such as bronchial or fallopian tubes (*17*).

The LDs are surrounded by elongated plaque gland cells that contain large quantities of spherical plaque vesicles, as visualized by Masson's trichrome staining (Fig. 1D). Higher-resolution imaging with STEM and focused ion beam scanning electron microscopy (FIB-SEM) revealed that cells are densely packed with clusters of vesicles, which line up between the epithelial cells ready for secretion (Fig. 1, F and H). Reconstruction of the FIB-SEM dataset further shows the three-dimensional (3D) distribution of the vesicles around the LD that are ready for release into the lumen (Fig. 1I and movie S2). Spectroscopic analysis with confocal Raman spectroscopy reveals that vesicles contain plaque proteins, identified by their Raman spectral fingerprint as described previously (*9*), which shows a prevalence of the aromatic amino acids tyrosine and/or DOPA (853, 1207, and 1615 cm^{-1}) and phenylalanine (1003 cm^{-1}) (Fig. 1, E and J, and figs. S1 and S2).

The epithelial cells surrounding the LDs also have a distinctive subcellular structure revealed by STEM and SEM (Fig. 1, G and H), which contains large cell nuclei and numerous mitochondria. These assignments were confirmed by Raman spectroscopy, which identified cell nuclei by the vibrational peaks of DNA (781 and 1577 cm^{-1}) and mitochondria by the resonant peaks of cytochrome c (748, 1126, and 1584 cm^{-1}) (Fig. 1J) (*18*–*21*). The large number of mitochondria are likely necessary to produce the energy needed for active cilia movement. Additionally, glycogen as a source of energy was detected by Raman spectroscopy based on a strong peak at 482 cm^{-1} (Fig. 1, E and J) (*18*, *19*).

As previously reported (*9*, *12*), Raman spectra acquired from the plaque secretory vesicles show no evidence of DOPA-metal coordination resonance peaks (Fig. 1J), although this is a prominent feature of the native plaque Raman spectrum (*10*). However, Raman spectroscopic imaging did identify the presence of DOPA-metal coordination outside of the vesicles in submicron spots surrounding the LDs (Fig. 1, E and J), as indicated by a highly distinctive set of intense Raman vibrations centered around 544, 592, and 640 cm^{-1} (catechol-metal interaction) and 1480 cm^{-1} (ring vibration), which result from the well-characterized resonance Raman enhancement of the vibrational modes associated with the



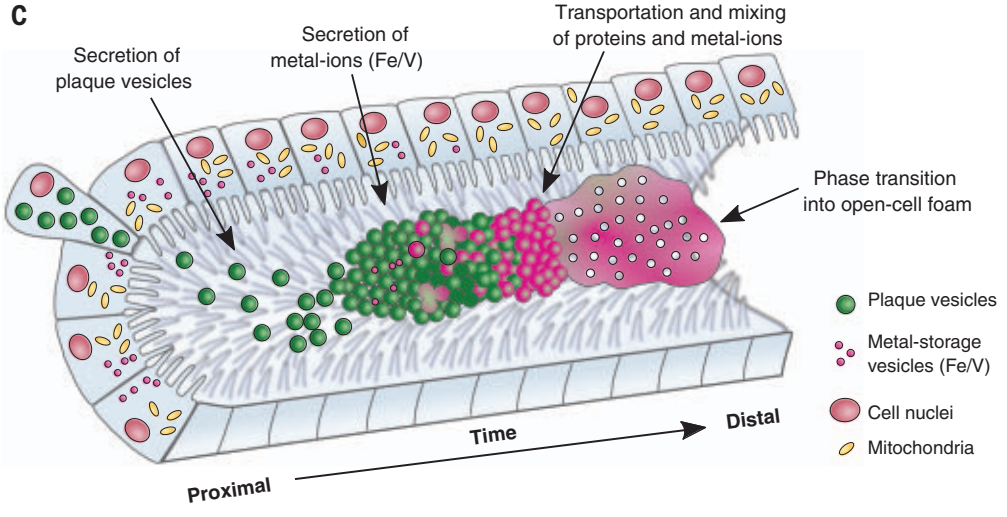


Fig. 4. Secretion of the MSPs in the LDs. (**A**) Raman image maps showing the distribution of plaque proteins in green and DOPA-metal coordination in pink at different stages of the secretion process. Panels I to III indicate increasing proximity to the distal depression. The boundary of the LD is highlighted with a white line. Corresponding average Raman spectra are shown below. (**B**) Raman spectral analysis of the induced plaque material from panel III in (A) showing the distribution of coordinated metal ions through the ratio of the DOPA-metal resonance peak (1487 cm^{-1}) to a protein peak (1446 cm^{-1}). Low concentrations are highlighted with cool colors (purple to green) and high concentrations with warm colors (yellow to red). Corresponding Raman spectra (averages of different ratios) are shown below. (**C**) Schematic model of the secretion process during plaque formation. The mussel secretes plaque vesicles and MSPs into the lumen of the LD microchannels where they are transported and mixed by active cilia movement. The Fe and V metal ions diffuse out of the MSPs and distribute throughout the dense protein phase, eventually forming pH-dependent DOPA-metal coordination bonds that mechanically fortify the porous plaque structure once it is released into seawater.

catechol-metal center (*3*, *22*). Although various metal ions, including Fe, V, Zn, and Cu, have been previously detected in *Mytilus* soft tissue using elemental analysis (*23*, *24*), this finding clearly localizes metal ions within the foot plaque gland, and, more specifically, surrounding the LDs.

To further investigate possible storage and secretion of metal ions, we used highly sensitive methods, including autometallography (AMG), micro–x-ray fluorescence (μXRF), and STEM with energy dispersive x-ray spectroscopy (STEM-EDS). AMG is an extremely sensitive histological staining method in which heavy metal ions are first transformed into metal sulfides, which then reduce silver ions in the presence of hydroquinone (*25*, *26*). Positive AMG staining was observed around the ventral groove (where the thread is formed) and directly around the LDs (Fig. 2A), confirming the presence of heavy metal ions in these regions. Because AMG is not metal-specific, μXRF imaging of 200-μm-thick dried mussel foot sections was performed to determine the identity of the metal ions. Although some impurities were apparent, consecutive serial sections show clear occurrence of iron and vanadium localized around the LDs in different mussel feet ($N = 3$) (Fig. 2B and fig. S3).

Localization of metal at higher spatial resolution was achieved by STEM-EDS of unstained foot sections and revealed the presence of vanadium and iron concentrated within bright (high electron density) submicron-sized particles surrounding the LDs, which we name here as metal storage particles (MSPs) (Fig. 2C). Based on their size and distribution, the MSPs are presumably the same submicron particles that were detected around the LDs using confocal Raman spectroscopic mapping (Figs. 1E and 2D). Furthermore, because of the presence of Raman resonance peaks specific to DOPA-metal coordination (Fig. 2E), these findings indicate that the metals in the MSPs are very likely associated with an organic component, coordinated with catechol groups in a bidentate fashion (*22*, *27*). This is further supported by the detection of nitrogen, oxygen, and carbon in the MSPs by STEM-EDS (Fig. 2C and fig. S4).

The LDs comprise a complex network of interconnected microchannels surrounded by the stockpiled building blocks (proteins and metals) required to build the plaque and equipped with the cilia needed to transport this cargo to its destination at the distal depression. To understand how these parts work together dynamically during plaque formation, we induced the plaque secretion process by injecting a 0.56 M KCl solution into the pedal ganglion at the base of the foot, as described previously (*9*, *16*). We then froze the secretion process at various time points and analyzed the tissue using histological imaging, FIB-SEM, and confocal Raman spectroscopic imaging. Histological imaging of induced feet shows that plaque vesicles are secreted into the LDs where they coalesce into a condensed fluid mass that stains consistently and homogeneously and that increases in size closer to the distal depression (Fig. 3A). FIB-SEM reconstruction reveals numerous vesicles being secreted through the cilia and mixing together in the lumen (Fig. 3B and movie S3). In the nascent induced plaque, a number of vesicles are intact, whereas the contents of other vesicles have coalesced into a fluid mass as described above, with some evidence of nanoporosity observed, reminiscent of the porous native plaque structure (*15*). Raman spectroscopic imaging of the coalescing mass in the LD lumen at early stages of secretion indicates a homogeneous composition with a spectrum nearly identical to that observed in the plaque vesicles (Fig. 4A).

Additionally, Raman imaging of induced plaques provides insights into secretion of the MSPs and their mixing with the plaque proteins (Fig. 4A). Although artificial thread induction is an unregulated process that deviates substantially from native thread formation, we observed different stages of this process that coincide with time after induction and position within the LD (i.e., closer to or further from the distal depression). At early secretion stages, distant from the distal depression, plaque proteins were present in the ducts, whereas MSPs were relegated to the surrounding tissue (Fig. 4A, panel I). At later stages, closer to the distal depression, MSPs appear to also be secreted into the lumen of the duct and mixed with the protein (Fig. 4A, panel II); however, at this stage, the metals are still localized in submicron spots and do not appear to bind to the surrounding plaque proteins. Finally, near the distal depression, DOPA-metal coordination was detected by Raman spectroscopy throughout most of the dense protein phase in the lumen, ostensibly because of mixing and diffusion of metal ions from the MSPs (Fig. 4A, panel III). As further support, the intensity of the DOPA-metal resonance Raman peaks relative to the protein peaks varies in different regions of the nascent plaque, with the appearance of hotspots in which the intensity of metal coordination decreases gradually, moving away from the hotspot (Fig. 4B).

Our study provides several important insights into the plaque formation process that are relevant to our biological understanding and its potential for bio-inspiration: (i) The Fe and V ions used to form the plaque are actively accumulated by mussels and stored in the foot separately from plaque proteins in small particles (MSPs) complexed with a catechol containing molecule. (ii) Plaque formation proceeds through a microfluidic-like process in which stockpiled precursor proteins and MSPs are deployed, transported, and actively mixed through a multiplex network of cilia-lined microchannels. (iii) As the secreted contents move through the microchannels, the MSPs release their metal ions, which apparently spread throughout the nascent plaque where they are coordinated by DOPA residues in the plaque proteins (Fig. 4C). Previous work shows that metal coordination provides cohesive interactions that facilitate the curing of the fluid into a solid adhesive under seawater conditions (*10*).

The discovery of the MSPs solves a long-standing mystery because it was previously unresolved how metal ions were added to the byssus—passively acquired from seawater or actively added by the mussel. Previous studies showed that metal ions could be removed from threads through EDTA chelation and then replaced by soaking threads in millimolar metal chloride solutions, supporting the passive hypothesis (*3*, *11*). However, it is important to mention that seawater concentrations of soluble Fe and V are in the range of 1 to 50 nM (*28*). In support of the active-uptake hypothesis, mussels are filter feeders that accumulate a broad range of metals present in the water column within both their soft tissue as well as their byssus (including Ag, Cd, Cr, Cu, Fe, Ni, Pb, Ti, V, and Zn), which led to the use of the byssus as a heavy metal pollution biomarker since the 1970s (*29*, *30*). An iron radiolabeling study from the 1970s provides the only compelling evidence that the metals in the byssus are first actively accumulated by the mussel in the soft tissue, although it stopped short of localizing iron specifically in the foot (*31*). Many invertebrates possess metal-containing granules (*32*), and membrane-bound metal particles containing Zn, Fe, and Ca were previously observed and isolated from the kidney tissue in *M. edulis* (*33*). The mixing of V- and Fe-enriched MSPs with the plaque proteins provides clear evidence that mussels leave nothing to chance, given the extremely low concentrations of these metals in seawater (*28*) and their crucial role in byssus mechanics (*10*, *11*). Furthermore, the detection of V in the MSPs is especially noteworthy because it is seldom used in biology in functional roles—especially not as a cross-linking element in materials (*34*). It has been recently shown that mussels may preferentially use DOPA-V cross-links to mechanically stabilize the outer protective cuticle of the byssus (*13*), and in vitro nanomechanical studies have shown that DOPA-V cross-linking can provide cohesive forces between byssus protein films that are nearly twofold greater than those with DOPA-Fe cross-linking (*35*). However, this metal preference has not yet been corroborated in the plaque. Future investigations will

focus on the functional relevance of V in the byssus (*34*).

Throughout biology, physiologically relevant metal ions (e.g., Fe, Zn, Cu) are almost always taken up, stored, transported, and regulated in living organisms in close association with biomolecules (e.g., ferritin, transferrin, metallothionein) (*28*, *29*). Similarly, we infer that storage of catechol-complexed Fe and V in the MSPs separately from the plaque proteins confers several advantages. For example, this precaution almost certainly circumvents premature cross-linking of the plaque proteins either through metal cross-linking at neutral to basic pH (*9*, *27*, *36*) or through covalent oxidative cross-linking at low pH (*37*) that would be expected to occur if they were not separated. The complexation of Fe and V during storage is also likely important because both are highly redox active and could potentially mediate unwanted side reactions in the cellular environment that lead to the production of harmful reactive oxygen species (*38*). Based on our current findings, it is unclear if the metals are stored as individually coordinated ions, as ion clusters, or perhaps as an amorphous mineral phase (as in ferritin). Nonetheless, our findings indicate that they are effectively partitioned from the plaque proteins and are able to be transferred to the plaque proteins during formation.

Considering previous studies, current insights into the microfluidic-like machinery and the spatiotemporal control of plaque assembly allow us to posit an updated model of plaque formation (Fig. 4C). During fabrication, plaque protein vesicles are secreted into the LDs where the contents of multiple vesicles coalesce, forming a dense fluid phase that flows through the ducts toward the distal depression. Later, MSPs are secreted and mixed with the plaque proteins, facilitated by active cilia movement and diffusion, with metal ions diffusing throughout the protein phase. Given these observations, we posit that the relative metal binding strength of the catechols in the MSPs vis-à-vis the plaque proteins is influenced by physicochemical conditions within the LDs (e.g., pH, salt content, or possibly chemical differences in the MSP catechols). Indeed, it is known that the plaque proteins are stored and secreted at pH 2 (*39*), a pH at which coordination of metal ions via catechols is weak. By contrast, the resonance Raman signal from MSPs clearly indicates bidentate chelation by DOPA (*22*), which is typically only observed at increased pH values (*27*, *36*). Thus, when the comparatively small MSPs are exposed to the acidic conditions in the fluid plaque secretion, the affinity of the catechols within the MSPs for the metal ions is likely reduced, leading to metal release and diffusion throughout the liquid protein phase, enhanced by cilia-driven mixing. When the pH is increased to the slightly basic pH of seawater (~pH 8), the dispersed metal ions are again strongly chelated by DOPA-enriched plaque proteins, such as mfp-2, forming cohesive coordination bonds that mechanically fortify the plaque structure. Previous findings are consistent with this sequence of events. For example, it was proposed that during plaque curing, the nascent plaque proteins undergo a phase transition from a fluid phase to form a micro- and nanoporous solid (*9*, *15*). It is unknown if this process is also mainly triggered by pH change, addition of metals, or other changes in the physicochemical environment. However, it was recently shown that freshly formed plaques placed immediately in acidic conditions do not form pores and are much softer (*40*), consistent also with the pH-dependent nature of the metal coordination process.

The insights from this study are relevant not only to the growing field of mussel-inspired polymers but also to the broader field of sustainable polymer production or even tissue engineering. For example, the prestorage of metal ions as complexed particles for gradual release may present a more effective means of mixing, which has remained a key processing challenge ever since the very first mussel-inspired metallopolymers were designed (*2*, *36*). More generally, understanding how the physicochemical environment can trigger a fluid-to-solid phase transformation may inspire the development of 3D printing techniques operating under environmentally friendly conditions. In particular, the spontaneous formation of a hierarchically porous solid from a liquid precursor is notable, and a deeper understanding of the physicochemical driving forces guiding this process is needed.

REFERENCES AND NOTES

1. B. P. Lee, P. B. Messersmith, J. N. Israelachvili, J. H. Waite, *Annu. Rev. Mater. Res.* **41**, 99–132 (2011).
2. E. Khare, N. Holten-Andersen, M. J. Buehler, *Nat. Rev. Mater.* **6**, 421–436 (2021).
3. M. J. Harrington, A. Masic, N. Holten-Andersen, J. H. Waite, P. Fratzl, *Science* **328**, 216–220 (2010).
4. E. Degtyar, M. J. Harrington, Y. Politi, P. Fratzl, *Angew. Chem. Int. Ed.* **53**, 12026–12044 (2014).
5. C. N. Z. Schmitt, Y. Politi, A. Reinecke, M. J. Harrington, *Biomacromolecules* **16**, 2852–2861 (2015).
6. C. H. Li, J. L. Zuo, *Adv. Mater.* **32**, e1903762 (2020).
7. D. G. DeMartini, J. M. Errico, S. Sjoestroem, A. Fenster, J. H. Waite, *J. R. Soc. Interface* **14**, 20170151 (2017).
8. J. H. Waite, *J. Exp. Biol.* **220**, 517–530 (2017).
9. T. Priemel, E. Degtyar, M. N. Dean, M. J. Harrington, *Nat. Commun.* **8**, 14539 (2017).
10. D. S. Hwang *et al.*, *J. Biol. Chem.* **285**, 25850–25858 (2010).
11. C. N. Z. Schmitt *et al.*, *J. R. Soc. Interface* **12**, 0466 (2015).
12. T. Priemel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7613–7621 (2020).
13. F. Jehle *et al.*, *Nat. Commun.* **11**, 862 (2020).
14. A. Tamarin, P. J. Keller, *J. Ultrastruct. Res.* **40**, 401–416 (1972).
15. E. Filippidi *et al.*, *J. R. Soc. Interface* **12**, 20150827 (2015).
16. A. Tamarin, P. Lewis, J. Askey, *J. Morphol.* **149**, 199–221 (1976).
17. A. E. Ivliev, P. A. C. 't Hoen, W. M. C. van Roon-Mom, D. J. M. Peters, M. G. Sergeeva, *PLOS ONE* **7**, e35618 (2012).
18. Z. Movasaghi, S. Rehman, I. U. Rehman, *Appl. Spectrosc. Rev.* **42**, 493–541 (2007).
19. C. Kallepitis *et al.*, *Nat. Commun.* **8**, 14843 (2017).
20. M. Okada *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 28–32 (2012).
21. K. Hamada *et al.*, *J. Biomed. Opt.* **13**, 044027 (2008).
22. I. Michaud-Soret, K. K. Andersson, L. Que Jr., J. Haavik, *Biochemistry* **34**, 5504–5510 (1995).
23. G. Azizi, M. Akodad, M. Baghour, M. Layachi, A. Moumen, *J. Mater. Environ. Sci.* **9**, 1170–1181 (2018).
24. J.-C. Amiard, R. Journel, H. Bacheley, *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **147**, 378–385 (2008).
25. G. Danscher, *Histochemistry* **71**, 1–16 (1981).
26. G. Danscher, M. Stoltenberg, *Prog. Histochem. Cytochem.* **41**, 57–139 (2006).
27. N. Holten-Andersen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2651–2655 (2011).
28. E. C. Theil, K. N. Raymond, in *Bioinorganic Chemistry*, I. Bertini, H. B. Gray, S. J. Lippard, J. S. Valentine, Eds. (University Science Books, 1994), pp. 1–35.
29. I. Marigómez, M. Soto, M. P. Cajaraville, E. Angulo, L. Giamberini, *Microsc. Res. Tech.* **56**, 358–392 (2002).
30. T. L. Coombs, P. J. Keller, *Aquat. Toxicol.* **1**, 291–300 (1981).
31. S. G. George, B. J. S. Pirie, T. L. Coombs, *J. Exp. Mar. Biol. Ecol.* **23**, 71–84 (1976).
32. B. E. Brown, *Biol. Rev. Camb. Philos. Soc.* **57**, 621–667 (1982).
33. S. G. George, T. L. Coombs, B. J. S. Pirie, *Biochim. Biophys. Acta, Gen. Subj.* **716**, 61–71 (1982).
34. D. C. Crans, J. J. Smee, E. Gaidamauskas, L. Yang, *Chem. Rev.* **104**, 849–902 (2004).
35. M. Mesko *et al.*, *Chem. Mater.* **33**, 6530–6540 (2021).
36. N. Holten-Andersen *et al.*, *J. Mater. Chem. B Mater. Biol. Med.* **2**, 2467–2472 (2014).
37. D. E. Fullenkamp, D. G. Barrett, D. R. Miller, J. W. Kurutz, P. B. Messersmith, *RSC Advances* **4**, 25127–25134 (2014).
38. P. Sharma, A. B. Jha, R. S. Dubey, M. Pessarakli, *J. Bot.* **2012**, 1–26 (2012).
39. N. R. Martinez Rodriguez, S. Das, Y. Kaufman, J. N. Israelachvili, J. H. Waite, *Biofouling* **31**, 221–227 (2015).
40. J. H. Bernstein, E. Filippidi, J. Herbert Waite, M. T. Valentine, *Soft Matter* **16**, 9339–9346 (2020).

