Why some innocent suspects confess to crimes *p. 1022*

Microbiota metabolism of drugs *pp. 1030 & 1055*

Hybrid nanofiltration membranes *pp. 1033 & 1057*

Science

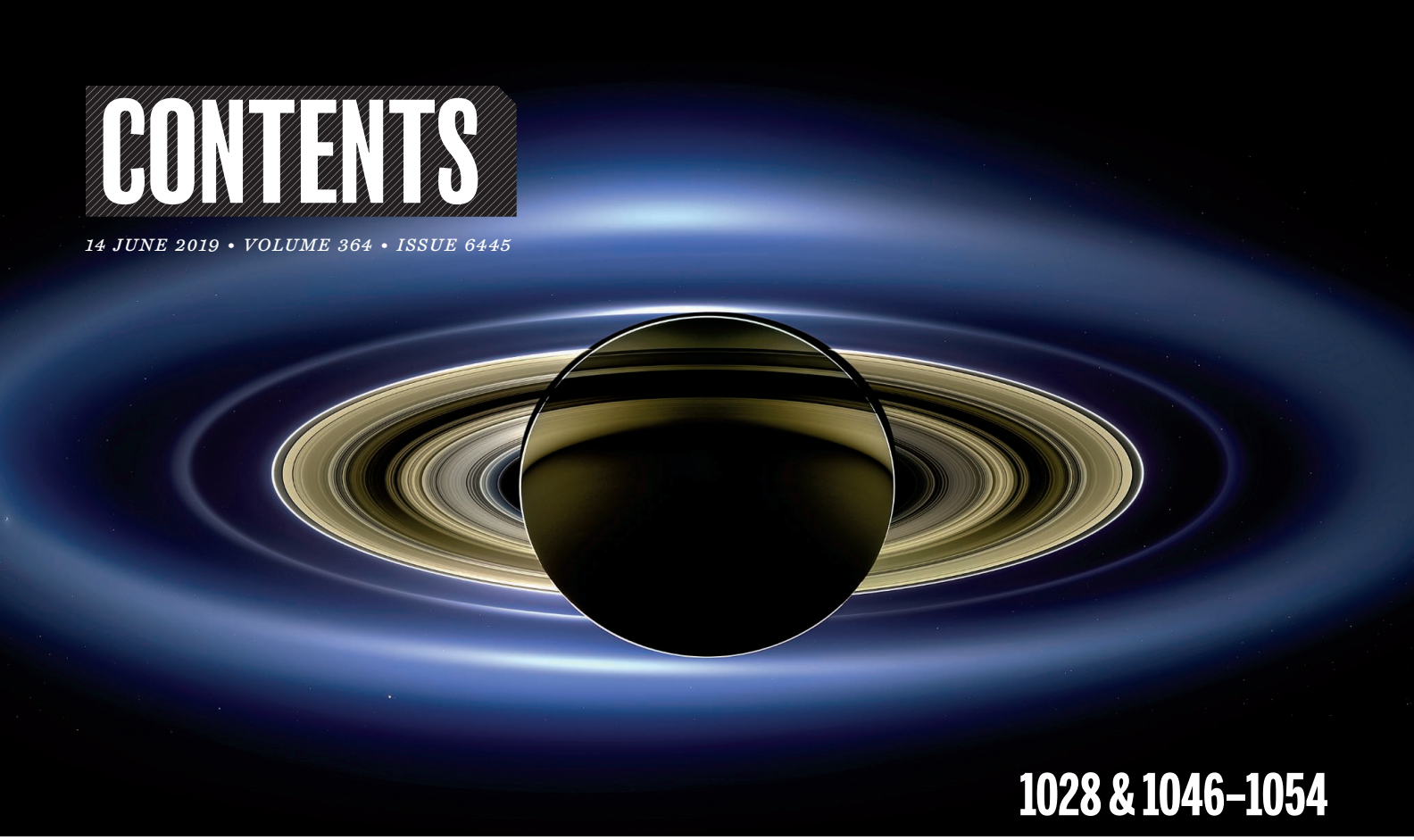
$15
14 JUNE 2019
sciencemag.org
AAAS

SATURN'S RINGS

Cassini gets in close *pp. 1028 & 1046–1054*

CONTENTS

14 JUNE 2019 • VOLUME 364 • ISSUE 6445



1028 & 1046–1054

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

PERSPECTIVES

POLICY FORUM

PHOTO: NASA/JPL-CALTECH/SSI

Published by AAAS

BOOKS *ET AL.*

LETTERS

RESEARCH

IN BRIEF

GAS GIANT PLANETS REVIEW

RESEARCH ARTICLES

RESEARCH ARTICLES

REPORTS

PHYSICS

DEPARTMENTS

ON THE COVER

The small moon Daphnis (radius 4 km) orbits within the Keeler gap in Saturn's A ring, seen in this image taken by the Cassini spacecraft. Tidal forces excite the nearby ring particles into a wavelike pattern. Gaps in the data at the top and bottom of the image have been artificially filled for illustrative purposes using measured properties of the rings. See pages 1028 and 1046–1054. *Image: NASA/JPL-Caltech/Space Science Institute/Tilmann Denk*

PHOTO: FEDERICO DICENTA

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. 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): $1971; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (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; bulk rate on request. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

Science AAAS

Editor-in-Chief Jeremy Berg, jberg@aaas.org

Executive Editor Monica M. Bradford

EDITOR, INSIGHTS Lisa D. Chong **EDITORS, RESEARCH** Valda Vinson, Jake S. Yeston **DEPUTY EDITORS** Julia Fahrenkamp-Uppenbrink(UK), Stella M. Hurtley(UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden(UK) **SR. EDITORS** Gemma Alderton(UK), Caroline Ash(UK), Brent Grocholski, Pamela J. Hines, Paula A. Kiberstis, Marc S. Lavine(Canada), Steve Mao, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon(UK), Keith T. Smith(UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Alana Warnke, Alice Whaley(UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Hilary Stewart(UK), Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements(UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock(UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand(Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery(UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Paul Voosen, Meredith Wadman **INTERN** Alex Fox **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile(Shanghai), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana(Barcelona), Emily Underwood, Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein(Editor), Katie Langin **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **PHOTO EDITOR** Emily Petersen **GRAPHICS EDITOR** Nirja Desai **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary

CEO - Executive Publisher Rush D. Holt

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman , Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy **CONTENT SPECIALIST** Jacob Hedrick **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **DIGITAL AD-OPS SPECIALIST** Patrick Gerrits **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **SENIOR DESIGNER** Kim Huynh **TRADE SHOW AND MEETINGS ASSOCIATE** Andrew Clamp

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **ASSOCIATE DIRECTOR, PRODUCT DEVELOPMENT AND SPJ** Hannah Heckner **SR. PRODUCT ASSOCIATE** Robert Koepke **DIGITAL PRODUCT STRATEGIST** Michael Hardesty **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGERS** Marco Castellan, Christos Skoutas **MANAGER, SYSTEMS AND OPERATIONS** Brian Holiahn **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

DIRECTOR, GLOBAL SALES Tracy Holmes **ASSOCIATE DIRECTOR, ADVERTISING SALES, US** Laurie Faraday **US EAST COAST AND MID WEST SALES** Glen Cox **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **PROJECT MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Kaoru Sasaki (Tokyo), Miyuki Tani(Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

---

MAIN HEADQUARTERS
Science/AAAS
1200 New York Avenue NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No.59 Zhongguancun St.
Haidian District, Beijing. 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
sciencemag.org/authors/science-information-authors

REPRINTS AND PERMISSIONS
sciencemag.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services

science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers

advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org

support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

EDITORIAL

Innovating a Green Real Deal

As the United States heads toward an election year in 2020, climate change promises to stay in the headlines. Increasing public concern about and desire to address the consequences of global warming drive this attention, along with the debate about a Green New Deal that promotes social justice alongside accelerated deep reductions in greenhouse gas emissions. Now is the time to translate these aspirations into action within the constraints of technical, cost, and social realities. This is optimization, not compromise—call it a Green Real Deal.

A Green Real Deal framework should be structured around a set of key characteristics. It must be science-based and analytically sound. It must be pragmatic in providing maximum optionality and flexibility, enabling a broad coalition to form. It must address all sectors of the economy, particularly those difficult to decarbonize such as transportation, industry, and agriculture. It must have a regional focus, because low-carbon solutions will necessarily be location dependent. And it must advance social equity and workforce development, avoiding stranded assets and stranded workers wherever possible and offering remedies for those most affected by the energy transition. By any name, these characteristics provide criteria for measuring Green Real Deal proposals.

What is needed to support this framework? A core component is innovation in technology, business models, and policy, and technology innovation is in many ways the key enabler among them. Without the development and deployment, at gigatonne scale, of breakthrough technologies that are affordable, the United States and the world will not reach net carbon neutrality. Breakthrough technology candidates could enable carbon direct removal, commodity-scale CO_2 utilization, biological and geological CO_2 storage at gigatonne scale, economic electricity storage at time scales from days to seasons, advanced nuclear fission and fusion energy, a hydrogen economy, full integration of information technology with the energy system, and advanced low-carbon fuels—and more.

Although we cannot expect these breakthrough technologies to have been developed and deployed at scale in the near term, innovation is nevertheless also crucial for decadal goals that put us on the pathway to mid-century deep decarbonization. The impacts of enormous cost reductions in solar energy, wind power, batteries, and light-emitting diodes are evident, but electricity emissions reductions are only 28% of the U.S. total. Major cost reductions, starting with an emphasis on energy efficiency, must be extended across all sectors of the economy. This depends on innovation that comes not only from research and development but also from the manufacturing and scaling experience gained from increasing deployment.

"The innovation agenda is also central to the social equity objective."

The innovation agenda is also central to the social equity objective. Mitigating climate change through technology and policy is critical for underserved communities, those hit first and worst by extreme weather. Also, most clean energy technologies remain expensive relative to incumbent technologies, and lowering energy prices through innovation is progressive in its impact across the income distribution. Further, platform technologies that support innovation, such as broadband, must be universally available.

Neither the deep decarbonization nor the social equity agendas are being pursued adequately today, but steps are readily available to start to remedy this. The innovation agenda is a good place to start. Today, there is much bipartisan discussion in Congress about substantially increasing the scale and scope of federal clean energy research, development, demonstration, and deployment (RDD&D) programs. The 2015 Mission Innovation commitment by 20 countries, led by the United States, to double clean energy investments provides a good benchmark. Such a commitment to boost federal clean energy RDD&D could and should encompass substantial support for regional innovation systems. Advancing an innovation initiative through Congress now will be an important step toward a Green Real Deal and perhaps awaken the kind of climate change accountability needed more broadly in 2020.

–Ernest J. Moniz

PHOTO: ENERGY FUTURES INITIATIVE

Ernest J. Moniz
was the 13th U.S. Secretary of Energy and is the CEO of Energy Futures Initiative, Washington, DC, USA. ejmoniz@energyfutures initiative.org

10.1126/science.aay3140

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

CONSERVATION

Deforestation spikes in Brazil as oversight ebbs

Deforestation in Brazil increased sharply in May, with 740 square kilometers of land newly cleared, according to the Real-time Deforestation Detection System (DETER), a satellite monitoring program run by the National Institute for Space Research in São José dos Campos. That's 35% more than in May 2018 and twice as much as in May 2017. Stricter regulations and monitoring led annual deforestation rates in the Brazilian Amazon to drop by 83% between 2004, when DETER operations began, and 2012, but they have been trending upward for the past 6 years because of political and economic pressures. DETER is a low-resolution alert system meant to guide law enforcement, but deforestation estimates based on its data track closely with official government numbers produced by another satellite monitoring program. The highest rates of forest clearing are typically recorded in June and July, when dry weather makes logging and burning easy. Many scientists fear that Brazilian President Jair Bolsonaro's loosening of environmental policies and enforcement will embolden loggers, farmers, and land developers to encroach further on the forest.

Fired researcher apologizes

#METOO | One week after Stanford University School of Medicine in Palo Alto, California, fired Jose Montoya, an expert on the parasitic disease toxoplasmosis who has also played a key role in elevating research on chronic fatigue syndrome (CFS), he issued a broad—but vague—apology for the conduct violations that led to the school's 30 May action. On 4 June, *The Stanford Daily* quoted an anonymous statement, which it said came from women affected by Montoya's actions, claiming there had been "extensive allegations of sexual misconduct, assault, and harassment" against him. In a statement subsequently released by a lawyer, Montoya, who is from Colombia, wrote, "The social norms in the U.S. are evolving and quite different than those from my culture and homeland. I did not sufficiently appreciate that difference." He added: "I have not been involved in any sexual or romantic relationships with employees, trainees, colleagues, or CFS team members."

Plant extinctions rank high

BIODIVERSITY | The first global analysis of its kind finds that twice as many plant species have disappeared as birds, mammals, and amphibians combined. Researchers reviewed published research, international databases, and museum specimens, finding 571 plant species that vanished in the past 250 years. For example, the Chile sandalwood tree in the South Pacific, which was exploited for its fragrant timber, was last seen in 1908. The number of extinct plant species is four times higher than the official listing kept by the International Union for the Conservation of Nature in Gland, Switzerland, the team reports this week in *Nature Ecology & Evolution*.

CRISPR alliance launched

DRUG DEVELOPMENT | Top CRISPR researchers at two University of California (UC) campuses have teamed up with pharmaceutical giant GlaxoSmithKline (GSK) to form a new laboratory in San Francisco that will jointly exploit the genome editor to find new drugs. UC Berkeley's Jennifer Doudna, a co-inventor of the powerful CRISPR tool, and UC San Francisco's Jonathan Weisman will select the academic talent to work in

PHOTO: FELIPE WERNECK/IBAMA/CC BY-SA

the new Laboratory for Genomics Research, which will identify new drug targets by using CRISPR to probe genome function. In all, GSK will fund 24 full-time UC researchers and contribute as many as 14 of its own employees. GSK has committed up to $67 million over 5 years for the unusual drug-screening effort. UC will own the intellectual property behind any new tools invented in the lab, while GSK will own the patents on the new drug targets.

Exoplanet-hunting tool debuts

ASTRONOMY | Researchers this week dedicated an instrument at the European Southern Observatory's Very Large Telescope in Chile to search for habitable exoplanets in the Alpha Centauri triple-star system, our nearest stellar neighbors. The system, 4 light-years away, contains a red dwarf star, Proxima Centauri, which has an Earth-like planet that scientists discovered in 2016; they want to know whether more are orbiting the system's two sunlike stars, Alpha Centauri A and B. The new instrument, New Earths in the AlphaCen Region, is designed to block most light from a star in order to capture the infrared light emitted by the warm surface of an orbiting planet. Astronomers can use those readings to determine the planet's temperature—and whether it could hold liquid water that could support life.

Europe boosts supercomputing

COMPUTING | The European Union will fund eight sites to host powerful new supercomputers as part of the €840 million European High-Performance Computing Joint Undertaking to support scientists and industry. They are: Sofia, Bulgaria; Ostrava, Czech Republic; Kajaani, Finland; Bologna, Italy; Bissen, Luxembourg; the Minho region in Portugal; Maribor, Slovenia; and Barcelona, Spain. The new machines, due by the end of 2020, will not be as fast as the "exascale" machines, capable of 1 billion billion calculations per second, being built in China and the United States. But some of them will be up to five times more powerful than the European Union's current top systems of the Partnership for Advanced Computing in Europe in Ixelles, Belgium. Three sites will get precursor-to-exascale machines, capable of executing more than 150 petaflops, or 150 million billion calculations per second; five others will get petascale machines, capable of executing at least 4 petaflops. The centers will support research in drug development, materials, weather forecasting, and other areas.

Ebola crosses into Uganda

INFECTIOUS DISEASE | The Ebola virus that has stubbornly persisted in the Democratic Republic of the Congo (DRC) since August 2018 has finally jumped the border, sickening a 5-year-old boy in Uganda. The Uganda Virus Research Institute in Entebbe confirmed the infection this week, the World Health Organization (WHO) announced. WHO has long feared that the DRC outbreak, which has sickened more than 2000 people there and killed about two-thirds of identified patients, would spread to neighboring countries. The boy, who came to Uganda from the DRC with his family, is now in an Ebola treatment facility. His case will likely trigger anew a discussion of whether WHO should declare a Public Health Emergency of International Concern (PHEIC) for the outbreak, the second largest since Ebola was recognized in 1976. A PHEIC would help WHO marshal more resources.

BY THE NUMBERS

74,000

Estimated plastic microparticles consumed by each American annually, on average, from fish, salts, water, and other sources. Health effects, if any, remain unclear (*Environmental Science & Technology*).

30%

U.S. Army personnel with ideal blood pressure, in 2012. Among U.S. civilian adults of similar ages, the figure was 55% (*Journal of the American Heart Association*).

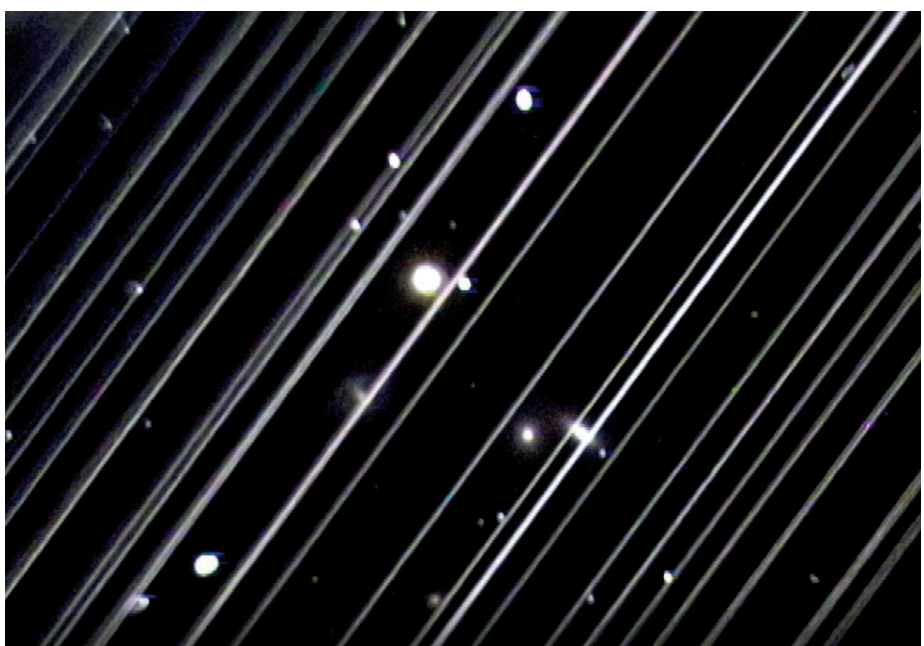
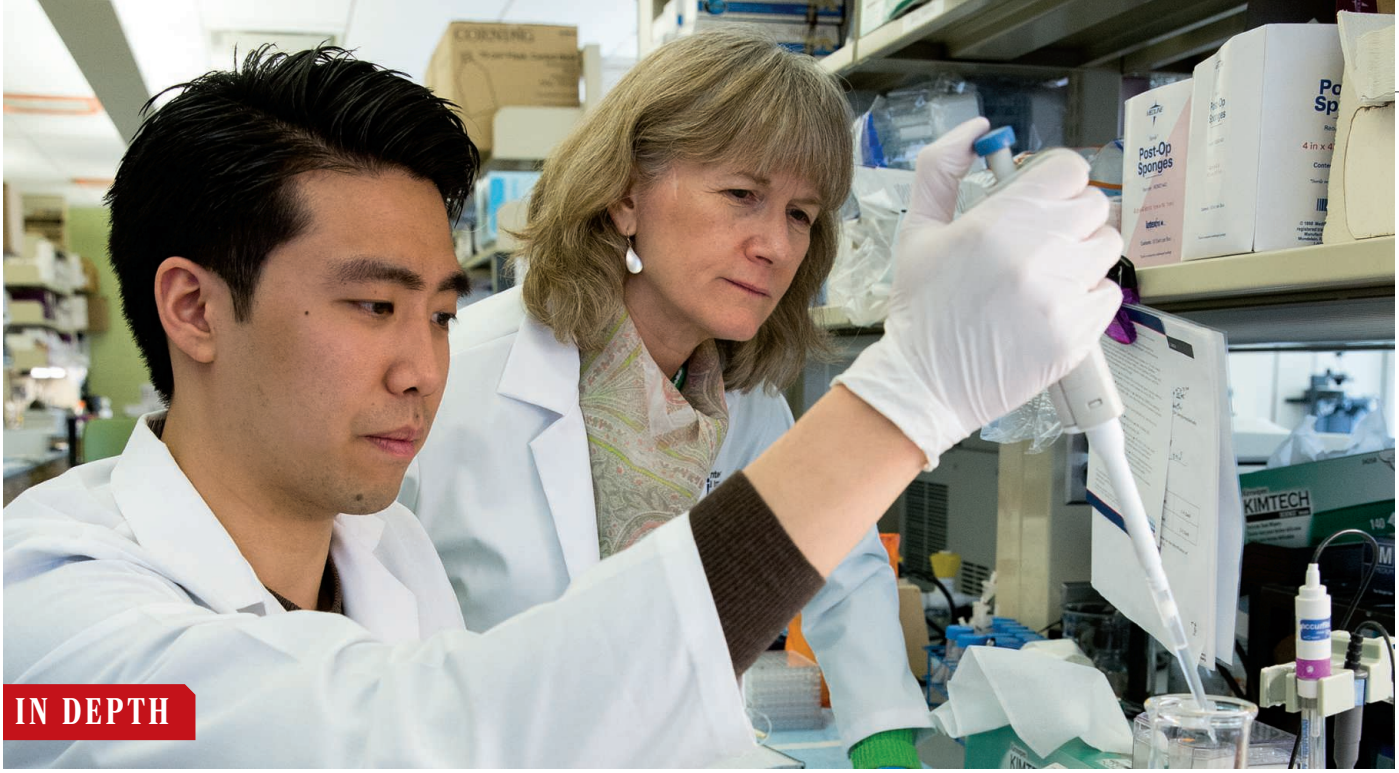


IMAGE: VICTORIA GIRGIS/LOWELL OBSERVATORY

ASTRONOMY

Scientists fear bright satellites

Two astronomical societies are worried that light reflected by large networks of satellites will spoil celestial observations. The concerns follow SpaceX's launch of 60 Starlink communication satellites, designed to bring internet service worldwide. Skywatchers soon saw their light trails, including, for example, in an image (above) made with a telescope at Lowell Observatory in Flagstaff, Arizona. The American Astronomical Society (AAS) in Washington, D.C., adopted a resolution of concern on 8 June, and the International Astronomical Union in Paris issued a similar statement on 3 June. As the satellites spread out and shift to higher orbits, the density and brightness of their tracks will decline. SpaceX, owned by entrepreneur Elon Musk, aims to send up 12,000 satellites in all, but Musk says it will try to limit their brightness. AAS said it hopes to work with SpaceX and other companies planning to launch mega-constellations to help them avoid negative effects on astronomy.



IN DEPTH

Renewals of two diabetes grants held by Megan Sykes (right) of Columbia University could be held up by a new federal policy scrutinizing fetal tissue research.

U.S. RESEARCH POLICY

Trump clamps down on fetal tissue research

Controversial policy could entangle funding proposals in lengthy new ethical reviews

By **Meredith Wadman** *and* **Jocelyn Kaiser**

Megan Sykes, an immunologist at Columbia University, has spent years using human fetal tissue to develop a mouse with a human-like immune system, which mimics how type 1 diabetes develops in humans. The tissue is donated after elective abortions, and the mice are testbeds for potential diabetes treatments. But last week, she learned that President Donald Trump, acting on a priority of advocacy groups opposed to abortion, had issued a new policy that is likely to cause lengthy delays the next time she seeks U.S. government grants for her work—and could even choke off federal funding for all studies that use fetal tissue. The policy "is incredibly disappointing," Sykes says, because it is a "politically motivated decree" that could derail numerous disease research efforts.

The new Trump policy, issued 5 June after a 9-month review led by officials at the Department of Health and Human Services (HHS), has three major components. One kills a long-standing contract between the National Institutes of Health (NIH) in Bethesda, Maryland, and the University of California (UC), San Francisco, under which the university used fetal tissue to develop humanized mice for HIV drug testing. Another ends research using fetal tissue conducted by any scientist directly employed by NIH. The third and widest-reaching provision adds a lengthy and uncertain step to NIH's process for awarding new or renewal grants to university scientists, such as Sykes, for studies that use human fetal tissue. It requires HHS to appoint a separate 14- to 20-member ethics advisory board to review each proposal that NIH reviewers have found worthy of funding. The review of up to 6 months will result in a funding recommendation to the HHS secretary, who can accept or reject the advice.

Enacting the new policy "was the president's decision .... to protect the dignity of human life," Judd Deere, deputy White House press secretary, told *Science*. It was applauded by antiabortion activists, whose lobbying prompted HHS to launch its review of U.S.-funded fetal tissue research in September 2018. "This is a major pro-life victory," said Marjorie Dannenfelser, president of the Susan B. Anthony List in Washington, D.C.

Many biomedical researchers were stunned, noting that the tissue, which would otherwise be discarded, has properties that make it valuable for research. It is less specialized than adult tissue, for instance, and readily adapts to new environments. "These new restrictions have no scientific or ethical basis and will roll back decades of consensus in the U.S., delaying the development of new treatments," said Doug Melton, president of the International Society for Stem Cell Research in Skokie, Illinois, and co-director of the Harvard Stem Cell Institute.

"The whole point here is to so wrap the research in red tape that it's impossible or at least unlikely to be feasible for many researchers," says bioethicist Alta Charo of the University of Wisconsin in Madison.