ACKNOWLEDGMENTS

We thank B. Kanngießer for helpful discussions and the staff at the Facility for Electron Microscopy Research (FEMR) (McGill) for their support in imaging. **Funding:** Natural Sciences and Engineering Research Council of Canada, NSERC Discovery Grant RGPIN-2018-05243 (M.J.H.); Canada Research Chair award, CRC Tier 2 950-231953 (M.J.H.); Max Planck Society (F.J. and L.B.); FQRNT Quebec Merit Fellowship for Foreign Students (T.P.). **Author contributions:** Conceptualization: T.P., M.J.H.; Methodology: T.P., F.J., I.M., P.Z., L.B.; Investigation: T.P., G.P., F.F., F.J., S.S., I.M., P.Z., L.B.; Visualization: T.P.; Funding acquisition: T.P., M.J.H.; Project administration: M.J.H.; Supervision: M.J.H.; Writing – original draft: T.P., M.J.H.; Writing – review and editing: T.P., F.J., P.Z., I.M., L.B., M.J.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9702
Materials and Methods
Figs. S1 to S4
References (*41*, *42*)
MDAR Reproducibility Checklist
Movies S1 to S3

12 April 2021; accepted 17 August 2021
10.1126/science.abi9702

POLYMER SCIENCE

Fracture, fatigue, and friction of polymers in which entanglements greatly outnumber cross-links

Junsoo Kim[1]†, Guogao Zhang[1,2]†, Meixuanzi Shi[1,3], Zhigang Suo[1]*

In gels and elastomers, the role of entanglements on deformation has been studied, but their effects on fracture, fatigue, and friction are less well understood. In this study, we synthesized polymers in which entanglements greatly outnumber cross-links. The dense entanglements enable transmission of tension in a polymer chain along its length and to many other chains. The sparse cross-links prevent the polymer chains from disentangling. These polymers have high toughness, strength, and fatigue resistance. After submersion in water, the polymers swell to equilibrium, and the resulting hydrogels have low hysteresis, low friction, and high wear resistance.

An elastomer consists of cross-linked polymer chains (*1*). The polymer chains fluctuate by thermal motion and can be stretched by a force. The cross-links connect the polymer chains, so that the elastomer recovers its shape when the force is removed. When the elastomer imbibes a solvent, the cross-links prevent the polymer chains from dissolving. The swollen elastomer is called a gel. Dense cross-links stiffen the elastomer and gel but embrittle them (*2*–*5*). This stiffness-toughness conflict is partially resolved in an interpenetrating polymer network where one network is grown within a second one (i.e., a double-network polymer) (*6*–*8*). For example, if one network has prestretched short chains, and the other has stretchable long chains (*6*), when subjected to a small stretch, neither network breaks and the short-chain network stiffens the material. With increasing stretch, the short-chain network will break at isolated spots, while the long-chain network transmits stress and elicits many other short chains to break. The distributed scission toughens the material but causes pronounced hysteresis. Toughness gained from hysteresis is ineffective in applications of cyclic load (*9*).

We resolve the stiffness-toughness conflict, in addition to achieving negligible hysteresis, by fabricating a single-network polymer in which all chains are long and cross-links are greatly outnumbered by entanglements (Fig. 1A). Entanglements function as slip links, which stiffen the polymer. Unlike cross-links, however, entanglements do not embrittle the polymer. When a highly entangled polymer is stretched, before a chain breaks, tension

[1]John A. Paulson School of Engineering and Applied Science, Kavli Institute for Nanobio Science and Technology, Harvard University, Cambridge, MA 02138, USA. [2]State Key Laboratory of Chemical Engineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China. [3]State Key Laboratory for Strength and Vibration of Mechanical Structures, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China.
*Corresponding author. Email: suo@seas.harvard.edu
†These authors contributed equally to this work.

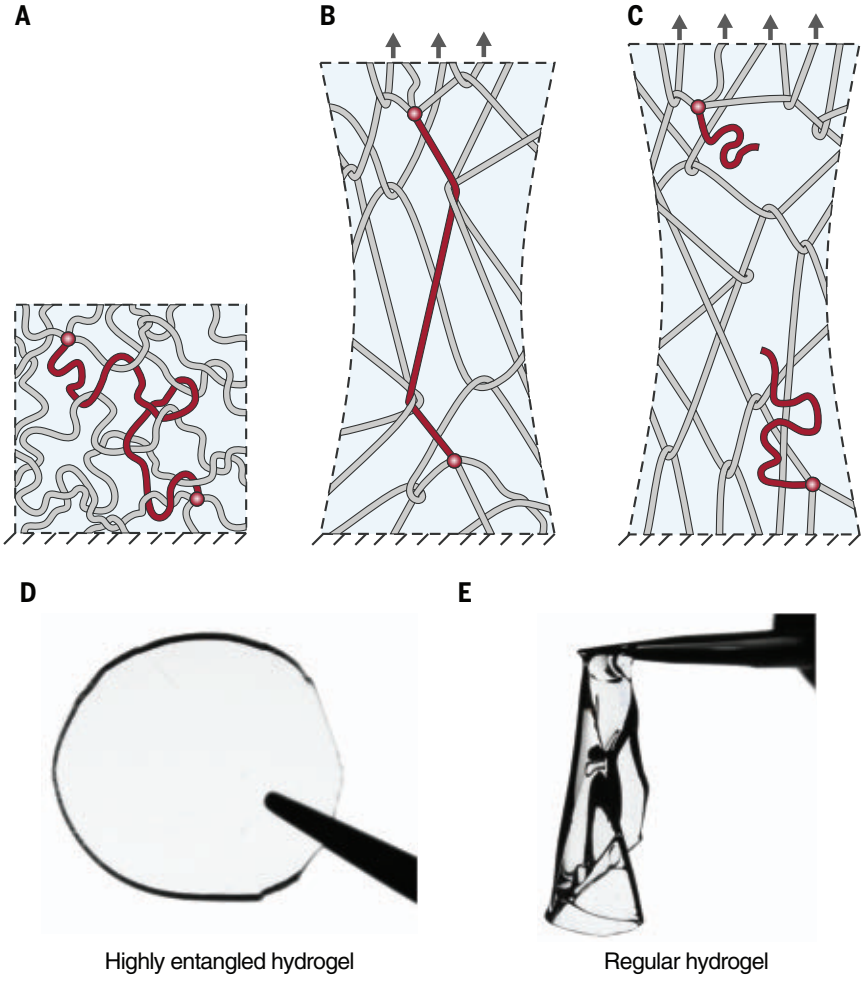


Fig. 1. A gel or elastomer in which entanglements greatly outnumber cross-links. (**A**) Each polymer chain has a large number of entanglements along its length and a cross-link (red dot) at each end. (**B**) A stretched polymer showing transmission of the tension to other chains. (**C**) A broken bond relaxes one chain and partly relaxes the entangled and cross-linked chains. Arrows indicate applied load. (**D**) Highly entangled hydrogel. (**E**) Regular hydrogel.

transmits along the chain and to many other chains through entanglements (Fig. 1B). When the chain breaks at a single covalent bond, the polymer dissipates the elastic energy in many chains, over long lengths (Fig. 1C). Unlike the double-network polymer, this single-network polymer achieves high toughness because all chains are long. Provided that the chains slip with low friction, the polymer has negligible hysteresis.

Whereas how entanglements affect stiffness has been appreciated since the inception of polymer science, how entanglements affect fracture, fatigue, and friction has not. In a polymer melt or solution, the entanglements provide elasticity on a short time scale and viscosity on a long time scale (*10*). In a cross-linked polymer, the entanglements cannot disentangle without chain scission, and they function as additional cross-links that stiffen the polymer (*11*, *12*). The stiffening effect of entanglements is pronounced when the cross-link density is low (*13*).

We ascertain that dense entanglements in hydrogels can be achieved using a precursor of typical ingredients but unusually low amounts of water, cross-linker, and initiator. Crowded monomers result in crowded polymers, which, as one might expect, densely entangle. For the precursor, let W be the water-to-monomer molar ratio, C be the cross-linker-to-monomer molar ratio, and I be the initiator-to-monomer molar ratio. To obtain a well-formed hydrogel, we fix $I/C = 0.4$, so that the initiating points that generate dangling chains are fewer than the cross-links (fig. S1).

We synthesize two polyacrylamide hydrogels using precursors of the same value of C, 3.2×10^{-5}, but different values of W, 2.0 and 25. After the as-synthesized hydrogels are submerged in water and swell to equilibrium, one hydrogel is turgid (Fig. 1D), but the other is flaccid (Fig. 1E and movie S1). The two hydrogels have the same cross-link density. If entanglements were absent, the two hydrogels would swell to the same equilibrium state. The difference in the mechanical behavior of the two hydrogels shows that they have different densities of entanglements. We call the former a highly entangled hydrogel, and the latter a regular hydrogel.

We study the conditions of W and C that form highly entangled hydrogels. Hydrogels of various values of W and C were synthesized and submerged in water. We measured the mass until the hydrogels swelled to equilibrium (fig. S2). The mass ratio of a fully swollen hydrogel to an as-synthesized hydrogel defines the swelling ratio (fig. S3). We plot the equilibrium value of the polymer-to-hydrogel mass ratio, ϕ, as a function of C and W (Fig. 2A and section 1 of the supplementary text in the supplementary materials). For a hydrogel made of a precursor of $W = 2.0$, in the absence

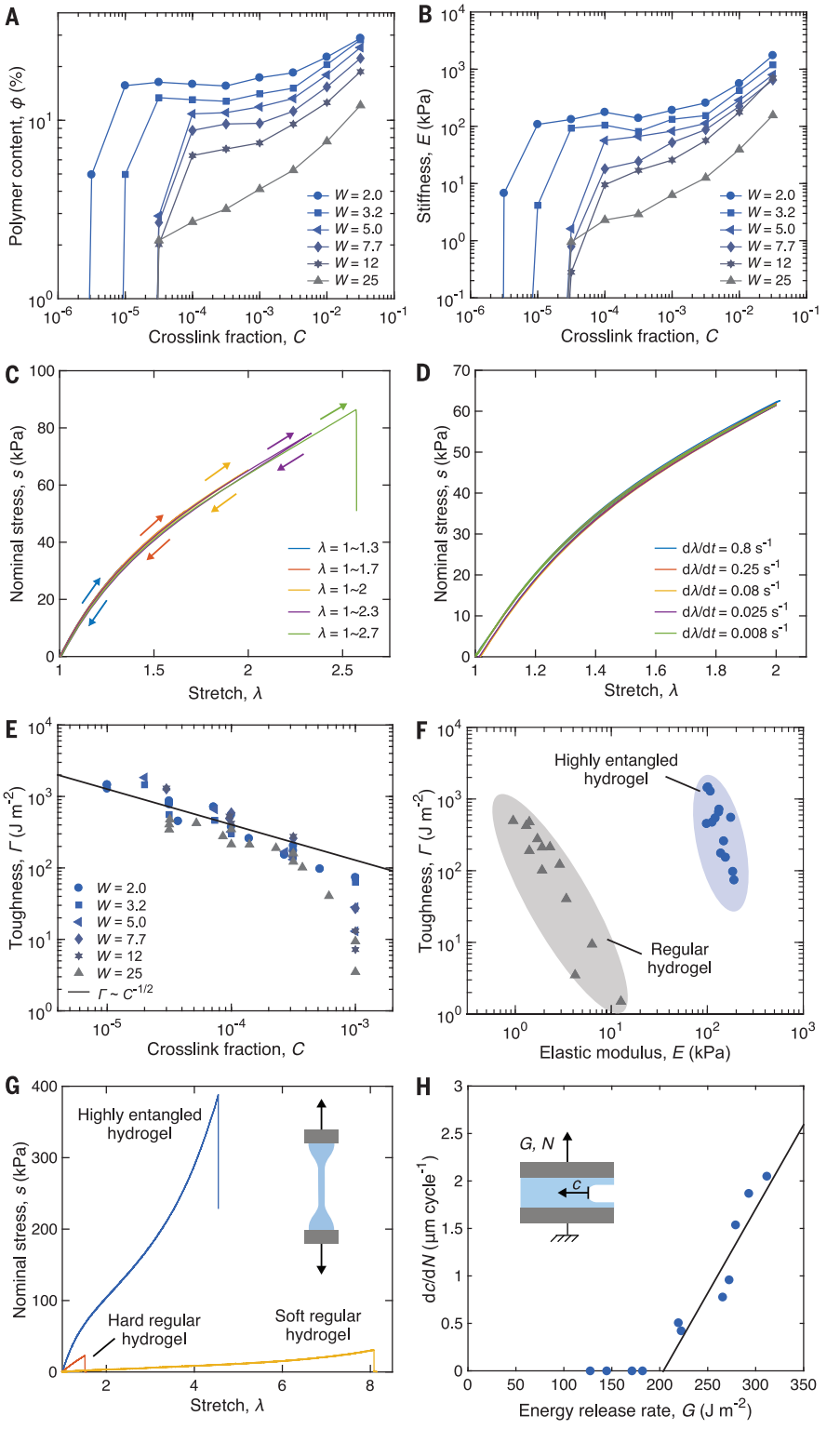


Fig. 2. Mechanical behavior of fully swollen hydrogels. (**A**) The polymer content, φ, of the fully swollen hydrogel. (**B**) The stiffness, E. (**C**) The hydrogel has negligible hysteresis. Arrows indicate loading and unloading. (**D**) The stress-stretch curve is rate-insensitive. (**E**) The toughness, Γ. Also plotted is the Lake-Thomas relation, $\Gamma \sim C^{-1/2}$. (**F**) Two families of hydrogels plotted in the stiffness-toughness plane. Regular hydrogels ($W = 25$) suffer from the stiffness-toughness conflict. Highly entangled hydrogels ($W = 2.0$) resolve this conflict and achieve both high stiffness and high toughness. (**G**) The stress-stretch curves of a highly entangled hydrogel, a hard regular hydrogel, and a soft regular hydrogel, under uniaxial tension. (**H**) The fatigue crack extension, c, per cycle, measured at various amplitudes of energy release rate. For (C), (D), (G), and (H), the highly entangled hydrogel was prepared using a precursor of $W = 2.0$ and $C = 1.0 \times 10^{-5}$.

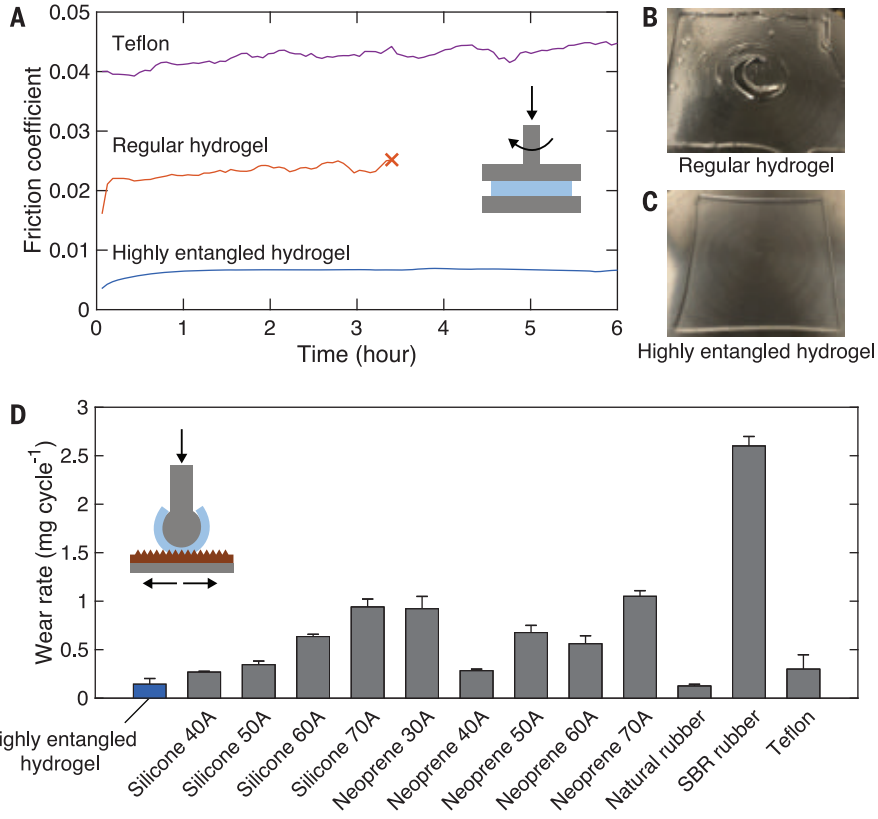


Fig. 3. Friction and wear. (**A**) The friction coefficients of a highly entangled hydrogel, regular hydrogel, and Teflon over time, tested in a rheometer (*29*). The normal pressure is 20 kPa, and the angular velocity is 1 rad/s. (**B**) The regular hydrogel ruptures after 3 hours. (**C**) The highly entangled hydrogel remains intact after 6 hours. (**D**) The wear rates of the highly entangled hydrogel and various materials. Hardness (durometer) is specified. Error bars indicate SEM; n = 10 for each group. SBR, styrene-butadiene rubber.

of cross-links, $C = 0$, chains dissolve in water, so that ϕ is vanishingly small. If C is nonzero but low, the hydrogel does not dissolve but is inelastic and squishy. At a critical value of $C \sim 1.0 \times 10^{-5}$, ϕ rises steeply. Aided by the dense entanglements, the critical C is low compared with values commonly used in making hydrogels. As C increases beyond the critical value, ϕ plateaus and is maintained by entanglements. The plateau lasts until $C \sim 10^{-3}$, beyond which ϕ increases gradually with C, indicating that cross-links are dense enough to prevail over entanglements. The plateau narrows as W increases and disappears when W exceeds about 5. For the same value of C, a precursor of a smaller W leads to a hydrogel of higher ϕ. The lower the value of W, the denser the entanglements and the lower the critical C. In a polymer network, the entropy of mixing drives swelling, but the entropy of elasticity drives deswelling (*14*). The entanglements function as additional cross-links and drive deswelling.

A similar plateau is also observed for stiffness (Fig. 2B). Such a plateau signifies the existence of entanglements and is used to estimate the density of entanglements (*15*). The stiffness of the hydrogel of $C = 1.0 \times 10^{-5}$ is similar to that of the hydrogel of $C = 1.0 \times 10^{-3}$. We assume that all cross-linkers in a precursor are incorporated into the polymer. Each cross-linker is an end of four chains, and each chain has two ends, so that the average number of monomers per chain is $(2C)^{-1}$. For the hydrogel of $C = 1.0 \times 10^{-5}$, each chain has, on average, $1/(2C) = 10^5/2$ monomers, but the measured stiffness indicates that the entanglements effectively shorten each chain to $10^3/2$ monomers. That is, each chain has entanglements equivalent to 10^2 cross-links.

The highly entangled hydrogel exhibits a high degree of elasticity. Under cyclic stretch to various amplitudes and a fixed rate, the hysteresis is negligible (Fig. 2C). The ratio of the dissipated energy (i.e., the area between the load and unload curves) to the applied work (i.e., the area under the load curve) is less than 1%. Furthermore, the stress-stretch curves are insensitive to the stretch rate over two orders of magnitude (Fig. 2D). The near-perfect elasticity results from several factors. First, although the cross-links are sparse, the dense entanglements aid in maintaining the network configuration when the polymer is stretched. Second, polymer chains are long and do not break before the sample fractures. Third, water has a low viscosity, such that the fully swollen hydrogel has low interchain friction. By comparison, the as-prepared hydrogels have low water content and exhibit pronounced hysteresis (fig. S4). The near-perfect elasticity is consistent with the measurements that the storage modulus is an order of magnitude higher than the loss modulus (fig. S5).

A highly entangled hydrogel behaves like a spring, but a double-network hydrogel behaves like a spring in parallel with a dashpot (*7*). When we hang masses on the highly entangled hydrogel and the double-network hydrogel and pluck, the former vibrates many cycles, but the latter damps to rest immediately (movie S2 and fig. S6A). A steel ball bounces on the highly entangled hydrogel but not on the double-network hydrogel (movie S3 and fig. S6B).

The highly entangled hydrogels exhibit high toughness (Fig. 2E). For hydrogels made from precursors of low C and W, cross-links are sparse, entanglements are dense, and toughness scales as $\Gamma \sim C^{-1/2}$, consistent with the prediction of the Lake-Thomas model (*3*). This consistency indicates that entanglements do not hinder the transmission of tension along the length of the long polymer chains. The toughness of 1460 J/m^2 is comparable to that of a double-network hydrogel (*16*). At small values of W, φ ranges from 8 to 16%. This modest change in φ is consistent with the observed weak dependence of toughness on W. For hydrogels made from precursors of high C and W, however, the cross-links are dense, the entanglements are sparse, and toughness falls substantially below the relation $\Gamma \sim C^{-1/2}$. The toughness observed here, 1 to 10 J/m^2, is consistent with values reported in the literature for regular polyacrylamide hydrogels (*7*).

Regular polymer networks suffer from the stiffness-toughness conflict (*4*, *5*). Stiffness scales with the cross-link density as $E \sim C$ (*2*), and toughness scales with the cross-link density as $\Gamma \sim C^{-1/2}$ (*3*), so that cross-links stiffen the polymers but embrittle them, and this also occurs in regular hydrogels. In the highly entangled hydrogels (W = 2.0), the entanglements stiffen polymers but do not embrittle them (Fig. 2F and fig. S7).

Entanglements also markedly strengthen polymers, as illustrated by comparing a highly entangled hydrogel with two types of regular hydrogels (Fig. 2G and section 2 of the supplementary text). The strength is 390 kPa for a highly entangled hydrogel (W = 2.0 and $C = 1.0 \times 10^{-5}$), 23 kPa for a hard regular hydrogel (W = 25 and $C = 1.0 \times 10^{-2}$), and 31 kPa for a soft regular hydrogel (W = 25 and $C = 3.2 \times 10^{-4}$). Entanglements strengthen a hydrogel for several reasons. First, entanglements constrain swell: φ = 16, 7.6, and 3.2% for the

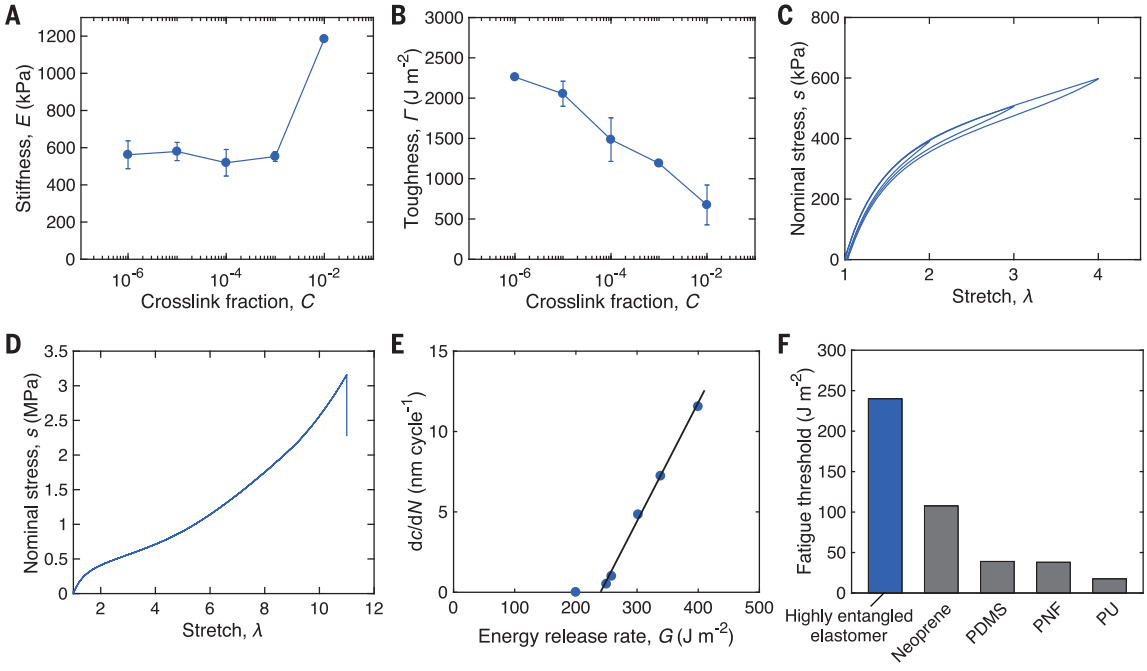


Fig. 4. A highly entangled elastomer. (**A**) At low C, the stiffness plateaus. (**B**) The toughness increases as C decreases. Error bars indicate SEM; $n = 3$ for each group. (**C**) Subject to load and unload, the highly entangled elastomer exhibits small hysteresis. (**D**) The stress-stretch curves of a highly entangled elastomer under uniaxial tension. (**E**) The fatigue threshold is ~240 J/m^2. (**F**) The highly entangled elastomer has a higher fatigue threshold than regular elastomers of similar stiffness. For (C) to (F), the elastomer is made with $C = 1.0 \times 10^{-5}$.

three hydrogels, respectively, so that the highly entangled hydrogel has more polymers to carry the load than the two regular hydrogels. Second, entanglements readily slip and enable tension to transmit in the chain along its length. Third, entanglements transmit tension between chains, and a highly entangled hydrogel has many more entanglements than a regular hydrogel. When a polymer is stretched, owing to statistical distribution, chains break at different times, which lowers the strength. The strength is increased when entanglements distribute tension both along a chain and between chains. Movie S4 shows that the highly entangled hydrogel lifts a weight at a nominal stress of 73 kPa but the two regular hydrogels do not.

Applications of polymers are often limited not by toughness under monotonic load but by fatigue threshold under cyclic load (*3*). For example, pure natural rubber has toughness of ~10^4 J/m^2 but a fatigue threshold of ~50 J/m^2. The toughness of natural rubber comes mainly from a dissipation process in the bulk (i.e., strain-induced crystallization), and the fatigue threshold comes from breaking chains across the crack plane (*3*). We measure the crack extension per cycle in a highly entangled hydrogel at various amplitudes of energy release rate (Fig. 2H). A linear regression of data estimates a fatigue threshold of ~200 J/m^2. This value is about four times that of natural rubber and about 20 times that of a regular hydrogel (*17*).

A highly entangled hydrogel is slippery and wear-resistant. When a hydrogel slides on a substrate, each polymer chain at the surface of the hydrogel is anchored to the polymer network on one end and mobile on the other end, provided that the polymer chain negligibly adsorbs to the substrate. These dangling and hydrophilic polymer chains stabilize a water-rich layer, which lubricates the surfaces (*18*). The friction decreases as the thickness of the water-rich layer increases, and the thickness scales with the size of the blob of the dangling chains. The highly entangled hydrogel has much longer polymer chains than the regular hydrogel, so that it has a lower friction coefficient. The friction coefficient of the highly entangled hydrogel is 0.0067, about three times lower than that of the regular hydrogel and six times lower than that of Teflon (Fig. 3A). Low friction, together with high toughness and fatigue threshold, leads to a low wear rate (*19*). The regular hydrogel ruptures after 3 hours of slide in a rheometer (Fig. 3B), but the highly entangled hydrogel remains intact after 6 hours of sliding (Fig. 3C). The highly entangled hydrogel has a lower wear rate than Teflon and most elastomers tested by us (Fig. 3D). A polyurethane elastomer and a silica-filled silicone have exceedingly low wear rates, which cannot be accurately determined in our experimental setup. These two materials, however, have friction coefficients much higher than that of the hydrogel.

The highly entangled hydrogels differ from existing hydrogels in other ways. For example, a sliding-ring hydrogel also distributes tension over long chains, but such a hydrogel has a low density of sliding rings and swells enormously, giving a low toughness of 20 J/m^2 and a low strength of 40 kPa (*20*, *21*). The stiffness-toughness conflict has also been resolved in polymers of microarchitectures (*22*–*24*). Unlike such polymers, a highly entangled hydrogel is synthesized in a single step, homogeneous above the scale of the mesh size, and optically transparent.

We next synthesize a highly entangled elastomer from a monomer, without solvent, at room temperature and extremely low C and I. We choose the monomer on the basis of the following considerations: (i) the monomer is a liquid of low viscosity, (ii) the resulting polymer is rubbery, and (iii) the resulting polymer has low entanglement molecular weight, as measured by a rheometer. To demonstrate, we choose ethyl acrylate as the monomer. It is a liquid of viscosity 0.55 mPa·s, even lower than water. Poly(ethyl acrylate) has a glass transition temperature of −21°C and a low entanglement molecular weight (*25*). We synthesize poly(ethyl acrylate) by using various values of C and a fixed value of $I/C = 0.1$. The polymer creeps below a critical value of $C \sim 10^{-6}$ but is highly elastic above it (fig. S8). As C increases beyond the critical value, stiffness first plateaus and then increases again after $C \sim 10^{-3}$ (Fig. 4A). The plateau confirms that these elastomers are highly entangled. The toughness reaches 2200 J/m^2 at $C = 10^{-6}$ (Fig. 4B). The high toughness is achieved even though the hysteresis is low (Fig. 4C). The low hysteresis can be further reduced by swelling the elastomer with an organic solvent (fig. S9). The highly entangled elastomer has a nominal strength of 3.2 MPa (Fig. 4D). This value corresponds to a true strength of 35 MPa, which is one order of magnitude higher than those of common unfilled acrylic elastomers (*26*). The highly entangled elastomer has a fatigue threshold of ~240 J/m^2 (Fig. 4E), which

is much higher than that of regular elastomers at a similar stiffness, including neoprene (*4*), polydimethylsiloxane (PDMS) (*27*), phosphonitrilic fluoroelastomer (PNF) (*27*), and polyurethane (PU) (*28*) (Fig. 4F). Below the fatigue threshold, the crack growth cannot be detected in the highly entangled elastomer after 200,000 cycles of stretch (fig. S10).

A regular polymer has a netlike topology, whereas a highly entangled polymer has a fabric-like topology. In the former, cross-links prevail over entanglements. In the latter, cross-links are outnumbered enormously by entanglements. The different topologies result in different properties. The highly entangled polymers are ideal load-bearing materials: They resolve the stiffness-toughness conflict and simultaneously achieve high toughness and low hysteresis. They are strong, fatigue-resistant, and transparent. Upon swelling, they have low friction and high wear resistance. We have illustrated the behavior of highly entangled polymers using two material systems. Using a similar approach, we have also synthesized and tested a polyacrylic acid hydrogel (fig. S11). Potential applications of highly entangled polymers include swell-tolerant gels, fatigue-resistant adhesives, low-friction coatings, and transparent ionic conductors.

REFERENCES AND NOTES

1. L. R. G. Treloar, *The Physics of Rubber Elasticity* (Oxford Univ. Press, 1975).
2. P. J. Flory, *Principles of Polymer Chemistry* (Cornell Univ. Press, 1953).
3. G. J. Lake, A. G. Thomas, *Proc. R. Soc. London Ser. A* **300**, 108–119 (1967).
4. A. K. Bhowmick, *J. Macromol. Sci. Part C Polym. Rev.* **28**, 339–370 (1988).
5. Y. Zhou *et al.*, *J. Appl. Mech.* **87**, 1–22 (2019).
6. J. P. Gong, Y. Katsuyama, T. Kurokawa, Y. Osada, *Adv. Mater.* **15**, 1155–1158 (2003).
7. J.-Y. Sun *et al.*, *Nature* **489**, 133–136 (2012).
8. E. Ducrot, Y. Chen, M. Bulters, R. P. Sijbesma, C. Creton, *Science* **344**, 186–189 (2014).
9. M. Baumgartner *et al.*, *Nat. Mater.* **19**, 1102–1109 (2020).
10. P. G. de Gennes, *J. Chem. Phys.* **55**, 572–579 (1971).
11. S. F. Edwards, *Proc. Phys. Soc.* **91**, 513–519 (1967).
12. M. Rubinstein, S. Panyukov, *Macromolecules* **35**, 6670–6686 (2002).
13. M. E. De Rosa, H. H. Winter, *Rheol. Acta* **33**, 220–237 (1994).
14. P. J. Flory, J. Rehner Jr., *J. Chem. Phys.* **11**, 521–526 (1943).
15. P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, 1979).
16. J. P. Gong, *Soft Matter* **6**, 2583–2590 (2010).
17. R. Bai, J. Yang, Z. Suo, *Eur. J. Mech. A Solids* **74**, 337–370 (2018).
18. J. P. Gong, *Soft Matter* **2**, 544–552 (2006).
19. A. H. Muhr, A. D. Roberts, *Wear* **158**, 213–228 (1992).
20. Y. Okumura, K. Ito, *Adv. Mater.* **13**, 485–487 (2001).
21. C. Liu *et al.*, *ACS Macro Lett.* **6**, 1409–1413 (2017).
22. Z. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 5967–5972 (2019).
23. C. Xiang *et al.*, *Mater. Today* **34**, 7–16 (2020).
24. S. Lin, J. Liu, X. Liu, X. Zhao, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10244–10249 (2019).
25. L. Andreozzi, V. Castelvetro, M. Faetti, M. Giordano, F. Zulli, *Macromolecules* **39**, 1880–1889 (2006).
26. R. M. B. Fernandes, L. L. Y. Visconte, R. C. R. Nunes, *J. Elastomers Plast.* **42**, 65–74 (2010).
27. A. N. Gent, R. H. Tobias, *J. Polym. Sci. Polym. Phys. Ed.* **20**, 2051–2058 (1982).
28. A. Cristiano, A. Marcellan, B. J. Keestra, P. Steeman, C. Creton, *J. Polym. Sci. B Polym. Phys.* **49**, 355–367 (2011).
29. J. P. Gong, G. Kagata, Y. Osada, *J. Phys. Chem. B* **103**, 6007–6014 (1999).

ACKNOWLEDGMENTS

The authors thank J. Steck, S. Hassan, and G. Nian for comments on the manuscript. **Funding:** This work was supported by MRSEC (DMR-2011754). J.K. acknowledges financial support from the Kwanjeong Educational Foundation. G.Z. acknowledges support from the Office of China Postdoctoral Council. M.S. is a visiting student at Harvard University supported by the China Scholarship Council. **Author contributions:** J.K. conceived of the project. J.K., G.Z., and Z.S. designed the study and interpreted the results. J.K., G.Z., and M.S. conducted the experiments. J.K. and Z.S. wrote the manuscript. Z.S. supervised the research. All authors discussed the result and commented on the manuscript.

Competing interests: Patents related to this work are planned.
Data and materials availability: All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg6320
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*30*–*32*)
Movies S1 to S4

19 January 2021; resubmitted 4 May 2021
Accepted 14 July 2021
10.1126/science.abg6320

CANCER MICROBIOME

Commensal bacteria promote endocrine resistance in prostate cancer through androgen biosynthesis

Nicolò Pernigoni[1,2]†, Elena Zagato[1,2]†‡, Arianna Calcinotto[1,2]†, Martina Troiani[1,2], Ricardo Pereira Mestre[3], Bianca Calì[1,2], Giuseppe Attanasio[1,2], Jacopo Troisi[4], Mirko Minini[1,2], Simone Mosole[1,2], Ajinkya Revandkar[1,2], Emiliano Pasquini[1,2], Angela Rita Elia[1,2], Daniela Bossi[1,2], Andrea Rinaldi[1,2], Pasquale Rescigno[5], Penny Flohr[6], Joanne Hunt[6], Antje Neeb[6], Lorenzo Buroni[6], Christina Guo[6], Jonathan Welti[6], Matteo Ferrari[3], Matteo Grioni[7], Josée Gauthier[8], Raad Z. Gharaibeh[8], Anna Palmisano[9,10], Gladys Martinetti Lucchini[11], Eugenia D'Antonio[3], Sara Merler[12,13], Marco Bolis[1,2,14], Fabio Grassi[15], Antonio Esposito[9,10], Matteo Bellone[7], Alberto Briganti[16,17], Maria Rescigno[18,19], Jean-Philippe Theurillat[1,2], Christian Jobin[8,20,21], Silke Gillessen[2,3], Johann de Bono[6], Andrea Alimonti[1,2,12,22]*

The microbiota comprises the microorganisms that live in close contact with the host, with mutual benefit for both counterparts. The contribution of the gut microbiota to the emergence of castration-resistant prostate cancer (CRPC) has not yet been addressed. We found that androgen deprivation in mice and humans promotes the expansion of defined commensal microbiota that contributes to the onset of castration resistance in mice. Specifically, the intestinal microbial community in mice and patients with CRPC was enriched for species capable of converting androgen precursors into active androgens. Ablation of the gut microbiota by antibiotic therapy delayed the emergence of castration resistance even in immunodeficient mice. Fecal microbiota transplantation (FMT) from CRPC mice and patients rendered mice harboring prostate cancer resistant to castration. In contrast, tumor growth was controlled by FMT from hormone-sensitive prostate cancer patients and *Prevotella stercorea* administration. These results reveal that the commensal gut microbiota contributes to endocrine resistance in CRPC by providing an alternative source of androgens.

Androgen deprivation therapy (ADT) remains the mainstay of treatment for patients with advanced prostate cancer. However, after an initial favorable response, patients often develop resistance to ADT with resulting tumor progression. Castration-resistant prostate cancer (CRPC) generally carries an unfavorable prognosis and new treatment strategies are needed (*1*). Although there is an urgent need for effective treatments for CRPC patients, therapies that delay the onset of CRPC in hormone-sensitive prostate cancer (HSPC) patients treated with ADT may also offer an alternative therapeutic strategy. The microbiota comprises the microorganisms that live in close contact with the host, usually with mutual benefit to one another (*2*). This relationship is described as symbiotic and is fundamental for the fitness of the host (*2*). Perturbations of this equilibrium can occur under pathological conditions, including cancer (*3*). Microbiota can directly affect tumor initiation through releasing of toxins (*4*–*6*) or can influence tumor cells through bacterial metabolites (*7*). In addition to this, the microbiota can contribute to tumor development through the promotion of inflammation (*8*) and shaping the tumor immune response (*9*, *10*). Increasing evidence shows that the microbiota is important for the antitumor activity of both chemotherapy (*11*, *12*) and immune checkpoint inhibitors (*13*–*16*), and modulation of the microbiota might enhance favorable treatment response in cancer patients. Previous findings in mouse models and human prostate tumor samples report the existence of a prostate microbiome that can support prostate tumor growth by promoting chronic

inflammation (*17*, *18*). In contrast, only a limited number of correlative studies have investigated the role of the gut microbiota in prostate cancer initiation and progression (*19*–*21*). Given the role played by the gut microbiota in cancer, an intriguing hypothesis is that the intestinal microbiota of patients suffering from prostate cancer could also participate in the host's hormone metabolism (*22*), thus affecting prostate cancer growth.

To address the impact of the intestinal microbiota on CRPC progression, we used two mouse models: the TRAMP-C1 allograft and the *Pten*$^{pc-/-}$ prostate conditional mouse models. In both models, surgical castration (CTX) is followed by a castration-sensitive phase [CS; in *Pten*$^{pc-/-}$ mice 4 weeks after CTX (i.e., 12 weeks of age); in TRAMP-C1 allograft mice 6 days after CTX] in which the tumor size shrinks because of androgen ablation, and a subsequent castration-resistant phase [CR; in *Pten*$^{pc-/-}$ mice 12 weeks after CTX (i.e., ≥20 weeks of age); in TRAMP-C1 allograft mice ≥10 days after CTX] in which the tumor becomes resistant to androgen ablation and starts to grow again (*23*). To deplete the intestinal microbiota, we treated these tumor-bearing mice with a cocktail of broad-spectrum antibiotics (*24*) (ABX) that reduced the fecal CFU counts by about 10 orders of magnitude (fig. S1, A and B). Microbiota ablation resulted in delayed tumor growth and improved survival in the TRAMP-C1 CTX context, but not in sham-operated animals (Fig. 1, A and B), with no impact on animal weight (fig. S1C). Microbiota ablation significantly reduced Ki67-positive prostate cancer cells in tumors of castration-resistant mice without altering the percentage of apoptotic cells as detected by cleaved caspase 3 (cC3) positivity (Fig. 1C and fig. S1, D and E). Also in the *Pten*$^{pc-/-}$ CTX model, microbiota ablation led to a reduction in prostate tumor volume as detected by magnetic resonance imaging (MRI) measurements of tumor volume over a time course, with no effect on sham-operated animals (Fig. 1, D and E). Histopathological evaluation of the anterior prostate (AP) lobe showed a clear reduction in tumor aggressiveness (percentage of glands with invasive area) in mice treated with ABX (fig. S1, F and G). Loss of intestinal microbiota led to a reduction of Ki67-positive cells in castration-resistant mice but did not affect cell proliferation in sham-operated animals (fig. S1H). Of note, ABX treatment of prostate epithelial tumor cells cultured in two dimensions did not impair tumor cell growth (fig. S1I). These data were validated in vivo in two additional CS mouse models of human prostate cancer: the LNCaP xenograft model and the LuCaP35 patient-derived xenograft (PDX) model. In both cases, ABX resulted in delayed onset of CRPC that improved survival (Fig. 1, F to I). Similar results were acquired in CTX mice treated with enzalutamide, a second-generation androgen receptor (AR) antagonist (*25*) (fig. S2, A and B). Overall, these data show that elimination of the intestinal microbiota affects prostate tumorigenesis in different mouse models, selectively in castrated mice.

We next hypothesized that CTX altered mouse intestinal microbiota. To address this point, we performed 16*S* ribosomal DNA (rDNA) sequencing on fecal DNA from sham-operated and castrated *Pten*$^{pc-/-}$ mice. We identified a compositional difference in the two cohorts (fig. S3A), with enrichment of specific microbiota species in both CR and CS mice (Fig. 1J and fig. S3B). Specifically, two species, *Ruminococcus gnavus* and *Bacteroides acidifaciens*, were particularly enriched in the fecal microbiota of CR *Pten*$^{pc-/-}$ mice (Fig. 1K). *R. gnavus* was also found enriched in fecal samples of castrated LNCaP mice, including those treated with enzalutamide (fig. S3, C to E). Intriguingly, bipolar androgen therapy in the TRAMP-C1 mouse model promoted tumor inhibition and decreased the abundance of Ruminococcaceae; this shows that changes in circulating host androgens affect gut microbiota composition (fig. S3, F to H).