A 1993 law formalized rules for using fetal tissue donated after elective abortions in U.S.-funded research. Last year, NIH spent $115 million on roughly 173 projects that rely on the tissue; about 160 were run by university scientists. One-third of the 173 grants focus on HIV/AIDS, many using humanized mice to probe, for example, how HIV hides out and evades the immune system, and what drugs might defeat it. Others tackle other infectious diseases, eye disease, and fetal development as well as toxic exposures during pregnancy.

NIH says its scientists are conducting just three projects affected by the new rules; all

PHOTO: MICHAEL DIVITO

will stop. "This decision is devastating. It effectively ends our studies looking into new approaches for an HIV cure," says Warner Greene, director of the Gladstone Institutes Center for HIV Cure Research in San Francisco. Greene is a partner in one of the projects, run by retrovirologist Kim Hasenkrug of NIH's Rocky Mountain Laboratories in Hamilton, Montana.

At universities, the policy allows existing projects to continue until their current NIH funding expires. Nearly half of these extramural grants will expire within the next 18 months, and scientists will need to apply for a renewal if they want to keep the work going. Grantees are now grappling with what the new review process might mean.

It has already caused at least one researcher to change course. HIV scientist Jerome Zack last week told colleagues at UC Los Angeles that he had decided to remove his work using fetal tissue to develop humanized mice from a renewal application, due at NIH in August, for a large grant supporting the university's long-standing Center for AIDS Research. "The grant covers way more than mouse work, it covers all HIV research on campus," he says. "I don't want to jeopardize that."

At Columbia, Sykes is worried about the one-third of her 15 staff who are funded through two NIH grants. She recently submitted a renewal proposal for one grant and planned to submit the other in July. HHS hasn't said when the policy will kick in. But when Sykes asked NIH officials how it might affect her proposals, the response "wasn't reassuring," she says.

Much could depend on whom the HHS secretary appoints to the ethics review boards. Under existing law governing HHS ethics boards, one-third to one-half of a board's members must be scientists, and each must include at least one theologian, one ethicist, one physician, and one attorney. The law "absolutely" would allow HHS to pack the boards with members who oppose abortion, Charo says.

Critics of the new policy also say it will undermine a goal of opponents of fetal tissue research: to find and encourage the use of alternatives. In December 2018, NIH Director Francis Collins noted that his agency was putting up to $20 million over 2 years into research on such alternatives. But scientists say that those alternatives need to be tested for validity against human fetal tissue itself. For the time being, Collins said in December 2018, fetal tissue would "continue to be the mainstay" for certain kinds of research.

The new rules could remove that mainstay. But Charo notes a new president could reverse the policy, which is not codified in law. ■

PATENT LAW

Draft bill reignites U.S. debate over patenting human genes

Lawmakers propose reversing Supreme Court decisions

By **Kelly Servick**

A congressional proposal to expand the range of inventions eligible for a U.S. patent is drawing protest from scientific societies and patient advocates. The draft bill, released last month by a bipartisan group of lawmakers, would reverse Supreme Court decisions that restricted patents on natural processes and products, including human genes.

In two hearings held last week by a subcommittee of the Senate Committee on the Judiciary, some representatives of the biotechnology industry and research universities applauded the proposal, saying it would boost innovation. But others were alarmed. The American Civil Liberties Union (ACLU) in Washington, D.C., and more than 100 other signatories released a letter warning that the proposed changes would stifle medical research and hinder patients' access to diagnostic tests. If it became law, the proposal "would result in a quagmire of patent claims and legal impediments to the normal scientific exchange," said Harold Varmus, a cancer biologist at Weill Cornell Medicine in New York City and former director of the National Institutes of Health.

> ***The bill "would result in a quagmire of patent claims ..."***
>
> **Harold Varmus**, Weill Cornell Medicine

The bill's Senate co-sponsors, Thom Tillis (R–NC) and Chris Coons (D–DE), say it will reduce frustration and confusion created by recent Supreme Court decisions. One, the landmark 2013 ruling in *Association for Molecular Pathology* (AMP) v. *Myriad Genetics,* established that human genes cannot be patented. Another, a 2012 decision in *Mayo Collaborative Services* v. *Prometheus Laboratories*, invalidated a patent on adjusting drug dosage based on levels of metabolites in a patient's blood because it relied on a "law of nature."

Critics say these rulings have caused investors to move their money to nations with stronger patent protections. The *Myriad* decision led to "a dead stop in research in the United States on isolated natural products," said attorney Sherry Knowles of Knowles Intellectual Property Strategies in Atlanta at the 4 June hearing. The bill would simply wipe out the rulings by specifying that the U.S. Patent and Trademark Office cannot consider "judicially created exceptions ... including 'abstract ideas,' 'laws of nature,' or 'natural phenomena'" in deciding whether an invention is patent eligible.

"It's a radical proposal," says Roger Klein, a physician and legal consultant in Cleveland, Ohio, and a member of the board of directors at AMP, a professional society that signed the ACLU letter. In the *Myriad* case, AMP challenged broad patents on two human genes held by the diagnostic company Myriad Genetics of Salt Lake City, which claimed exclusive rights to tests for cancer-associated mutations in those genes.

The proposal probably wouldn't revive the fight over the ownership of single human genes—and that "was never the intent" of the effort, Tillis told *Science*. Arti Rai, a patent law expert at Duke University in Durham, North Carolina, notes that because researchers have already published extensive descriptions of the human genome, many claims to specific genes would no longer be considered "novel," as U.S. patent law requires.

But the bill could help developers of newer genetic technologies win and protect patents.

By overturning the *Mayo* decision, the bill would allow patent claims on the function of specific genetic variants, Klein says. For example, a company could patent a test that uses a specific mutation to predict whether a person will respond to a drug. He fears allowing such patents would make it "extraordinarily difficult ... to put [that information] to use in patient care."

The bill would also allow patents on polygenic risk scores, the complex analyses of many genes to estimate disease risk. And it could make newly discovered genes from nonhuman species patent eligible. Coons and Tillis believe language in the bill stipulating what makes an invention sufficiently "useful" to be patentable would keep naturally occurring genes and pure laws of nature off limits. But, Tillis says, "If we get sufficient input from legal scholars and others that we may not be clear on that, then we'll draft language that makes it very clear." ■

ARCHAEOLOGY

Mountain high: oldest clear signs of pot use

THC levels in braziers show psychoactive marijuana use along ancient Silk Road

By **Andrew Lawler**

Today, more than 150 million people regularly smoke cannabis, making it one of the world's most popular recreational drugs. But when and where humans began to appreciate the psychoactive properties of weed has been more a matter of speculation than science. Now, a team led by archaeologists Yang Yimin and Ren Meng of the Chinese Academy of Sciences in Beijing reports clear physical evidence that mourners burned cannabis for its intoxicating fumes on a remote mountain plateau in Central Asia some 2500 years ago.

The study, published this week in *Science Advances*, relies on new techniques that enable researchers to identify the chemical signature of the plant and even evaluate its potency. "We are in the midst of a really exciting period," says team member Nicole Boivin of the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena, Germany. The paper is part of a wider effort to track how the drug spread along the nascent Silk Road, on its way to becoming the global intoxicant it is today.

Cannabis, also known as hemp or marijuana, evolved about 28 million years ago on the eastern Tibetan Plateau, according to a pollen study published in May. A close relative of the common hop found in beer, the plant still grows wild across Central Asia. More than 4000 years ago, Chinese farmers began to grow it for oil and for fiber to make rope, clothing, and paper.

Pinpointing when people began to take advantage of hemp's psychoactive properties has proved tricky. Archaeologists had made claims of ritual cannabis burning in Central Asian sites as far back as 5000 years ago. But new analyses of those plant remains by other teams suggest that early cannabis strains had low levels of tetrahydrocannabinol (THC), the plant's most powerful psychoactive component, and so lacked mind-altering properties. One academic who works in Central Asia said he and colleagues tried to smoke and eat wild varieties—but got no buzz.

Archaeologists have spotted signs of ancient cannabis use from western China to the Caucasus.

The cannabis burned 2500 years ago at the Jirzankal cemetery, 3000 meters high in the Pamir Mountains in far western China, was different. Excavations there have uncovered skeletons and wooden plates, bowls, and Chinese harps, as well as wooden braziers that held burning material. All are typical of the Sogdians, a people of western China and Tajikistan who generally followed the Persian faith of Zoroastrianism, which later celebrated the mind-expanding properties of cannabis in sacred texts. At Jirzankal, glass beads typical of Western Asia and silk from China confirm the long-distance trade for which the Sogdians became famous, and isotopic analysis of 34 skeletons showed that nearly a third were migrants. Radiocarbon analysis put the burials at about 500 B.C.E.

The wooden braziers were concentrated in the more elite tombs. Yang's and Ren's team ground bits of brazier into powder and applied gas chromatography and mass spectrometry to identify chemical compounds left behind. They found unusually high levels of THC compared with typical wild cannabis, although much less than in today's highly bred plants. The cannabis was apparently burned in an enclosed space, so mourners almost certainly inhaled THC-laced fumes, the authors say, making this the earliest solid evidence of cannabis use for psychoactive purposes.

The region's high altitude could have stressed the cannabis, creating plants naturally high in THC, says co-author Robert Spengler, also of MPI-SHH. "It is quite likely that people came across cannabis plants at higher elevations that were naturally producing higher THC levels," he says. But humans may also have intervened to breed a more wicked weed, he adds.

"The methods are convincing, and the data are unambiguous regarding early use of cannabis as a psychoactive substance," says Tengwen Long, an environmental scientist at the University of Nottingham in the United Kingdom who has researched cannabis origins. But Megan Cifarelli, an art historian at Manhattanville College in Purchase, New York, who has studied ancient drug use, notes the aromatic fumes might also have had another purpose: to mask the smell of a putrefying corpse.

Yang's and Ren's team thinks cannabis use was restricted to elites until potent pot began to spread across Central Asia through the Silk Road linking China with Iran. In 440 B.C.E., the Greek historian Herodotus wrote that the nomadic Scythians, who controlled vast areas from Siberia to Eastern Europe, made tents and heated rocks in order to inhale hemp vapors that made them "shout for joy." And Andrei Belinski, an archaeologist based at the heritage museum in Stavropol, Russia, in 2013 began to excavate a nearby 2400-year-old Scythian tomb that held gold vessels bearing residues of both opium and cannabis, supporting the idea that elites used the drug first.

Ancient artwork and textual references from Syria to China (*Science*, 20 April 2018, p. 249) hint at even earlier cannabis drug use, and the new analytical methods could soon provide concrete evidence of this, says Michael Frachetti, an archaeologist at Washington University in St. Louis, Missouri. But it's already clear that the ancient Silk Road trafficked in more than spices, grains, and ideas. "Crops weren't just about food," he says. "They were also about making contact with another world." ■

Ancient people put cannabis leaves and hot stones in this brazier, and likely inhaled the resulting smoke.

CREDITS: (MAP) N. DESAI/*SCIENCE*; (PHOTO) XINHUA WU

Better wind data on Hurricane Michael might have shown it would hit the Florida Panhandle.

PHOTO: JABIN BOTSFORD/*THE WASHINGTON POST* VIA GETTY IMAGES

METEOROLOGY

Satellites see hurricane winds despite military signal tweaks

Data from CYGNSS satellites could sharpen storm forecasts

By **Paul Voosen**

A mission to probe winds deep inside hurricanes, where most satellites cannot see and few aircraft venture, is showing signs of success despite an unexpected obstacle linked to tensions in the Middle East.

A constellation of eight microsatellites has harvested data that—if folded into the National Oceanic and Atmospheric Administration's (NOAA's) weather models—could have sharpened forecasts of several recent hurricanes, including Michael, a category-5 storm that hit Florida last year. "We're finally getting stuff that really looks useful," says Frank Marks, who leads hurricane researchers exploring the data at NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML) in Miami, Florida. But the progress was hard-won for scientists on NASA's $157 million Cyclone Global Navigation Satellite System (CYGNSS), who discussed early results at a meeting last week, just as another Atlantic hurricane season kicked off.

With its flotilla of satellites crisscrossing the tropical oceans, CYGNSS can see through the thick clouds of cyclones. The satellites collect radio signals beamed from standard GPS beacons after they bounce off the ocean's surface. The reflections are influenced by sea's roughness, which depends on wind speed. But a month after launch in December 2016, the team noticed the GPS signals were wavering. "We assumed they are constant," says Christopher Ruf, CYGNSS's principal investigator and an atmospheric scientist at the University of Michigan in Ann Arbor. "And they're not."

The U.S. military runs the GPS system, and in January 2017, it began to boost the radio power on 10 of its GPS satellites as they passed over a broad region centered on northern Syria. The power boosts, which can thwart jamming, have recurred without warning, each lasting several hours. "It's an opaque situation, obviously, because it's a classified military situation," Ruf says. The swings don't interfere with other scientific uses of GPS. But they threw off the constellation's measurements of high winds by 5 meters a second or more—the difference between a category-2 and category-3 hurricane.

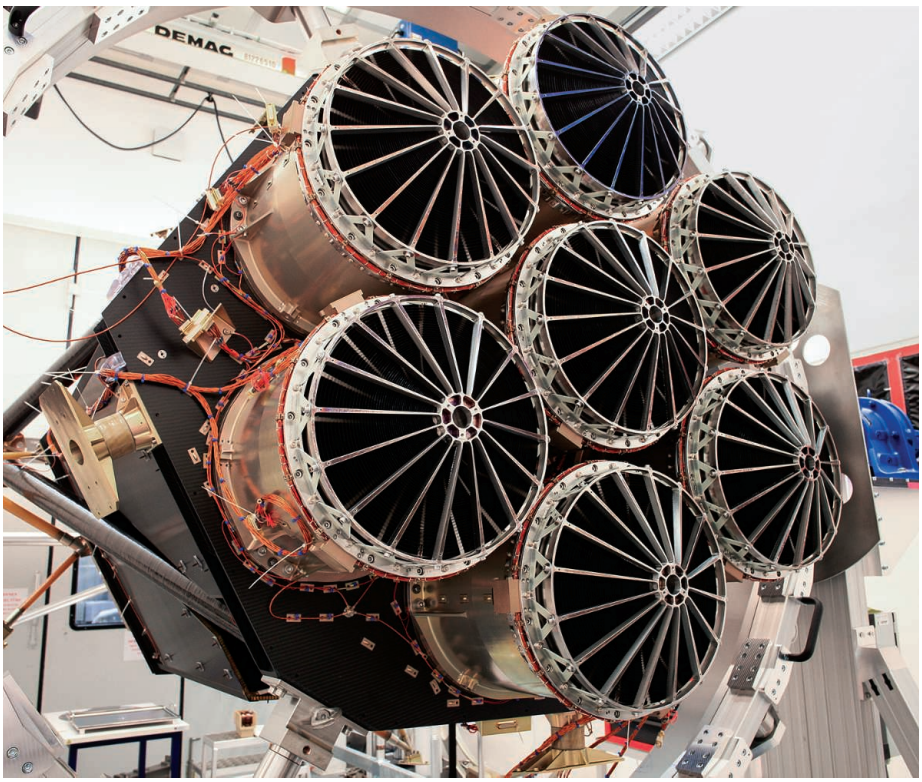
After 2 years of work, the CYGNSS team has compensated by reprogramming its satellites on the fly. The satellites carry large antennas to catch reflected GPS signals, but they also have small antennas that receive direct GPS signals, for tracking time and location. The team repurposed the small antennas to measure the signal strength of the GPS satellites, making it possible to correct the wind speed measures. "It works," Ruf says. "We've been testing it for a number of months."

Even before that fix, the wind data were good enough to improve some hurricane forecasts, says Bachir Annane, an atmospheric scientist at AOML. In the case of Michael, NOAA's forecast models failed, Annane says: They predicted it would track too far west, close to Alabama rather than Florida, and underestimated its ferocious winds. When he reran the models with CYGNSS winds, Annane found that the storm's forecasted early track and its intensity stayed closer to reality. The wind data would have improved track forecasts for two other recent hurricanes, Harvey and Irma, as well, he says.

The satellites are also giving scientists a view of the winds underlying the Madden-Julian oscillation, a large cluster of storms that periodically forms in the Indian Ocean and marches around the equator, influencing global weather (*Science*, 2 October 2015, p. 22). "Seeing under the rain was a big draw for us," says Eric Maloney, an atmospheric scientist at Colorado State University (CSU) in Fort Collins, because scientists have long debated what fuels the storms. Last week at the CYGNSS meeting, Bohar Singh, an atmospheric scientist who works with Maloney at CSU, described evidence from CYGNSS that persistent winds boost ocean evaporation under a 3000-kilometer-wide set of rainstorms, sustaining them. That finding could help scientists forecast how the storm belt will change in a warmer climate, Maloney says.

After a few tweaks, CYGNSS can now look at land, too. Its antennas are capturing signs of soil moisture, says Clara Chew, a remote sensing hydrologist at the University Corporation for Atmospheric Research in Boulder, Colorado. Although not as capable as a single dedicated satellite, CYGNSS's multiple satellites make more frequent measurements, which could help it monitor flood risks and track how different soils retain rain. "You can start to quantify how long the soil remembers," Chew says.

NOAA scientists hope the new GPS fix will unleash the microsatellites' full potential for looking into storms, perhaps revealing new insights into why some hurricanes suddenly intensify. NOAA isn't likely to start using the CYGNSS data in its routine forecasts, Marks says. The satellites don't belong to the weather agency, and they are unlikely to last more than 7 years before failing. But he thinks their success against the odds could help persuade NOAA to launch its own wind-monitoring constellation. ■



A German survey instrument on Spektr-RG contains seven x-ray telescopes, each with 54 nested mirrors.

PHOTO: © P. FRIEDRICH/MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS

ASTRONOMY

X-ray telescope keeps Russia's space science hopes alive

Spektr-RG, designed to study the effect of mysterious dark energy on galaxies, will be nation's only space observatory

By **Daniel Clery**

Russia's beleaguered space science program is hoping for a rare triumph this month. Spektr-RG, an x-ray satellite to be launched on 21 June from Kazakhstan, aims to map all of the estimated 100,000 galaxy clusters that can be seen across the universe. Containing as many as 1000 galaxies and the mass of 1 million billion suns, the clusters are the largest structures bound by gravity in the universe. Surveying them should shed light on the evolution of the universe and the nature of the dark energy that is accelerating its expansion.

First proposed more than 30 years ago as part of a Soviet plan for a series of ambitious "great observatories" along the lines of NASA's Hubble Space Telescope, Spektr-RG fell victim to cost cutting in cash-strapped, post-Soviet Russia. But the roughly €500 million satellite, which will carry German and Russian x-ray telescopes, was reborn early last decade with a new mission. Its original goal was to scan the sky for interesting x-ray sources, such as supermassive black holes gorging on infalling material; now, by mapping galaxy clusters, it would find out what makes the universe tick. The new goal meant further delays. "There have been many ups and downs," says Peter Predehl, leader of the team at the Max Planck Institute for Extraterrestrial Physics (MPE) in Garching, Germany, that built one of the satellite's two telescopes. "Whenever we thought we were out of the woods, a new one came along."

Spektr-RG was born in the late 1980s. Glasnost was encouraging Soviet researchers to collaborate with Western colleagues, and studies of SN 1987A, the nearest supernova in modern times, had demonstrated the power of x-rays for tracing such violent events. Rashid Sunyaev of Moscow's Space Research Institute (IKI) proposed an x-ray observatory to orbit above Earth's atmosphere, which blocks x-rays. The 6-ton mission soon bristled with five telescopes and involved 20 institutes in 12 countries, including the United States. But after the collapse of the Soviet Union, Roscosmos, Russia's space agency, struggled to keep its Mir space station aloft and contribute to the growing International Space Station (ISS). "They told us the spacecraft was too large for Russia, too ambitious," says Sunyaev, now at the Max Planck Institute for Astrophysics in Garching. "It just died."

Resurrection began in 2003 with plans for a smaller mission that would carry a U.K.-built all-sky x-ray monitor and MPE's x-ray survey telescope, called ROSITA—which had been destined for the ISS but was grounded by the Challenger space shuttle disaster. The new impetus was cosmology. Studies of distant supernovae in the 1990s had revealed that the expansion of the universe is accelerating. Researchers wanted to know more about dark energy, the mysterious force that was causing it, and whether it varied in space or over time. Galaxy clusters are among the best indicators, says x-ray astronomer Andrew Fabian of the Institute of Astronomy (IoA) in Cambridge, U.K. "Clusters are the most massive objects in the universe, the pinnacle of galaxy formation, and are very sensitive to cosmological models."

They are best seen in x-rays because the gaps between galaxies are filled with gas that is heated to millions of degrees as the galaxies jostle together in a cluster. By mapping the clusters, Spektr-RG "will study the evolution of the structure of the universe," says Esra Bulbul, of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, who recently joined the MPE team.

The challenge was to boost the capabilities of the existing ROSITA telescope, which could only garner up to 10,000 galaxy clusters. Discussions led to a €90 million "extended" eROSITA, paid for by MPE and the German Aerospace Center, DLR. It is an array of seven identical telescopes with five times the effective collecting area of the original instrument. Russia and Germany signed an agreement in 2007 with launch penciled in for 2012.

But mission development was not smooth. The U.K. instrument failed to win funding and was replaced with a Russian telescope that will complement eROSITA by detecting scarcer "hard" x-rays. Though harder to collect, those higher energy photons are particularly useful for seeing the supermassive black holes at galactic cen-

ters, because they pierce the clouds of gas and dust that shroud them.

Making the mirrors for eROSITA posed another challenge. Because x-rays would penetrate a flat telescope mirror, focusing them requires cylindrical mirrors that gather the photons in glancing, low-angle reflections off inner surfaces. Each of eROSITA's seven scopes contains 54 gold-plated cylindrical mirrors, nested inside one another, that must be shaped precisely to bring the photons to a focus. Making them proved so hard that the MPE team had to fire its main contractor part way through. "It almost killed us," Predehl says.

A decision to site the telescope at a stable, gravitationally balanced point beyond the moon, outside the shelter of Earth's magnetic field, meant electronics had to be hardened against solar radiation. Incompatibility between the German and Russian electronics delayed the launch, as did problems with the spacecraft's communications system and a change in launch rocket.

Now that Spektr-RG is finally ready, expectations are high. "It's going to be revolutionary in terms of numbers," says IoA astronomer George Lansbury, taking x-ray studies into "the big data regime."

It may also be a rare high point for Russia's great observatories program. Previously, only one has made it into orbit: 2011's Spektr-R, a radio astronomy mission that fell short of expectations and could not be revived after malfunctioning earlier this year (*Science*, 29 July 2011, p. 512).

Astronomers may face a long wait for Spektr-RG's successors: the ultraviolet telescope Spektr-UV and Spektr-M, a millimeter-wave radio telescope. Spektr-UV has survived moments of near-death, most recently in 2014 when Russia's annexation of Ukraine's Crimean peninsula caused major Ukrainian partners to withdraw. The mission is now slated for a 2025 launch, but, Sunyaev says, some collaborators, including a German team supplying a spectrograph, have dropped out. Spektr-M, which would come next, is not yet fully funded, he says. And in the meantime, rival telescopes launched by other countries may scoop up the science the Russian missions aim to do.

"Russia is doing as much as possible with the budget available," says Spektr-RG chief Mikhail Pavlinsky of IKI. He notes that Roscosmos's lean budget, worth $20.5 billion over 10 years, faces multiple demands. Russia is building the landing system for the European ExoMars rover, due to launch next year, and like other countries it hopes to return to the moon, with the Luna 25 lander in 2021. For Russia's astrophysicists, Pavlinksy says, "It means slow progress." ■

CLINICAL TRIALS

Experimental drug holds off type 1 diabetes

Two weeks of therapy delays disease for average of 2 years

By **Jennifer Couzin-Frankel**

Culminating a 33-year odyssey, scientists this week reported a milestone in type 1 diabetes: evidence that it can be held off. At the American Diabetes Association meeting in San Francisco, California, and in *The New England Journal of Medicine*, researchers reported that 2 weeks of an experimental drug delayed disease by an average of about 2 years in young people at very high risk.

"These data are the first to show it is possible to prevent the progression of type 1 diabetes," says Lucienne Chatenoud at Hôpital Necker-Enfants Malades in Paris, who wasn't involved in this study.

Type 1 diabetes is an autoimmune disease that requires constant blood sugar monitoring and can lead to heart disease and kidney failure. Years before diagnosis, the sentries of the immune system, T cells, attack β cells in the pancreas. But those insulin-secreting cells remain largely intact, offering a window for intervening.

> ***"... it is possible to prevent the progression of type 1 diabetes."***
>
> **Lucienne Chatenoud**, Hôpital Necker-Enfants Malades

Decades ago, Kevan Herold, an endocrinologist now at Yale University, turned to an antibody drug designed by Jeffrey Bluestone, an immunologist now at the University of California, San Francisco. It shuts down activated T cells by targeting a molecule on the cells' surface called CD3. In the 1990s, they and a French team co-led by Chatenoud showed that anti-CD3 could prevent or reverse diabetes in a mouse model. Small clinical trials in newly diagnosed people also looked promising.

But two large trials in new-onset patients proved disappointing. "That was devastating," Bluestone says of the results, announced in 2010. Still, he, Herold, and others stayed hopeful, believing those trials used doses that were too low or included participants who didn't have the autoimmune form of diabetes.

Herold convinced a diabetes clinical trials network called TrialNet to break new ground: Support a study of the drug in people who didn't yet have diabetes. Starting in 2011, the trial recruited those with an estimated 75% chance of getting diabetes in the next 5 years, based on unstable blood sugar and antibodies in their blood that indicated attacks on the pancreas.

Forty-four volunteers received teplizumab, as the drug Bluestone helped design is now called, intravenously every day for 14 days. Another 32 got a placebo. The difference was stark. In the treatment group, the median time to a diabetes diagnosis was just over 4 years; in the placebo group, it was 2 years. Forty-three percent of those who got teplizumab developed diabetes after 5 years versus 72% of controls. Side effects were mild and resolved within weeks.

"To gain 2 years of an insulin-free life … that's significant," says Mark Atkinson, a pathologist at the University of Florida in Gainesville. "You have to think of mom or dad having 2 years less of getting up at night" to check their child's blood sugar, he says, along with a potentially lower risk of complications.

The big question now is what's next. Carla Greenbaum, chair of TrialNet and an endocrinologist at Benaroya Research Institute in Seattle, Washington, hopes researchers will probe whether other immune therapies can also delay disease. She also wonders whether a second dose of teplizumab would extend its efficacy.

Given the striking results, a larger placebo-controlled prevention trial of teplizumab might be hard to justify. And although the volunteers in Herold's trial all had a first-degree relative with diabetes, at least 85% of patients don't—so only widespread screening would reach everyone at risk. "Will the public even participate?" Atkinson wonders.

Still, Herold hopes his study marks a turning point. As he was finalizing the data, he noticed that the trial's first volunteer had disappeared. Herold found out later that the young man had gotten teplizumab. "I called him up and said, 'What's going on?'" Herold recalls. Not much, the volunteer admitted; he had forgotten to stay in touch. "That's terrific," Herold thought. Forgetting is what someone with diabetes can't do—so for this man, who is disease free, it meant everything. ■

THE CONFESSION

A psychologist has shown how police questioning can get innocent people to condemn themselves

By **Douglas Starr**

Saul Kassin has studied interrogations by observing them and simulating them in the lab.

At 16, Huwe Burton confessed to killing his mother. He was still in shock from discovering her body when New York City police began to interrogate him. After 3 hours of being threatened and cajoled, he told the police what they wanted to hear. He soon recanted, knowing he was innocent and hoping the justice system would clear him.

Burton was convicted of second-degree murder in 1991 and received a sentence of 15 years to life.

After 20 years in prison, he was released on parole, but he never could shake the stigma of the conviction. Attorneys from several organizations worked for more than a decade to clear him. They produced facts that contradicted the confession and showed evidence of prosecutorial misconduct. But for the Bronx District Attorney's Office, Burton's confession outweighed all other evidence; after all, who would admit to a crime they did not commit? Finally, last summer Burton's attorneys brought in Saul Kassin, a psychologist at the John Jay College of Criminal Justice in New York City who is one of the world's leading experts on interrogation.

"I went in prepared to make a 15-minute presentation, but the attorneys started asking some really good questions," Kassin says. "Before you knew it, we had a discussion that lasted almost 2 1/2 hours."

Kassin explained that false confessions are not rare: More than a quarter of the 365 people exonerated in recent decades by the nonprofit Innocence Project had confessed to their alleged crime. Drawing on more than 30 years of research, Kassin told the legal team how standard interrogation techniques combine psychological pressures and escape hatches that can easily cause an innocent person to confess. He explained how young people are particularly vulnerable to confessing, especially when stressed, tired, or traumatized, as Burton was.

Kassin's presentation helped open the prosecutors' eyes to the emerging science of interrogation and false confession. Six months later, on 24 January, Judge Steven Barrett of the Bronx Supreme Court vacated Burton's 3-decade-old conviction, citing such work as the basis of his decision. "Having Dr. Kassin come in and give a master class on the science of false confessions was a turning point," says Steven Drizin, co-director of the Center on Wrongful Convictions at Northwestern University in Chicago, Illinois, who led the team that pursued Burton's exoneration.

Although scores of people have been cleared of false confessions since DNA evidence entered U.S. courtrooms, the Burton case was the first time someone had been exonerated on the basis of the scientific analysis of interrogation. As such, it marks the coming of age of research that is profoundly affecting the justice system. Confessions are being questioned as never before—not just by defense lawyers, but by lawmakers and some police departments, which are reexamining their approach to interrogation.

Kassin is part of a cadre of scientists who have flipped conventional wisdom about confessions—and about the perception of truth. His cleverly designed experiments have probed the psychology that leads to false confessions. In more recent work, he has shown how a confession, true or not, can exert a powerful pull on witnesses and even forensic examiners, shaping the entire trial.

"Saul Kassin is one of the godfathers of the innocence movement," says Rebecca Brown, policy director of the Innocence Project in New York City. Drizin has his own metaphor: "If there was a Mount Rushmore to the study of false confessions, Dr. Kassin's face would be on it."

> "... confessions that look real can actually be false, even if they're corroborated by informants and forensic science."
>
> **Saul Kassin**, John Jay College of Criminal Justice

CONFESSIONS HAVE ALWAYS BEEN the "gold standard" indicator of guilt, even though some proved spectacularly misleading. For example, a man who had admitted to a murder in 1819 narrowly escaped hanging when his supposed victim was found living in New Jersey. The first scientific red flag came from Hugo Münsterberg, a renowned Harvard University psychologist, who in 1908 warned about "untrue confessions ... under the spell of overpowering influences." But it took several shocking false confession cases in the late 1980s and the introduction of DNA evidence to the justice system for the extent of wrongful convictions to emerge—and with it how often false confessions played a role.

Kassin was not surprised, having spent years studying police interrogation techniques. In person he projects a kind of affable intensity, with piercing brown eyes and a conversational style that lends urgency to even a casual chat. Raised in a working-class neighborhood of New York City, he got his bachelor's degree at Brooklyn College in New York (tuition: $53 per semester) and his Ph.D. at the University of Connecticut in Storrs, both in psychology. As a postdoc at the University of Kansas in Lawrence, he studied how juries make decisions and was struck by the power of a confession to practically guarantee a guilty verdict.

He also began to wonder how often those confessions were genuine, after he learned about the Reid interrogation technique, the near-universal method taught to police. Its training manual—now in its fifth edition—was first published in 1962 by John Reid, a former Chicago detective and lie detector expert, and Northwestern University law professor Fred Inbau. "I was horrified," Kassin says. "It was just like Milgram's obedience studies, but worse."

Stanley Milgram, a psychologist at Yale University and one of Kassin's heroes, had conducted studies in the 1960s in which subjects were encouraged to give electric shocks to other subjects who were not learning their lessons quickly enough. The volunteers, who didn't know the shocks they gave were fake, were disturbingly willing to inflict pain when someone in authority told them to.

A Reid interrogation looks different at first. It starts with a behavioral assessment, in which the officer asks questions—some irrelevant and some provocative—while watching for signs of deception, such as looking away, slouching, or crossing the arms. If the suspect is thought to be lying, the investigator moves on to phase two, the formal interrogation. Now, they amp up the questioning—repeatedly accusing the suspect, insisting on hearing details, and ignoring all denials. Meanwhile, the investigator offers sympathy and understanding, minimizing the moral (but not legal) dimension of the crime and easing the path to confession. (Example: "This never would have happened if she didn't dress so provocatively.")

That phase, with an authority figure applying psychological pressure, reminded Kassin of Milgram's infamous experiments. But whereas Milgram got someone to "harm" another person, the Reid technique gets people to harm themselves by admitting guilt. Kassin suspected that the pressure might sometimes lead to false confessions.

To find out, he decided in the early 1990s to model the Reid technique in the lab, with student volunteers. In what Kassin called the computer crash paradigm, he had students take rapid-fire dictation on computers. He warned them that the system had a glitch and that hitting the Alt key would trigger a crash. That part was a fib: The computers were programmed to crash regardless of which keys were hit. The experimenter then accused the students of hitting the Alt key.

PHOTO: (OPPOSITE PAGE) DREW GURIAN

PHOTO: CLARENCE DAVIS/NEW YORK DAILY NEWS/GETTY IMAGES

At first, none confessed. Then, Kassin added variables based on what he and other researchers had learned about actual police interrogation tactics. Sometimes, for example, police falsely tell a suspect they have witnesses to the crime—causing a suspect to doubt their own version of events. (Under U.S. law, police are permitted to lie.) In one of the most striking examples, Marty Tankleff, a Long Island teenager, came to breakfast one morning in 1988 to find his parents stabbed on the kitchen floor, his mother dying and his father in a coma. Detectives thought Tankleff was not sufficiently grief-stricken, so he became their prime suspect. After hours of getting nowhere, a detective said he had called Tankleff's father at the hospital and that the injured man said Tankleff had committed the crime. (In truth, his father died without regaining consciousness.) Shocked beyond reason, Tankleff confessed. He spent 19 years in prison before the culprits were found and he was set free.

Kassin could never simulate that kind of trauma in the lab, but he could set up a variation of the computer crash experiment in which a confederate claimed to have seen the student hit the wrong key. Those students confessed at more than double the rate of students paired with witnesses who said they hadn't seen anything. Under some circumstances, nearly every student facing a false witness confessed.

Some students ended up believing they really had caused the crash, coming up with explanations such as, "I hit the wrong key with the side of my hand." So deeply had they internalized their guilt that some refused to believe Kassin when he told them the truth.

Another detective told Kassin that during an interrogation, he didn't actually lie about the evidence in hand, but said he expected new, potentially incriminating evidence to come in. For example, an interrogator might tell a suspect that they were waiting for lab results on DNA from the crime scene. You might think that doing so would get the innocent to deny the crime more vehemently because they expected the results to absolve them. Kassin, however, had interviewed exonerated men who said the prospect of new evidence had a surprising effect. Some confessed just to get out of the stressful situation, figuring that the evidence would later clear them. "They think their innocence is their ticket out of there," he says.

Kassin tested such police "bluffs" in a variation of the computer crash experiment. This time, in addition to accusing the students, the experimenter said that all the keystrokes had been recorded on the server and would soon be examined. The false confession rate soared. Postexperiment questionnaires revealed that many of the bluffed students, like the men Kassin had interviewed, signed a confession to get out of the room and assumed they'd later be cleared. In that sense, Kassin says, belief in one's innocence and faith in the justice system can themselves be risk factors.

SOCIAL SCIENTISTS WORLDWIDE have repeated variations of the computer crash experiments, with similar results. But critics have questioned Kassin's findings because the "crimes" his subjects were charged with could have been simple acts of carelessness, committed unwittingly, and because confessing bore no serious consequences. Joseph Buckley, president of John E. Reid & Associates Inc. in Chicago, the company that copyrighted the Reid technique in the early 1960s, adds that Kassin's studies lack validity because they were not conducted using professional interrogators. Buckley says false confessions occur only when interrogators don't closely follow procedures. In a January report, Buckley said the Reid technique isn't meant to force a confession. Instead, he wrote, its goal "is to create an environment that makes it easier for a subject to tell the truth."

Work by other researchers has answered some of those criticisms. Social psychologist Melissa Russano at Roger Williams University in Bristol, Rhode Island, designed an experiment in which volunteers were asked to solve a set of logic problems—some working in groups and some alone. The researchers stipulated that under no circumstances should anyone assist the students working alone. Beforehand, however, a few students were coached to become visibly upset. That prompted some of their classmates to help, in violation of the rules.

In those experiments, the helpers could not have committed the "crime" without

PHOTO: GREGG VIGLIOTTI/THE NEW YORK TIMES

Huwe Burton falsely confessed to killing his mother. Nearly 30 years passed before he was exonerated.

knowing, and confessing carried some consequence because cheating violated the college's honor code. But, just as Kassin found, accusatory questioning often provoked false confessions. Russano also tested another component of standard interrogations—the "minimization" technique that lowers the emotional barrier to confessing. She and colleagues would say things such as, "You probably didn't realize what a big deal this was." That technique increased false confession rates by 35%.

Other researchers, including Gísli Guðjónsson, a former Icelandic detective who became an eminent psychologist at King's College London, have shown how some individuals are especially susceptible to such pressure. Factors such as mental impairment, youth, and substance addiction make people quicker to doubt their own memory and, under pressure, to confess, Guðjónsson found. Law professor Richard Leo of the University of San Francisco in California reported that fewer than 20% of U.S. suspects invoke their Miranda rights against self-incrimination, perhaps hoping to appear cooperative. He and social psychologist Richard Ofshe, then at the University of California, Berkeley, also described "persuaded" confessions in which a suspect, worn down by hours of interrogation, goes into a fugue and begins to believe their own guilt. The problem is especially pronounced among adolescents like Burton, who are both impressionable and cowed by authority.

Much of the Reid technique involves watching for verbal and nonverbal signs of deception, something many police investigators think they are skilled at doing. Kassin put that confidence to the test more than a decade ago. He recruited the best liars he could find—a group of prisoners at a Massachusetts penitentiary. For a small fee he asked half to tell the truth of their crimes on video and the other half to lie, saying they had committed someone else's crime. He showed the videos to college students and police. Neither group did particularly well at truth detection (the average person is right about half the time), but the students performed better than the police. Yet the police felt more certain about their conclusions. "That's a bad combination," Kassin says. "Their training makes them less accurate and more confident at the same time."

A POSTER IN KASSIN'S OFFICE at John Jay College shows 28 faces: men, women, adults, adolescents, white, black, Hispanic. "Look at how many different types of people there are—all of humanity," Kassin says. "And what they have in common is that they all gave false confessions. There's no one kind of person who can give a false confession. It can happen to anybody."

Kassin has helped many of them. Defense lawyers and human rights organizations around the world often call on him to analyze confessions or testify about the nature of interrogation—sometimes as a paid consultant or witness, sometimes pro bono. One face on the poster belongs to Amanda Knox, the U.S. college student studying in Italy who was coerced into confessing to the murder of her roommate. Kassin's reports to Italian courts were involved in getting her freed. He testified for John Kogut, a Long Island man who after an 18-hour interrogation falsely confessed to raping and murdering a 16-year-old girl. DNA evidence had won Kogut's release after he spent 18 years in prison, but prosecutors retried him on the basis of the confession. Kassin's 2005 testimony helped exonerate Kogut.

Then there was Barry Laughman, a man with the mental capacity of a 10-year-old, who in 1987 confessed to raping and murdering an elderly neighbor after police falsely told him they found his fingerprints at the scene. After his confession, the police disregarded all other evidence. Neighbors who offered alibis for Laughman were told they must be mistaken. His blood was type B, but the only blood at the crime scene was type A. So the forensic expert proposed a novel theory: that bacterial degradation could have changed the blood type from B to A. Laughman spent 16 years in prison until DNA evidence finally cleared him. (Kassin later testified when Laughman sued the state.)

To Kassin, Laughman's case showed that confession doesn't just trump other evidence, but can corrupt it as well. After a confession, alibis are recanted, witnesses

change stories, police ignore exculpatory evidence, and forensic scientists reinterpret material. In Huwe Burton's case, for example, police had caught a neighbor with a history of violence driving the dead mother's stolen car, but they did not consider him a suspect because Burton had confessed.

The magnitude of the effect emerged in 2012, when Kassin and colleagues published an analysis of 59 false confession cases from the Innocence Project. Forty-nine of those also involved other mistakes, such as eyewitness errors and mistaken forensics—a far higher proportion than in nonconfession cases. In 30 of those cases, the confession was the first piece of evidence collected. In other words, once the police had a confession, all the other evidence lined up to support it. That has an ironic effect: Even when confessions have turned out to be false, appeals courts have ruled that the other evidence is strong enough to support the conviction, Kassin says. "The courts completely missed out that the other evidence was corrupted."

"There's no one kind of person who can give a false confession. It can happen to anybody," says Saul Kassin, who keeps a photo gallery of innocent people convicted after false confessions in his office.

Other groups have shown experimentally how a narrative can shape forensic evidence. One dramatic example came in 2011, when U.K. psychologist Itiel Dror and U.S. DNA expert Greg Hampikian tested the people you would least expect to be affected by bias—DNA specialists. Dror and Hampikian obtained the printed DNA results from a rape case in which a man was found guilty. The original genetic analysts had been told that police had a suspect in custody; the forensic experts then determined that the suspect's DNA was part of the crime scene sample. To see whether knowledge of the arrest caused bias, Dror and Hampikian gave the printouts to 17 experts unconnected with the case and told them nothing about the suspect. Only one of them matched the suspect's DNA to the crime sample. Such findings support the increasingly popular idea that all forensic science should be "blinded"—conducted without any knowledge about the suspects.

Sometimes a confession will override even untainted DNA evidence. In the infamous "Central Park Five" case dramatized in a new Netflix series, five teenagers in 1989 confessed after hours of interrogation to brutally beating and raping a female jogger in New York City. They quickly recanted, and none of the DNA recovered from the victim was theirs. Yet two juries convicted them after the prosecutor explained away the contradiction. She came up with a theory that a sixth unidentified accomplice had also raped the victim and was the only person to ejaculate. (The "unindicted co-ejaculator" theory has been used in other wrongful convictions as well.) Thirteen years later, the man whose DNA matched the sample—a convicted serial rapist and murderer serving a life sentence—confessed that he alone had committed the crime.

How could such an injustice occur? Kassin published a study in 2016 in which he simulated the situation with mock jury experiments. When presented with a simple choice between a confession and DNA, people would choose DNA. But if the prosecutor offered a theory as to why the DNA contradicted the confession, the juries overwhelmingly sided with the confession—an insight, he says, into the power of story to influence judgment.

CHANGE IS COMING. By 2010, the evidence about how interrogations can go wrong had become so compelling that Kassin and several colleagues from the United States and United Kingdom wrote an American Psychological Association white paper warning about the risk of coercion. They suggested several reforms, such as prohibiting lying by police, limiting interrogation time, recording all interrogations from start to finish, and eliminating the use of minimization. They also said the practice of seeking confessions was so inherently damaging that it might be necessary to "completely reconceptualize" the tactic and come up with something new.

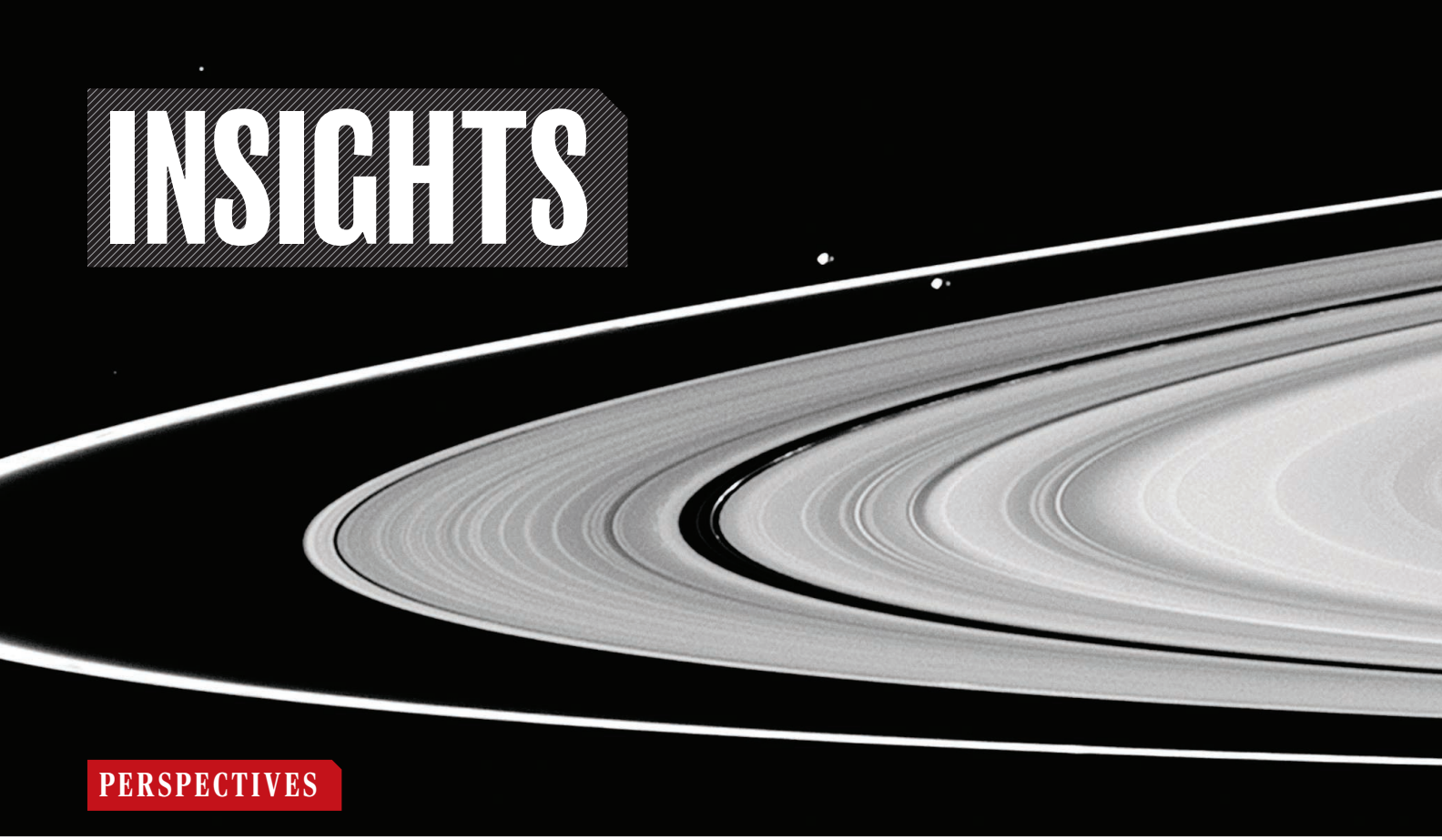
One model comes from England, where police did away with their Reid-style interrogation system in the early 1990s after several false conviction scandals. Police there now use a system designed to identify deception based not on visible signs of emotional stress, but on "cognitive load," which can lead liars to stumble as they try to keep their stories straight. English police conduct the kind of open-ended interviews that journalists might use and are encouraged not to go after confessions. Several other countries including New Zealand and Australia, along with parts of Canada, have adopted the new method. They also record the entire interrogation to make the process transparent, something that 25 U.S. states have also adopted.

Two years ago, one of the largest U.S. interrogation trainers, Chicago-based Wicklander-Zulawski & Associates Inc., stopped teaching accusatory interviews and embraced the nonconfrontational methods Kassin and his colleagues advocate. The company was influenced by the proliferation of research and a desire to minimize false confessions, says Dave Thompson, vice president of operations. "We realized there's a better way to talk to people today than the way we talked to people 20 or 30 years ago."

Kassin sees progress, too. In March, he spoke to a group that until recently might have been hostile to his message: 40 district attorneys from around the country who want to learn to avoid wrongful convictions. "My point with them was that they are going to be fooled—that confessions that look real can actually be false, even if they're corroborated by informants and forensic science," he says. "I wanted to let them know that alarm bells should go off when they see a confession case." ■

Douglas Starr is a journalist in Boston.

PHOTO: DREW GURIAN



PLANETARY SCIENCE

The origin of Saturn's rings and moons

Cassini data constrain the age and history of the giant planet's rings

By **Shigeru Ida**

Saturn's rings consist of vast numbers of small icy particles that frequently collide with each other. Within Saturn's Roche limit (that is, closer to Saturn's center than twice Saturn's physical radius R_S), the icy particles disperse and form rings, whereas outside the Roche limit, they grow through pairwise collisions to form moons. The formation of Saturn's rings and moons should thus be closely related. On pages 1054, 1053, and 1052 of this issue, respectively, Tiscareno *et al.* (*1*), Buratti *et al.* (*2*), and Iess *et al.* (*3*) report results from the Cassini mission's ring-grazing orbit and Grand Finale observations, which reveal the detailed structure of the rings and associated moons. The results strongly suggest that Saturn's rings are much younger than Saturn itself and provide important clues to the origin of the rings and moons.

Saturn has a total of eight mid-size and large moons (diameters greater than 200 km), which orbit the planet in coplanar and circular orbits. The diameters of the mid-size moons, from Mimas to Rhea, increase monotonically from 400 km to 1500 km (see the figure). Titan, the largest satellite with a diameter of 5150 km, has a separate orbit from the inner moons at $r = 20R_S$. On the other hand, small, irregular-shaped moons such as Pandora, Prometheus, and Epimetheus (ring moons with diameters of 10 to 100 km, see the images) exist in the outermost part of the rings near the Roche limit. Estimated bulk densities show that the large moons have a range of different rock-to-ice ratios, whereas the ring particles are mostly composed of water ice.

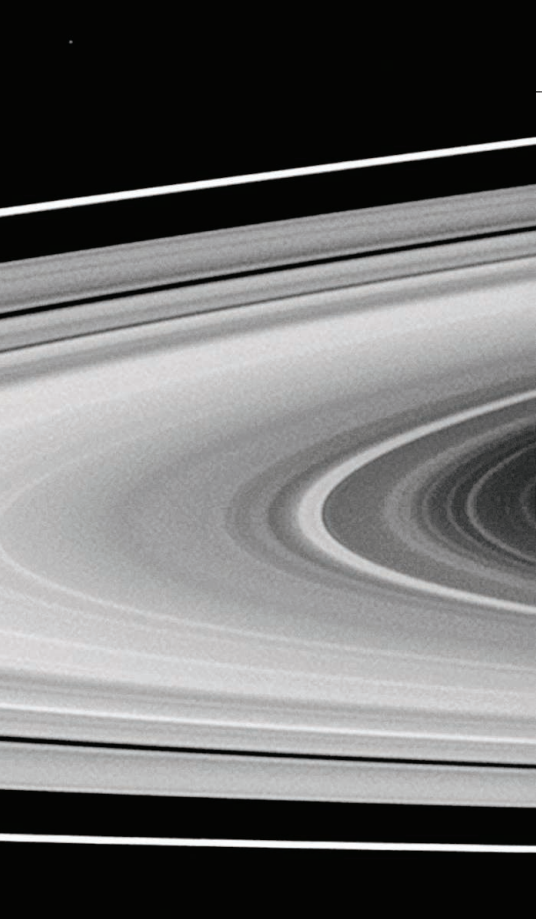
Two main models have been proposed to explain the formation of Saturn's moons (see the figure). They may have accreted either from small, kilometer-sized bodies in a circumplanetary gas disk (model 1) (*4*, *5*) or from icy particles that diffused out from a massive ring (model 2) (*6*, *7*). According to another recent model, they may have reaccreted from debris generated by collisions between first-generation moons that formed by either of the two main models (*8*).