Given the reported dependence of the immune response on the composition of the gut microbiota (*9*–*11*, *15*), we analyzed the systemic and tumor-infiltrating immune cell populations in castrated mice but found that ABX treatment induced only minimal variations in circulating level of cytokines (fig. S4A), percentage of tumor-infiltrating immune subsets (fig. S4B), and percentages of immune cell subsets in other organs (fig. S4, C to G). In confirmation, ABX treatment was also effective in TRAMP-C1 allograft mice treated with an anti-Ly6G depleting antibody targeting tumor-infiltrating myeloid cells (fig. S5A) and in NOD-SCID mice lacking T, B, and NK cells (fig. S5B). Overall, these data show that the intestinal microbiota is altered in CRPC and sustains tumor growth without regulating systemic and local immune responses.

To better dissect the functional role of the intestinal microbiota in sustaining CRPC growth, we performed fecal microbiota transplantation (FMT) experiments in TRAMP-C1 allograft recipients with feces from mice that had become CR from previous treatment (CR FMT) or from wild-type mice (HD FMT) and followed the evolution of the tumor volume upon castration. To avoid competition with the endogenous microbiota, we pretreated recipient mice with ABX for 7 days before CTX and

Table 1. Patients' characteristics.

Clinical data (at time of swab)	HSPC (*n* = 19)	CRPC (*n* = 55)
Age (years; mean ± SD)	69.84 ± 11.45	69.03 ± 11.51
Prostate-specific antigen (PSA) (μg/liter; mean ± SD)	14.03 ± 28.51	176.89 ± 399.24
Neutrophil/lymphocyte ratio (NLR) (mean ± SD)	2.7 ± 1.05	3.78 ± 2.66
Gleason score at diagnosis (mean ± SD)	7.43 ± 2.12	8.06 ± 1.63
Overall survival (months; mean ± SD)	9.7 ± 3.12	8.8 ± 2.98
Enzalutamide/abiraterone treatment	4/19	22/55
Docetaxel/cabazitaxel treatment	0/19	5/55
Metastasis	12/19	51/55

[1]Institute of Oncology Research, Oncology Institute of Southern Switzerland, CH-6500 Bellinzona, Switzerland. [2]Faculty of Biomedical Sciences, Università della Svizzera Italiana, CH-1011 Lugano, Switzerland. [3]Medical Oncology Unit, Oncology Institute of Southern Switzerland, Ente Ospedaliero Cantonale, 6500 Bellinzona, Switzerland. [4]Theoreo Srl, Montecorvino Pugliano, 84090 Salerno, Italy. [5]Candiolo Cancer Institute, FPO-IRCCS, Candiolo, 10060 Turin, Italy. [6]Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, UK. [7]Cellular Immunology Unit, IRCCS Ospedale San Raffaele, 20100 Milan, Italy. [8]Department of Medicine, Division of Gastroenterology, Hepatology and Nutrition, University of Florida College of Medicine, Gainesville, FL, USA. [9]Experimental Imaging Center, San Raffaele Scientific Institute, 20100 Milan, Italy. [10]School of Medicine, University Vita–Salute San Raffaele, Milan, Italy. [11]Clinical Microbiology Unit, Ente Ospedaliero Cantonale, EOC, CH-6500 Bellinzona, Switzerland. [12]Department of Medicine, Venetian Institute of Molecular Medicine, University of Padova, Padova, Italy. [13]Section of Oncology, Department of Medicine, University of Verona, 37134 Verona, Italy. [14]Computational Oncology Unit, Department of Oncology, Istituto di Ricerche Farmacologiche "Mario Negri" IRCCS, 20156 Milano, Italy. [15]Institute for Research in Biomedicine, Faculty of Biomedical Sciences, Università della Svizzera Italiana, 6500 Bellinzona, Switzerland. [16]Division of Oncology, Unit of Urology, URI, IRCCS Ospedale San Raffaele, Milan, Italy. [17]University Vita–Salute San Raffaele, Milan, Italy. [18]IRCCS Humanitas Research Hospital, 20089 Rozzano, Milan, Italy. [19]Department of Biomedical Sciences, Humanitas University, 20090 Pieve Emanuele, Milan, Italy. [20]Department of Infectious Diseases and Immunology, University of Florida College of Medicine, Gainesville, FL, USA. [21]Department of Anatomy and Cell Biology, University of Florida College of Medicine, Gainesville, FL, USA. [22]Department of Health Sciences and Technology, Eidgenössische Technische Hochschule (ETH) Zürich, Zurich, Switzerland.
***Corresponding author. Email: andrea.alimonti@ior.usi.ch**
†These authors contributed equally to this work.
‡Present address: INGM, Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi," Milan, Italy, and Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy.

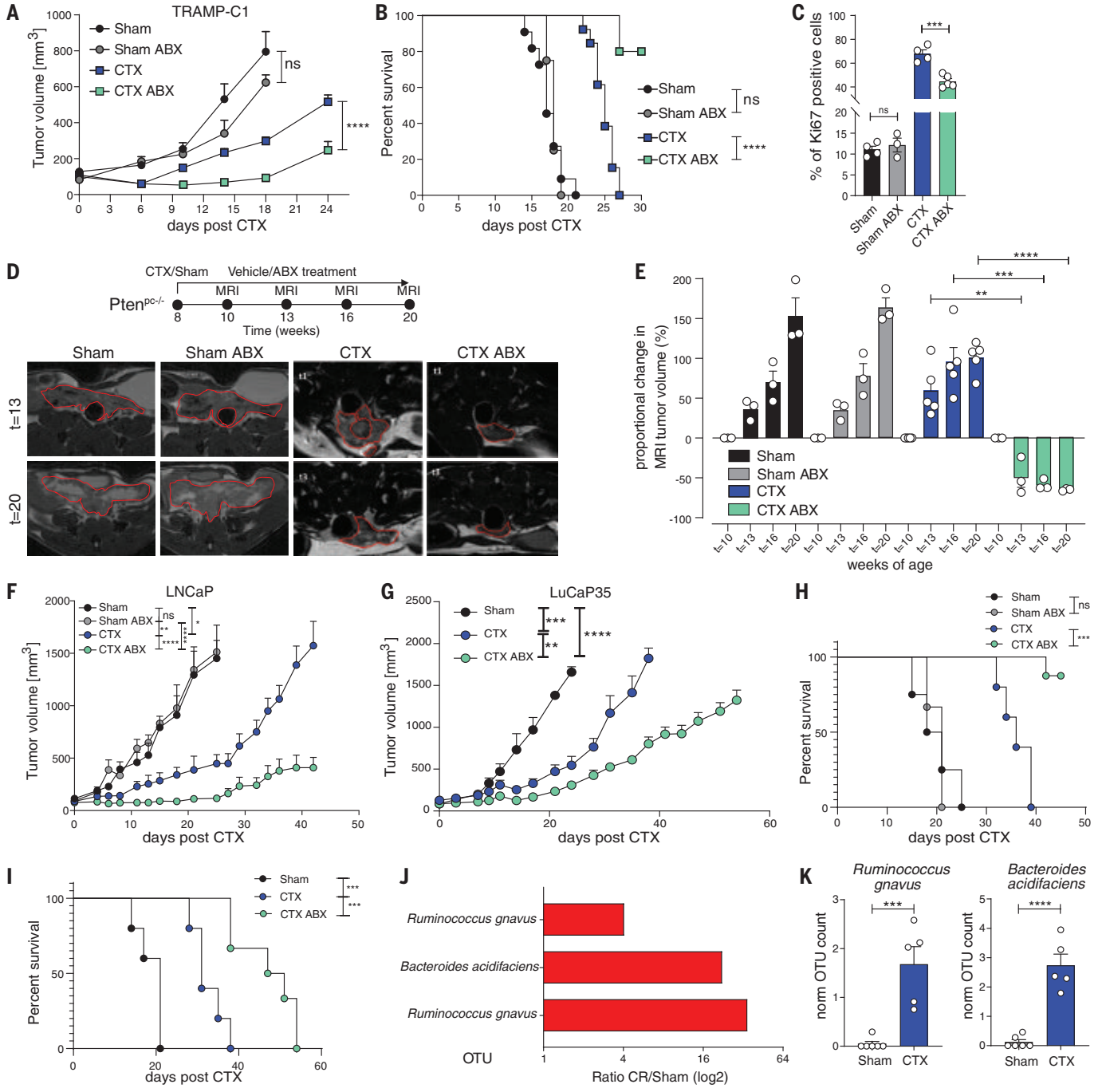


Fig. 1. Depletion of the intestinal microbiota in castrated but not in sham-operated mice affects CRPC growth. C57BL6/N mice were challenged subcutaneously with 2.5×10^6 TRAMP-C1 cells. When tumors became palpable, mice were either castrated (CTX) or sham-operated (sham). After surgery, mice were either administered normal drinking water or antibiotic cocktail (ABX: neomycin 1 g/liter, ampicillin 1 g/liter, vancomycin 0.5 g/liter, and metronidazole 2 mg per mouse every other day). (**A**) Tumor volume (sham, n = 14; sham ABX, n = 4; CTX, n = 20; CTX ABX, n = 10). (**B**) Survival curve. (**C**) Percentage of Ki67-positive cells (sham, n = 4; sham ABX, n = 3; CTX, n = 4; CTX ABX, n = 5). *Pten*$^{pc-/-}$ mice were sham-operated or castrated at 9 weeks of age and either left untreated or treated with ABX; prostatic tumor volume was monitored through serial magnetic resonance imaging (MRI) over a time course. (**D**) Experimental scheme and representative MRIs of sham, sham ABX, CTX, or CTX ABX mouse at 13 and 20 weeks of age. (**E**) Waterfall plot representing proportional change in tumor size for sham (n = 3), sham ABX (n = 3), CTX (n = 5), and CTX ABX (n = 3) *Pten*$^{pc-/-}$ mice. NRG mice [non-obese diabetic (NOD)–Rag1(null)–γ chain(null) mice] were challenged with LNCaP or LuCaP35, castrated or sham-operated when the tumor was palpable, and either left untreated or treated with ABX. (**F** and **H**) Tumor volume (F) and survival curve (H) of LNCaP allograft (sham, n = 4; sham ABX, n = 3; CTX, n = 5; CTX ABX, n = 7). (**G** and **I**) Tumor volume (G) and survival curve (I) of LuCaP35 allograft (sham, n = 5; CTX, n = 6; CTX ABX, n = 6). The composition of the CS and CR microbiota was analyzed through 16*S* rDNA sequencing of the feces of sham (n = 6) and CR (n = 5) *Pten*$^{pc-/-}$ mice. (**J**) Waterfall plot representing the abundance of the significantly different OTUs identified at species level as a ratio CR/CS. (**K**) Normalized OTU counts relative to *R. gnavus* and *B. acidifaciens* (sham, n = 6; CR, n = 5). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$ [two-way analysis of variance (ANOVA) and Sidak's multiple-comparison test in (A), (F), and (G); Mantel-Cox test in (B), (H), and (I); unpaired two-sided Student's t test in (C), (E), and (K)]; NS, not significant. Data in (A), (C), (E), (F), (G), and (K) are means ± SEM.

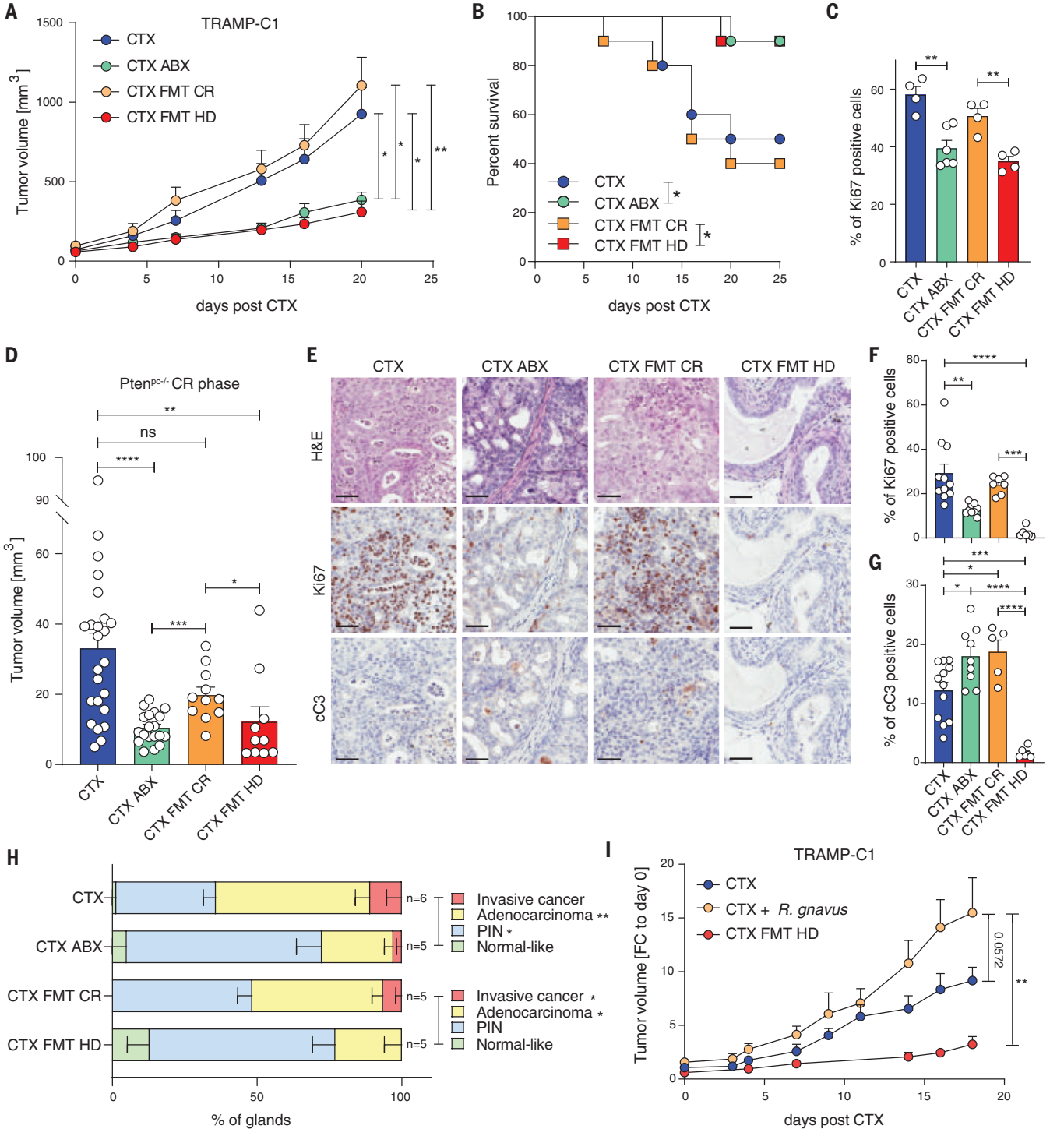


Fig. 2. Fecal microbiota transplantation from CRPC mice supports tumor growth in castrated recipient mice. TRAMP-C1 allograft mice were treated with ABX for 7 days before CTX, then either left untreated (n = 9), treated with ABX (n = 9), or received FMT from HD (n = 8) or CR (n = 9) mouse donors. (**A** to **C**) Tumor volume (A), survival curve (B), and immunohistochemical quantification of Ki67 staining (C) (CTX, n = 4; CTX ABX, n = 6; CTX FMT CR, n = 4; CTX FMT HD, n = 4). *Pten*$^{pc-/-}$ mice were treated with ABX for 7 days before CTX, then either left untreated, treated with ABX, or received FMT from CR or HD donors. (**D**) AP tumor volume at 20 weeks of age (CTX, n = 24; CTX ABX, n = 19; CTX FMT CR, n = 11; CTX FMT HD, n = 10). (**E**) Representative hematoxylin and eosin (H&E), Ki67, and cC3 immunohistochemical staining in AP lobes of *Pten*$^{pc-/-}$ mice at 20 weeks of age. Scale bar, 45 μm. (**F** and **G**) Quantification of Ki67-positive cells (CTX, n = 11; CTX ABX, n = 9; CTX FMT CR, n = 7; CTX FMT HD, n = 6) (F) and cC3-positive cells (CTX, n = 12; CTX ABX, n = 9; CTX FMT CR, n = 5; CTX FMT HD, n = 6) (G). (**H**) Histopathological score of anteroposterior lobes of CTX (n = 6), CTX ABX (n = 5), CTX FMT CR (n = 5), and CTX FMT HD (n = 5) *Pten*$^{pc-/-}$ mice at 20 weeks of age. PIN, prostatic intraepithelial neoplasia. (**I**) Tumor volume of TRAMP-C1 allograft mice treated with ABX for 7 days before CTX, then either left untreated (n = 8) or treated with *R. gnavus* (n = 5) via oral gavage. CTX FMT HD was used as internal control. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 [two-way ANOVA and Sidak's multiple-comparison test in (A); Mantel-Cox test in (B); unpaired two-sided Student's t test in (C), (H), and (I); Mann-Whitney test in (D); one-way ANOVA and Tukey's multiple-comparison test in (F) and (G)]. Data in (A), (C), (D), (F), (G), (H), and (I) are means ± SEM.

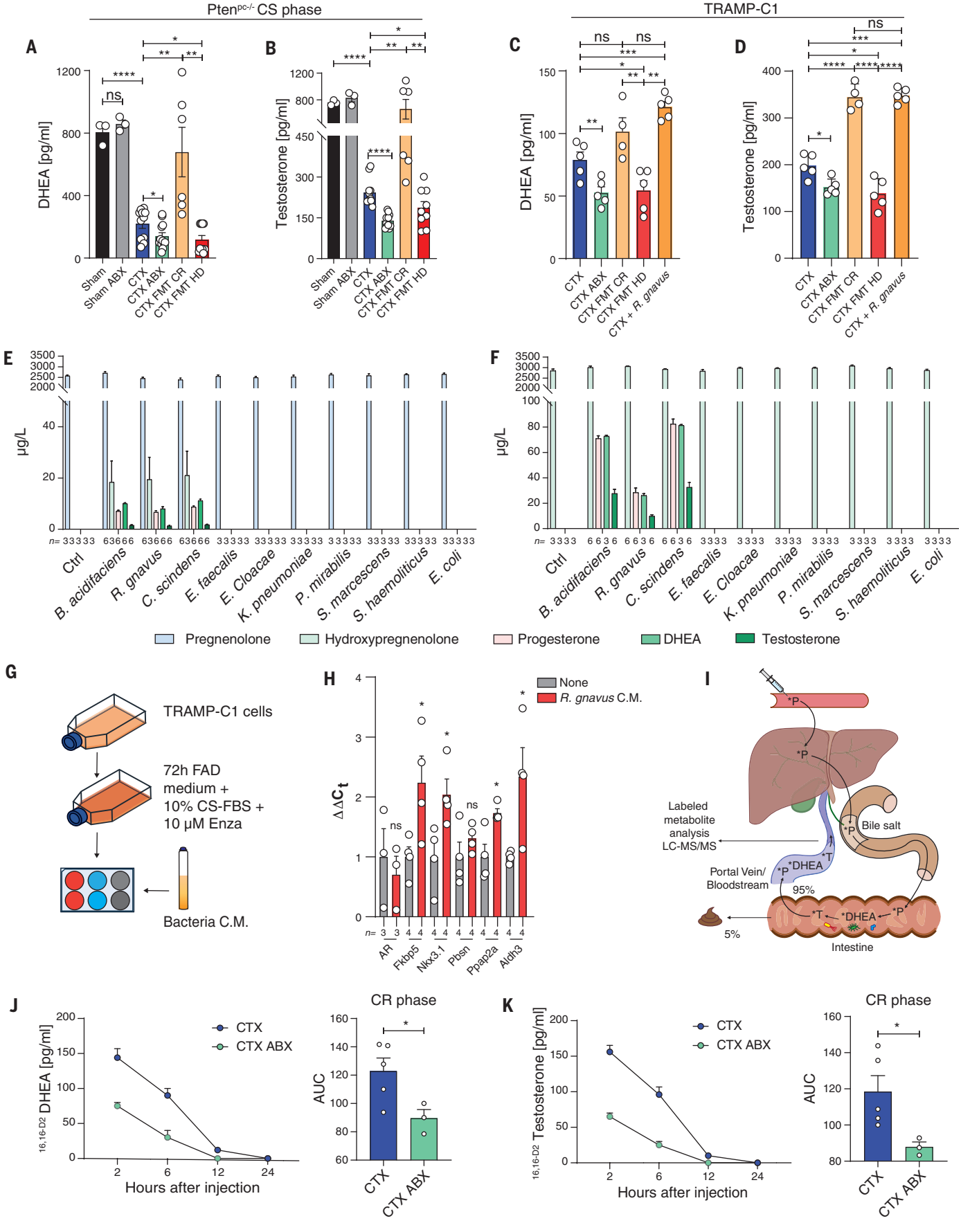

Pten pc-/- CS phase
TRAMP-C1
A
B
C
D
DHEA [pg/ml]
Testosterone [pg/ml]
Sham
Sham ABX
CTX
CTX ABX
CTX FMT CR
CTX FMT HD
CTX + R. gnavus
E
F
µg/L
Ctrl
B. acidifaciens
R. gnavus
C. scindens
E. faecalis
E. Cloacae
K. pneumoniae
P. mirabilis
S. marcescens
S. haemoliticus
E. coli
Pregnenolone
Hydroxypregnenolone
Progesterone
DHEA
Testosterone
G
TRAMP-C1 cells
72h FAD medium + 10% CS-FBS + 10 µM Enza
Bacteria C.M.
H
None
R. gnavus C.M.
ΔΔCt
AR
Fkbp5
Nkx3.1
Pbsn
Ppap2a
Aldh3
I
Labeled metabolite analysis LC-MS/MS
Bile salt
Portal Vein/ Bloodstream
95%
5%
Intestine
J
K
CTX
CTX ABX
Hours after injection
CR phase
AUC

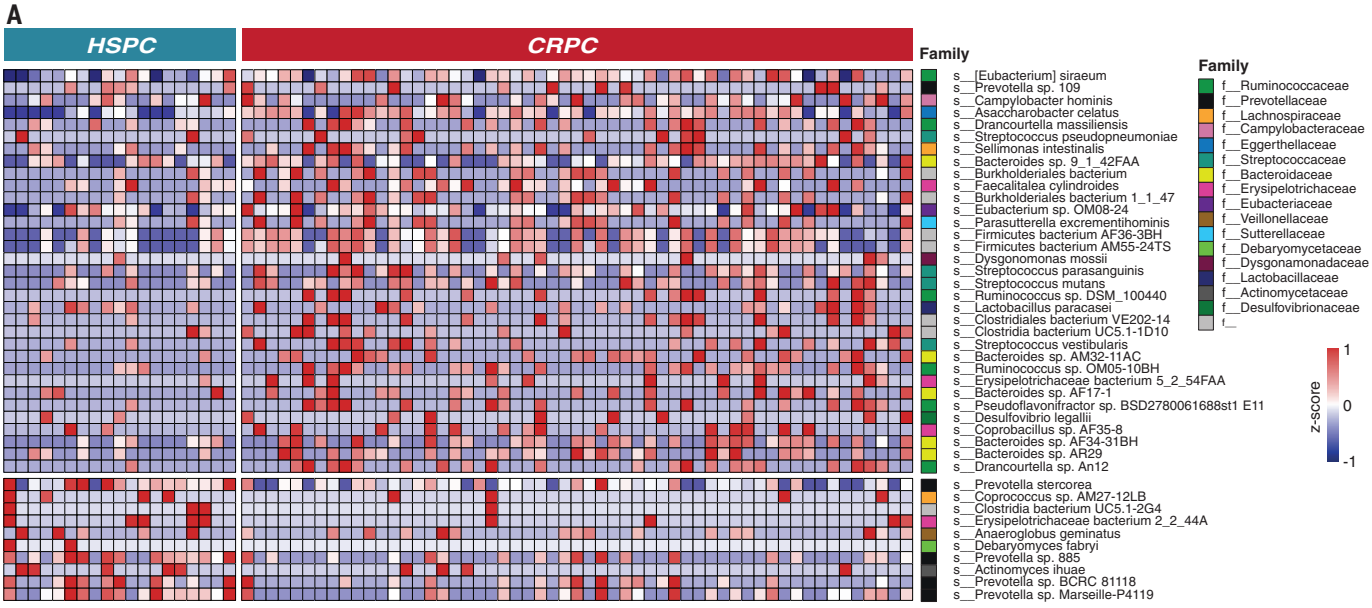
Fig. 3. The CRPC microbiota is enriched in bacterial species that produce androgens and affect tumor cell growth. (**A** and **B**) Targeted metabolomic analysis was performed in sera of *Pten*$^{pc-/-}$ mice CTX, CTX ABX, CTX FMT CR, and CTX FMT HD at 13 weeks of age. Levels of DHEA (A) and testosterone (B) (sham, *n* = 3; sham ABX, *n* = 3; CTX, *n* = 11; CTX ABX, *n* = 12; CTX FMT CR, *n* = 6; CTX FMT HD, *n* = 10). (**C** and **D**) Targeted metabolomic analysis was performed in sera of TRAMP-C1 mice CTX, CTX ABX, CTX FMT CR, and CTX FMT HD 18 days after castration. Levels of DHEA (C) and testosterone (D) (CTX, *n* = 5; CTX ABX, *n* = 5; CTX FMT CR, *n* = 4; CTX FMT HD, *n* = 5; CTX + *R. gnavus*, *n* = 5). (**E** and **F**) *B. acidifaciens*, *R. gnavus*, *C. scindens*, *E. faecalis*, *E. cloacae*, *K. pneumoniae*, *P. mirabilis*, *S. marcescens*, *S. haemoliticus*, and *E. coli* were cultured for 48 hours in the presence of the indicated androgen precursors in anaerobic conditions and analyzed for production of androgens or androgen precursors through LC-MS/MS. Androgen pathway intermediates in bacterial conditioned media (CM) were quantified after treatment with pregnenolone (E) and hydroxypregnenolone (F); Ctrl, no bacteria. (**G**) Experimental scheme. TRAMP-C1 cells in full androgen deprivation (FAD) were treated with CM of *R. gnavus* preincubated with pregnenolone or TSB for 48 hours and analyzed with quantitative reverse transcription polymerase chain reaction (qRT-PCR). CS-FBS, charcoal-stripped FBS; Enza, enzalutamide. (**H**) qRT-PCR analyses of the indicated genes. *Pten*$^{pc-/-}$ mice treated or not with ABX were injected i.v. with 75 ng of ^{D}pregnenolone. Serum was sampled at 2, 6, 12, and 24 hours after the injection and analyzed for the presence of deuterated androgen molecules. (**I**) Experimental scheme: Deuterated pregnenolone was injected i.v. and its conversion in downstream metabolites monitored through LC-MS/MS. (**J** and **K**) Time-course quantification and area under curve (AUC) of ^{D}DHEA (CTX, *n* = 5; CTX ABX, *n* = 3) (J) and ^{D}testosterone (CTX, *n* = 5; CTX ABX, *n* = 3) (K). **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001 [one-way ANOVA and unpaired two-sided Student's *t* test in (A) to (D); unpaired two-sided Student's *t* test in (H), (J), and (K)]. Data in (A), (B), (C), (D), (E), (F), (H), (J), and (K) are means ± SEM.

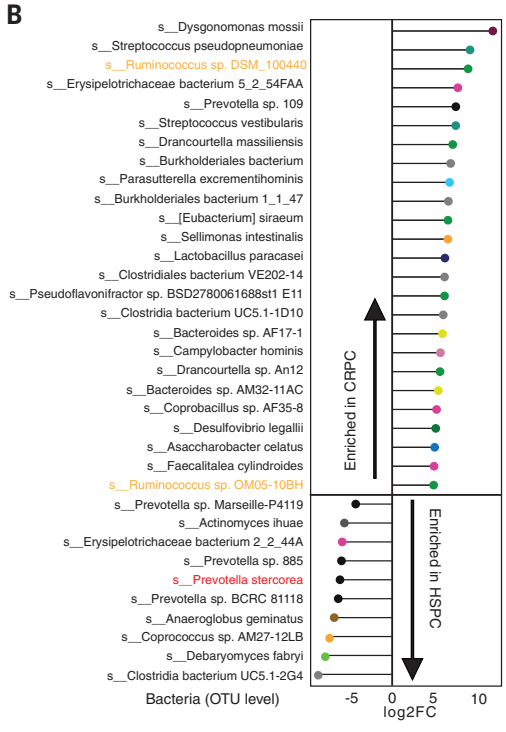
FMT (*24*) (fig. S6A). FMT successfully engrafted in the hosts (fig. S6, B and C). Whereas CR FMT resulted in the rapid emergence of CRPC, HD FMT controlled tumor growth to levels comparable to ABX treatment (Fig. 2A). Notably, CR FMT also had a significant impact on survival (Fig. 2B). CR FMT induced an increase in tumor cell proliferation as measured by Ki67 staining, whereas HD FMT was associated with decreased Ki67 staining (Fig. 2C and fig. S6D). Both treatments did not alter the percentage of apoptotic (cC3-positive) tumor cells (fig. S6E). These data were also validated in CTX *Pten*$^{pc-/-}$ mice treated with either CR or HD FMT (Fig. 2, D to G, and fig. S6F). Histopathological evaluation of AP lobes showed a significant reduction in the frequency of adenocarcinoma and invasive cancer in HD FMT–treated CTX *Pten*$^{pc-/-}$ mice as compared to those receiving CR FMT (Fig. 2H). In the FMT setting, we also did not find major differences in the relative abundance of tumor-infiltrating immune cell populations (fig. S6, G to J). Because the metagenomic analysis of the CR microbiota had revealed specific enrichment in *R. gnavus* and *B. acidifaciens*, we tested the functional impact of these two species in promoting CRPC growth by performing colonization experiments. TRAMP-C1 allograft mice were treated with ABX for 7 days prior to CTX to eliminate the competing endogenous microbiota. After CTX, mice were either left untreated or administered *R. gnavus* or *B. acidifaciens* orally every other day. Administration of *R. gnavus* increased tumor growth relative to untreated animals (Fig. 2I). Overall, these data show that the murine CR microbiota and *R. gnavus* are able to sustain tumor growth, whereas HD FMT delays the onset of CRPC.

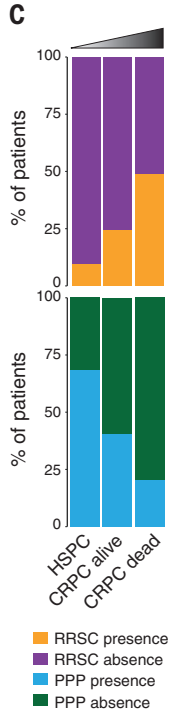
Because the intestinal microbiota is known to affect the host's metabolome (*26*), we performed untargeted metabolomic analyses of sera from *Pten*$^{pc-/-}$ CTX mice treated with or without ABX. This revealed different metabolomic profiles in the two cohorts (fig. S7A). Interestingly, we detected a significant reduction in circulating dehydroepiandrosterone (DHEA) and testosterone in microbiota-depleted animals, even if the abundance of the upstream metabolite pregnenolone was not altered (fig. S7B). In line with these findings, the expression levels of AR target genes were significantly reduced in animals devoid of intestinal microbiota (fig. S7C). Targeted metabolomic analysis in *Pten*$^{pc-/-}$ mice confirmed that both DHEA and testosterone were decreased in CTX mice treated with ABX, whereas in intact mice ABX treatment did not alter androgen levels (Fig. 3, A and B). Moreover, CR FMT and *R. gnavus* administration in TRAMP-C1 mice resulted in increased circulating DHEA and testosterone levels when compared to HD FMT treatment (Fig. 3, C and D).

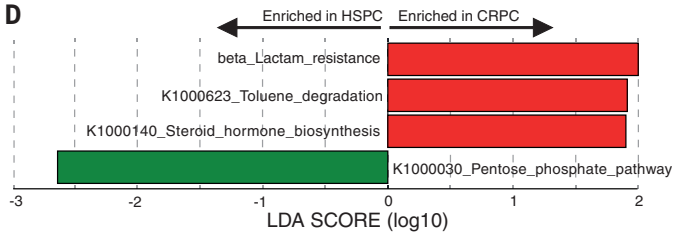
We next assessed whether the bacteria found enriched in the gut of CR mice were capable of synthesizing androgenic steroids. A bacterial strain isolated from the human microbiota, *Clostridium scindens*, can convert glucocorticoids into androgens (*27*, *28*). We therefore hypothesized that *R. gnavus* and *B. acidifaciens* may have a similar metabolic capability. To further explore this, we cultured these two strains in the presence of a panel of metabolites of the androgen biosynthesis pathway: cholesterol (the progenitor metabolite of the androgen synthesis pathway), pregnenolone, hydroxypregnenolone, cortisol, and aldosterone (fig. S7D). We analyzed the ability of bacterial species to convert in vitro these compounds into other intermediates of the pathway by using liquid chromatography with tandem mass spectrometry (LC-MS/MS) (fig. S7E). Interestingly, we found that only *R. gnavus* and *B. acidifaciens* converted pregnenolone and hydroxypregnenolone, but not other metabolites, into downstream metabolites of the pathway, including DHEA and testosterone (Fig. 3, E and F, and fig. S7F). As negative controls, we tested seven bacterial strains indigenous to the gut microbiota (*Enterococcus faecalis*, *Enterobacter cloacae*, *Klebsiella pneumoniae* 27, *Proteus mirabilis*, *Serratia marcescens*, *Staphylococcus haemoliticus*, *Escherichia coli*). We then tested the ability of conditioned media (CM) of *R. gnavus* incubated with pregnenolone to activate the AR signaling on TRAMP-C1 cells cultured in full androgen deprivation (FAD) conditions (Fig. 3G). The CM of *R. gnavus* incubated with culture broth alone was used as a control. Expression of AR target genes highlighted the ability of *R. gnavus* CM, cultured in the presence of pregnenolone, to stimulate the transcription of AR target genes relative to control CM (Fig. 3H).

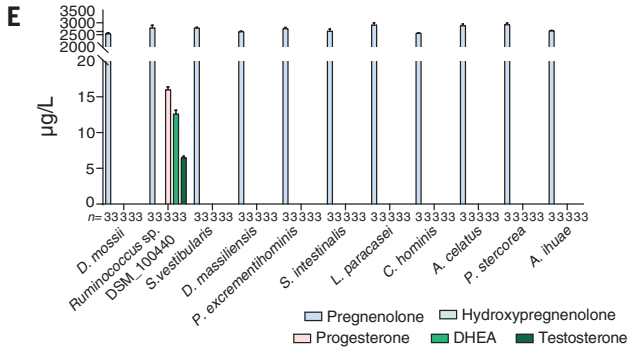
We next hypothesized that in CTX mice, circulating pregnenolone could reach the gut through the enterohepatic circulation, where it is metabolized to DHEA and testosterone by the gut microbiota (Fig. 3I). In CTX mice and CTX patients, cholesterol can be metabolized into pregnenolone by the adrenal glands (*29*). We therefore tested this hypothesis in vivo by injecting deuterated (^{D}) pregnenolone intravenously (i.v.) into CTX *Pten*$^{pc-/-}$ mice either treated or untreated with ABX to deplete the intestinal microbiota (fig. S8A). The abundance of ^{D}pregnenolone that reached the intestine (and by extension the microbiota) after the injection was comparable in both groups (fig. S8B). The conversion of this substrate into the downstream deuterated metabolites was monitored in the blood over a time-course experiment and detected with LC-MS/MS. We observed that in CTX mice, ablation of the microbiota resulted in a significant reduction of circulating ^{D}DHEA and ^{D}testosterone, particularly at 2 and 6 hours after injection, relative to control mice both in CR (Fig. 3, J and K) and in CS phases (fig. S8, C and D). To corroborate these data, we assessed whether microbiota depletion by ABX was still effective in a prostate tumor model lacking the AR (PC3) and in a model insensitive to castration (LUCaP145.2) and found that in these contexts, ABX treatment was ineffective (fig. S9, A to D). These data were further validated in LUCaP145.2 mice castrated and treated with enzalutamide. In these mice, androgen deprivation induced a significant expansion of fecal *R. gnavus* and *B. acidifaciens* (fig. S9, E

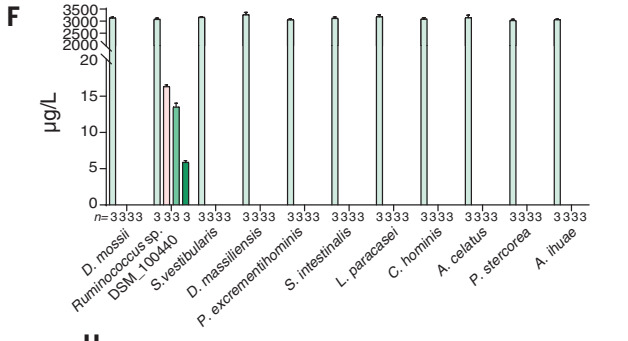

A
HSPC
CRPC
Family
s__[Eubacterium] siraeum
s__Prevotella sp. 109
s__Campylobacter hominis
s__Asaccharobacter celatus
s__Drancourtella massiliensis
s__Streptococcus pseudopneumoniae
s__Sellimonas intestinalis
s__Bacteroides sp. 9_1_42FAA
s__Burkholderiales bacterium
s__Faecalitalea cylindroides
s__Burkholderiales bacterium 1_1_47
s__Eubacterium sp. OM08-24
s__Parasutterella excrementihominis
s__Firmicutes bacterium AF36-3BH
s__Firmicutes bacterium AM55-24TS
s__Dysgonomonas mossii
s__Streptococcus parasanguinis
s__Streptococcus mutans
s__Ruminococcus sp. DSM_100440
s__Lactobacillus paracasei
s__Clostridiales bacterium VE202-14
s__Clostridia bacterium UC5.1-1D10
s__Streptococcus vestibularis
s__Bacteroides sp. AM32-11AC
s__Ruminococcus sp. OM05-10BH
s__Erysipelotrichaceae bacterium 5_2_54FAA
s__Bacteroides sp. AF17-1
s__Pseudoflavonifractor sp. BSD2780061688st1 E11
s__Desulfovibrio legallii
s__Coprobacillus sp. AF35-8
s__Bacteroides sp. AF34-31BH
s__Bacteroides sp. AR29
s__Drancourtella sp. An12
s__Prevotella stercorea
s__Coprococcus sp. AM27-12LB
s__Clostridia bacterium UC5.1-2G4
s__Erysipelotrichaceae bacterium 2_2_44A
s__Anaeroglobus geminatus
s__Debaryomyces fabryi
s__Prevotella sp. 885
s__Actinomyces ihuae
s__Prevotella sp. BCRC 81118
s__Prevotella sp. Marseille-P4119
Family
f__Ruminococcaceae
f__Prevotellaceae
f__Lachnospiraceae
f__Campylobacteraceae
f__Eggerthellaceae
f__Streptococcaceae
f__Bacteroidaceae
f__Erysipelotrichaceae
f__Eubacteriaceae
f__Veillonellaceae
f__Sutterellaceae
f__Debaryomycetaceae
f__Dysgonamonadaceae
f__Lactobacillaceae
f__Actinomycetaceae
f__Desulfovibrionaceae
f__
z-score
1
0
-1

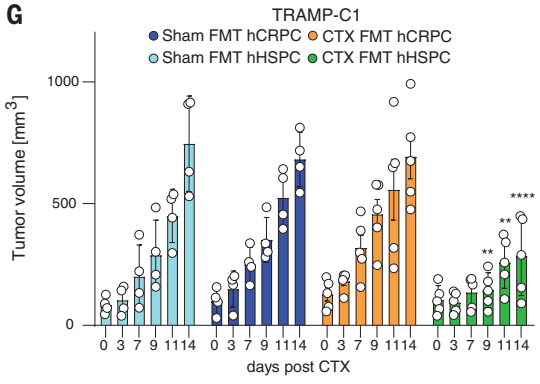

B
s__Dysgonomonas mossii
s__Streptococcus pseudopneumoniae
s__Ruminococcus sp. DSM_100440
s__Erysipelotrichaceae bacterium 5_2_54FAA
s__Prevotella sp. 109
s__Streptococcus vestibularis
s__Drancourtella massiliensis
s__Burkholderiales bacterium
s__Parasutterella excrementihominis
s__Burkholderiales bacterium 1_1_47
s__[Eubacterium] siraeum
s__Sellimonas intestinalis
s__Lactobacillus paracasei
s__Clostridiales bacterium VE202-14
s__Pseudoflavonifractor sp. BSD2780061688st1 E11
s__Clostridia bacterium UC5.1-1D10
s__Bacteroides sp. AF17-1
s__Campylobacter hominis
s__Drancourtella sp. An12
s__Bacteroides sp. AM32-11AC
s__Coprobacillus sp. AF35-8
s__Desulfovibrio legallii
s__Asaccharobacter celatus
s__Faecalitalea cylindroides
s__Ruminococcus sp. OM05-10BH
s__Prevotella sp. Marseille-P4119
s__Actinomyces ihuae
s__Erysipelotrichaceae bacterium 2_2_44A
s__Prevotella sp. 885
s__Prevotella stercorea
s__Prevotella sp. BCRC 81118
s__Anaeroglobus geminatus
s__Coprococcus sp. AM27-12LB
s__Debaryomyces fabryi
s__Clostridia bacterium UC5.1-2G4
Enriched in CRPC
Enriched in HSPC
Bacteria (OTU level)
-5
0
5
10
log2FC


C
100
75
50
25
0
% of patients
% of patients
HSPC
CRPC alive
CRPC dead
RRSC presence
RRSC absence
PPP presence
PPP absence


D
Enriched in HSPC
Enriched in CRPC
beta_Lactam_resistance
K1000623_Toluene_degradation
K1000140_Steroid_hormone_biosynthesis
K1000030_Pentose_phosphate_pathway
-3
-2
-1
0
1
2
LDA SCORE (log10)


E
µg/L
n=
D. mossii
Ruminococcus sp. DSM_100440
S.vestibularis
D. massiliensis
P. excrementihominis
S. intestinalis
L. paracasei
C. hominis
A. celatus
P. stercorea
A. ihuae
Pregnenolone
Hydroxypregnenolone
Progesterone
DHEA
Testosterone