Saturn formed through accretion of H/He gas from the solar nebula onto a rock/ice core. During accretion, a gas disk is likely to have formed around Saturn through which gas flowed onto Saturn's surface. In the disk, icy and rocky grains condensed. Model 1 assumes that the grains accumulated into kilometer-sized bodies, which occasionally collided with each other to form moons in the disk. Some moons could migrate through interaction with the gas disk into the interior of the Roche limit. Because the density difference between ice and rock means that the Roche limit radius is larger by ~40% for ice than for rock, the icy mantle of the migrating moon(s) may have been stripped first, providing icy particles for the rings (*9*). Numerical simulations of model 1 have been able to reproduce both Titan and the rings (*5*, *9*), but not the inner moons.

On the other hand, a compilation of century-long ground observations and Cassini data has shown that Saturn's tidal torque is causing the inner moons' orbits to expand

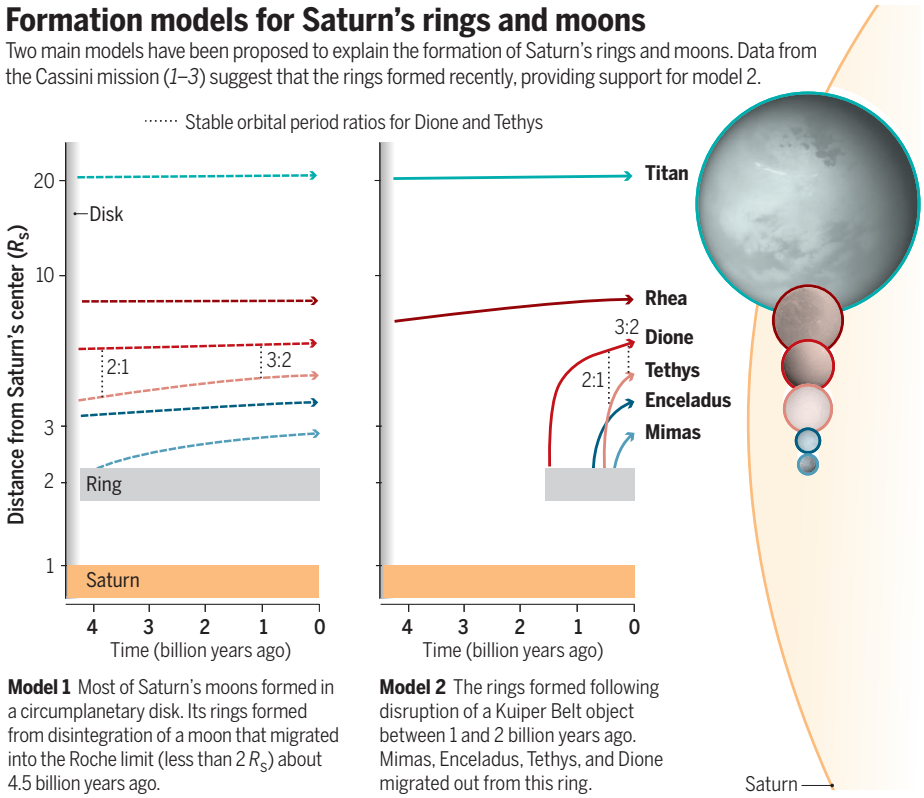
Earth-Life Science Institute, Tokyo Institute of Technology, 152-8550 Tokyo, Japan. Email: ida@elsi.jp

PHOTO: NASA/JPL/SPACE SCIENCE INSTITUTE



Formation models for Saturn's rings and moons

Two main models have been proposed to explain the formation of Saturn's rings and moons. Data from the Cassini mission (*1*–*3*) suggest that the rings formed recently, providing support for model 2.



Model 1 Most of Saturn's moons formed in a circumplanetary disk. Its rings formed from disintegration of a moon that migrated into the Roche limit (less than 2 R_S) about 4.5 billion years ago.

Model 2 The rings formed following disruption of a Kuiper Belt object between 1 and 2 billion years ago. Mimas, Enceladus, Tethys, and Dione migrated out from this ring.

Data from the Cassini mission (1-3) suggest that Saturn's rings and small moons, such as Pandora and Prometheus, formed in the past two billion years.

rapidly (*10*). The results require the inner mid-size moons—Mimas, Enceladus, Tethys, and Dione—to have formed no more than ~1 to 2 billion years before the present. This is a challenge for model 1, which assumes that both the rings and the moons formed at the time of Saturn's formation, 4.5 billion years ago. It is possible, however, that Saturn's tidal torque was weaker in the past, consistent with formation of the rings and moons at 4.5 billion years (*11*). The latest observations from the Cassini mission provide further constraints on when the rings and moons formed.

On the basis of precise gravity measurements, Iess *et al.* estimate the mass of the rings as ~0.4 times the mass of the moon Mimas. This value is consistent with a theoretical prediction based on radial diffusion of the ring (*12*). Suppose that the ring was initially massive. In the massive ring, gravitational interactions among the ring particles may create a density wake structure; the wakes cause rapid radial diffusion of the ring and the ring mass decreases with time. Because the wakes become weak as the mass decreases, the diffusion slows down, which means that the ring mass must have been similar to its current small mass for most of its history. Close-up observations of the ring reported by Tiscareno *et al.* show a hint of the wakes, but no definitive evidence. They found that visible fine structures in the ring are dominated by particle properties (such as regolith character, bulk porosity, and particle size distributions), rather than by the density of the ring particle distribution. More detailed analysis of the complex ring structure is needed to constrain the ring evolution.

If the diffusion due to the wakes is responsible for the ring evolution, the overall mass of the current ring does not constrain the formation time and initial mass of the ring. Cassini in situ measurements of the interplanetary dust flux over 10 years show that the properties of the light reflected by the rings are not consistent with interplanetary dust contamination for as long as 4.5 billion years. Iess *et al.* suggest that to be consistent with the observed clean rings, the rings must be as young as ~10 to 100 million years. Buratti *et al.* argue that ice particles and water vapor from geysers on Enceladus accrete on the ring moons and probably also the ring particles, hiding past contamination and making them appear younger than they are. Nonetheless, the Cassini observations suggest that the rings are comparatively young.

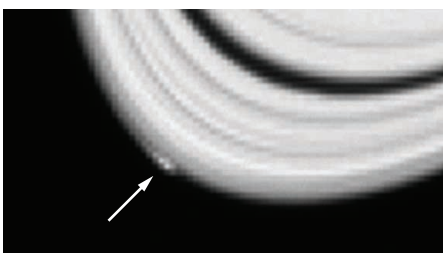
Young rings are more consistent with model 2, provided that the rings are older than the mid-sized moons. Whereas the circumplanetary disk in model 1 existed only during Saturn's formation, the massive ring in model 2 can be formed any time by tidal disruption of a Kuiper Belt object that coincidentally entered the interior of the Roche limit, although the probability of such an encounter would decrease with time (*13*).

In model 2, if the impacting object has a rocky core and an icy outer layer, its icy layer would have been selectively stripped (*14*). Precession due to Saturn's ellipticity would have resulted in the formation of a debris torus, and collisions between the debris particles would have ground down the icy debris particles, eventually resulting in the formation of a geometrically thin ring in Saturn's equatorial plane through collision damping (*14*). Because the Roche limit for rock is smaller than that for ice, rocky chunks can survive in the rings. The rocky chunks can then accrete icy particles, consistent with the detailed structure of ring moons reported by Buratti *et al.* These ring moons eventually diffuse out from the Roche limit and grow through mutual collisions to form mid-size moons (*7*). The moons' orbits expand as a result of Saturn's tidal torque.

Model 2 cannot fully explain the orbital configurations of the mid-size moons. According to orbital simulations, Tethys and Dione should have become trapped in an orbital configuration where the ratio of their orbital periods was 2:1 or 3:2 (called a 2:1 or 3:2 resonance) during the tidal orbital expansion to their current orbits (see the figure), but this is inconsistent with their current orbits (*15*). To avoid the trapping, Cuk *et al.* (*8*) proposed that the orbits of tidally expanding

moons became unstable, resulting in mutual collisions that caused them to disintegrate, and that the current moons reaccreted ~1 billion years ago. This could also account for the high crater density on Mimas's surface, even if Mimas is young. Another possibility to avoid the trapping is the temporary excitation of orbital eccentricity through gravitational perturbations from the original massive ring. The effect of exciting the eccentricity of Enceladus's orbit is to avoid the Dione/Tethys trapping (*15*).

Another problem for model 2 is the low probability of an encounter with a Kuiper Belt object less than 2 billion years ago. Such an encounter would have been very rare except during the Solar System formation stage at ~4.5 billion years ago and the Late Heavy Bombardment era at ~4 billion years ago (*13*). Detailed astronomical observations from the Gaia spacecraft have started to identify past close encounters between the Solar System and other stars in the Milky Way galaxy; these data will help to constrain the frequency of encounters between Saturn and Kuiper Belt objects. A clear answer to the long-standing question of when and how Saturn's rings formed has not yet been obtained, but the Cassini data provide important pieces of the puzzle. ■



A small object, called Peggy, could be a new moon forming on the edge of Saturn's A-ring.

REFERENCES AND NOTES

1. M. S. Tiscareno *et al.*, *Science* **364**, 1054 (2019).
2. B. J. Buratti *et al.*, *Science* **364**, 1053 (2019).
3. L. Iess *et al.*, *Science* **364**, 1052 (2019).
4. I. Mosqueira, P. R. Estrada, *Icarus* **163**, 232 (2003).
5. R. M. Canup, W. R. Ward, *Nature* **441**, 834 (2006).
6. S. Charnoz *et al.*, *Icarus* **216**, 535 (2011).
7. A. Crida, S. Charnoz, *Science* **338**, 1196 (2012).
8. M. Ćuk, L. Dones, D. Nevorny, *Astrophys. J.* **820**, 97 (2016).
9. R. M. Canup, *Nature* **468**, 943 (2010).
10. V. Lainey *et al.*, *Icarus* **281**, 286 (2017).
11. J. Fuller, J. Luan, E. Quataert, *Mon. Not. R. Astron. Soc.* **458**, 3867 (2016).
12. J. Salmon, S. Charnoz, A. Crida, A. Brahic, *Icarus* **209**, 771 (2010).
13. S. Charnoz, A. Morbidelli, L. Dones, J. Salmon, *Icarus* **199**, 413 (2009).
14. R. Hyodo, S. Charnoz, K. Ohtsuki, H. Genda, *Icarus* **282**, 195 (2017).
15. A. Nakajima, S. Ida, J. Kimura, R. Brasser, *Icarus* **317**, 570 (2019).

ACKNOWLEDGMENTS

I thank Y. Sekine and R. Hyodo at ELSI for helpful discussions.

10.1126/science.aaw3098

MICROBIOLOGY

Gut microbes metabolize Parkinson's disease drug

A gut bacterial pathway that degrades the drug levodopa is identified and can be inhibited

By **Cora O'Neill**[1,2,3]

The trillions of microorganisms that form the gut microbiota contain a treasure trove of enzymes. These directly modify and metabolize dietary components, drugs, and toxins that humans ingest. Although this is often beneficial, the gut microbiota can modify drug bioavailability and efficacy (*1*, *2*). Levodopa (L-dopa), the major drug treatment for Parkinson's disease, displays highly variable and interindividual responses with reduced efficacy over time. On page 1055 of this issue, Maini Rekdal *et al.* (*3*) identify a two-step gut microbial enzymatic pathway that degrades L-dopa to dopamine and then to *m*-tyramine, thus potentially limiting drug availability in patients. Moreover, they identify a small molecule that blocks this L-dopa–metabolizing bacterial pathway, with the aim of increasing L-dopa availability in Parkinson's disease patients.

Globally, 10 million people are living with Parkinson's disease, and L-dopa has been the primary treatment for over 50 years (*4*). Dopamine becomes progressively depleted in the brains of Parkinson's disease patients, causing motor impairment, including tremor, rigidity, and slowness of movement. L-dopa, the precursor of dopamine, is generally taken orally. The drug is absorbed in the small intestine, and unlike dopamine, crosses the blood-brain barrier into the brain, where it is converted to dopamine by decarboxylation. Dopamine replacement alleviates the motor symptoms of Parkinson's disease, but does not prevent Parkinson's disease progression. L-dopa is prematurely decarboxylated to dopamine before it gets to the brain (particularly in the gut), and this limits drug bioavailability and causes gastrointestinal problems (*4*). Counteracting this effect is achieved by coadministration of peripheral human decarboxylase inhibitors (*4*). However, L-dopa bioavailability still varies considerably between patients, and efficacy wanes over time, necessitating increased doses with unpredictable fluctuating motor responses and adverse side effects (*4*, *5*).

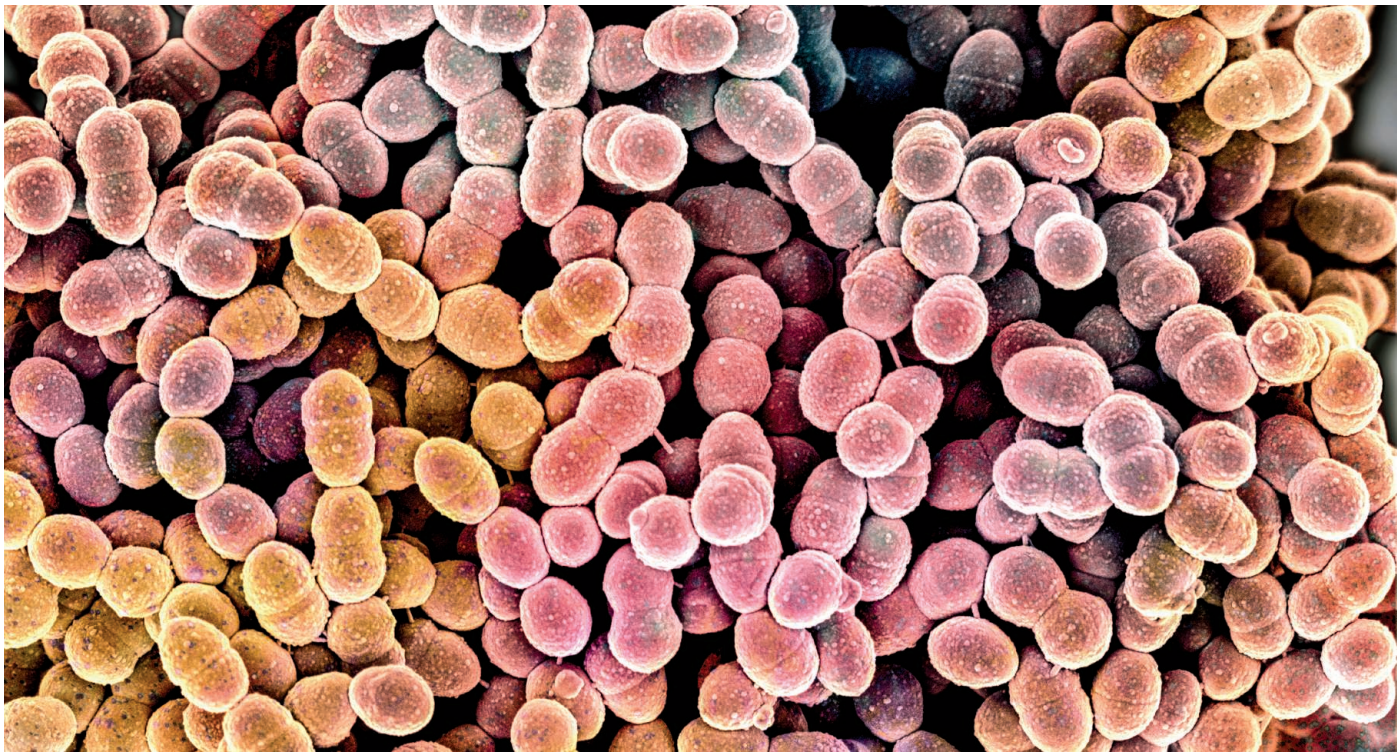
Gastrointestinal dysfunction, including constipation, delayed gastric emptying, and small intestine bacterial overgrowth, are key components of Parkinson's disease, and these impair L-dopa intestinal absorption and drug responses (*5*). Gut dysfunction occurs before motor symptoms and worsens as Parkinson's disease advances. Indeed, an origin of Parkinson's disease in the gut has been proposed (*5*, *6*). The gut microbiome is altered in patients with Parkinson's disease (*5*, *6*), and this might underlie gut-brain axis pathophysiology (*5*–*7*) as well as limit L-dopa therapies (*5*). Remarkably, it has been known since 1971 that gut microbes metabolize L-dopa to dopamine and *m*-tyramine in Parkinson's disease (*8*), but identifying which microbes and enzymes are responsible for this two-step pathway remained to be discovered.

Maini Rekdal *et al.* used integrated interdisciplinary approaches to identify the L-dopa metabolic pathway. For the first step, L-dopa conversion to dopamine, there were some clues. Tyrosine decarboxylase, present predominantly within the gut bacteria of the *Enterococcus* and *Lactobacillus* genera (*9*–*11*), was reported to also decarboxylate dopamine (*9*, *11*). By screening established human microbiome datasets, Maini Rekdal *et al.* showed that most tyrosine decarboxylase (*tdc*) genes were present within these genera, particularly in *Enterococcus*, consistent with other recent findings (*12*). Moreover, *E. faecalis* species were found to be the most efficient strains at decarboxylating L-dopa. These strains were shown to prefer tyrosine, but decarboxylate L-dopa and tyrosine simultaneously and with maximal activity at low pH, similar to the acid environment in the small intestine where L-dopa is absorbed (*3*, *12*). Both studies also showed that *E. faecalis tdc* gene inactivation obliterates L-dopa decarboxylation (*3*, *12*).

Finding the gut microbial species and enzymes that dehydroxylate dopamine to *m*-tyramine (the second step) was more challenging because no bacterial species was known to have this capability. Maini Rekdal *et al.* used elegant multidisciplinary approaches and identified this activity to be specific to

[1]*School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.* [2]*APC Microbiome Ireland, University College Cork, Cork, Ireland.* [3]*Cork Neuroscience Centre, University College Cork, Cork, Ireland. Email: c.oneill@ucc.ie*

PHOTO: NASA/JPL-CALTECH/SPACE SCIENCE INSTITUTE



The gut bacterium *Enterococcus faecalis* (scanning electron micrograph shown) degrades the Parkinson's disease drug levodopa, perhaps limiting its availability in patients.

PHOTO: NANO CREATIVE/SCIENCE SOURCE

Eggerthella lenta species and close relatives in *Actinobacterial* genera. The authors characterized a distinct dopamine-inducible dehydroxylation enzyme that removes the para hydroxyl group of dopamine to produce *m*-tyramine. Although all *Eg. lenta* strains examined were dopamine-inducible, less than 50% dehydroxylated dopamine. Intriguingly, this was due to the presence of a single-nucleotide polymorphism (SNP) that resulted in an Arg^{506} to Ser substitution that inactivated the enzyme. This highlights an underappreciated role for SNPs in gut microbial function, and the importance of probing metabolic function, rather than assigning similar functions to gut bacterial strains.

What do these findings mean for people with Parkinson's disease? Maini Rekdal *et al.* show that the enzymes that degrade L-dopa occur in microbiomes from human stool samples and that L-dopa degradation occurs with considerable variation in people with and without Parkinson's disease. They show that L-dopa degradation can be predicted predominantly by microbial *tdc* gene expression and *E. faecalis* abundance in stool samples, and also by the presence of Arg^{506} in *Eg. lenta*. Furthermore, recent studies show that higher amounts of *tdc* in stool from Parkinson's disease patients correlate with increasing L-dopa dosage and disease duration (*12*).

Do human decarboxylase inhibitors, such as carbidopa, block the microbial enzymes? It has been shown in culture that this is not the case (*12*), and Maini Rekdal *et al.* confirmed this result in Parkinson's disease patient stool samples. To potentially counteract L-dopa degradation, Maini Rekdal *et al.* identified a small-molecule inhibitor, α-fluoromethyltyrosine (AFMT), that specifically inhibited microbial L-dopa decarboxylase activity, including in Parkinson's disease patient microbiotas. AFMT shows potential to block degradation of L-dopa by *E. faecalis* in mice. Moreover, blood plasma L-dopa concentrations were higher in rats when L-dopa and carbidopa were coadministered with non–L-dopa–metabolizing *E. faecalis* mutants compared with active strains (*12*). Together, these findings indicate that blocking bacterial L-dopa decarboxylase activity in patients with Parkinson's disease, with knowledge of the abundance of this enzyme in an individual, could personalize and potentially improve L-dopa therapies. Substantial information is available on the regulation of tyrosine decarboxylase and the *tdc* operon in *E. faecalis* (*10*, *13*), providing complementary avenues to understanding and modulating this enzyme's activity in Parkinson's disease.

The identification of specific gut bacteria and enzymes involved in drug metabolism, such as L-dopa, is a vital, but underexplored, research area. An increasing number of drugs have been identified to be metabolized by gut microbes, enabling their activation, inactivation, and sometimes toxicity. In turn, drugs can sculpt and regulate gut microbial composition (*1*, *2*). The findings that similar biochemical pathways occur in the human brain and gut microbes, as shown for L-dopa and dopamine, highlight an intricate metabolic cross-talk between gut microbiome and human brain metabolism. Deciphering the extent to which potential alterations in such gut microbiota–brain metabolism contributes to brain health, the development of Parkinson's disease, and Parkinson's disease therapeutics is an important research avenue. It is now crucial to further investigate whether measuring and inhibiting gut microbial tyrosine and L-dopa decarboxylase activity could predict and potentially improve L-dopa efficacy and treatment outcomes for people with Parkinson's disease. ■

REFERENCES AND NOTES

1. P. Spanogiannopoulos, E. N. Bess, R. N. Carmody, P. J. Turnbaugh, *Nat. Rev. Microbiol.* **14**, 273 (2016).
2. G. Clarke *et al.*, *Pharmacol. Rev.* **71**, 198 (2019).
3. V. Maini Rekdal *et al.*, *Science* **364**, eaau6323 (2019).
4. S. Fahn, *Mov. Disord.* **30**, 4 (2015).
5. R. F. Pfeiffer, *Curr. Treat. Options Neurol.* **20**, 54 (2018).
6. M. Lubomski *et al.*, *J. Neurol.* 10.1007/s00415-019-09320-1 (2019).
7. E. Sherwin, T. G. Dinan, J. F. Cryan, *Ann. N. Y. Acad. Sci.* **1420**, 5 (2018).
8. M. Sandler, B. L. Goodwin, C. R. Ruthven, D. B. Calne, *Nature* **229**, 414 (1971).
9. H. Blashcko, *Biochim. Biophys. Acta* **4**, 130 (1950).
10. N. Connil *et al.*, *Appl. Environ. Microbiol.* **68**, 3537 (2002).
11. H. Zhu *et al.*, *Sci. Rep.* **6**, 27779 (2016).
12. S. P. van Kessel *et al.*, *Nat. Commun.* **10**, 310 (2019).
13. M. Perez *et al.*, *Appl. Microbiol. Biotechnol.* **99**, 3547 (2015).

ACKNOWLEDGMENTS

Financial support was provided by Science Foundation Ireland (grant no. SFI/12/RC/2273).

10.1126/science.aax8937

CONDENSED MATTER PHYSICS

Catching structural transitions in liquids

Femtosecond pulse x-ray diffraction reveals structural changes between liquid states

By Feng Rao[1], Wei Zhang[2], Evan Ma[3]

Phase-change random-access memory is a key ingredient for nonvolatile memory and neuro-inspired computing devices (*1*, *2*). It exploits the ability of chalcogenide phase-change materials (PCMs) to rapidly switch between logic states "0" (glassy) and "1" (crystalline). The "0" state is reached through fast quenching of the melted PCM (cooling at a rate that retains the amorphous atomic arrangement in the liquiid). In this process, the diffusivity and viscosity change by ~17 orders of magnitude, the majority of which happens over a narrow temperature range (*1*). What is happening inside the cooling liquid to make this possible has remained elusive because characterization that requires nanoseconds of measurement time is intercepted by crystallization. On page 1062 of this issue, Zalden *et al.* (*3*) overcame this challenge by using femtosecond pulse x-ray diffraction and captured a liquid-liquid transition (LLT) in PCM liquids during superfast cooling.

The finding of Zalden *et al.* has implications for better understanding PCM-based device operation. Although the "0" state allows data storage for a very long time at room temperature (300 K), it only takes nanoseconds to transform into the "1" state, when the glassy solid is heated to above its glass transition temperature (T_g) (for example, up to 700 K). This echoes the drastic jump in the dynamics of the supercooled liquid PCM over a narrow temperature window. A plausible explanation is that there is a LLT with an underlying structural transformation.