F
µg/L
n=
D. mossii
Ruminococcus sp. DSM_100440
S.vestibularis
D. massiliensis
P. excrementihominis
S. intestinalis
L. paracasei
C. hominis
A. celatus
P. stercorea
A. ihuae


G
TRAMP-C1
Sham FMT hCRPC
CTX FMT hCRPC
Sham FMT hHSPC
CTX FMT hHSPC
Tumor volume [mm3]
days post CTX

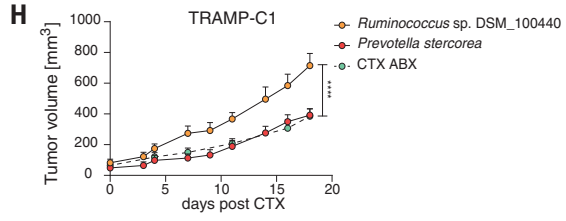

H
TRAMP-C1
Ruminococcus sp. DSM_100440
Prevotella stercorea
CTX ABX
Tumor volume [mm3]
days post CTX

Fig. 4. Metagenomic analysis of human CRPC fecal samples identifies bacterial species that produce androgens and promote castration resistance in gut-humanized mice. Rectal swabs were collected from HSPC (n = 19) and CRPC patients (n = 55) and shotgun whole-genome metagenome (WGM) sequencing was performed. (**A**) Heat map representing differentially abundant bacterial species (false discovery rate < 0.01) between HSPC and CRPC. (**B**) Waterfall plot showing bacterial species significantly enriched in the two cohorts [species with log(FC) > 4]. (**C**) Frequency of patients with detectable unfavorable species (RRSC; *Ruminococcus* sp. DSM_100440, *Ruminococcus* sp. OM05_10BH, *S. vestibularis*, *C. bacterium* VE202_14) or favorable species (PPP; *Prevotella* sp. BCRC_81118, *Prevotella* sp. Marseille_P4119, *Prevotella* sp. 885) according to clinical status (HSPC, CRPC alive, or CRPC dead patients). (**D**) Linear discriminant analysis (LDA) score of top KEGG pathways enriched in HSCP and CRPC of CH cohort. *D. mossii*, *Ruminococcus* sp. DSM_100440, *S. vestibularis*, *D. massiliensis*, *P. excrementihominis*, *S. intestinalis*, *L. paracasei*, *C. hominis*, *A. celatus*, *P. stercorea*, and *A. ihuae* were cultured for 48 hours in the presence of diverse androgen precursors in anaerobic conditions and analyzed for production of androgens or androgen precursors through LC-MS/MS. (**E** and **F**) Quantification of androgen pathway intermediates in bacterial conditioned media after pregnenolone (E) and hydroxypregnenolone (F) incubation. (**G**) Tumor volume of TRAMP-C1 mice that were treated with ABX for 7 days before CTX or sham operation and then with FMT from either hHSPC or hCRPC donors (sham hHSPC, n = 4; sham hCRPC, n = 4; CTX hCRPC, n = 5; CTX hCRPC, n = 5). (**H**) Tumor volume of TRAMP-C1 mice treated with ABX for 7 days before CTX, then treated via oral gavage with *Ruminococcus* sp. DSM_100440 (n = 6), *P. stercorea* (n = 5), or ABX (n = 9). $^{**}P < 0.01$, $^{****}P < 0.0001$ [two-way ANOVA and Sidak's multiple-comparison test in (G) and (H)]. Data in (E), (F), (G), and (H) are means ± SEM.

and F). However, the expansion of these bacteria had no impact on the growth of these androgen-insensitive tumors. Together, these data show that species enriched in the murine CR microbiota can participate in androgen metabolism in vitro and in vivo, and that this metabolic activity sustains androgen-dependent prostate tumor growth.

To validate the translational potential of our findings, we analyzed the gut microbiota of two patient cohorts with HSPC (n = 19) and metastatic CRPC (n = 55) (Table 1, fig. S10, A and B, and table S1). We performed shotgun whole-genome metagenome (WGM) sequencing of rectal swabs, leading to >16 million short DNA sequence reads per sample. These analyses highlighted a specific microbial signature in CRPC patients, with 33 species specifically enriched in CRPC and 10 species enriched in HSPC microbiota (Fig. 4A). We did not observe overall differences in the diversity indices of the two cohorts (fig. S10C). Interestingly, in the CRPC cohort, we observed an enrichment of members of the *Ruminococcus* and *Bacteroides* genera as compared to HSPC patients, mirroring the analysis performed in the murine model (fig. S10, D and E). Among the species enriched in the hCRPC microbiota, we found five species belonging to the *Bacteroides* genus and two species belonging to the *Ruminococcus* genus, whereas in the hHSPC microbiota we found four commensal species belonging to the *Prevotella* genus (Fig. 4, A and B). The commensal bacteria that were most significantly associated with a poor clinical outcome (defined on patient survival from the day of the swab, for 14 months), independent of disease stage, were *Ruminococcus* sp. DSM_100440, *Ruminococcus* sp. OM05_10BH, *Streptococcus vestibularis*, and *Clostridiales bacterium* VE202_14 (collectively RRSC) (fig. S11, A to D), whereas a more favorable clinical outcome was associated with *Prevotella* spp. BCRC_81118, Marseille_P4119, and 885 (collectively PPP) (fig. S11, E to G). Moreover, patients with the simultaneous presence of the RRSC species (termed an "unfavorable" fingerprint) had poorer survival relative to the patients where the four species were not simultaneously present (fig. S11H). In contrast, patients with the concomitant presence of the PPP species (termed a "favorable" fingerprint) were associated with improved survival (fig. S11I). Only a minority of the surviving patients at follow-up (HSPC 9.21% and CRPC alive 24.32%) had this unfavorable fingerprint, with the favorable microbiota fingerprint associating with a greater number of surviving patients after 14 months of follow-up (Fig. 4C and fig. S11J). In the CRPC patient cohort, *Ruminococcus* sp. DSM_100440 and *C. bacterium* VE202_14 were associated with a poor clinical outcome ($P < 0.01$), whereas *Prevotella* sp. 885 was linked to a favorable prognosis ($P < 0.1$). Next, we stratified mCRPC patients according to different ADT regimens (abiraterone versus enzalutamide) and we found that patients treated with enzalutamide but not abiraterone presented an expansion of the Ruminococcaceae family (fig. S12, A and B). KEGG pathway analysis performed on the gut microbiota of HSPC and CRPC patients showed that the steroid hormone biosynthesis pathway was enriched in the microbiota of patients with CRPC (Fig. 4D).

To assess whether the species enriched in the gut microbiota of the CRPC patients were capable of synthesizing androgens, we performed in vitro analysis of nine bacterial species enriched in the CRPC microbiota (*Dysgonomonas mossii*, *Ruminococcus* sp. DSM_100440, *S. vestibularis*, *Drancourtella massiliensis*, *Parasutterella excrementihominis*, *Sellimonas intestinalis*, *Lactobacillus paracasei*, *Campylobacter hominis*, *Asaccharobacter celatus*) and two species enriched in the HSPC microbiotas (*Prevotella stercorea*, *Actinomyces ihuae*). These bacteria were incubated with a panel of metabolites of the androgen biosynthesis pathway and the conversion to downstream metabolites was monitored, following the same experimental scheme used for the murine species (fig. S7E). Only *Ruminococcus* sp. DSM_100440, enriched in CRPC microbiota, had the ability to transform pregnenolone and hydroxypregnenolone into downstream androgenic steroids, including DHEA and testosterone (Fig. 4, E and F). *Ruminococcus* spp. DSM_100440 and/or OM05_10BH were more abundant in fecal samples of patients having higher serum testosterone levels (>10 ng/dl, 3/4, 75%) [χ^2 (1, n = 15) = 2.7841, P = 0.095205] (fig. S12C).

We next used the CM of *Ruminococcus* sp. DSM_100440 incubated with pregnenolone to treat two patient-derived organoids (PDOs), CP50 and CP50C, that are sensitive and insensitive (due to the presence of ARsv) to androgen deprivation, respectively (*30*) (fig. S13A). *Ruminococcus* sp. DSM_100440 CM stimulated the transcription of AR target genes in CP50 but did not do so in CP50C (fig. S13, B and C). Intriguingly, abiraterone, a selective inhibitor of CYP17A1, inhibited the bacterial conversion of pregnenolone in DHEA and testosterone (fig. S13, D to F). RNA sequencing of *R. gnavus* treated with pregnenolone showed up-regulation of 22 genes [log(FC) > 3.5, where FC is fold change relative to vehicle], some of which share high sequence homology with human CYP17 (fig. S13, G to I). These data support the existence of a bacterial enzyme that synthesizes androgenic steroids; however, further investigation is needed to identify the bacterial enzyme(s) responsible for the steroid biosynthesis. To examine whether human CRPC microbiota can promote tumor growth in vivo, we generated a mouse model where the gut microbiota of tumor-bearing mice were human-colonized by performing FMT experiments with feces from hHSPC or hCRPC patients. TRAMP-C1 mice pretreated with ABX were subjected or not to CTX and received FMT, and tumor growth was then monitored (Fig. 4G and fig. S14A). In line with the results obtained with murine FMT, hHSPC FMT limited tumor growth when compared to hCRPC FMT in CTX but not in sham-operated mice (Fig. 4G). In fact, hHSPC FMT reduced intratumoral expression of the AR target gene *Fkbp5* (fig. S14B). Notably, *Ruminococcus* sp. DSM_100440 could also reverse the efficacy

of ABX in castrated TRAMP-C1 allograft mice when compared to *P. stercorea* (enriched in HSPC) (Fig. 4H). As observed in the other studied mouse models, administration of *Ruminococcus* sp. DSM_100440 increased the circulating levels of DHEA and testosterone in recipient mice (fig. S14, C and D). Overall, these data provide a microbial blueprint of hCRPC microbiota and identify microbial species that alone or in combination affect prostate cancer outcome.

The gut microbiota is a recognized player in host fitness (*2*), modulating numerous bioactive molecules in the intestine, blood, and various extraintestinal organs, and may affect many cancer types through different mechanisms (*3*, *8*, *11*–*15*). However, its role in prostate cancer has remained underexplored. Some studies have reported altered fecal microbiota in prostate cancer patients, but the mechanisms through which the microbiota affects tumor growth have not been directly addressed (*19*–*21*). By using different mouse models of prostate cancer including PDXs, we show that androgen deprivation (CTX, CTX+Enza) drives the expansion of a peculiar intestinal microbiota, and that the gut microbiota affects CRPC growth by contributing to the host's androgen metabolism. Our findings show that these particular species can contribute to androgen metabolism, prostate cancer growth, endocrine treatment resistance, and disease outcome.

ADT is the standard first-line therapeutic strategy in patients with lethal prostate cancer, and numerous mechanisms of resistance to this treatment have been described, including increased AR expression, AR splicing (*31*), activation of aberrant cell signaling, recruitment of myeloid-derived suppressor cells (*23*) or plasma cells (*32*), and paracrine factors secreted by stromal cells as well as lineage plasticity (*33*–*35*). In patients treated with ADT, optimal castration is reached when patients' testosterone plasma level is reduced below 50 ng/dl. However, clinical evidence demonstrates that patients having androgen levels below 32 ng/dl have better outcomes than patients with testosterone levels between 32 and 50 (*36*). Thus, subtle variations in plasma androgen levels can affect the prognosis of prostate cancer patients. Our findings, if further validated in the clinic, may provide novel opportunities for the therapy of prostate cancer patients. Indeed, we have shown that FMT with hormone-sensitive microbiota or administration of *P. stercorea* can decrease androgens levels in CTX mice and delay the onset of CRPC. FMT has become the first-line therapy against *Clostridium difficile* infection (*37*), with clinical trials also demonstrating its efficacy in ulcerative colitis (*38*, *39*), and a number of ongoing clinical trials are studying this therapeutic strategy further. However, translation to transformative clinical trials in prostate cancer using FMT could be challenging, because HSPC patients become CRPC several years after starting ADT (*40*).

Moreover, we have identified in CRPC patients a fecal bacterial signature associated with decreased overall patient survival. If validated in prospective and larger clinical trials evaluating antibiotic and bacterial consortium transplantation strategies, this signature might be used as a minimally invasive biomarker to identify patients that could benefit from microbiota manipulation strategies.

REFERENCES AND NOTES

1. T. Chandrasekar, J. C. Yang, A. C. Gao, C. P. Evans, *Transl. Androl. Urol.* **4**, 365–380 (2015).
2. G. Eberl, *Mucosal Immunol.* **3**, 450–460 (2010).
3. W. S. Garrett, *Science* **348**, 80–86 (2015).
4. J. P. Nougayrède *et al*., *Science* **313**, 848–851 (2006).
5. A. Cougnoux *et al*., *Gut* **63**, 1932–1942 (2014).
6. J. C. Arthur *et al*., *Science* **338**, 120–123 (2012).
7. E. Zagato *et al*., *Nat. Microbiol.* **5**, 511–524 (2020).
8. A. D. Kostic *et al*., *Cell Host Microbe* **14**, 207–215 (2013).
9. C. Gur *et al*., *Immunity* **42**, 344–355 (2015).
10. A. Calcinotto *et al*., *Nat. Commun.* **9**, 4832 (2018).
11. S. Viaud *et al*., *Science* **342**, 971–976 (2013).
12. N. Iida *et al*., *Science* **342**, 967–970 (2013).
13. M. Vétizou *et al*., *Science* **350**, 1079–1084 (2015).
14. A. Sivan *et al*., *Science* **350**, 1084–1089 (2015).
15. B. Routy *et al*., *Science* **359**, 91–97 (2018).
16. V. Gopalakrishnan *et al*., *Science* **359**, 97–103 (2018).
17. Y. Feng *et al*., *BMC Genomics* **20**, 146 (2019).
18. I. Cavarretta *et al*., *Eur. Urol.* **72**, 625–631 (2017).
19. K. S. Sfanos *et al*., *Prostate Cancer Prostatic Dis.* **21**, 539–548 (2018).
20. M. A. Liss *et al*., *Eur. Urol.* **74**, 575–582 (2018).
21. D. M. Golombos *et al*., *Urology* **111**, 122–128 (2018).
22. H. Neuman, J. W. Debelius, R. Knight, O. Koren, *FEMS Microbiol. Rev.* **39**, 509–521 (2015).
23. A. Calcinotto *et al*., *Nature* **559**, 363–369 (2018).
24. F. Fransen *et al*., *Immunity* **43**, 527–540 (2015).
25. M. P. Menon, C. S. Higano, *Curr. Oncol. Rep.* **15**, 69–75 (2013).
26. A. Visconti *et al*., *Nat. Commun.* **10**, 4505 (2019).
27. J. Winter *et al*., *J. Lipid Res.* **25**, 1124–1131 (1984).
28. J. M. Ridlon *et al*., *J. Lipid Res.* **54**, 2437–2449 (2013).
29. E. A. Mostaghel *et al*., *Clin. Cancer Res.* **25**, 426–439 (2019).
30. J. Welti *et al*., *Clin. Cancer Res.* **24**, 3149–3162 (2018).
31. H. M. L. Kallio *et al*., *Br. J. Cancer* **119**, 347–356 (2018).
32. M. Ammirante, J. L. Luo, S. Grivennikov, S. Nedospasov, M. Karin, *Nature* **464**, 302–305 (2010).
33. S. Y. Ku *et al*., *Science* **355**, 78–83 (2017).
34. P. Mu *et al*., *Science* **355**, 84–88 (2017).
35. Z. Zhang *et al*., *Cancer Cell* **38**, 279–296.e9 (2020).
36. S. Dason, C. B. Allard, J. Tong, B. Shayegan, *Can. Urol. Assoc. J.* **7**, E263–E267 (2013).
37. G. Cammarota *et al*., *Gut* **66**, 569–580 (2017).
38. S. Paramsothy *et al*., *Lancet* **389**, 1218–1228 (2017).
39. P. Moayyedi *et al*., *Gastroenterology* **149**, 102–109.e6 (2015).
40. S. Gillessen *et al*., *Eur. Urol.* **77**, 508–547 (2020).

ACKNOWLEDGMENTS

We thank the members of Alimonti lab, IOR/IRB institute for instrument and core animal facility (Bellinzona), and all the collaborators for their help, meticulous analysis of the data, and critical input; the Clinical Microbiology Unit, in particular S. Favero; and D. Westaby for contribution to clinical specimen collection. **Funding:** Supported by ERC consolidator (683136) and Swiss Cancer League (KFS4267-08-2017) grant, Dr. Josef Steiner Foundation, Swiss Card-Onco-Grant of Alfred and Annemarie von Sick grant, Helmut Horten, SNSF (310030_176045), PCUK (RIA15-ST2-018), PCF challenge award (20CHAL04), Lega Ticinese Conto il Cancro, and Fondazione Leonardo. Also supported by a Prostate Cancer Foundation Young Investigator 2019 award (A.C.); the Prostate Cancer Foundation through a PCFYI award and PTCRC SEE PROS ONCOLOGIA–FPRC 5 PER MILLE–MS 2017 (P.R.); University of Florida Health Cancer Center Funds and University of Florida Department of Medicine Gatorade Fund (C.J.); and Prostate Cancer UK, the Movember Foundation through the London Movember Centre of Excellence (CEO13_2-002), the John Black Charitable Foundation and Prostate Cancer Foundation (18CHAL06), Cancer Research UK (Centre Programme grant), Experimental Cancer Medicine Centre grant funding from Cancer Research UK and the Department of Health, and Biomedical Research Centre funding to the Royal Marsden (J.d.B.). J.d.B. is a National Institute for Health Research (NIHR) Senior Investigator. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. We thank all foundations for their support. **Author contributions:** A.C. and A.A. conceived the project; A.A., N.P., E.Z., and A.C. designed the experiments and interpreted the results; N.P., E.Z., A.C., G.A., B.C., A.R.E., D.B., and E.P. performed the in vivo and in vitro experiments; M.T. performed bioinformatic analyses with support from M.Bo. and J.-P.T.; A.Ri. and M.Bo. performed bacteria RNA sequencing and analysis; M.T., N.P., and A.A. interpreted shotgun metagenomic sequencing analyses; J.T. performed metabolomics analyses; M.M. performed the RT-qPCR experiment; E.Z., N.P., A.C., and B.C. performed FACS analyses; S.Mo. performed immunohistochemical experiments; A.Re. performed gland scoring; R.P.M., P.R., P.F., J.H., M.F., E.D.A., A.B., and S.Me. selected and provided human samples and helped with patients' data; A.N., L.B., and J.W. performed in vitro PDO experiments; M.G. and M.Be. performed treatments in the MRI facility; A.P. and A.E. performed MRI analysis and analyzed the results; J.G., R.Z.G., and C.J. performed and analyzed 16*S* rRNA sequencing; G.M.L. provided the microbiology facility at the EOC microbiology unit; J.-P.T., F.G., and M.R. contributed to data discussion and interpretation; S.G. and J.d.B. supervised human experiments and helped in clinical data interpretation; E.Z., N.P., A.C., and A.A. wrote the paper. **Competing interests:** J.T. is employed by the company Theoreo Srl, which is involved in the development and marketing of metabolomics-based diagnostic tests. S.G. is a member of the advisory board, steering committee, or speakers' bureau of Orion, Janssen, Amgen, Merck Sharp & Dohme, Tolero Pharmaceuticals, Astellas Pharma, Bayer, Roche, Pfizer, Telixpharma, BMS, and AAA International. J.d.B. has served on advisory boards and received fees from companies including Amgen, Astra Zeneca, Astellas, Bayer, Bioxcel Therapeutics, Boehringer Ingelheim, Cellcentric, Daiichi, Eisai, Genentech/Roche, Genmab, GSK, Harpoon, Janssen, Merck Serono, Merck Sharp & Dohme, Menarini/Silicon Biosystems, Orion, Pfizer, Qiagen, Sanofi Aventis, Sierra Oncology, Taiho, Terumo, and Vertex Pharmaceuticals. J.d.B. is an employee of the ICR, which has received funding or other support for his research work from AZ, Astellas, Bayer, Cellcentric, Daiichi, Genentech, Genmab, GSK, Janssen, Merck Serono, Merck Sharp & Dohme, Menarini/Silicon Biosystems, Orion, Sanofi Aventis, Sierra Oncology, Taiho, Pfizer, and Vertex, and which has a commercial interest in abiraterone, PARP inhibition in DNA repair–defective cancers, and PI3K/AKT pathway inhibitors (no personal income). J.d.B. is inventor, with no financial interest, on patent 8,822,438 submitted by Janssen that covers the use of abiraterone acetate with corticosteroids. N.P., E.Z., A.C., M.T., and A.A. are inventors on patent application with filing number 102021000021974 submitted by the Institute of Oncology Research (IOR) that covers the use of antibiotics, FMT, and probiotics in prostate cancer therapy. **Data and materials availability:** All data are available in the main text or the supplementary materials. 16*S* sequencing data have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under bioproject ID PRJNA748258. Whole-genome metagenomics data from human fecal swab have been deposited in NCBI SRA under bioproject ID PRJNA749645. Bacterial RNA-seq data have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE180863.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf8403
Materials and Methods
Figs. S1 to S14
Tables S1 to S3
References (*41*–*55*)

23 November 2020; resubmitted 9 June 2021
Accepted 8 September 2021
10.1126/science.abf8403

TOPOLOGICAL OPTICS

Topological phonon-polariton funneling in midinfrared metasurfaces

S. Guddala[1,2], F. Komissarenko[1,2], S. Kiriushechkina[1], A. Vakulenko[1], M. Li[1,2,3], V. M. Menon[2,3], A. Alù[4,3,1], A. B. Khanikaev[1,2,3]*

Topological photonics offers enhanced control over electromagnetic fields by providing a platform for robust trapping and guiding of topological states of light. By combining the strong coupling between topological photons with phonons in hexagonal boron nitride (hBN), we demonstrate a platform to control and guide hybrid states of light and lattice vibrations. The observed topological edge states of phonon-polaritons are found to carry nonzero angular momentum locked to their propagation direction, which enables their robust transport. Thus, these topological quasiparticles enable the funneling of infrared phonons mediated by helical infrared photons along arbitrary pathways and across sharp bends, thereby offering opportunities for applications ranging from Raman and vibrational spectroscopy with structured phonon-polaritons to directional heat dissipation.

Topologically nontrivial phases of matter have been affecting our fundamental understanding of the physics of quantum systems. Successfully transferred to the realms of classical physics, topological phases have been realized in a variety of photonic (*1*, *2*), acoustic, and mechanical systems (*3*, *4*). Among classical platforms, topological photonics represents the first example of the successful realization of different topological phases (*2*). The advantage of classical systems—electromagnetic systems in particular—enabled by their engineerability, not only facilitates the testing of some of the most peculiar theoretical models of topological physics but also expands the domain of topological phases into previously unthinkable regimes, such as topological lasers (*5*–*7*), non-Hermitian systems (*8*, *9*), and nonlinear topological phases (*10*–*12*).

Another advantage of photonic topological systems lies in the ability of electromagnetic fields to interact with a broad range of excitations, from magnons and spin waves to phonons, plasmons, and excitons, which gives rise to the formation of half-light half-matter quasiparticles known as polaritons (*13*). This opens new realms for topological multiphysics, where topological phases can be transferred from one type of physical domain to another or emerge as the result of multiphysics interactions (*14*–*16*). Recent work has shown that interactions of light with excitons may lead to the formation of topological exciton-polaritons (*17*–*20*).

Here, we demonstrate a topological polaritonic system in which midinfrared (mid-IR) topological photonic states are strongly coupled with in-plane vibrations (phonons) in hBN. Our system consists of a metasurface supporting a spin-Hall–like (Z_2) topological photonic phase (*21*) with a (few-layer) hBN on top of it (Fig. 1A). The metasurface is formed by an array of triangular holes (*22*) in silicon-on-sapphire (SOS) substrate. This metasurface geometry is known to undergo a topological transition from a trivial to a nontrivial phase (Fig. 1B) when the design of the unit cell changes from shrunken [triangles within the unit cell are closer to one another than the triangles in the nearby unit cells (Fig. 1A, light blue hexagon)] to expanded [triangles in the unit cell are farther from one another compared with those in the nearby unit cells (Fig. 1A, red hexagon)]. The existence of topological photonic edge states has been extensively studied

[1]Department of Electrical Engineering, Grove School of Engineering, City College of the City University of New York, New York, NY 10031, USA. [2]Department of Physics, City College of New York, New York, NY 10031, USA. [3]Physics Program, Graduate Center of the City University of New York, New York, NY 10016, USA. [4]Photonics Initiative, Advanced Science Research Center, City University of New York, New York, NY 10031, USA.
*Corresponding author. Email: akhanikaev@ccny.cuny.edu

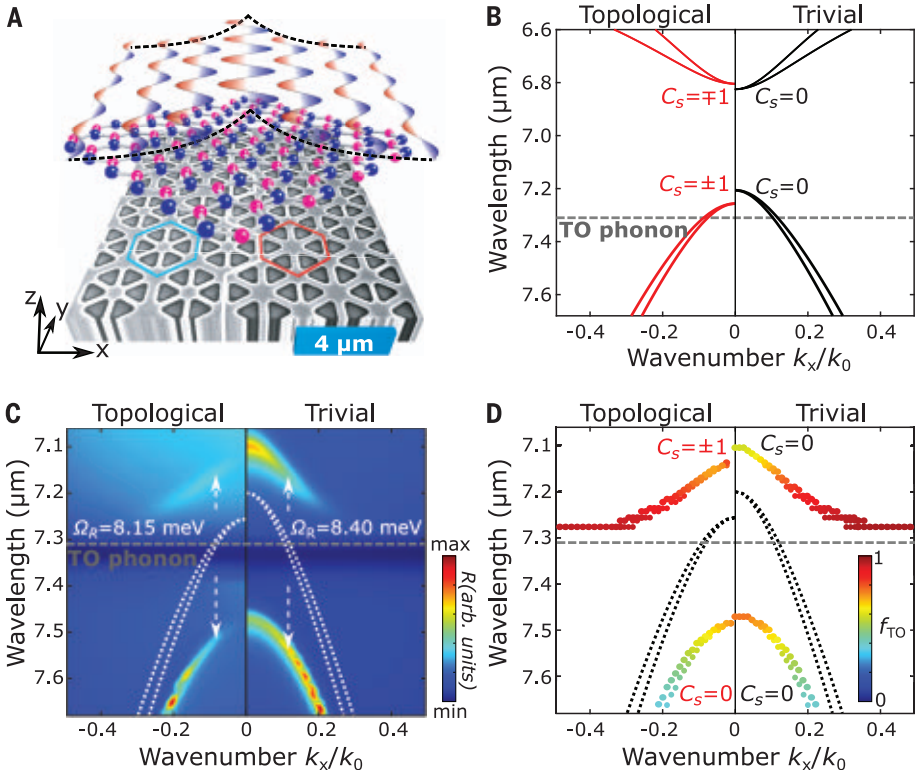


Fig. 1. Phonon-polaritons in heterostructures formed by a topological mid-IR metasurface loaded by an hBN layer. (**A**) Schematic of the phonon-polariton system with extruded scanning electron microscopy (SEM) image of the metasurface with a few-layer hBN on top. Light blue and red hexagons indicate unit cells of trivial and topological domains, respectively, and the armchair-shaped white line shows the domain wall. Color waves show transverse phonon oscillations in hBN as a result of topological edge polaritons. (**B**) Photonic band structure of topological and trivial mid-IR metasurfaces. A dashed line indicates the spectral position of the TO phonon mode in hBN. (**C**) Numerical reflectance resulting from bulk phonon-polaritons in topological and trivial metasurfaces. Ω_R between the arrows indicates the Rabi splitting between the upper and lower polaritons for the two cases. (**D**) Spectral position and fraction (f_{TO}) of the phononic component in bulk polaritons. Dotted lines in (C) and (D) show the photonic bands without hBN.

in such a system at optical frequencies (*22*, *23*). Here, we tailor this metasurface to support bulk modes in the mid-IR frequency range (near ω_{TO} = 1368 cm^{-1}, which corresponds to the wavelength $\lambda \approx 7.31$ μm, where ω_{TO} is the transverse optical (TO) phonon resonance frequency of hBN) (Fig. 1B) to enable their hybridization with TO phonons in hBN (*24*–*26*).

The addition of the hBN layer on top of the metasurface leads to coupling between photonic and phononic modes, giving rise to the formation of hybrid polaritonic states. The condition of strong coupling, required for the observation of topological phonon-polaritons (*20*), can be achieved when the hBN thickness exceeds ~5 nm. From this point forward, we consider the case of the 15-nm-thick hBN. To confirm the strong coupling regime, we performed full-wave numerical calculations. The calculated reflectance in the vicinity of the lower bulk photonic bands and the phonon line is shown in Fig. 1C, which evidences the formation of lower and upper phonon-polaritons with a Rabi splitting (Ω_R) of ~8.15 meV and 8.40 meV (supplementary note 1) (*27*) for topological and trivial cases, respectively. We then extracted the polaritonic band structure by tracing the maxima of energy stored in the system, shown in Fig. 1D, which confirms the avoided crossing behavior. The color in Fig. 1D indicates the fraction of the phonon component in the polaritonic mode f_{TO}, which can be directly related to the Hopfield coefficient of the mode (supplementary note 2) (*27*). It is worth noting here that our system represents a polaritonic system, and thus, it should be differentiated from recent demonstrations of topological phases for pure mechanical modes (*28*–*30*). TO phonons in hBN are characterized by vanishing group and energy velocities at small phonon wave numbers; therefore, lattice vibrations do not directly contribute to the energy transfer of the reported polaritonic modes (supplementary note 4) (*27*). The transport of vibrations is thus mediated by the electromagnetic field strongly coupled to phonons through the polarization of hBN.

An important consequence of the strong coupling is the transfer of topological invariant from the photonic mode to the upper polaritonic mode (*20*), which is ultimately responsible for the formation of topologically robust helical edge phonon-polaritons. A calculation of the topological invariants shows that, for the case of a TO phonon crossing the lower photonic (EM) bulk mode [$\omega_{TO} < \omega_{EM}(k = 0)$], the upper polariton acquires spin-Chern numbers $C_s = \pm 1$ for the two pseudospins (supplementary note 3) (*27*)—a scenario realized in the experiment described below. For the case of the TO phonon line above the lower photonic bulk mode, the spin-Chern number is transferred to the lower polariton instead.

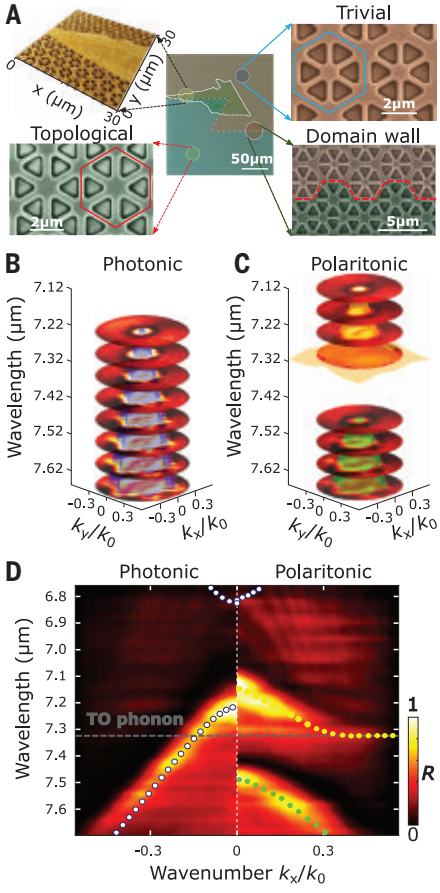


Fig. 2. Experimental evidence of topological phonon-polariton formation. (**A**) SEM images of metasurface with trivial (orange-shaded) and topological (green-shaded) domains separated by an armchair-shaped domain wall. Zoom-ins show enlarged regions with hBN obtained by atomic force microscopy (AFM), with topological and trivial domains, and the domain wall. (**B**) Raw data: Experimental constant-frequency contours imaged from the region without hBN and representing slices of lower photonic bands in Fig. 1, B and D. The blue semitransparent surface shows calculated dispersion. (**C**) Raw data: Experimental constant-frequency contours corresponding to the region with hBN and corresponding to the bulk polaritonic bands in Fig. 1, C and D. The semitransparent bands show calculated dispersions of lower (green) and upper (yellow) bulk phonon-polaritons. (**D**) Slices of a 3D Fourier tomogram showing frequency bands along the *x* direction. Left and right halves correspond to photonic (no hBN) and polaritonic (with hBN) cases. The dotted lines show simulated photonic (white) and polaritonic (yellow and green) band dispersions. The trivial domain case with bright dipolar bands is shown in (B) through (D). The topological domain data are provided in fig. S4.

To experimentally confirm strong coupling between photonic and phononic modes, and at the same time to enable the observation of topological edge phonon-polaritons, a sample characterized by a zigzag-shaped domain wall was fabricated by patterning a 1-μm SOS substrate (Fig. 2A). A 15-nm hBN layer was transferred on top to cover half of the zigzag (Fig. 2A), which enabled a direct comparison of topological photonic and polaritonic regimes.

A home-built mid-IR imaging system (fig. S1) was used for both real-space and Fourier plane imaging of the sample (supplementary materials, materials and methods section) (*27*). This system allows the extraction of a sequence of constant frequency contours (a Fourier tomogram), with Fig. 2, B and C, illustrating the case of a trivial metasurface without and with hBN, respectively, near the polaritonic modes. In agreement with our numerical results, the addition of the 15-nm hBN layer alters the band structure as a result of the strong coupling, leading to the formation of the upper polariton, which asymptotically approaches the TO phonon line, and of the lower polariton, which asymptotically approaches the lower photonic band. The polariton formation is even more obvious from the slices of the three-dimensional (3D) band structure taken in the k_x-direction (Fig. 2D), which compares photonic and polaritonic cases side by side.

We then confirmed the presence of helical polaritonic edge states and their robustness. We collected both Fourier plane and real-space images of the structure under linear and circularly polarized illumination in the regions with and without hBN at the domain wall. As a reference, the numerically calculated reflectivity spectra from the region near the domain wall (Fig. 3A) clearly show edge states in both photonic and polaritonic regimes. Extraction of the spectral position of the edge bands for both cases is then done by tracing the stored energy maximum (Fig. 3B).

The top panel of Fig. 3C shows experimental Fourier space images along the domain walls obtained for left circularly polarized (LCP) excitations. These images reveal the one-way character and helical nature of the edge states for both photonic (left) and polaritonic (right) cases. The bottom panel shows the spectral position of the edge states extracted from the reflectivity spectra. In agreement with bulk-boundary correspondence and our theoretical predictions in Fig. 3, A and B, the edge states asymptotically approach the upper polaritonic band that carries a nonzero spin-Chern number. The excellent agreement between experimental results and the simulations in Fig. 3, A and B, implies that our numerical results can be used to explore otherwise inaccessible properties of the edge states. Specifically, we can conclude that the edge polaritons carry a substantial phononic component, which reaches 50% at $k_x = 0$. By directly calculating transverse angular momentum (fig. S3), we also deduce that the edge phonon-polaritons inherit the helical nature of the photonic modes (supplementary note 5) (*27*).

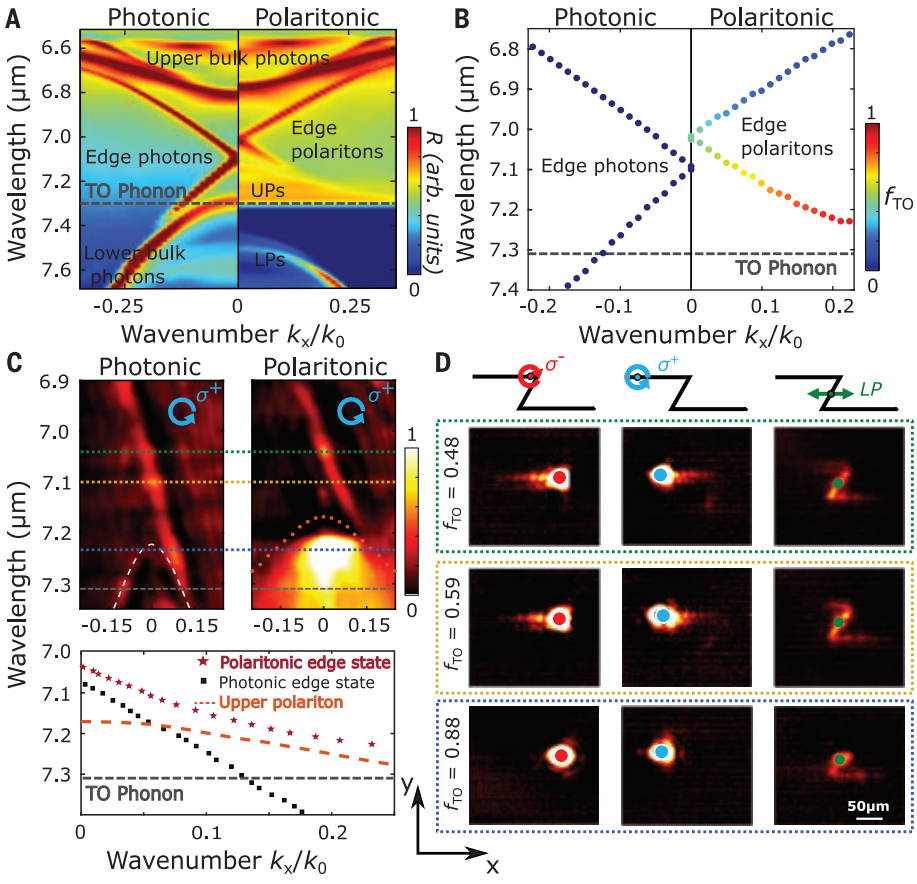


Fig. 3. Topological edge phonon-polaritons. (**A**) Numerical reflectivity spectra of topological domain wall revealing photonic and polaritonic edge states. UPs, upper polaritons; LPs, lower polaritons. (**B**) Numerically calculated spectral position and phononic fractions of the edge polaritons. (**C**) (Top) Experimental Fourier plane images revealing dispersion of edge states. (Bottom) Dispersion of edge states extracted from reflectivity spectra. (**D**) Unidirectional (left and middle panels) and bidirectional (right panels) excitation of the edge states by circularly and linearly polarized (LP) laser beam, respectively. The beam is focused on different locations of the domain wall (indicated by the top insets) for three different frequencies [color-coded dashed lines indicate the frequencies in (C)] for increasing phononic fraction f_{TO}. Cross-polarized collection suppresses the laser spot in the right panels. The gray dashed lines in (A) and (C) indicate the spectral position of the TO phonon mode. The white dashed line and brown dotted line show the calculated dispersion of bulk photonic and upper polaritonic bands, respectively.

The helical character of edge phonon-polaritons in real space and their robustness were confirmed in a series of measurements with excitation at different locations (indicated in the top inset of Fig. 3D) along the zigzag of the domain wall and for different frequencies. First, edge polaritons were excited with a circularly polarized beam, and we observed unidirectional propagation (left and middle panels in Fig. 3D) of the modes. Second, linearly polarized light was used to excite the edge state in the middle of the zigzag, exactly at the boundary of the hBN flake (Fig. 2A). This configuration allowed us to observe polaritonic and photonic edge states at once. We can see that, like the photonic edge states, the edge phonon-polaritons exhibit topological resilience concerning sharp bending, with the only difference of shorter propagation length because of the absorption in hBN. Nonetheless, the propagation distance of the edge polaritons reaches ~80 μm for k_x = 0 (upper row in Fig. 3D), despite the 50% phonon fraction (compared with ~150 μm for the purely photonic scenario). As expected, for the frequencies below the bulk upper polariton (lower row in Fig. 3C), the edge mode is observed only in the region without hBN, again confirming that the polaritonic edge state originates from the bulk upper polariton carrying a topological charge. More details on the comparison of propagation features of the edge states can be found in fig. S5 (supplementary section 7) (*27*).

Our results reveal a topological polaritonic phase with resilient helical boundary states representing half-light half-phonons trapped to a topological mid-IR metasurface integrating a few-layer hBN. Our work provides an opportunity to manipulate phonons by endowing them with topological characteristics through their hybridization with topological light. Topological phonon-polaritons pave the way for robust and unidirectional heat sinks and can lay the foundation for Raman spectroscopy with structured vibrations carrying angular momentum locked to their propagation direction.

REFERENCES AND NOTES

1. L. Lu, J. D. Joannopoulos, M. Soljačić, *Nat. Photonics* **8**, 821–829 (2014).
2. T. Ozawa *et al.*, *Rev. Mod. Phys.* **91**, 015006 (2019).
3. Z. Yang *et al.*, *Phys. Rev. Lett.* **114**, 114301 (2015).
4. S. D. Huber, *Nat. Phys.* **12**, 621–623 (2016).
5. B. Bahari *et al.*, *Science* **358**, 636–640 (2017).
6. G. Harari *et al.*, *Science* **359**, eaar4003 (2018).
7. M. A. Bandres *et al.*, *Science* **359**, eaar4005 (2018).
8. D. Leykam, K. Y. Bliokh, C. Huang, Y. D. Chong, F. Nori, *Phys. Rev. Lett.* **118**, 040401 (2017).
9. S. Weimann *et al.*, *Nat. Mater.* **16**, 433–438 (2017).
10. D. Smirnova, D. Leykam, Y. Chong, Y. Kivshar, *Appl. Phys. Rev.* **7**, 021306 (2020).
11. L. J. Maczewsky *et al.*, *Science* **370**, 701–704 (2020).
12. S. Mukherjee, M. C. Rechtsman, *Science* **368**, 856–859 (2020).
13. D. N. Basov, M. M. Fogler, F. J. García de Abajo, *Science* **354**, aag1992 (2016).
14. V. Peano, C. Brendel, M. Schmidt, F. Marquardt, *Phys. Rev. X* **5**, 031011 (2015).
15. T. Karzig, C. E. Bardyn, N. H. Lindner, G. Refael, *Phys. Rev. X* **5**, 031001 (2015).
16. A. V. Nalitov, D. D. Solnyshkov, G. Malpuech, *Phys. Rev. Lett.* **114**, 116401 (2015).
17. M. Milićević *et al.*, *2D Mater.* **2**, 034012 (2015).
18. S. Klembt *et al.*, *Nature* **562**, 552–556 (2018).
19. W. Liu *et al.*, *Science* **370**, 600–604 (2020).
20. M. Li *et al.*, *Nat. Commun.* **12**, 4425 (2021).
21. L. H. Wu, X. Hu, *Phys. Rev. Lett.* **114**, 223901 (2015).
22. S. Barik *et al.*, *Science* **359**, 666–668 (2018).
23. N. Parappurath, F. Alpeggiani, L. Kuipers, E. Verhagen, *Sci. Adv.* **6**, eaaw4137 (2020).
24. S. Dai *et al.*, *Science* **343**, 1125–1129 (2014).
25. M. Autore *et al.*, *Light Sci. Appl.* **7**, 17172 (2018).
26. J. D. Caldwell *et al.*, *Nat. Rev. Mater.* **4**, 552–567 (2019).
27. See the supplementary materials.
28. R. Süsstrunk, S. D. Huber, *Science* **349**, 47–50 (2015).
29. J. P. Mathew, J. del Pino, E. Verhagen, *Nat. Nanotechnol.* **15**, 198–202 (2020).
30. H. Ren *et al.*, Topological phonon transport in an optomechanical system. arXiv:2009.06174 [cond-mat.mes-hall] (2020).
31. S. Guddala *et al.*, Topological phonon-polariton funneling in mid-infrared metasurfaces, version 2, Zenodo (2021); https://doi.org/10.5281/zenodo.5524881.