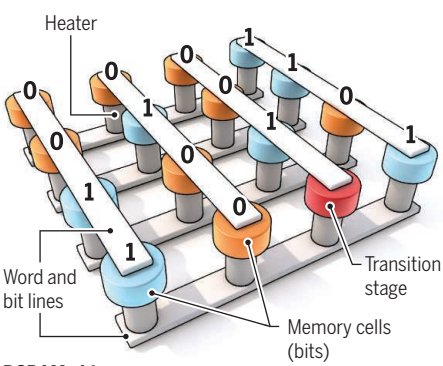
Zalden *et al.* used laser pulses (each 50 fs in duration) to heat a thin film of a PCM [either AgIn-doped Sb_2Te (AIST) or $Ge_{15}Sb_{85}$] above its melting temperature (T_m) and then monitored the atomic structure with femtosecond x-ray diffraction as the PCM was undergoing rapid quenching to avoid crystallization. The short time interval of the x-ray pulses ensured accurate tracking of atomic structure with minimal atomic motion. The radii ratio ($R = r_2/r_1$) of the first and second coordination shells [which refers to a central atom and the nearest atoms around it (first shell) and second nearest neighbors (second shell)] increased as the liquid quenched, displaying an inflection point in the slope of the R curve at ~660 and ~610 K for AIST and $Ge_{15}Sb_{85}$, respectively, which is well below the T_m of each PCM. The inflection point is referred to as T_{LL}, the temperature of a LLT. This structural transformation was attributed to oscillations in bond length known as Peierls distortion. This results in long and short bonds, starting from an octahedral-like atomic environment (see the figure). Zalden *et al.* noted a pre-peak in the x-ray scattering spectra when approaching T_{LL}, indicating periodic correlations in real space at ~6.0 Å. This corresponds to twice the typical bond length (~3.0 Å) in AIST and $Ge_{15}Sb_{85}$. Such periodic characteristics are consistent with long-short bonding pairs having an ~180° bonding angle. First-principles molecular dynamics simulations indicate that a pseudo–band gap opens up that accompanies the increasing Peierls distortion as electrons become more localized between atoms. Zalden *et al.* associate their observed pronounced increase in Peierls distortion to a LLT, akin to the case in phosphorous in which polymerization causes a LLT with changes in electronic properties (*4*).

For AIST, ultrafast differential scanning calorimetry (*5*) and laser-reflectivity measurements (*6*) suggested a fragile-to-strong liquid transition during cooling. For a "strong" liquid (such as silica), its viscosity follows Arrhenius behavior, with a nearly constant activation energy for viscous flow from the T_g to T_m. For a "fragile" liquid (such as o-terphenyl), the apparent activation energy is rather high (near the T_g) but decreases as the temperature increases (*7*). A fragile-to-strong liquid crossover could explain the high atomic mobility at elevated temperatures for fast crystallization and yet sluggish diffusion at room temperature for long-term data retention. Zalden *et al.* explicitly associated the LLT implicated by Peierls distortion with the fragile-to-strong liquid crossover.

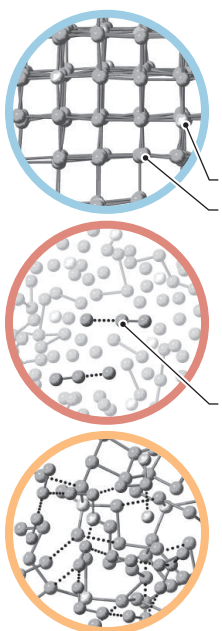
Challenging issues remain with regard to Peierls distortion as the sole structural origin of the fragile-to-strong liquid crossover. Although the LLT identified is expected to render an increment in the energy barrier for viscous flow, the actual escalation in activation energy should be assessed to compare with the known data (*5*, *6*). Also, it seems possible that Peierls distortion can appear at a temperature much higher than the temperature window found for fragile-to-strong

Phase-change memory

A memory device can use a phase-change material (such as $Ge_{15}Sb_{85}$) to switch quickly between logic states "1" (crystalline) and "0" (amorphous). These states are programmed through reversible phase transformation.



PCRAM chip
Memory cells (blue, orange) made of phase-change material are arranged in a crossbar phase-change random access memory (PCRAM) array.



GRAPHIC: C. BICKEL/*SCIENCE*

[1]*College of Materials Science and Engineering, Shenzhen University, Shenzhen 518060, China.* [2]*Center for Advancing Materials Performance from the Nanoscale, State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China.* [3]*Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: fengrao@szu.edu.cn; wzhang0@mail.xjtu.edu.cn; ema@jhu.edu*

crossover (*5*, *6*). If the amplitude of Peierls distortion increases gradually, there could be additional structural mechanisms that set off the much sharper rise of the apparent activation energy of viscosity-diffusivity at the "knee" observed in the Angell plot of some PCMs (*5*, *8*). For example, there could be structural heterogeneities at or beyond medium range that cause dynamical heterogeneities resembling glass-like behavior. The role of more flexible long bonds, and the possible cross-linking organization into an extended network, should be examined. First-principles molecular dynamics simulations have shown an *R* parameter less temperature dependent than the experimental finding. One possible reason is the rapid cooling rate and the inadequate equilibration time at each temperature step in simulations. This discrepancy is compounded by a lack of atomistics data on the unconventional dynamics in supercooled PCM liquids. An in-depth understanding may become manageable by using the interatomic potentials developed with machine-learning methods (*9*).

Femtosecond x-ray diffraction and absorption spectroscopies can be used to probe into liquid states in the already commercialized $Ge_2Sb_2Te_5$ (*8*) and the recently designed $Sc_{0.2}Sb_2Te_3$ (*2*) PCMs, in which nucleation, rather than growth (as in AIST and $Ge_{15}Sb_{85}$), plays a dominant role for nanoseconds operation. In general, LLTs appear to occur often, in systems from water to PCMs to metallic glass–forming liquids (*10*). However, LLTs are difficult to confirm experimentally—not to mention the lack of a systematic understanding as to which material systems exhibit them, and when, and why. In all cases, LLTs should feature a clear structural indicator. The approach of Zalden *et al.* opens a new avenue for interrogating the complex and swift structural changes in highly dynamic liquids. This should ultimately aid in optimizing the performance of PCM-based devices. ■

REFERENCES AND NOTES

1. W. Zhang *et al.*, *Nat. Rev. Mater.* **4**, 150 (2019).
2. F. Rao *et al.*, *Science* **358**, 1423 (2017).
3. P. Zalden *et al.*, *Science* **364**, 1062 (2019).
4. Y. Katayama, *J. Non-Cryst. Solids* **312-314**, 8 (2002).
5. J. Orava *et al.*, *Adv. Funct. Mater.* **25**, 4851 (2015).
6. M. Salinga *et al.*, *Nat. Commun.* **4**, 2371 (2013).
7. S. Wei *et al.*, *Acta Mater.* **129**, 259 (2017).
8. J. Orava *et al.*, *Nat. Mater.* **11**, 279 (2012).
9. F. C. Mocanu *et al.*, *J. Phys. Chem. B* **122**, 8998 (2018).
10. C. P. Royall, S. R. Williams, *Phys. Rep.* **560**, 1 (2015).

ACKNOWLEDGMENTS

F.R. thanks the National Natural Science Foundation of China (61622408), the Major Provincial Basic Research Program of Guangdong (2017KZDXM070), and the Science and Technology Foundation of Shenzhen (JCYJ20180507182248605, JCYJ20170302150053136). W. Z. thanks the National Natural Science Foundation of China (61774123). E. M. is supported at Johns Hopkins University by the U.S. Department of Energy, Basic Energy Sciences, Division of Materials Science and Engineering, under grant DE-FG02-16ER46056.

10.1126/science.aax6333

GRAPHIC: N. DESAI/*SCIENCE*

MEMBRANES

Scaling up nanoporous graphene membranes

Practical desalination membranes with nanoporous graphene will need new morphologies

By **Baoxia Mi**

Suitably engineered graphene-based materials could potentially be used to make water-separation membranes for applications such as desalination. Graphene-based materials with water-permeable pores can be made by creating either nanopores in graphene monolayers (see the figure, top) (*1*, *2*) or two-dimensional (2D) channels that form between nanosheets of graphene oxide (GO) (see the figure, middle) (*3*). Both approaches face challenges for scaling to practical membrane sizes on the meter scale. The former requires creating a high density of subnanometer pores (*4*) on a defect-free monolayer graphene sheet that has high out-of-plane mechanical strength (*5*), and materials meeting all these requirements have largely been limited to micrometer-scale lateral dimensions. On page 1057 of this issue, Yang *et al.* (*6*) report production of nanoporous graphene on the centimeter scale that can reject between 85 and 97% of the salt from saltwater.

The difference in difficulty of scaling up the two approaches can be best appreciated by noting that unlike the challenging procedures for nanoporous graphene synthesis, centimeter-scale stacked GO membranes were readily made for initial bench-scale studies. Synthesis methods for GO have existed for decades, and it can be mass-produced through chemical oxidization and ultrasonic exfoliation of graphite (*7*). Thus, GO has proven less complicated and more economical to scale up.

The hydroxyl, carboxyl, and epoxide functional groups on GO can be chemically modified to help control the size of the 2D channel (the interlayer spacing) and produce membranes with different separation capabilities (*3*, *8*). However, these functional groups also create the major problem with the stacked GO membrane—swelling in aqueous solutions (*9*)—which often results in low salt rejection (fraction of salt removed), typically in the range from 30 to 80%. Physical confinement—for example, with a polymer—can increase salt rejection to 97% (*10*). However, this strategy is feasible for centimeter-scale GO membranes, but may be more complicated to achieve in large-scale applications.

Unlike GO membranes, nanoporous monolayer graphene membranes are not inherently tolerant to defects. As the graphene membrane area increases, it becomes more prone to defects that can create openings

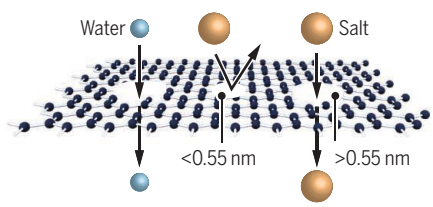
Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA. Email: mib@berkeley.edu

Toward practical graphene membranes

Combining two competing approaches may allow fabrication of meter-scale desalination membranes with two-dimensional materials.

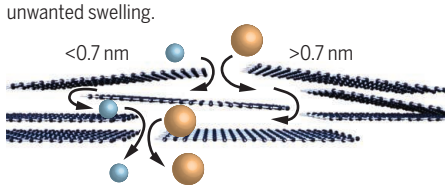
Through the pores

Salt molecules (both cations and anions) are rejected by nanopores in monolayer graphene that pass water. Yang *et al.* report centimeter-scale membranes of this type with low defect densities and high salt rejection.



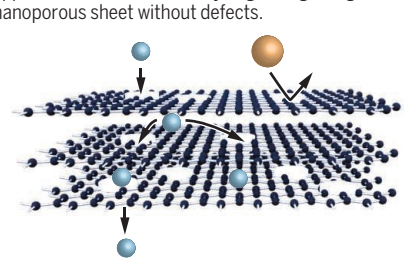
Around the stacks

In graphene oxide membranes, separation is achieved by water moving through channels between nanosheets faster than salt. Such membranes are prone to unwanted swelling.



Through and around

Meter-scale nanoporous graphene membranes could potentially be made by stacking smaller sheets. This approach avoids the difficulty of growing a large nanoporous sheet without defects.



larger than the required nanopore dimension. These defects cause a deterioration in separation capability (essentially leakage), and their presence is why most previous studies could only demonstrate reasonable desalination performance with nanoporous graphene membranes on a micrometer scale. Yang *et al.* overcome this problem by using chemical vapor deposition (CVD) to produce centimeter-scale graphene with high mechanical strength. Subsequent oxygen plasma etching created relatively uniform nanopores (diameters in the range of 0.3 to 1.2 nm), which are likely decorated by oxygenated functional groups. This study brings the membrane area to the centimeter scale, which is large enough to be tested in a bench-scale membrane system.

Nonetheless, there are many challenges to extending this approach to create the meter-scale membranes needed for the membrane modules used in the desalination industry. Although the pore-size distribution is impressively narrow, it needs to be further narrowed to improve the salt rejection to >99% for seawater desalination. Atomistic simulations indicate that absolute salt rejection can be achieved when the pore diameter is <0.55 nm (*2*). Also, the synthesis technique faces substantial challenges to increase the membrane area to the meter scale because larger graphene sheets almost inevitably have larger-than-desired pores and more defects.

Scaling up the membrane area while finely controlling the functional groups on nanopore edges poses another challenge. The presence of certain functional groups (such as hydroxyl groups) decreases salt rejection, likely because of the formation of hydrogen bonds with ions that lower the free-energy barrier for ions to pass through the pores. Other functional groups (such as charged carboxylate groups) can increase salt rejection through charge effects (*2*).

A promising strategy for addressing the above scaling-up problems would be to make stacked nanoporous graphene (SNG) membranes from nanosheets in a manner similar to the fabrication of GO membranes. As a few layers of nanoporous graphene sheets are stacked, the functional groups on the edge of nanopores will work as spacers to keep interlayer channels open. Water will not only pass through the in-plane nanopores of each layer but also permeate through the interlayer channels (see the figure, bottom). The SNG membrane should have better scalability and performance because it combines several advantages of the nanoporous monolayer graphene membrane and the stacked GO membrane. It should have higher selectivity because it integrates multiple rejection mechanisms, including the gating effect (size exclusion and charge repulsion) of in-plane nanopores, the gating effect at the entrance of the 2D channel between neighboring porous graphene sheets, and the hindered diffusion of ions and molecules caused by the narrowness of the 2D channel. Stacking multiple layers of nanoporous graphene sheets would allow the defects and large pores in each individual sheet to be shielded by the neighboring sheets and minimize the ability of ions and molecules to bypass the membrane.

The SNG membrane would have a slightly lower water flux than the nanoporous monolayer graphene membrane (*2*) because of multilayer stacking, but would maintain a higher flux than the stacked GO membrane (*11*) because the stacking of porous graphene sheets would result in a less tortuous transport path than that in the stacked GO. The SNG membrane would also be less subjected to swelling than stacked GO membranes and have a less compromised nanostructure and performance. It should be much easier and cheaper to scale up SNG membranes to a meter-scale area by stacking centimeter-scale or even smaller nanoporous graphene sheets by using widely available coating techniques such as layer-by-layer assembly, vacuum filtration, and electrospray, rather than by synthesizing individual meter-scale nanoporous graphene sheets.

The challenges of scaling up 2D materials for high-efficiency membrane-based desalination call for in-depth exploration of creative ideas. The large family of related 2D materials could be used to overcome some of the technical problems facing the graphene-based membrane technology. For example, a 2D zeolite nanosheet with uniform pores of ~0.5 nm oriented normal to the plane could be an ideal material for water-salt separation. Molybdenum disulfide (MoS_2) has stronger van der Waals interaction between nanosheets and could be used to control the swelling of stacked graphene-based membranes (*12*) by forming a stacked porous graphene-MoS_2 composite structure. ■

REFERENCES AND NOTES

1. M. E. Suk, N. R. Aluru, *J. Phys. Chem. Lett.* **1**, 1590 (2010).
2. D. Cohen-Tanugi, J. C. Grossman, *Nano Lett.* **12**, 3602 (2012).
3. R. K. Joshi *et al.*, *Science* **343**, 752 (2014).
4. L. Wang *et al.*, *Nat. Nanotechnol.* **12**, 509 (2017).
5. P. R. Kidambi *et al.*, *ACS Appl. Mater. Interfaces* **10**, 10369 (2018).
6. Y. Yang *et al.*, *Science* **364**, 1057 (2019).
7. D. R. Dreyer *et al.*, *Chem. Soc. Rev.* **39**, 228 (2010).
8. B. Mi, *Science* **343**, 740 (2014).
9. S. Zheng *et al.*, *ACS Nano* **11**, 6440 (2017).
10. J. Abraham *et al.*, *Nat. Nanotechnol.* **12**, 546 (2017).
11. R. R. Nair *et al.*, *Science* **335**, 442 (2012).
12. Z. Wang *et al.*, *Nano Lett.* **17**, 7289 (2017).

ACKNOWLEDGMENTS

The author is supported by NSF award nos. CBET-1565452 and CBET-1706059.

10.1126/science.aax3103

MICROBIOLOGY

A marine chemical defense partnership

A flavobacterium protects a green alga and sea slug from predation

By **Samantha Mascuch** *and* **Julia Kubanek**

A multitude of biotic and abiotic factors influence marine ecosystem function. On page 1056 of this issue, Zan *et al.* (*1*) report the molecular basis of a tripartite marine symbiosis in which bacterial production of a chemical defense molecule contributes to coral reef community structure. This example serves as an important reminder of the vital "invisible" contributions of microbes to coral reef ecology. Ultimately, microbial determinants of ecosystem structure may impact biodiversity with implications for coral reef stability in the face of environmental perturbations (*2*).

The biodiversity of the oceans is evident in coral reefs, kelp forests, and salt marshes, as well as in the swarming world of microscopic marine animals and single-celled algae found in a drop of seawater. Many benefits of marine biodiversity to human society are widely recognized, including the provision of food, oxygen (through photosynthesis), and new medicines. Increasingly, the complex interactions that occur between members of ecological communities are also being explored and appreciated for their contributions to biodiversity.

Symbiotic bacteria can live on marine plants and animals or, in the case of endosymbionts, inside their cells and tissues and contribute to biological complexity, directly supporting their hosts' survival. Benefits to marine hosts from symbiotic bacteria include nutrition (provision of carbon-based energy sources through photosynthesis or chemosynthesis), acquisition of new sensory modalities (e.g., magnetoreception) (*3*), light production (i.e., bioluminescence), and protection against pathogens or

School of Biological Sciences and School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA. Email: julia.kubanek@biosci.gatech.edu

predators. Bacteria that protect their hosts against natural enemies typically do so by producing specialized chemicals, called natural products. These are of interest not only because of their ecological functions but also for their potential medicinal value. These bacteria and the chemicals they produce are an additional layer of diversity that supports ecosystem function.

Interactions between hosts and their associated microorganisms (microbiota) can take many forms. Some hosts harbor complex microbiota consisting of numerous microorganisms, as observed in the human gut and in many, although not all, marine sponges. In other hosts, including many insect species and the green alga *Bryopsis* sp. studied by Zan *et al.*, microbiotas consist of fewer species. Paradoxically, these simpler microbiotas represent an especially promising source of chemical diversity in cases where bacteria have become specialized, evolving in tandem and becoming codependent with their hosts. These host-adapted bacteria synthesize structurally distinct natural products, integral to the symbiotic relationship, which may not be found elsewhere.

In the early days of marine natural products research, it was predicted that many medicinally valuable chemicals discovered from marine organisms, such as sponges and sea squirts, were actually of microbial origin, because of their structural resemblance to known bacterial products. Testing this hypothesis, however, has been challenging because laboratory cultivation of symbiotic microbes is often unsuccessful. With the advent of advanced metagenomic and bioinformatic technologies, identification of the genetic source of marine natural products became possible. Now, examples abound of structurally diverse molecules associated with marine animals that are definitively known to be produced by bacteria (*4*–*12*).

Zan *et al.* employed an integrated metagenomic, analytical chemistry, and bioinformatic approach to uncover the origins of the natural product kahalalide F. This cytotoxic chemical is of interest both for its defensive ecological role on coral reefs and for its medicinal potential; it was previously evaluated in phase II clinical trials as a possible treatment for melanoma, hepatocellular carcinoma, non–small cell lung cancer, and psoriasis (*13*). The authors discovered that kahalalide F is produced by a previously unknown Flavobacteriaceae endosymbiont (the uncultured "*Candidatus* Endobryopsis kahalalidefaciens") of the green alga *Bryopsis* sp.

On tropical Pacific coral reefs, kahalalide F protects *Bryopsis* sp. from being eaten by herbivorous fishes (*14*, *15*) (see the figure). Despite the general effectiveness of kahalalide F as a feeding deterrent, a specialist predator of *Bryopsis* sp., the sea slug *Elysia rufescens,* consumes the alga and sequesters the defensive chemical—but does not maintain its bacterial producer—and thereby acquires its own protection against predation (*1*, *14*, *15*). In this way, a single molecule moves through the ecosystem, from bacterium to alga to slug, increasing the survival of multiple organisms and contributing to the higher-order structure of the reefscape.

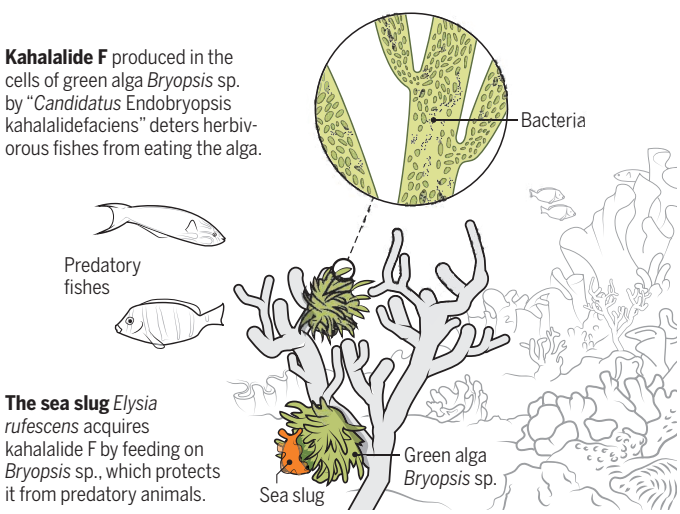
"*Ca.* E. kahalalidefaciens" resides inside algal cells and relies on *Bryopsis* sp. for the metabolic building blocks necessary for kahalalide production. In return, "*Ca.* E. kahalalidefaciens" devotes a substantial portion of its transcriptional activity to the manufacture of these chemicals, which no other organism is known to make. Analysis of the genome and transcriptome of "*Ca.* E. kahalalidefaciens" by Zan *et al.* provided insight into the origins of chemical diversity within the kahalalide structural class as well as clues about the potential evolutionary trajectories of distinct biosynthetic gene clusters, each of which produces a different kahalalide structure.

Multiple biosynthetic pathways for the production of structural analogs of kahalalide F have evolved in "*Ca.* E. kahalalidefaciens" through genomic duplication, divergence, and interpathway exchange. Zan *et al.* propose that differential expression of these biosynthetic pathways offers a window into the selection processes acting on them. The authors therefore inferred a hierarchy of likely ecological importance among related kahalalides, whereby valuable pathways are highly expressed and pathways with lower expression levels are predicted to be less important and, in several cases, in the process of being evolutionarily lost. Accordingly, the biosynthetic pathway for the most potent chemical defense, kahalalide F, was found to be the most highly expressed kahalalide pathway. The findings permit a new hypothesis to be considered for this ecological system: That bacterial molecular evolution involves continual introduction of chemical diversity and subsequent retention of biosynthetic pathways encoding molecules that enhance algal (and bacterial) adaptation and survival.

Zan *et al.* have demonstrated that the same symbiotic associations that have been characterized in marine invertebrates, whereby bacteria produce bioactive chemicals, also occur in a eukaryotic marine alga. The reduced genome, reliance on host primary metabolism, investment in secondary metabolism (to produce kahalalides), and localization in algal cells support the role of the flavobacterium "*Ca.* E. kahalalidefaciens" as a *Bryopsis* sp. endosymbiont, even in the absence of its cultivation. Ultimately, the discovery by Zan *et al.* reinforces the foundational roles of microbes in marine ecosystems. ■

Microbial deterrent promotes survival

In the Hawaiian coral reef system, the marine bacterium "*Candidatus* Endobryopsis kahalalidefaciens" produces the chemical defense molecule kahalalide F, which promotes survival of algae and sea slugs.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. J. Zan *et al.*, *Science* **364**, eaaw6732 (2019).
2. K. S. McCann, *Nature* **405**, 228 (2000).
3. C. L. Monteil *et al.*, *Nat. Microbiol.* 10.1038/s41564-019-0432-7 (2019).
4. M. M. Schofield, S. Jain, D. Porat, G. J. Dick, D. H. Sherman, *Environ. Microbiol.* **17**, 3964 (2015).
5. S. Sudek *et al.*, *J. Nat. Prod.* **70**, 67 (2007).
6. E. W. Schmidt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7315 (2005).
7. R. Ueoka *et al.*, *Nat. Chem. Biol.* **11**, 705 (2015).
8. S. M. Gromek *et al.*, *Front. Microbiol.* **7**, 1342 (2016).
9. M. S. Gil-Turnes, M. E. Hay, W. Fenical, *Science* **246**, 116 (1989).
10. P. M. Flatt *et al.*, *Mar. Biol.* **147**, 761 (2005).
11. J. Piel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16222 (2004).
12. Z. Lin *et al.*, *Chem. Biol.* **20**, 73 (2013).
13. J. Gao, M. T. Hamann, *Chem. Rev.* **111**, 3208 (2011).
14. M. T. Hamann, P. J. Scheuer, *J. Am. Chem. Soc.* **115**, 5825 (1993).
15. M. A. Becerro, G. Goetz, V. J. Paul, P. J. Scheuer, *J. Chem. Ecol.* **27**, 2287 (2001).

10.1126/science.aax8964

POLICY FORUM

INTELLECTUAL PROPERTY

Science fiction: Fictitious experiments in patents

Prophetic examples may unnecessarily distort understanding

By **Janet Freilich**[1] *and* **Lisa Larrimore Ouellette**[2]

Although it may surprise scientists, one can receive a patent in many jurisdictions without implementing an invention in practice and demonstrating that it works as expected. Instead, inventors applying for patents are allowed to include predicted experimental methods and results, known as prophetic examples, as long as the examples are not written in the past tense (*1*–*3*). Allowing untested inventions to be patented may encourage earlier disclosures about new ideas and provide earlier certainty regarding legal rights—which may help small firms acquire financing to bring their ideas to market. Yet granting patents too early may also discourage researchers from doing the work to bring ideas to fruition (*4*, *5*). Even if allowing untested inventions to be patented is desirable, we think prophetic examples deserve closer scrutiny, and clearer labeling, because of the likelihood that they are unnecessarily confusing—particularly to scientists, many of whom read patents but are unlikely to appreciate that not all the claims are based on actual data.