ACKNOWLEDGMENTS

Funding: The work was supported by the ONR award N00014-21-1-2092, NSF grants DMR-1809915 and OMA-1936351, the DARPA Nascent program, and the Simons Collaboration on Extreme Wave Phenomena. A.A. acknowledges support by ONR award N00014-19-1-2011 and a Vannevar Bush Faculty Fellowship. **Author contributions:** A.B.K. and S.G. conceived the project. S.G., F.K., S.K., and A.V. performed the experiments. A.B.K., M.L., S.G., A.V., and S.K. performed theoretical studies. A.B.K., A.A., and V.M.M. supervised the research. A.B.K. and S.G. wrote the manuscript with input from all coauthors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are deposited at Zenodo (*31*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj5488
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (*32*–*39*)

19 May 2021; accepted 31 August 2021
10.1126/science.abj5488

Caltech

Faculty Opening in Chemical Engineering

Division of Chemistry and Chemical Engineering
California Institute of Technology
https://applications.caltech.edu/jobs/chemical

The Division of Chemistry and Chemical Engineering at the California Institute of Technology invites applications for a **faculty position** at any academic level in the area of chemical engineering. Candidates with strong commitments to research and teaching excellence are encouraged to apply. For appointments at the assistant professor level, the term of the initial appointment is four years and is contingent upon completion of all requirements for a Ph.D. in chemical engineering or in a related field. Reappointments beyond the initial term is contingent upon successful review conducted prior to the commencement of the fourth year. Exceptionally well-qualified applicants at the tenured level will also be considered.

Interested candidates should apply at **https://applications.caltech.edu/jobs/chemical**. Applications should include a curriculum vita, publication list, teaching statement, a description of proposed research, and three letters of recommendation. Applicants should submit a diversity and inclusion statement that discusses past and/or anticipated contributions to improving diversity, equity, and inclusion in the areas of research, teaching, and/or outreach. Applications should be received by **November 15, 2021**.

Caltech has implemented a mandatory vaccination policy effective **October 15, 2021**, requiring COVID-19 vaccinations for all employees (with Institute-approved exemptions for religious or medical contraindications only). At present, an individual is deemed "fully vaccinated" when that person has received, at least 14 days prior, either the second dose in a two-dose COVID-19 vaccine series (e.g., Pfizer or Moderna) or a single-dose COVID-19 vaccine (e.g., Johnson & Johnson). As a condition of employment, employees must submit official COVID-19 vaccine documentation. More information may be found at **https://together.caltech.edu**.

We are an equal opportunity employer, and all qualified applicants will receive consideration for employment without regard to age, race, color, religion, sex, sexual orientation, gender identity, national origin, disability status, protected veteran status, or any other characteristic protected by law.

RUTGERS
School of Arts and Sciences

FACULTY POSITION IN ORGANIC CHEMISTRY

The Department of Chemistry and Chemical Biology (CCB) at Rutgers, The State University of New Jersey (New Brunswick) is seeking outstanding candidates for a tenure-track appointment at the **assistant professor** level.

Our search will focus on candidates with expertise in organic chemistry, preferably with an emphasis on design and synthesis of functional polymers or biomaterials, broadly defined, but we will consider excellent candidates at all ranks and in all areas of organic chemistry, including but not limited to molecular synthesis, medicinal chemistry, and chemical biology.

CCB faculty are expected to build and sustain a world-class research program, have a significant record of scholarship in fundamental research, and excel in teaching and service roles. The successful applicant will be housed in our new 145,000 sf, state-of-the-art chemistry building and dedicated Presidential Diversity Funds may be available to support the accelerated recruitment of candidates from underrepresented groups.

How to Apply: Applicants should submit the following to **https://jobs.rutgers.edu/postings/143027**

(1) cover letter and detailed CV; **(2)** four to six-page research development statement; **(3)** one-page teaching statement; and **(4)** contact information for three references. Questions about this search should be directed to **chemchair@chem.rutgers.edu**. Review of applications will begin on **October 15, 2021** and will continue until the position is filled.

Rutgers University is an EEOC/AA employer

John Innes Centre
Unlocking Nature's Diversity

Independent Research Fellowships Leading to Tenured Faculty Positions at the John Innes Centre

The John Innes Centre (JIC), is a world leading centre of excellence in plant and microbial sciences based on the Norwich Research Park, UK. We are inviting applications from outstanding researchers who either hold, or wish to apply for Independent Research Fellowships (such as a **UKRI Future Leaders Fellowship**, or a **Royal Society University Research Fellowship**). Shortlisted candidates will be invited to give a seminar at the Fellows Conference, which will be held **07 February 2022**. During the conference, you will be able to discuss your proposals, the development of your group and your future career plans in depth with JIC Faculty in one-to-one meetings.

After the conference, we will select and mentor outstanding candidates in writing Fellowship applications and/or offer the opportunity to move existing Fellowships to the JIC. Candidates who win Fellowships will be considered for transfer onto tenure-track at 3 years, and if transferred, for tenure at 5 years. Considerable additional resources will be provided to Fellows by the Centre. For further information, please contact Prof. Matt Hutchings: **matt.hutchings@jic.ac.uk**.

Further details and particulars can be found at **https://www.jic.ac.uk/vacancies/**

To apply: please e-mail a 2-page summary of your research plan, a copy of your CV and arrange for three letters of recommendation to be e-mailed to **fellows@jic.ac.uk** by **Friday 03 December 2021**. Before applying, please read our Privacy Notice **https://jic.link/IRF-Privacy**.

The John Innes Centre is a registered charity (No223852) grant-aided by the Biotechnology and Biological Sciences Research Council and is an Equal Opportunities Employer and supports flexible working.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!

ScienceCareers.org/booklets

Tenure-track Faculty Positions in Biochemistry

Department of Chemistry and Biochemistry, College of Sciences and Mathematics
Auburn University

The Department of Chemistry and Biochemistry at Auburn University invites applications for two nine-month, tenure-track faculty positions at the rank of **Assistant** or **Associate Professor** in Biochemistry. Candidates in all areas of biochemistry are encouraged to apply. Bioinorganic spectroscopy, natural product biosynthesis, -omics, and structural biology are areas that align well with departmental strengths and ambitions. Successful candidates are expected to develop vigorous, externally funded research programs. Duties also include teaching at both the undergraduate and graduate levels. Thus, excellent written and interpersonal communication skills are required. A Ph.D. in Chemistry, Biochemistry, or a closely related field and at least one year of postdoctoral experience are required at the time employment begins. Candidates must be eligible to work in the United States at the time employee begins and continue working legally for the proposed term of employment.

Review of applications will begin **October 31, 2021**, and will continue until the positions are filled. Please submit a cover letter, curriculum vitae, statements of research plans and teaching philosophy, information for three professional references, and an inclusion statement describing how their past and/or planned contributions to teaching, research, and/or service will serve to advance the mission of the College of Sciences and Mathematics in creating an inclusive environment. Please see **https://www.auburn.edu/cosam/about/mission-vision-oied.htm** for more information.

At the College of Sciences and Mathematics, we are committed to cultivating culturally competent professionals. Our department values diversity and is a welcoming and inclusive community of 28 research and teaching faculty and over 80 graduate students. Additional information about the Department of Chemistry and Biochemistry is available at: **http://www.auburn.edu/chemistry**. As an R1, Doctoral University and one of the nation's premier land, sea, and space grant institutions, Auburn University is understanding of and sensitive to the family needs of faculty, including dual-career couples. For more information, please visit: **http://www.auburn.edu/academic/provost/facultyjobs/**.

To apply for this position, please visit: **https://www.auemployment.com/postings/25261**.

Auburn University is an EEO/Vet/Disability Employer and committed to building an inclusive and diverse community.

UT Southwestern Medical Center

TENURE-TRACK POSITION

The Department of Physiology invites outstanding scientists with Ph.D., M.D., or equivalent degrees to apply for tenure-track faculty positions at the level of **Assistant Professor**.

Candidates who bring innovative approaches to the study of any under-explored/unexplored questions broadly related to physiology are encouraged to apply. The scientific excellence of the candidates is more important than the specific area of research. These positions are part of the continuing growth of the Department at one of the country's leading academic medical centers. They will be supported by significant laboratory space, competitive salaries, state-of-the-art core facilities and exceptional start-up packages. The University of Texas Southwestern Medical Center is the scientific home to six Nobel Prize laureates and many members of the National Academy of Sciences and Institute of Medicine. UT Southwestern conducts more than 5,800 research projects annually totaling more than $524 million. Additional information about the Department of Physiology can be found at **http://www.utsouthwestern.edu/education/medical-school/departments/physiology/index.html**.

Information regarding careers can be found at:
https://jobs.utsouthwestern.edu/.

Applicants should submit a CV, a brief statement of current and proposed research, and a summary of your two most significant publications describing the importance of the work (100-150 words each). Please arrange to have three letters of recommendation sent on his/her behalf. All items should be submitted to: **http://academicjobsonline.org/ajo/jobs/19221**. Completed applications will be reviewed starting **November 1, 2021**. You may email questions to **ron.doris@utsouthwestern.edu**.

UT Southwestern Medical Center is an Equal Opportunity/Affirmative Action Employer. Women, minorities, veterans and individuals with disabilities are encouraged to apply.

UT Southwestern Medical Center

FACULTY POSITION – VIROLOGY

The Department of Microbiology at UT Southwestern Medical Center is seeking new faculty members in virology at the **Assistant Professor** (tenure track) level via a multi-year rolling search. The appointee will be expected to develop a front-rank, competitive, independent research program on a medically relevant viral pathogen(s). Some preference may be given to candidates working on emerging/re-emerging viruses or those that represent global threats.

An attractive start-up package, competitive salary, and generous laboratory space are available to conduct research in a highly collaborative, supportive, and dynamic environment.

Candidates also will be considered for UT Southwestern's $2.15M Endowed Scholars (start-up) Program Candidates should have a Ph.D. and/or M.D. degree with 3+ years of postdoctoral experience and an exceptional publication record.

Please send a cover letter, C.V., contact information for three letters of recommendation, and a brief summary of future research to:

VirologySearchCommittee@utsouthwestern.edu.

UT Southwestern Medical Center is committed to an educational and working environment that provides equal opportunity to all members of the University community. As an equal opportunity employer, UT Southwestern prohibits unlawful discrimination, including discrimination on the basis of race, color, religion, national origin, sex, sexual orientation, gender identity, gender expression, age, disability, genetic information, citizenship status, or veteran status. To learn more, please visit: ***https://jobs.utsouthwestern.edu/why-work-here/diversity-inclusion/.***

By **Samantha J. Butler**

The confidence to question

It was almost business as usual when I rose from the audience to ask a question after a research talk. I was nervous, as I often am in these situations. Since the beginning of my career, my brain's fear center in the amygdala has been determined to warn me that no good will come from speaking up. Over the years I've learned to project confidence, even though I still question the validity of my question right up to the moment I start to talk. But I also know a penetrating question can be invaluable. I hope to receive such questions when I give a talk, and I teach my students to lean into them. So, I asked the presenter to put her data slides back up and drilled into them for a couple minutes. I posed an alternative hypothesis; the presenter pushed back. Others had questions, so I sat down. It was a tense but healthy debate, I thought.

But this exchange didn't happen at a scientific meeting. It was at an informational session at my child's elementary school, which is run by my university. The speaker, a colleague in the education department, had presented data showing pupils' math test scores had increased, which the school attributed to its teaching method emphasizing storytelling and group discussion. I had countered that, because the school did not keep statistics on which children were tutored—something that my casual observations suggested was increasingly common—it was impossible to determine whether its teaching method or the tutoring was responsible for the test score gains.

I had hoped I was helping the school and the research group. That's not how they saw it. Within days of the meeting, the principal called me into her office to tell me that my behavior was "a threat to the democratic values of the school." I was asked to pledge my loyalty to the school's leadership and educational methods. When I refused, I was removed from the school's board of advisers.

My amygdala had been right all along. I felt burning shame. I was also confused: I believed I shared the school's values, so how was my speaking up a threat? I questioned my abilities to advocate for my child and worried this incident would impact him at school. These doubts seeped into my professional life as well. In faculty meetings, I began to monitor how colleagues responded to me and tried to keep my questions as concise as possible. During a meeting about how to navigate diversity concerns, I teared up unexpectedly, suddenly overwhelmed with fear about the consequences of speaking up. Was it better to remain silent?

"Speaking up can take an emotional toll—but staying silent would be so much worse."

It was unnerving to find myself questioning a hard-won capability that is central to my success as a scientist. In graduate school, I lacked the confidence to ask questions in most public settings, but I greatly admired a student who did so at the end of every seminar. As a postdoc, inspired by brilliant, fearless friends who asked the most devastating questions, I finally mastered my anxieties enough to begin to do so myself. I was commended for my critical eye, which gave me confidence as I started my own laboratory.

Yet what made me good at my job had now made me an unacceptably confrontational mother. My professional and personal roles seemed to be pushing me in opposite directions, leaving me feeling stuck in the middle.

With some time and support, I realized I needed to master my amygdala again. A school mother I didn't know well confided that she admired what I had said, which added to my confidence that I had done the right thing. When I agonized to a colleague that perhaps I had spoken too stridently, she fiercely defended my right to speak critically.

Slowly, a tide turned in my mind. I have worked hard at generating and asking probing questions, and it's a valuable skill. Yes, speaking up can take an emotional toll—but staying silent would be so much worse. ■

Samantha J. Butler is a professor at the University of California, Los Angeles. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

HARVARD UNIVERSITY

Tenure-Track Assistant Professor in Chemistry

Department of Chemistry and Chemical Biology
Faculty of Arts and Sciences
Harvard University

Candidates are invited to apply for a tenure-track **assistant professorship** in the open field of chemistry and chemical biology. The appointment is expected to begin on **July 1, 2022**. The tenure-track professor will be responsible for teaching at the undergraduate and graduate levels. We are seeking candidates who have an outstanding research record and a strong commitment to undergraduate and graduate teaching. Doctorate required by expected start date.

Please submit the following materials through the ARIeS portal **(https://academicpositions.harvard.edu/postings/10543)**: **1)** Cover letter; **2)** Curriculum Vitae; **3)** Teaching/advising statement (describing teaching philosophy and practices); **4)** Research statement; **5)** Statement describing efforts to encourage diversity, inclusion, and belonging, including past, current, and anticipated future contributions in these areas; and, **6)** Names and contact information of 3-5 referees, who will be asked by a system-generated email to upload a letter of recommendation once the candidate's application has been submitted. Three letters of recommendation are required, and the application is considered complete only when at least three letters have been received. At least one letter must come from someone who has not served as the candidate's undergraduate, graduate, or postdoctoral advisor. Applications must be submitted no later than **October 15, 2021**.

Harvard is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, disability status, protected veteran status, or any other characteristic protected by law.

HARVARD UNIVERSITY

Tenured Professor in Chemistry

Department of Chemistry and Chemical Biology
Faculty of Arts and Sciences
Harvard University

Candidates are invited to apply for a tenured professorship in the field of theory. The appointment is expected to begin on **July 1, 2022**. The tenured professor will be responsible for teaching at the undergraduate and graduate levels. We are seeking candidates who have an outstanding research record and a strong commitment to undergraduate and graduate teaching. Doctorate required by expected start date.

Please submit the following materials through the ARIeS Portal **(https://academicpositions.harvard.edu/postings/10545): 1)** Cover letter; **2)** Curriculum Vitae; **3)** Teaching/Advising statement (describing teaching philosophy and practices); **4)** Research statement; **5)** Statement describing efforts to encourage diversity, inclusion, and belonging, including past, current, and anticipated future contributions in these areas; and, **6)** Names and contact information of 3-5 referees, who will be asked by a system-generated email to upload a letter of recommendation once the candidate's application has been submitted. Three letters of recommendation are required, and the application is considered complete only when at least three letters have been received. At least one letter must come from someone who has not served as the candidate's undergraduate, graduate, or postdoctoral advisor. **Deadline:** Applications will be accepted until the position is filled.

Harvard is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, disability status, protected veteran status, or any other characteristic protected by law.

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

UW-Madison has two open positions for faculty members in the areas of **planetary astrochemistry** and **observational exoplanet spectroscopy**. These positions are the second and third in a campus-wide cluster hiring initiative on Origins of Life, Artificial Life, and Astrobiology (OoLALA; **https://astrobio.wisc.edu/**).

Successful candidates will contribute to the cluster through their research, teaching, and interaction with the vibrant OoLALA community and the Wisconsin Center for Origins Research (WiCOR), a new interdisciplinary research initiative in the UW-Madison Astronomy Department. We invite candidates for these positions to consult the position vacancy listings (**PVL 241623** and **PVL 241624**) at **http://jobs.wisc.edu.**

Applicants must have a Ph.D. in Chemistry, Geosciences, Atmospheric Sciences, Astronomy, Planetary Science, or a related field at the time of appointment and a record of peer-reviewed research publications in the field. We are especially interested in candidates who are committed to building an equitable and inclusive educational environment, and who can bring new perspectives on recruiting, mentoring, and educating students from diverse backgrounds.

COLUMBIA | VAGELOS COLLEGE OF PHYSICIANS AND SURGEONS

2022 Schaefer Research Scholars Program Awards

The Vagelos College of Physicians and Surgeons (VP&S) is pleased to announce the *2022 Schaefer Research Scholars Program Awards.* The *Awards,* made possible through a bequest from Dr. Ludwig Schaefer, are made annually to four research scientists who have distinguished themselves in human physiology, as broadly defined, and whose current work is of outstanding merit. The proposed research must have the potential to illuminate the field. Two awards are made to research scientists residing or working in North or South America, and two awards are made to research scientists residing or working outside North or South America. Each award consists of a $50,000 cash prize and up to $200,000 in direct research support.

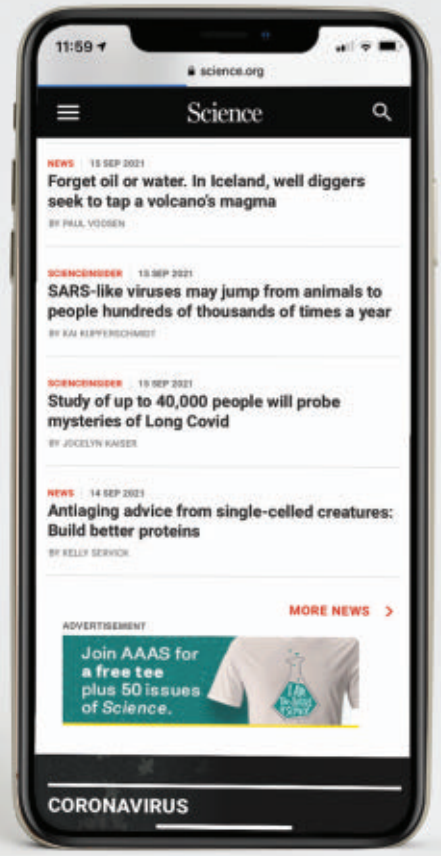
Applications must include a cover sheet; a research proposal (one page); a research budget (not to exceed $200,000 in total direct costs) delineated by cost category (salary, fringe, supplies, etc.) for one year (7/1/2022–6/30/2023); a curriculum vitae (not to exceed 10 pages); and a page summarizing applicant's research support. Internal candidates must obtain a nomination letter from their Department Chair. External candidates must present letters from the Dean or equivalent in their home institution as well as from the Columbia University Irving Medical Center collaborator, if applicable.

Nomination deadline: November 22, 2021, at 5:00 p.m. (EST)
To apply, visit https://www.ps.columbia.edu/schaefer
and submit all required materials.
Awardees will be notified in February 2022.

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