The U.S. Patent and Trademark Office (USPTO) formally recognized prophetic examples in 1981, but the practice is considerably older. A patent application need only contain sufficient information that a skilled researcher in the field would recognize as credibly demonstrating how to make and use the invention (*6*). Prophetic examples are one way to help satisfy this legal standard for inventions that have not yet been demonstrated to work (*2*). Although prophetic examples that are close variations on actual experiments are preferable, many prophetic examples appear to be entirely hypothetical predictions. Preliminary research suggests that these examples are particularly prevalent in chemistry and biology; an estimated 17% of examples in U.S. patents in these fields are prophetic, and almost one-quarter of U.S. patents in these fields have at least one prophetic example—making prophetic examples a commonplace feature (for examples, see the box) (*7*).

Because of concerns about awarding patents to unproven inventions, prophetic examples are viewed with greater skepticism in Europe (*8*), Canada (*9*), Japan (*10*), and China (*11*). However, because patents with the same contents are often filed in multiple regions, prophetic examples originating in U.S. applications will often be present in applications filed in other jurisdictions. Further, because patent offices and examiners in those countries commonly read and cite patents from other jurisdictions, countries skeptical of prophetic examples still feel their effects.

PROPHETIC EXAMPLES MAY BE CONFUSING

Contrary to the assertions of some patent scholars that scientists never read patents, survey evidence shows that many researchers do look to the patent literature for general research, to browse information about cutting-edge technologies, and to learn how other researchers solved particular problems (*12*). Training on how to search patents is even provided in some undergraduate science classes (*12*). But the usefulness of patents as a source of technical information is diminished if scientific readers are unable to distinguish actual data from predicted results.

Although scientists read patents—and therefore also read prophetic examples—the verb tense rule that distinguishes these predicted results from actual data is unlikely to be familiar to the average scientist. Most patent drafters do not seek to intentionally mislead readers, but they are writing for a legal audience and using conventions that may be unclear to nonlegal readers. Prophetic examples are confusing because they mimic real experiments, particularly by including excessive detail—for example, age of the hypothetical patient ("a 46-year-old woman…")—and specific, numerical results ("blood pressure is reduced within 3 hours…"). Some preliminary work suggests that of 100 randomly selected patents with only prophetic examples—that is, no actual data—that were cited in a scientific article or book for a specific proposition, 99 were not cited in a way that made clear that the cited information was prophetic (*7*). To the contrary, these prophetic patents were cited with phrases such as "[d]ehydration reaction in gas phase has been carried out over solid acid catalysts" (*7*), suggesting that prophetic examples mislead scientist readers.

Prophetic examples may also be confusing to other readers who are unfamiliar with the tense rule, such as investors seeking to accurately evaluate complex technologies. Causing further misunderstanding, the subtlety of prophetic examples may literally be lost in translation for patent applications that must be translated into different languages because they are filed in international jurisdictions. To be sure, quantifying the cost of this confusion would be challenging, especially because most confused scientists, investors, and patent examiners are likely unaware of the problem. But given the lack of a corresponding benefit, there seems to be no reason to perpetuate the practice. Nothing in patent law requires early-stage ideas to be described in a way that might confuse these different audiences by mimicking factual experiments; prophetic examples could be signaled more clearly or avoided altogether.

WHY USE PROPHETIC EXPERIMENTS?

To explore whether benefits for patentees from prophetic examples can be obtained through less confusing patent-drafting methods, we interviewed professional patent prosecutors who write U.S. patents. As described in the supplementary materials, we identified prophetic examples as those written in the present or future tense. We then contacted a randomly selected sample of patent prosecutors in the fields of chemistry and biology who, in patent applications filed between 2011 and 2013, either never used prophetic examples or used prophetic examples in more than half of applications filed. We conducted 26 interviews, with a yield rate of 67%.

Prosecutors who use prophetic examples consistently explained that such examples make clear how an inventor expects an idea to work in scenarios for which there is not time or money to test before the desired patent filing deadline. Because patents can cover all variations of an invention described in enough detail for others to make and use without undue experimentation, prophetic examples with predicted

[1]*Fordham University School of Law, New York, NY, USA.*
[2]*Stanford Law School, Stanford University, Stanford, CA, USA.*
Email: jfreilich1@fordham.edu; ouellette@law.stanford.edu

results can extend the patent's legal protection. For example, if an inventor has made a particular protein in her laboratory but also believes that the protein will work similarly if certain amino acids are switched, a prosecutor can draft prophetic examples with the alternate sequences and a prediction of the expected outcome. These examples help the inventor obtain patent coverage beyond the specifics of what has been done in the laboratory, including to block competitors from working on similar technologies.

Interviewed prosecutors generally acknowledged the possibility that scientists reading prophetic examples might be unable to correctly interpret the verb tense rule, although they emphasized the legality of the practice and their duty to obtain the strongest possible patent for their clients. They also agreed, however, that an equally strong patent could be obtained with prophetic examples that were explicitly labeled as predictions that had not been carried out. Interviewees who do not use prophetic examples argued that there is no legal reason to present these predictions in the form of fictitious experiments with specific results rather than in more general terms; to the contrary, prophetic examples carry some legal risk, such as if the example turns out to be inoperative. Prosecutors were particularly wary of using prophetic examples in patent applications that would be filed internationally, given the greater skepticism of these examples in certain countries.

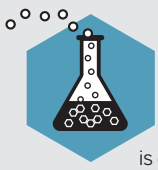
The only benefit to patentees that would be reduced by requiring greater clarity seems to be the benefit that comes from confusion. For example, several prosecutors suggested that prophetic examples could illustrate a technology's promise to potential investors, who might not be able to distinguish between prophetic examples and experiments actually conducted. This potential confusion was considered a benefit to patentees, but this benefit does not seem worth preserving.

MORE CLARITY, LESS CONFUSION

The benefits flowing from prophetic examples exist because some patent systems recognize and allow the use of hypothetical experiments and data. Within these legal systems, prosecutors, patent examiners, and courts can already identify prophetic examples through the tense rule, so requiring a more explicit distinction between prophetic and nonprophetic examples would have no legal impact; prophetic examples would continue to be recognized and rewarded as such, just with lower risk of confusion.

Patently Prophetic

The present tense used in patents for a chemical synthesis, a medical procedure, and a medical device suggests that the procedures likely had not actually been conducted at the time of filing a patent application.



U.S. Patent No. 3,931,205

2.5 g of 2-(5H-[1]benzopyrano[2,3-b]pyridin-7-yl)acrylic acid is dissolved in 20 ml of 0.5 N aqueous sodium hydroxide solution, and 1 g of Raney nickel is added. The solution is stirred in a hydrogen stream at ordinary pressure and temperature until absorption of 230 ml of hydrogen is attained. The Raney nickel is removed by filtration, and the filtrate is neutralized with hydrochloric acid. The resulting crystalline precipitate is filtered off, washed with water, and recrystallized from aqueous dioxane to give 1.8 g of 2-(5H-[1]benzopyrano[2,3-b]pyridin-7-yl) propionic acid melting at 183°–184°C.

U.S. Patent No. 6,869,610

A 46-year-old woman presents with pain localized at the deltoid region due to an arthritic condition. The muscle is not in spasm, nor does it exhibit a hypertonic condition. The patient is treated by a bolus injection of between about 50 units and 200 units of intramuscular botulinum toxin type A. Within 1 to 7 days after neurotoxin administration the patient's pain is substantially alleviated. The duration of significant pain alleviation is from about 2 to about 6 months.

U.S. Patent No. 7,291,497

Each patch [for drawing and sampling 0.1 ml of blood for vancomycin] consists of two parts. ... Micro-needles automatically draw small quantities of blood painlessly. A mechanical actuator inserts and withdraws the needle ... mak[ing] several measurements after the patch is applied. ... Needles are produced photolithographically in molds at [the Stanford Nanofabrication Facility]. ... Blood flows through the micro-needles into the blood reservoir. ...

The impact of clarifying prophetic examples would also be felt outside the legal systems that allow the practice. Scientists previously unable to distinguish or unaware of the distinction between prophetic and real experiments would gain more information and clarity. Investors using patents as a source of information about new technologies would find such information clearer and more useful. And international patent offices wrongly interpreting prophetic examples when tenses are lost in translation would be able to avoid such errors.

What should be done? A simple and effective solution is to require that prophetic examples in new patent applications be clearly labeled, perhaps with a heading such as "hypothetical experiment" or an introductory phrase such as "it is expected that these experiments would provide these results." In the United States, for example, this change could be implemented by the USPTO along with its other rules for patent formatting. The USPTO already requires that prophetic examples be labeled (by avoiding the past tense), so our proposal does not add a labeling requirement; it merely makes an existing requirement more effective. Further, patent drafters should be encouraged to be mindful of clarity and avoid potentially confusing phrases and details.

Just because some patents are not based on actual results does not mean they need to be confusing. Scientists regularly write grant applications in a way that makes clear what preliminary data they have already acquired and what the expected goal of the proposed project is. Perhaps this is an area in which the patent system could learn from the scientific community. ■

REFERENCES AND NOTES

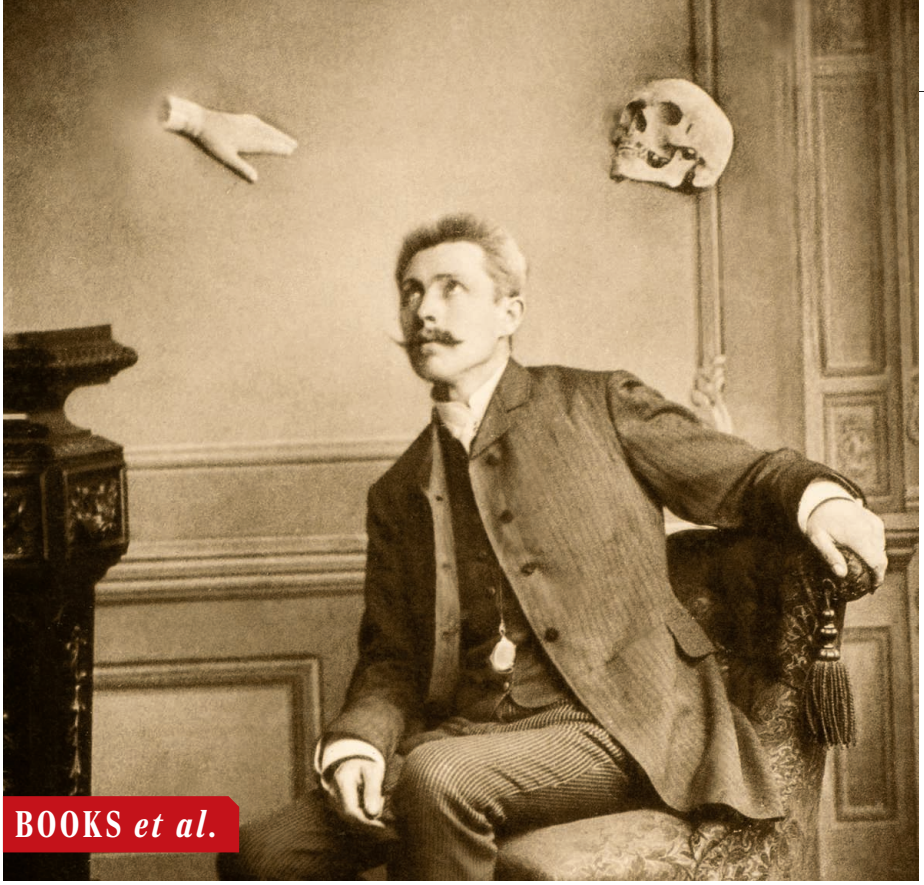
1. USPTO, *Manual of Patent Examining Procedure*, §§608.01(p), 2164.02 (2018).
2. R. P. Merges, J. F. Duffy, *Patent Law and Policy* (Carolina Academic Press, ed. 7, 2017).
3. T. Decision, 0792/00, Boards of Appeal of the European Patent Office (2002).
4. L. L. Ouellette, *Vanderbilt Law Rev.* **69**, 1825 (2016).
5. M. A. Lemley, *Boston Univ. Law Rev.* **96**, 1171 (2016).
6. M. Adelman *et al.*, *Global Issues in Patent Law* (Thomson Reuters, 2011), pp. 152–171.
7. J. Freilich, *SSRN* 10.2139/ssrn.3202493 (2018).
8. Decision T 0210/02, Boards of Appeal of the European Patent Office (2004).
9. Canadian Intellectual Property Office, *Manual of Patent Office Practice*, §12.04.03a (2017).
10. Japan Patent Attorneys Association, *Questions and Answers Regarding Japanese Patent Practice* (2012), Q77.
11. State Intellectual Property Office of the People's Republic of China, *Guidelines for Patent Examination* (2010).
12. L. L. Ouellette, *Nat. Biotechnol.* **35**, 421 (2017).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1036/suppl/DC1

10.1126/science.aax0748

GRAPHICS: A. KITTERMAN/*SCIENCE*



BOOKS *et al.*

A floating hand appears to transfix magician Jacoby-Harms, an effect created with double exposures.

PHOTO: F. A. DAHLSTRÖM/*THE SPECTACLE OF ILLUSION* (PUBLISHED BY D.A.P.)/COURTESY OF THE WELLCOME LIBRARY, LONDON

PSYCHOLOGY

Mind tricks

Magic and mysticism reveal cognitive shortcuts with implications beyond entertainment

By **Clive Wilkins** *and* **Nicola S. Clayton**

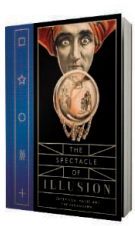
Published to coincide with the Wellcome Collection's exhibition "Smoke and Mirrors: The Psychology of Magic," *The Spectacle of Illusion* is a delightful and informative roller coaster that explores our fascination with magic, the paranormal, and the psychology of cognitive illusions. Author Matt Tompkins—who is both a psychologist and a magician—makes a detailed analysis of magicians and the responses they elicit in unsuspecting spectators, with some intriguing thoughts about magic, its close alignment with the paranormal, and psychologists' enduring fascination with psychical research.

Jean Eugène Robert-Houdin, the father of modern magic, famously argued that the magician is not a performer of juggling tricks but an actor playing the role of someone with supernatural powers. The French use "magicien" to refer to the supernatural component and "prestidigateur" to describe a magician with nimble fingers. This distinction obfuscates the fact that what is perceived to be magic takes place not in the hands of the magician but in the mind of the spectator. James Randi astutely summarizes, "As a magician, I was able to see two things very clearly: a) how people can be fooled, and b) how they fooled themselves, and the second is far more important than the first."

The Spectacle of Illusion is laid out in a series of acts that explore the transformation of magic from early mesmeric and spiritualist phenomena and master magicians to psychical research and parapsychological investigators. Although a short chapter, the book's culmination—"Act five: The psychology of illusion"—is the highlight, providing a coherent analysis of how the mind is tricked when asked to see, to reason, and to remember. Here, Tompkins delves into all the attendant ramifications of cognitive roadblocks and biases and explores our propensity to believe in more than we perceive.

From a scientific perspective, both magic effects and other violations of expectancy activate comparable regions of the prefrontal cortex. As such, magic has the potential to explore the subjective experience of what we see, when we remember, and how we reason.

The reviewers are at the Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK. Email: cw567@cam.ac.uk



The Spectacle of Illusion
Matthew L. Tompkins
D.A.P., 2019. 224 pp.

Magicians, we learn, discovered a number of key features of cognition long before they were studied by psychologists. Richard Hodgson and Samuel John Davey, for example, would not have considered themselves to be experimental psychologists, yet their 1887 paper "The Possibilities of malobservation and lapse of memory from a practical point of view" (*1*) described what later became known as the "reconstructive nature of memory" (*2*). Likewise, the "princess card trick"—first described by Thomas Nelson Downs in 1909 (*3*) and invented by magician Henry Hardin—relies on a cognitive blip, known as "change blindness," in which spectators fail to detect changes to a scene when the changes are accompanied by visual disruption. This phenomenon was not examined in the psychology literature until 1997 (*4*).

Psychologists have also contributed to our understanding of magic. Here, Tompkins turns to the work of Alfred Binet on the psychology of prestidigitation, sometimes known as sleight of hand (*5*). Although best known for his work on IQ tests, in the paper to which Tompkins alludes, Binet demonstrates that photography can effectively destroy the psychological power of a magic effect by stripping away the artifice and other elements that create some cognitive illusions. For this reason, close-up magicians often request that audiences not film their performances.

The Spectacle of Illusion elegantly interweaves the expertise and perspectives of both psychologists and magicians to offer a powerful account of how metacognitive paradoxes work through illusions of omission and commission. There is still much to unpack about the nature of human cognitive processes but no reason to believe we won't do so. As Tompkins writes, "No psychologist can claim that science has been able to fully describe how a human mind can construct conscious experience. But just because something must remain unexplained…does not mean it is unexplainable." ■

REFERENCES AND NOTES

1. S. J. Davey, *Proc. Soc. Psych. Res.* **4**, 381 (1887).
2. F. C. Bartlett, *Remembering a Study in Experimental and Social Psychology* (Cambridge Univ. Press, 1932).
3. T. Nelson Downs, *The Art of Magic* (Downs-Edwards, 1909).
4. R. A. Rensink *et al.*, *Psychol. Sci.* **8**, 368 (1997).
5. A. Binet, *The Psychology of Prestidigitation, Annual Report of the Board of Regents of the Smithsonian Institution, USA*, M. Nichols, transl. (Smithsonian Institution, 1894).

10.1126/science.aax8553

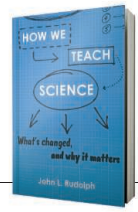
EDUCATION

Much ado about method

Past efforts to reshape American science education offer lessons for future reformers

By **Christopher J. Phillips**

How We Teach Science
What's Changed, and Why It Matters
John L. Rudolph
Harvard University Press, 2019. 316 pp.



Compared with reading, writing, and arithmetic, science is a relative newcomer to the primary and secondary school curriculum, emerging only in the late 19th century. Nevertheless, proponents of the subject have established it as central to what an educated person needs to know, not least because of the promise of good jobs in scientific fields.

Even if nearly every school district in the United States now treats it as a required subject, there has been almost no consensus on what science classes should entail. Some have claimed that the subject should be taught as a single methodology, presenting the scientific method as a fixed number of discrete steps. Others emphasize it as a disparate collection of techniques—some inductive, others deductive, and divided up into specific disciplinary approaches. Teachers have disagreed on whether it is best taught through textbooks or laboratory experiments, as a set of conclusions and facts, or as a mode of inquiry. Most contemporary scientists would agree that there is no single method for doing science, but beyond that, there has not been much to agree on.

John Rudolph's *How We Teach Science* traces the different strands and debates in American science education over the past 130 years. He focuses the book on influential thinkers and reformers, from Edwin Hall, John Dewey, and G. Stanley Hall in the early years to James Bryant Conant, Joseph J. Schwab, and F. James Rutherford in the latter.

Whatever the differences among its protagonists, *How We Teach Science* reveals a set of interlocutors conducting what was essentially a single conversation about science education for well over a century. When Rutherford set out to reform science education in the 1980s, for example, he did so having already worked on Harvard's Project Physics textbooks in the 1960s and citing the Conant-era *General Education in a Free Society* from 1945 as a primary influence. In turn, Conant wrote extensively about how to "correct" Dewey and Hall's earlier ideas of the nature of science.

There is also continuity in another sense. Rudolph comes down harshly on the fate of science education reform, arguing that every reform effort has essentially failed. In doing so, he echoes the National Science Foundation's inquiry in the 1970s, which concluded that teachers were expensive to retrain and consistently resistant to new pedagogical methods; that pupils continued to use textbooks slavishly, despite reformers' efforts; that "methods" were seen as fixed steps rather than as guides to arriving at reliable conclusions, however they were presented; and that disciplines were repeatedly reduced to sets of facts and truths rather than modes of systematic inquiry (*1*).

Two teens work together to assemble a science fair project in 1960.

Rudolph's book centers on a key shift that took place in the years after World War II, when the teaching of science was transformed from a practice that was thought of as being good for students to one that was thought of as being good for science itself. Science class was no longer a space for transforming hearts and minds but one for mass-producing researchers, laboratory workers, and theorists.

This is historically accurate, although it is dependent on an American, post–Civil War framing. (Rudolph dodges the claim made by Michael S. Teitelbaum and others that warnings of ongoing scientific shortages were used as rhetorical cover for overproducing scientists in order to lower labor costs.) Expanding the picture would reveal much older tensions as well as different contexts in which science might have been taught as either transformative of the self, useful in its applications, or relevant for public engagement. Given the book's emphasis on American schools, Rudolph's choice is reasonable, but it fails to give the reader a broader sense of why learning science has mattered across different countries and eras.

Rudolph also avoids the intersection of pedagogy and participation. Women, for example, have long been explicit targets for science education, whether botany in the 19th century or computing in the 20th, even though that has never translated into equality among the ranks of professional scientists.

Consistent with his title, Rudolph is instead laser focused on the "how" of the science classroom—how its practice varies across time, how its meaning is debated by reformers, and how its role in education shifts as schools themselves change. And for that discussion there is no better guide. Rudolph, a professor at University of Wisconsin–Madison as well as a former science teacher and editor of *Science Education*, has previously written definitive accounts of both Sputnik-era and Progressive-era science reforms.

How We Teach Science may end on a pessimistic note, given the litany of past failures to reform science education, but it isn't a fatalistic book. After all, the task at hand—to develop a scientifically informed public—is as urgent and important in our age of autonomous vehicles and CRISPR as it was in the industrial and atomic ages. ■

The reviewer is at the Department of History, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: cjp1@cmu.edu

REFERENCES AND NOTES

1. S. L. Helgeson *et al.*, *The Status of Pre-College Science, Mathematics, and Social Science Education: 1955-1975* vol. I, science education (U.S. Government Printing Office, 1977).

10.1126/science.aax2483

LETTERS

Edited by **Jennifer Sills**

Inclusive chimpanzee conservation

In their Report "Human impact erodes chimpanzee behavioral diversity" (29 March, p. 1453), H. S. Kühl *et al.* find that chimpanzees in modified landscapes show low behavioral diversity and propose the establishment of "chimpanzee cultural heritage sites" to safeguard behavioral variation. We are concerned that their conclusion propagates a view that some populations are not worth conserving.

At sites with highest modification, such as agricultural landscapes where people and chimpanzees share spaces entirely, human activities are driving chimpanzee behavioral flexibility and diversification, including novel behaviors (*1*, *2*). However, extinction risk in these areas is higher as a result of population isolation and anthropogenic-driven mortality (*3*). Conservationists also have incomplete data; not all chimpanzee behaviors in the wild are demonstrably socially learned and transmitted—and therefore cultural.

We agree that it is important to protect the behavioral diversity of culturally rich wildlife (*4*), but such populations should not always be given blanket priority over those living in closer contact with humans. We must ensure that apes living across the anthropogenic continuum are given a fighting chance.

Kimberley J. Hockings[1]* and Matthew R. McLennan[1,2]
[1]Centre for Ecology and Conservation, University of Exeter, Cornwall, UK. [2]The Bulindi Chimpanzee and Community Project, Hoima, Uganda.
*Corresponding author.
E-mail: k.hockings@exeter.ac.uk

REFERENCES AND NOTES

1. K. J. Hockings *et al.*, *Trends Ecol. Evol.* **30**, 215 (2015).
2. S. Krief *et al.*, *PLOS One* **9**, e109925 (2014).
3. Arcus Foundation, *State of the Apes: Industrial Agriculture and Ape Conservation* (Cambridge University Press, 2015).
4. P. Brakes *et al.*, *Science* **363**, 1032 (2019).

10.1126/science.aax5543

Response

We agree with Hockings and McLennan that, ideally, all wild chimpanzee populations would be equally protected. We do not presume to suggest that culture should be the sole basis for chimpanzee conservation efforts. Rather, we advocate a more integrative approach to conservation in which behavioral diversity of populations is another variable that can be described and quantified and can ultimately contribute to the protection of wildlife (*1*). Cultural and behavioral diversity should be prioritized alongside population size, genetic diversity, and demographic viability.

The proximity of chimpanzee populations to human activity is just one factor in conservation priorities.

Hockings and McLennan are correct that chimpanzees exhibit remarkable behavioral flexibility when living in human-dominated landscapes. To take advantage of new foraging opportunities and increase their chance for survival in fragmented and degraded habitats, chimpanzees have incorporated agricultural crops into their diet (*2*) and nest construction (*3*), as well as adapted their activity patterns to be more nocturnal (*4*). However, such behaviors represent evolutionarily novel traits compared with the behaviors we investigated, which have previously demonstrated cultural (*5*, *6*) or population variation [e.g., (*7*–*9*)]. Of course, larger comparative datasets that include chimpanzees living in human-dominated landscapes will make possible the study of fundamental mechanisms of behavioral and cultural diversification.

We agree with Hockings and McLennan that it is difficult to know the true cultural repertoire of all chimpanzee communities and that, when faced with limited information, we should err on the side of caution

PHOTO: © SHAH ANUP/ANIMALS ANIMALS

when considering conservation actions. However, if we want to protect chimpanzee cultural heritage, which is to protect historically significant traits originating from the past, then we must protect populations in pristine, less-disturbed habitats.

Hjalmar S. Kühl[1,2]*, Christophe Boesch[1,3], Lars Kulik[1], Fabian Haas[1], Mimi Arandjelovic[1], Paula Dieguez[1], Gaëlle Bocksberger[1], Anthony Agbor[1], Samuel Angedakin[1], Emmanuel Ayuk Ayimisin[1], Mattia Bessone[1], Gregory Brazzola[1], Rebecca Chancellor[4,5], Heather Cohen[1], Charlotte Coupland[1], Emmanuel Danquah[6], Tobias Deschner[1], Dervla Dowd[3], Annemarie Goedmakers[7], Anne-Céline Granjon[1], Josephine Head[1], Daniela Hedwig[8,9], Veerle Hermans[10], Sorrel Jones[1,11,12], Jessica Junker[1], Kevin E. Langergraber[13], Juan Lapuente[1], Kevin Lee[1,13], Manuel Llana[14], Sergio Marrocoli[1], Rumen Martin[1], Maureen S. McCarthy[1], Amelia C. Meier[1], David Morgan[15], Mizuki Murai[1], Emily Neil[1], Emma Normand[3], Lucy Jayne Ormsby[1], Liliana Pacheco[14], Alex Piel[16], Sebastien Regnaut[3], Aaron Rundus[5], Crickette Sanz[17], Fiona Stewart[16], Nikki Tagg[10], Virginie Vergnes[3], Adam Welsh[1], Erin G. Wessling[1,2], Jacob Willie[10,18], Roman M. Wittig[1,19], Yisa Ginath Yuh[1], Kyle Yurkiw[1], Ammie K. Kalan[1]*

[1]Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [2]German Centre for Integrative Biodiversity Research, Halle-Leipzig-Jena, 04103 Leipzig, Germany. [3]Wild Chimpanzee Foundation, 04103 Leipzig, Germany. [4]Department of Anthropology & Sociology, West Chester University, West Chester, PA 19382, USA. [5]Department of Psychology, West Chester University, West Chester, PA 19382, USA. [6]Department of Wildlife and Range Management, Faculty of Renewable Natural Resources, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [7]Chimbo Foundation, Amstel 49, 1011 PW Amsterdam, Netherlands. [8]The Aspinall Foundation, Port Lympne Wild Animal Park, Hythe, Kent, UK. [9]Elephant Listening Project, Bioacoustics Research Program, Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA. [10]Centre for Research and Conservation, Royal Zoological Society of Antwerp, B-2018 Antwerp, Belgium. [11]Royal Holloway, University of London Egham Hill, Egham Surrey TW20 0EX, UK. [12]Royal Society for the Protection of Birds, Sandy SG19 2DL, UK. [13]School of Human Evolution and Social Change & Institute of Human Origins, Arizona State University, Tempe, AZ 85281, USA. [14]Instituto Jane Goodall España, Station Biologique Fouta Djallon, Dindéfélo, Région de Kédougou, Senegal. [15]Lester E. Fisher Center for the Study and Conservation of Apes, Lincoln Park Zoo, Chicago, IL 60614, USA. [16]School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK. [17]Department of Anthropology, Washington University in Saint Louis, St. Louis, MO 63130, USA. [18]Terrestrial Ecology Unit, Ghent University, 9000 Ghent, Belgium. [19]Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, Abidjan 01, Cote d'Ivoire.
*Corresponding author. Email: kuehl@eva.mpg.de (H.S.K.); ammie_kalan@eva.mpg.de (A.K.K.)

REFERENCES AND NOTES

1. P. Brakes *et al.*, *Science* **363**, 1032 (2019).
2. K. J. Hockings *et al.*, *Trends Ecol. Evol.* **30**, 215 (2015).
3. M. S. McCarthy, J. D. Lester, C. B. Stanford, *Int. J. Primatol.* **38**, 321 (2017).
4. N. Tagg *et al.*, *Am. J. Phys. Anthropol.* **166**, 510 (2018).
5. A. Whiten *et al.*, *Behaviour* **138**, 1481 (2001).
6. A. Whiten, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7790 (2017).
7. H. S. Kühl *et al.*, *Sci. Rep.* **6**, 22219 (2016).
8. C. Boesch *et al.*, *Am. J. Primatol.* **79**, e22613 (2017).
9. J. Pruetz, P. Bertolani, *PaleoAnthropology* **2009**, 252 (2009).

10.1126/science.aax6339

Romanian carnivores at a crossroads

In October 2016, in an unexpected move, the Romanian government provisionally suspended the hunting of brown bears and wolves, shaking the decades-old wildlife management system of regulated hunting (*1*, *2*). This decision provided an opportunity to reset Romanian wildlife management and conservation and to implement science-based management. A year later, a new management system was implemented, allowing only the removal of problem animals at the epicenter of human-wildlife conflict, based on a case-by-case approval process (*3*). However, there is increasing pressure, fueled by a politically charged climate and a negative campaign focused on wildlife damage, to substantially reduce the abundance of large carnivores (*4*, *5*). The role of scientific evidence is still missing from the discussions about protecting Europe's largest large-carnivore populations, which is not an uncommon situation for wildlife management systems (*6*). Romania should seize this opportunity to enhance the scientific knowledge about carnivore ecology and human-carnivore coexistence (*7*) and enact science-based policy.

Past management of large carnivores in Romania has been based on biologically unrealistic data collected through a flawed methodology lacking scientific oversight. As in other countries, misguided strategies have highlighted data gaps and the need to infuse science into large-carnivore management (*8*–*10*). In recent years, Romania has been building up the science around carnivore population ecology and habitat conservation (*7*, *11*, *12*), yet nationwide data critical for management, such as population size or demographic structure, are still missing. The absence of reliable biological and ecological data makes it difficult to prioritize the best management and conservation measures to promote long-term viability while alleviating human-wildlife conflict.

Resource managers, hunters, environmental groups, and citizens should work together with Romanian and foreign academics to set up a long-term large-carnivore research program, which should be backed by the Romanian government and the European Union. Such a program could combine regional initiatives (*7*) with national-level monitoring (*4*). It could address long-term ecological and social science questions as well as immediate needs for solving human-carnivore conflict and enabling science-based policy. Transparent science accepted by all parties could be the catalyst for Romania to reconcile its own large-carnivore conservation strategies. We are at a crossroads. Romania can either serve as an example of human-carnivore coexistence in the European Union or become a cautionary tale of politics driving wildlife management and miss an opportunity to safeguard Europe's last wild frontier.

Viorel Popescu[1,2], Mihai Pop[2,3], Silviu Chiriac[4], Laurentiu Rozylowicz[2]

[1]Biological Sciences, Ohio University, Athens OH 45701, USA. [2]Centre for Environmental Research (CCMESI), University of Bucharest, Bucharest, Romania. [3]Association for Biodiversity Conservation, Focsani, Romania. [4]Vrancea Environmental Protection Agency, Focsani, Romania.
*Corresponding author.
Email: popescu@ohio.edu

REFERENCES AND NOTES

1. Romanian Government, "Emergency Ordinance 57" (2007); http://legislatie.just.ro/Public/DetaliiDocument/83289 [in Romanian].
2. V. Salvatori *et al.*, *Wild. Biol.* **8**, 3 (2002).
3. Romanian Ministry of Environment, "Order 1169" (2017); http://legislatie.just.ro/Public/DetaliiDocument/192825 [in Romanian].
4. Romanian Ministry of Environment, "Order 625" (2018); http://legislatie.just.ro/Public/DetaliiDocumentAfis/202150 [in Romanian].
5. Romanian Parliament, "Proposal 260" (2019); www.cdep.ro/pls/proiecte/upl_pck2015.proiect?cam=2&idp=17797 [in Romanian].
6. K. A. Artelle *et al.*, *Sci. Adv.* **4**, eaao0167 (2018).
7. T. Hartel *et al.*, *Cons Biol.* 10.1111/cobi.13334 (2019).
8. V. D. Popescu, K. A. Artelle, M. I. Pop, S. Manolache, L. Rozylowicz, *J. Appl. Ecol.* **53**, 1248 (2016).
9. K. A. Artelle *et al. PLOS One* **8**, e78041 (2013).
10. L. E. Ogden, *BioSci.* **67**, 691 (2017).
11. B. Cristescu, C. Domokos, K. J. Teichman, S. E. Nielsen, *PeerJ.* **7**, e6549 (2019).
12. M. I. Pop, R. Iosif, I. Miu, L. Rozylowicz, V. D. Popescu, *Anim. Conserv.* **21**, 352 (2018).

10.1126/science.aax6742

TECHNICAL COMMENT ABSTRACTS

Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds"

Martin Bulla, Jeroen Reneerkens, Emily L. Weiser, Aleksandr Sokolov, Audrey R. Taylor, Benoît Sittler, Brian J. McCaffery, Dan R. Ruthrauff, Daniel H.

Catlin, David C. Payer, David H. Ward, Diana V. Solovyeva, Eduardo S. A. Santos, Eldar Rakhimberdiev, Erica Nol, Eunbi Kwon, Glen S. Brown, Glenda D. Hevia, H. River Gates, James A. Johnson, Jan A. van Gils, Jannik Hansen, Jean-François Lamarre, Jennie Rausch, Jesse R. Conklin, Joe Liebezeit, Joël Bêty, Johannes Lang, José A. Alves, Juan Fernández-Elipe, Klaus Michael-Exo, Loïc Bollache, Marcelo Bertellotti, Marie-Andrée Giroux, Martijn van de Pol, Matthew Johnson, Megan L. Boldenow, Mihai Valcu, Mikhail Soloviev, Natalya Sokolova, Nathan R. Senner, Nicolas Lecomte, Nicolas Meyer, Niels Martin Schmidt, Olivier Gilg, Paul A. Smith, Paula Machín, Rebecca L. McGuire, Ricardo A. S. Cerboncini, Richard Ottvall, Rob S. A. van Bemmelen, Rose J. Swift, Sarah T. Saalfeld, Sarah E. Jamieson, Stephen Brown, Theunis Piersma, Tomas Albrecht, Verónica D'Amico, Richard B. Lanctot, Bart Kempenaers

Kubelka *et al.* (Reports, 9 November 2018, p. 680) claim that climate change has disrupted patterns of nest predation in shorebirds. They report that predation rates have increased since the 1950s, especially in the Arctic. We describe methodological problems with their analyses and argue that there is no solid statistical support for their claims.

Full text: dx.doi.org/10.1126/science.aaw8529

Response to Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds"

Vojtěch Kubelka, Miroslav Šálek, Pavel Tomkovich, Zsolt Végvári, Robert P. Freckleton, Tamás Székely

Bulla *et al.* dispute our main conclusion that the global pattern of nest predation is disrupted in shorebirds. We disagree with Bulla *et al.*'s conclusions and contest the robustness of their outcomes. We reaffirm our results that provide clear evidence that nest predation has increased significantly in shorebirds, especially in the Arctic.

Full text: dx.doi.org/10.1126/science.aaw9893

ERRATA

Erratum for the Report "Conformationally supple glucose monomers enable synthesis of the smallest cyclodextrins" by D. Ikuta *et al.*, *Science* **364,** eaay0378 (2019). Published online 17 May 2019; 10.1126/science.aay0378

Erratum for the Report "Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry" by D. S. Chorev *et al.*, *Science* **364,** eaax7485 (2019). Published online 26 April 2019; 10.1126/science.aax7485

Cite as: M. Bulla *et al.*, *Science* 10.1126/science.aaw8529 (2019).

Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds"

Martin Bulla[1,2,3]*†, Jeroen Reneerkens[2,4]†, Emily L. Weiser[5]†, Aleksandr Sokolov[6], Audrey R. Taylor[7], Benoît Sittler[8,9], Brian J. McCaffery[10], Dan R. Ruthrauff[11], Daniel H. Catlin[12], David C. Payer[13], David H. Ward[11], Diana V. Solovyeva[14], Eduardo S. A. Santos[15], Eldar Rakhimberdiev[4,16], Erica Nol[17], Eunbi Kwon[12], Glen S. Brown[18], Glenda D. Hevia[19], H. River Gates[20], James A. Johnson[21], Jan A. van Gils[2], Jannik Hansen[22], Jean-François Lamarre[23], Jennie Rausch[24], Jesse R. Conklin[4], Joe Liebezeit[25], Joël Bêty[26], Johannes Lang[9,27], José A. Alves[28,29], Juan Fernández-Elipe[30], Klaus-Michael Exo[31], Loïc Bollache[32], Marcelo Bertellotti[19], Marie-Andrée Giroux[33], Martijn van de Pol[34], Matthew Johnson[35], Megan L. Boldenow[36], Mihai Valcu[1], Mikhail Soloviev[16], Natalya Sokolova[6], Nathan R. Senner[37], Nicolas Lecomte[38], Nicolas Meyer[9,32], Niels Martin Schmidt[22,39], Olivier Gilg[9,32], Paul A. Smith[40], Paula Machín[30], Rebecca L. McGuire[41], Ricardo A. S. Cerboncini[42], Richard Ottvall[43], Rob S. A. van Bemmelen[44], Rose J. Swift[45], Sarah T. Saalfeld[21], Sarah E. Jamieson[46], Stephen Brown[47], Theunis Piersma[2,4], Tomas Albrecht[48,49], Verónica D'Amico[19], Richard B. Lanctot[21]†, Bart Kempenaers[1]*†

[1]Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany. [2]NIOZ Royal Netherlands Institute for Sea Research, Department of Coastal Systems and Utrecht University, 1790 AB Den Burg, Texel, Netherlands. [3]Faculty of Environmental Sciences, Czech University of Life Sciences, 16521 Prague, Czech Republic. [4]Conservation Ecology Group, Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, 9700 CC Groningen, Netherlands. [5]Upper Midwest Environmental Sciences Center, U.S. Geological Survey, La Crosse, WI 54603, USA. [6]Arctic Research Station, Institute of Plant and Animal Ecology, 629400 Labytnangi, Russia. [7]Department of Geography and Environmental Studies, University of Alaska, Anchorage, AK 99508, USA. [8]Nature Conservation and Landscape Ecology, University of Freiburg, 79106 Freiburg, Germany. [9]Arctic Ecology Research Group (GREA), F-21440 Francheville, France. [10]Yukon Delta National Wildlife Refuge, U.S. Fish and Wildlife Service, Grand View, WI 54839, USA. [11]Alaska Science Center, U.S. Geological Survey, Anchorage, AK 99508, USA. [12]Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA 24061, USA. [13]Natural Resource Sciences, National Park Service, Anchorage, AK 99501, USA. [14]Institute of Biological Problems of the North, FEB RAS, Magadan 685000, Russia. [15]BECO do Departamento de Zoologia, Rua do Matão, Universidade de São Paulo, 05508-090 São Paulo, Brazil. [16]Department of Vertebrate Zoology, Lomonosov Moscow State University, 119234 Moscow, Russia. [17]Biology Department, Trent University, Peterborough, ON K9J 7B8, Canada. [18]Wildlife Research and Monitoring, Ministry of Natural Resources and Forestry, Peterborough, ON K9L 1Z8, Canada. [19]Grupo de Ecofisiología Aplicada al Manejo y Conservación de Fauna Silvestre, Centro para el Estudio de Sistemas Marinos (CESIMAR)-CCT CONICET-CENPAT, 9120 Puerto Madryn, Argentina. [20]Pacific Flyway Program, National Audubon Society, Anchorage, AK 99501, USA. [21]Migratory Bird Management, U.S. Fish and Wildlife Service, Anchorage, AK 99503, USA. [22]Department of Bioscience, Aarhus University, 4000 Roskilde, Denmark. [23]Science & Technology Program, Polar Knowledge Canada, Cambridge Bay, NU X0B 0C0, Canada. [24]Canadian Wildlife Service, Environment and Climate Change Canada, P.O. Box 2310, Yellowknife, NT X1A 2P7, Canada. [25]Audubon Society of Portland, Portland, OR 97210, USA. [26]Department of Biology and Center for Northern Studies, University of Quebec, Rimouski, QC G5L 3A1, Canada. [27]Clinic for Birds, Reptiles, Amphibians and Fish/Working Group for Wildlife Biology, Giessen University, 35392 Giessen, Germany. [28]DBIO & CESAM-Centre for Environmental and Marine Studies, Department of Biology, University of Aveiro, 3810-193 Aveiro, Portugal. [29]South Iceland Research Centre, University of Iceland, Fjolheimar IS-800 Selfoss & IS-861 Gunnarsholt, Iceland. [30]Aptdo. Correos 32, 5480 Candeleda, Spain. [31]Institute of Avian Research "Vogelwarte Helgoland," 26386 Wilhelmshaven, Germany. [32]Laboratoire Chrono-environnement, Université de Franche-Comté, UMR 6249 CNRS-UFC, F-25000 Besançon, France. [33]Faculty of Sciences, Université de Moncton, Moncton, NB E1A 3E9, Canada. [34]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6708PB Wageningen, Netherlands. [35]Plumas National Forest, USDA Forest Service, Quincy, CA 95971, USA. [36]Biology and Wildlife Department, University of Alaska, Fairbanks, AK 99775, USA. [37]Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA. [38]Department of Biology, Uninversité de Moncton, Moncton, NB E1A 3E9, Canada. [39]Arctic Research Centre, Aarhus University, 8000 Aarhus C, Denmark. [40]National Wildlife Research Centre, Environment and Climate Change Canada, Ottawa, ON K1S 5B6, Canada. [41]Arctic Beringia Program, Wildlife Conservation Society, Fairbanks, AK 99709, USA. [42]Departamento de Zoologia, Universidade Federal do Paraná, 81531-980 Curitiba, Brazil. [43]Frostavallsvägen 325, S-24393 Höör, Sweden. [44]Wageningen Marine Research, 1976CP IJmuiden, Netherlands. [45]Cornell Lab of Ornithology and Department of Natural Resources, Cornell University, Ithaca, NY 14850, USA. [46]Centre for Wildlife Ecology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. [47]Shorebird Recovery Program, Manomet Inc., P.O. Box 545, Saxtons River, VT 05154, USA. [48]Institute of Vertebrate Biology, Czech Academy of Sciences, 60300 Brno, Czech Republic. [49]Faculty of Science, Charles University, 128 44 Prague, Czech Republic.

*Corresponding author. bulla.mar@gmail.com (M.B.); b.kempenaers@orn.mpg.de (B.K.)

†These authors formed the core team behind this work. The remaining authors are listed alphabetically according to their first name.

Kubelka *et al.* (Reports, 9 November 2018, p. 680) claim that climate change has disrupted patterns of nest predation in shorebirds. They report that predation rates have increased since the 1950s, especially in the Arctic. We describe methodological problems with their analyses and argue that there is no solid statistical support for their claims.

Climate change affects organisms in a variety of ways (*1*–*4*), including through changes in interactions between species. Kubelka *et al.* (*5*) reported that a specific type of trophic interaction, namely depredation of shorebird nests, has increased globally over the past 70 years. The authors state that their results are "consistent with climate-induced shifts in predator-prey relationships." They also claim that the historical perception of a latitudinal gradient in nest predation, with the highest rates in the tropics, "has been recently reversed in the Northern Hemisphere, most notably in the Arctic." They conclude that "the Arctic now represents an extensive ecological trap for migrating birds, with a predicted negative impact on their global population dynamics." These conclusions have far-reaching implications for evolutionary and population ecology, as well as for shorebird conservation and related policy decisions (*6*). Therefore, such claims require robust evidence, strongly supported by the data. Here, we dispute this evidence.

First, Kubelka *et al.* graphically show nonlinear, spatiotemporal variation in predation rates (their figure 2, A and B, and figure 3) and suggest that in recent years, predation has strongly increased in North temperate and especially Arctic regions, but less so in other areas. However, they only statistically test for linear changes in predation rates over time for all regions combined, and for each geographical region (their table S2) or period (before and after 2000; their table S6) separately. To substantiate their conclusions, they should have presented statistical evidence for an interaction between region/latitude and year/period on predation rate. Moreover, their analyses control for spatial autocorrelation but fail to model non-independence of data from the same site (pseudo-replication).

Using the data of Kubelka *et al.*, we ran a set of mixed-effect models, structurally reflecting their results depicted in their figure 2, A and B, and figure 3, but including location as a random factor (Table 1) (*7*). These analyses show (i) that much of the variation in nest predation rate is explained by study site (>60%, compared to species: <5%), implying a reduced effective sample size; (ii) that all regions—except the South temperate—show similar predation rates; and (iii) that nest predation rates increase over time similarly across all geographical areas (Fig. 1, A to F). Linear models without interaction terms are much better supported than nonlinear models with interactions (Table 1), indicating that predation rates in the Arctic are not increasing any faster than elsewhere (Fig. 1, B, C, E, and F). Thus, these results provide no evidence that the rate at which nest predation increased over time varies geographically.

Second, for the period under study, not only the climate has changed, but also the research methods. Hence, it remains unclear whether nest predation rates have indeed increased over time and if so, why. Kubelka *et al.* used the Mayfield method (*8*, *9*) to calculate daily nest predation rates as the number of depredated nests divided by "exposure" [the total time (in days) all nests were observed]. However, 59% of the 237 populations they used lacked information on exposure. They circumvented this problem by estimating exposure based on the description of nest search intensity in the respective studies (*10*). The key question is when nests were found. Kubelka *et al.* decided that in 114 populations, nests were found such that 60% of the nesting period (egg laying and incubation combined) was "observed" (B = 0.6; nests searched once or twice a week). For 14 populations they used B = 0.9 (nests searched daily or found just after laying), and for 11 populations they used B = 0.5 (assuming nests found midway during the nesting period). However, the choice of B value remains subjective (*7*), and for 38% of the 128 populations where Kubelka *et al.* used $B > 0.5$, we found no information in the reference to suggest that this was appropriate. This issue is not trivial, because using higher B values (i.e., assuming that nests were found earlier than they actually were) overestimates exposure and hence underestimates nest predation rates.

The proportion of populations with estimated exposure declines over time (*7*), particularly after 2000 and especially in the Arctic (Fig. 1G). The timing of the decline coincides with Kubelka *et al.*'s definition of historic and recent data and with the suggested exponential rise of predation in the Arctic (their figure 2, A and B, and figure 3, A and B). Indeed, the results are sensitive to variation in estimated exposure during the "historic period" (Fig. 1H). Although Kubelka *et al.* correctly state that the estimated and true predation rates are highly correlated [using studies with quantitative information on exposure; see supplementary materials of (*5*)], the true rate is typically underestimated for the higher B values they used (Fig. 1I). Given these issues, the main result—the apparent increase in daily nest predation rate over time, especially in the Arctic—may simply be an artifact. To further assess the robustness of the change in predation rate over time, we used only populations where nest predation rates were calculated on the basis of known exposure (N = 98). These analyses reduced the effect of year by ~50% (*7*) and resulted in weak, nonsignificant linear trends (Fig. 1, C and F), which suggests that there is little evidence for changing predation rates.

Finally, we note that nest searching effort and frequency of nest visits likely increased in recent years as researchers learned how best to obtain accurate estimates of nest survival (*11*–*13*). Researchers also intensified their activities (e.g., capturing adults to band, tag, and collect samples and placing monitoring equipment near nests, which may increase the predation rate) (*14*, *15*). Thus, an increase in the quality of data reporting as well as increased research activity around nests may have further induced a time-

dependent bias in estimates with an underestimation of true predation rates in the historic data (see above), and perhaps an overestimation in the contemporary data.

In summary, reanalysis of the data of Kubelka *et al.*, evaluation of the quality and interpretation of the published data used, and considerations about changes in research methods over the past 70 years lead us to conclude that there is no robust evidence for a global disruption of nest predation rates due to climate change. We argue that their claim that the Arctic has become an ecological trap for breeding shorebirds is untenable.

REFERENCES

1. G.-R. Walther, E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J.-M. Fromentin, O. Hoegh-Guldberg, F. Bairlein, Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002). doi:10.1038/416389a Medline
2. N. Jonzén, A. Lindén, T. Ergon, E. Knudsen, J. O. Vik, D. Rubolini, D. Piacentini, C. Brinch, F. Spina, L. Karlsson, M. Stervander, A. Andersson, J. Waldenström, A. Lehikoinen, E. Edvardsen, R. Solvang, N. C. Stenseth, Rapid advance of spring arrival dates in long-distance migratory birds. *Science* **312**, 1959–1961 (2006). doi:10.1126/science.1126119 Medline
3. S. J. Thackeray, P. A. Henrys, D. Hemming, J. R. Bell, M. S. Botham, S. Burthe, P. Helaouet, D. G. Johns, I. D. Jones, D. I. Leech, E. B. Mackay, D. Massimino, S. Atkinson, P. J. Bacon, T. M. Brereton, L. Carvalho, T. H. Clutton-Brock, C. Duck, M. Edwards, J. M. Elliott, S. J. G. Hall, R. Harrington, J. W. Pearce-Higgins, T. T. Høye, L. E. B. Kruuk, J. M. Pemberton, T. H. Sparks, P. M. Thompson, I. White, I. J. Winfield, S. Wanless, Phenological sensitivity to climate across taxa and trophic levels. *Nature* **535**, 241–245 (2016). doi:10.1038/nature18608 Medline
4. D. Zurell, C. H. Graham, L. Gallien, W. Thuiller, N. E. Zimmermann, Long-distance migratory birds threatened by multiple independent risks from global change. *Nat. Clim. Chang.* **8**, 992–996 (2018). doi:10.1038/s41558-018-0312-9 Medline
5. V. Kubelka, M. Šálek, P. Tomkovich, Z. Végvári, R. P. Freckleton, T. Székely, Global pattern of nest predation is disrupted by climate change in shorebirds. *Science* **362**, 680–683 (2018). doi:10.1126/science.aat8695 Medline
6. C. E. Studds, B. E. Kendall, N. J. Murray, H. B. Wilson, D. I. Rogers, R. S. Clemens, K. Gosbell, C. J. Hassell, R. Jessop, D. S. Melville, D. A. Milton, C. D. T. Minton, H. P. Possingham, A. C. Riegen, P. Straw, E. J. Woehler, R. A. Fuller, Rapid population decline in migratory shorebirds relying on Yellow Sea tidal mudflats as stopover sites. *Nat. Commun.* **8**, 14895 (2017). doi:10.1038/ncomms14895 Medline
7. M. Bulla, J. Reneerkens, E. L. Weiser, R. B. Lanctot, B. Kempenaers, Supporting information for Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds." Open Science Framework, https://osf.io/x8fs6/.
8. H. Mayfield, Nesting success calculated from exposure. *Wilson Bull.* **73**, 255–261 (1961).
9. H. Mayfield, Suggestions for calculating nest success. *Wilson Bull.* **87**, 456–466 (1975).
10. A. J. Beintema, Inferring nest success from old records. *Ibis* **138**, 568–570 (1996). doi:10.1111/j.1474-919X.1996.tb08084.x
11. S. J. Dinsmore, G. C. White, F. L. Knopf, Advanced techniques for modeling avian nest survival. *Ecology* **83**, 3476–3488 (2002). doi:10.1890/0012-9658(2002)083[3476:ATFMAN]2.0.CO;2
12. T. L. Shafer, F. R. Thompson III, Making meaningful estimates of nest survival with model-based methods. *Stud. Avian Biol.* **34**, 84–95 (2007).
13. J. J. Rotella, Modeling nest-survival data: Recent improvements and future directions. *Stud. Avian Biol.* **34**, 145–148 (2007).
14. N. Verboven, B. J. Ens, S. Dechesne, Effect of investigator disturbance on nest attendance and egg predation in Eurasian oystercatchers. *Auk* **118**, 503–508 (2001). doi:10.1642/0004-8038(2001)118[0503:EOIDON]2.0.CO;2
15. B. W. Meixell, P. L. Flint, Effects of industrial and investigator disturbance on Arctic-nesting geese. *J. Wildl. Manage.* **81**, 1372–1385 (2017). doi:10.1002/jwmg.21312

ACKNOWLEDGMENTS

We thank all who contributed to the original studies on predation, shared data, helped to find the original data, or helped in preparing this comment [see (*7*) for details]. **Funding:** Supported by the Max Planck Society (B.K.), an EU Horizon 2020 Marie Curie individual fellowship (4231.1 SocialJetLag; M.B.), the Czech University of Life Sciences (CIGA: 2018421; M.B.), a Netherlands Polar Programme grant (NWO 866.15.207; J.R. and T.P.), and the U.S. Fish and Wildlife Service (R.B.L.). **Author contributions:** M.B., J.R., R.B.L., and B.K. initiated this work; M.B., J.R., E.L.W., D.V.S., R.B.L., and B.K. extracted and evaluated the data; most authors collected data contained in the primary sources; E.L.W. investigated the differences between newly extracted and original data; M.B. conducted the statistical analyses and drew the figure with help from M.V. and E.L.W. and input from J.R., R.B.L., and B.K.; B.K. and M.B. wrote the manuscript with input from J.R., E.L.W., and R.B.L. and finalized it with input from all authors. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data, code, methods, results, figures, and tables associated with this comment are freely available from Open Science Framework (*7*).

6 February 2019; accepted 29 May 2019
Published online 14 June 2019
10.1126/science.aaw8529

Table 1. Comparison of models explaining spatiotemporal variation in daily nest predation rate using the original Kubelka *et al.* data. Letters and results in bold refer to panels in Fig. 1; A and D are the models reflecting figures 2A and 3A in (*5*). Each model is fitted with maximum likelihood and controlled for number of nests in a given population (ln-transformed) and for multiple populations at a given site or for a given species, using site and species as random intercepts. Daily predation rate (dependent variable) was ln-transformed after adding 0.01 [following (*5*)]. Predictors are Year (mean year of the study), Hemisphere (Northern versus Southern), Latitude (degrees), Geographical Area (Arctic, North temperate, North tropics, South tropics, South temperate), and Period [historic (1944–1999) versus recent (2000–2016)]. Models that include Period (instead of Year) are not supported by the data [less likely than the best model by factors of 69 to 320, as indicated by the evidence ratio (model weight of the first-ranked model relative to that of the given model, i.e., how many times the first-ranked model is more likely than the given model)]. Models including the interaction between time and geographical region/latitude do not improve the model fit or are much less supported by the data than are models without the interaction. See (*7*) for model output and analyses of total predation rates. Note that we used quadratic or third-order polynomial terms to mimic the relationships depicted in Kubelka *et al*.'s figures (*5*). Number of parameters denotes number of model parameters without the random effects. ΔAIC is the difference in Akaike information criterion between the first-ranked model (AIC = 349.8) and the given model. Model probability refers to Akaike weight (w_i), the weight of evidence (probability) that a given model is the best-approximating model.

Model	Predictors	Number of parameters	ΔAIC	Model probability	Evidence ratio
	Year + Hemisphere + Latitude (absolute)	5	0.00	0.26	1
E	**Year + Latitude (3rd polynomial)**	**6**	**0.05**	**0.25**	**1.02**
	Year + Geographical Area	7	0.51	0.2	1.29
B	**Year (quadratic) + Geographical Area**	**8**	**1.43**	**0.13**	**2.04**
	Year × Hemisphere × Latitude (absolute)	9	2.74	0.07	3.92
	Year × Latitude (3rd polynomial)	9	2.78	0.06	4.08
	Year × Geographical Area	11	6.31	0.01	23.36
A	**Year (quadratic) × Geographical Area**	**16**	**6.43**	**0.01**	**24.89**
D	**Period × Latitude (3rd polynomial)**	**9**	**8.48**	**0**	**69.26**
	Period × Hemisphere × Latitude (absolute)	9	9.66	0	124.9
	Period + Hemisphere + Latitude (absolute)	5	10.30	0	175.3
	Period + Latitude (3rd polynomial)	6	11.50	0	319.7

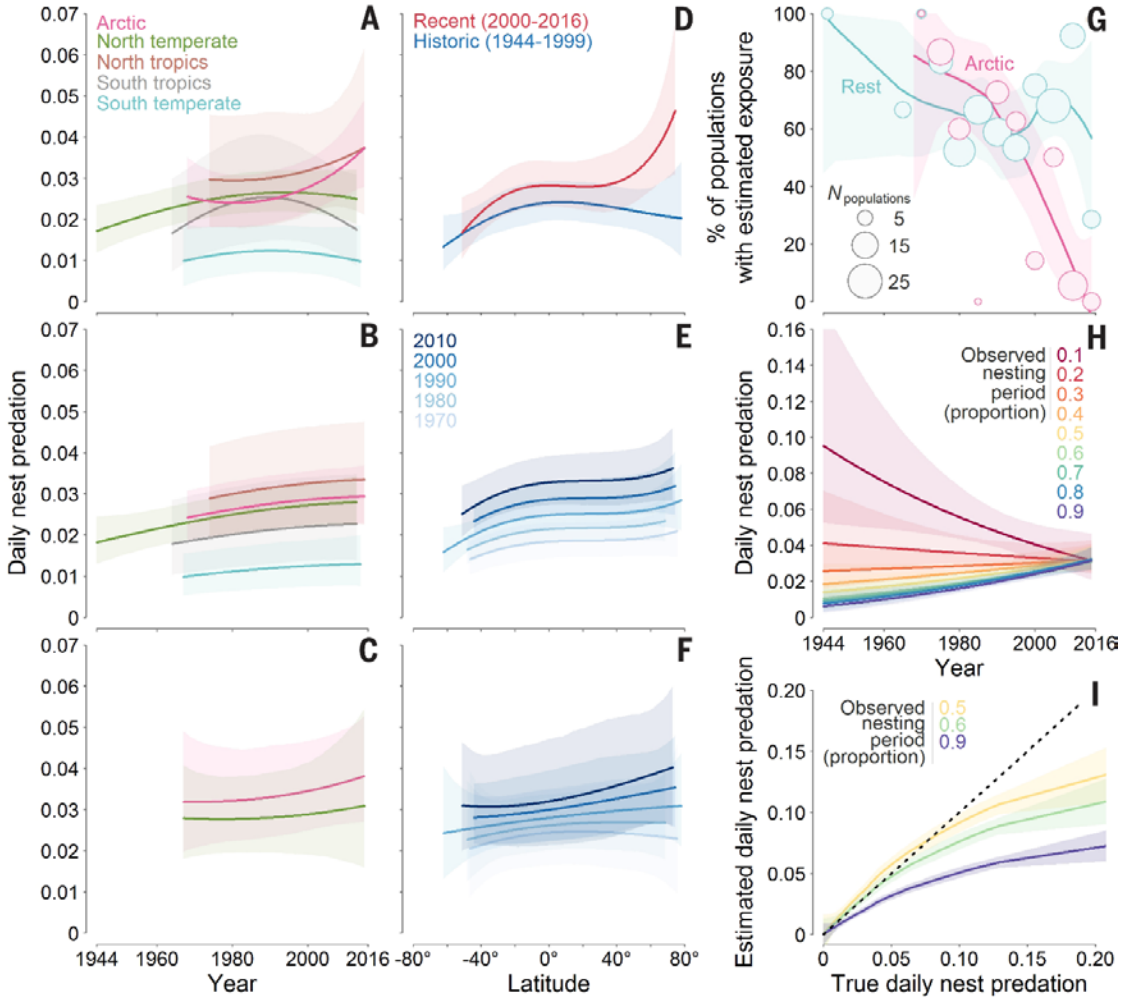


Fig. 1. Spatiotemporal variation in daily nest predation rates of shorebirds. (**A** to **C**) Predation rate in relation to year for different geographical regions: with interaction and using all populations (A), without interaction and using all populations (B), with interaction and using only the 88 populations with known exposure from the Arctic and North temperate region (C). The model behind (A) is less supported by the data than the model behind (B) by a factor of ~18 (Table 1). (**D** to **F**) Predation rate in relation to latitude for different periods: with interaction (period as two-level factor) and using all populations (D), without interaction (year as continuous variable) and using all populations (E), with interaction and using only the 98 populations with known exposure (F). The model behind (D) is less supported than the model behind (E) by a factor of ~70 (Table 1). In (A) to (F), lines and shaded areas represent model predictions with 95% confidence interval (CI) based on posterior distribution of 5000 simulated values. Note the weak ($P > 0.64$) temporal increase in (C) [estimate = 0.08 (95% CI, −0.07 to 0.2) from a linear model without interaction] and (F) [estimate = 0.06 (95% CI, −0.09 to 0.17)]. See Table 1 for model description and comparison and (*7*) for details. (**G**) Temporal change in the percentage of populations in which exposure was estimated [following (*10*)] to calculate predation rate. Note the sharp decline in the Arctic relative to the other regions [see (*7*) for overall and region-specific changes]. Circles represent data for 5-year intervals. (**H**) Modeled changes in predation rate over time assuming different values of *B* (proportion of nesting period observed; higher values indicate nests found sooner after egg laying) for populations with unknown exposure and year <2000 (leaving the original estimates for all remaining populations). This exercise explores the sensitivity of the results to using older studies where the stage at which nests were found is less certain. (**I**) Relation between true and estimated predation rate for different values of *B* [$N = 65$ populations, as in (*5*)]. The dashed line indicates a slope of 1 (i.e., estimated values equaling true values). In (G) and (I), lines and shaded areas represent locally estimated scatterplot smoothing with 95% CI; in (H), lines and shaded areas represent model predictions with 95% CI based on posterior distribution of 5000 simulated values.

Cite as: V. Kubelka *et al.*, *Science* 10.1126/science.aaw9893 (2019).

Response to Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds"

Vojtěch Kubelka[1,2,3]*, Miroslav Šálek[4], Pavel Tomkovich[5], Sold Végvári[6], Robert P. Freckleton[7], Tomás Székely[1,2,8]*

[1]Department of Evolutionary Zoology and Human Biology, University of Debrecen, H-4032 Debrecen, Hungary. [2]Milner Centre for Evolution, University of Bath, Claverton Down, Bath BA2 7AY, UK. [3]Department of Biodiversity Research, Global Change Research Institute, Czech Academy of Sciences, 603 00 Brno, Czech Republic. [4]Faculty of Environmental Sciences, Czech University of Life Sciences, 165 21 Prague, Czech Republic. [5]Zoological Museum, Moscow MV Lomonosov State University, Moscow 125009, Russia. [6]Department of Conservation Zoology, University of Debrecen, H-4032 Debrecen, Hungary. [7]Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. [8]State Key Laboratory of Biocontrol and College of Ecology and Evolution, Sun Yat-sen University, Guangzhou 510275, China.

*Corresponding author. Email: kubelkav@gmail.com (V.K.); t.szekely@bath.ac.uk (T.S.).

Bulla *et al.* dispute our main conclusion that the global pattern of nest predation is disrupted in shorebirds. We disagree with Bulla *et al.*'s conclusions and contest the robustness of their outcomes. We reaffirm our results that provide clear evidence that nest predation has increased significantly in shorebirds, especially in the Arctic.

In our study (*1*) we showed significant increases in daily nest predation rate in ground-nesting shorebirds that were especially strong in the Arctic. Bulla *et al.* (*2*) raise four concerns about our findings. Here, we highlight statistical and methodological problems in Bulla *et al.*'s analyses.

First, Bulla *et al.* argue that we should have presented evidence of a statistically significant interaction between region/latitude and year/period on daily nest predation rate in order to demonstrate that the temporal trend of increasing predation rate varies spatially. This criticism ignores that the analyses both we and they reported are based on a logarithmically transformed response variable [$\log(x + 0.01)$: transformation applied based on model diagnostics]. The presence or absence of interactions based on transformed data is difficult to interpret. For example, log-transformation means that on the arithmetic scale, the effects of model predictors are multiplicative, which implies interactive effects.

This issue is demonstrated in Fig. 1. The data suggest strong temporal increases in daily nest predation (Fig. 1A) (*3*) with an evident difference in responses between the Northern and Southern Hemispheres. The lines in Fig. 1 are based on a model fitted to the log-transformed data and confirm the presence of both temporal and geographic effects. Diagnostic analysis confirms that the geographic difference is clearly evident in the data, becoming larger through time (Fig. 1B). The simple model reported in Fig. 1A captures this increase, despite the lack of an interaction term, because of the data transformation. However, Bulla *et al.* failed to recognize this and were therefore unable to detect the patterns shown in Fig. 1.

Figure 1C shows an example of the complexity underneath this overall global pattern: Differences between the Arctic and the nearest region (North temperate) are small on average, but recently daily nest predation has risen notably in the Arctic (Fig. 1D). Our published analyses and visualizations were designed to explore the nature of such interactions by examining the spatiotemporal patterns in more detail. Figures 2 and 3 in (*1*) demonstrate these spatially variable effects on which our conclusions are based and include mean trends with associated confidence intervals.

Second, Bulla *et al.* suggest that our models were pseudo-replicated, although they acknowledge that we corrected for spatial autocorrelation. In our spatial models (*1*, *4*–*6*), spatial dependency is modeled as a random variance in which all points from the same location have the same (maximal) covariance. Our spatial term also accounts for possible covariance between sites that are close to each other, which the models of Bulla *et al.* do not (*7*).

Third, Bulla *et al.* raised concerns about "observation time" (which they refer to as *B*). *B* represents the proportion of incubation and egg-laying periods for which successful nests in a particular population are followed by researchers. These values are used to estimate nest exposure using Beintema's method (*8*); this converts "apparent predation" to daily nest predation values. Alternatively, if the exposure and number of depredated nests are given, then daily nest predation may be computed directly (*1*, *9*, *10*). Bulla *et al.* argue that the increase in daily nest predation over time

may be an artifact of using Beintema's methodology (*8*). We do not agree with Bulla *et al.* for three reasons:

1) For 56 populations where data using both approaches were available, the estimated values were not statistically different from the directly computed ones [see supplementary material of (*1*) and Fig. 2A]. The only trend suggested is a slight but nonsignificant underestimation of daily nest predation using Beintema's conversion after year 2000 (Fig. 2B) that is opposite to the concern raised by Bulla *et al.*

2) The simulations performed by Bulla *et al.* include observation times (B) from 0.1 to 0.9 [figure 1H in (*2*)] with the artifactual increases in predation rate for historic data when B is very low (0.1 to 0.5). However, B should in principle vary from 0.5 of the incubation and egg-laying period in studies with random nest search to approximately 0.9 when all nests are found during the egg-laying period as a result of intensive research (*8*, *11*). Bulla *et al.* computed an average $B = 0.65$ for more than 10,000 shorebird nests at 16 Arctic sites and do not report $B < 0.5$ at any site or in any population (*12*). Lower values ($B < 0.5$) might be plausible only in exceptional circumstances (e.g., when low nest search intensity is restricted to the late breeding period only). However, no estimate of B in our dataset was lower than 0.5 (*1*, *13*); therefore, Bulla *et al.*'s simulations using $B < 0.5$ for our data are unjustified.

3) Contrary to Bulla *et al.*, we find that the global temporal increase in nest predation is significant when daily nest predation was either calculated directly or converted from "apparent predation" [table S3 in (*1*); Fig. 2C] (*14*).

Finally, Bulla *et al.* hypothesize that increased research intensity over the years could have influenced the temporal increase in nest predation. To address this issue, we scored the research intensity for the data used in (*1*); however, we found no evidence that intensive research may have elevated nest predation rates (Fig. 2D). Therefore, notwithstanding the possible negative effect of disturbance through more intensive study, it is unlikely that research intensity would have contributed to the temporal increase in nest predation we reported (*1*).

Taken together, these findings provide further evidence that the global pattern of nest predation was disrupted in shorebirds, and this conclusion remains unaffected by the concerns raised by Bulla *et al.* (*2*). The key results of (*1*) are straightforward, supported by simple models and data: There is a strong temporal increase in nest predation rates at a global scale, with significant spatial variation. Our conclusions are reinforced by a detailed study of five shorebird species that reported increased nest predation for all five species between 1948 and 2006, using 200 data points from six European countries (*15*). Our results remain robust, and together they appear to reveal a significant, previously underappreciated global threat to shorebirds.

REFERENCES AND NOTES

1. V. Kubelka, M. Šálek, P. Tomkovich, Z. Végvári, R. P. Freckleton, T. Székely, Global pattern of nest predation is disrupted by climate change in shorebirds. *Science* **362**, 680–683 (2018). doi:10.1126/science.aat8695 Medline
2. M. Bulla *et al*., Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds". *Science* **364**, eaaw8529 (2019). doi:10.1126/science.aaw8529
3. V. Kubelka *et al*., Data, R codes, and supporting information for Response to Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds". Open Science Framework, https://osf.io/46bt3/.
4. W. Jetz, R. P. Freckleton, Towards a general framework for predicting threat status of data-deficient species from phylogenetic, spatial and environmental information. *Philos. Trans. R. Soc. B* **370**, 20140016 (2015). doi:10.1098/rstb.2014.0016 Medline
5. R. P. Freckleton, W. Jetz, Space versus phylogeny: Disentangling phylogenetic and spatial signals in comparative data. *Proc. R. Soc. B* **276**, 21–30 (2009). doi:10.1098/rspb.2008.0905 Medline
6. V. Kubelka *et al*., Data and R codes from "Global pattern of nest predation is disrupted by climate change in shorebirds". Dryad (2018), doi:10.5061/dryad.45g90h4.
7. The large number of random group levels (111 species and 152 localities) used by Bulla *et al*., relative to the number of observations (237), as well as the large proportion of groups represented by just single observations (59 of 111 species and 114 of 152 localities), raise concerns about the robustness of their approach—specifically, the identifiability of the random and residual terms when the fixed effects vary at the level of localities.
8. A. J. Beintema, Inferring nest success from old records. *Ibis* **138**, 568–570 (1996). doi:10.1111/j.1474-919X.1996.tb08084.x
9. H. Mayfield, Nesting success calculated from exposure. *Wilson Bull.* **73**, 255–261 (1961).
10. H. F. Mayfield, Suggestions for calculating nest success. *Wilson Bull.* **87**, 456–466 (1975).
11. Observation time (B) varying between 0.52 and 0.67 for six Dutch shorebird species can be derived from Beintema's work (*5*). $B > 0.5$ was found for all investigated shorebirds, including cryptic common snipe (*Gallinago gallinago*), which is well known for the difficulty of nest finding.
12. M. Bulla, J. Reneerkens, E. L. Weiser, R. B. Lanctot, B. Kempenaers, Supporting information for Comment on "Global pattern of nest predation is disrupted by climate change in shorebirds." Open Science Framework, https://osf.io/x8fs6/.
13. Bulla *et al*.'s claim that "for 38% of the 128 populations where Kubelka *et al*. used $B > 0.5$, we found no information in the reference to suggest that this was appropriate" is not substantiated in their comment; see information on data collection and research intensity in relevant source publications referenced in (*1*).
14. Bulla *et al*.'s treatment of random effects, as we note earlier, is problematic because of the preponderance of groups represented by single observations: When we removed groups represented by single observations and reran the analysis [figure 1F in (*2*)], the temporal trend was indeed statistically significant using the approach of Bulla *et al*. (*3*).
15. M. Roodbergen, B. van der Werf, H. Hötker, Revealing the contributions of reproduction and survival to the Europe-wide decline in meadow birds: Review and meta-analysis. *J. Ornithol.* **153**, 53–74 (2012). doi:10.1007/s10336-011-0733-y

ACKNOWLEDGMENTS

We thank the people who provided us with unpublished data or additional information to already published articles or helped us with gray literature and useful contacts; see (*1*) for details. D. Childs and A. J. Beintema helped us with their critical comments on the issues discussed within the article. **Funding:** Supported by ÉLVONAL-KKP 126949 of the Hungarian government (V.K. and T.S.); Ministry of

Education, Youth and Sports of the Czech Republic grant LO1415 (V.K.); Ministry of Education, Youth and Sports of the Czech Republic, Kontakt II LH 13278 (M.Š.); MSU Zoological Museum grant AAAA-A16-116021660077-3 (P.T.); and Royal Society Wolfson Merit Award and grant NKFIH-2558-1/2015 (T.S.). **Author contributions:** All authors conceived the study; V.K. collected the data with help from all co-authors; and V.K., R.P.F., and T.S. analyzed the data and wrote the paper with input from all co-authors. **Competing interests:** The authors declare that they have no competing interests.

9 March 2019; accepted 29 May 2019
Published online 14 June 2019
10.1126/science.aaw9893

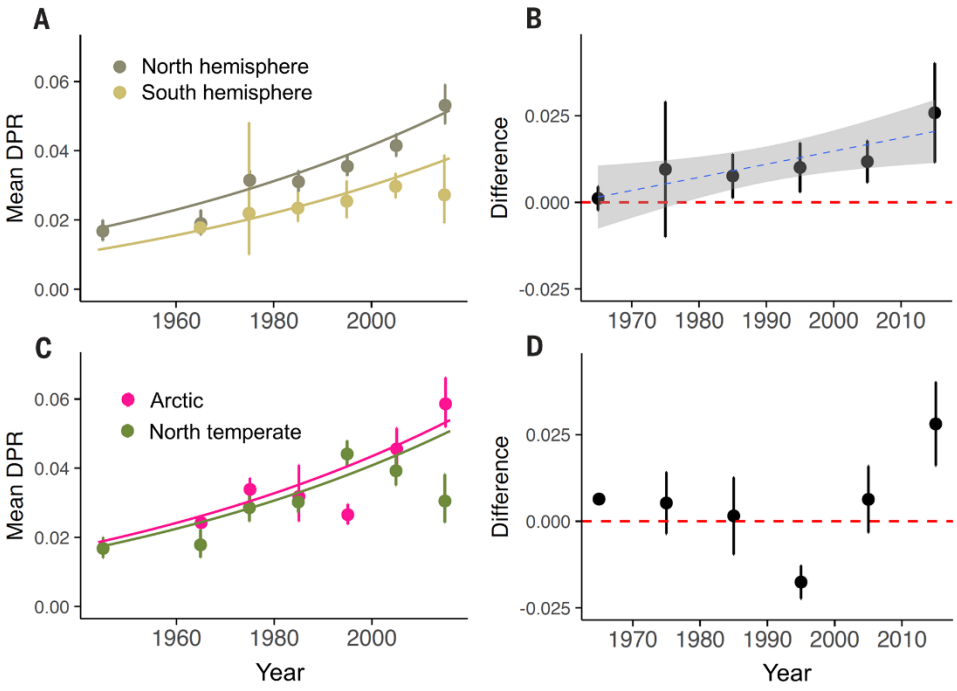


Fig. 1. Temporal and spatial variation in daily nest predation rate (DPR) at a global scale. We fitted the simplest model [log(DPR + 0.01) as a linear function of latitude and year]; both latitude and year are highly statistically significant. This is a simpler model than any used by Bulla *et al*. (*2*), but notably has a lower Akaike information criterion (AIC) than the models they considered. Note that although the model does not include an interaction term on the log scale, the lines diverge on the arithmetic scale, indicating geographically divergent outcomes. This applies equally to the models of Bulla *et al*. We also provide comparison with raw data as decadal averages (±SEM) in all graphs; see (*3*) for model details. (**A**) Fitted values (lines) for Northern and Southern Hemispheres compared with data. (**B**) Difference of DPR in the Northern relative to the Southern Hemisphere; the blue dashed line with 95% confidence intervals indicates a significant increase in this difference over the period of study. (**C**) Comparison of Arctic and North temperate regions. (**D**) Difference of DPR in the Arctic relative to the North temperate region. See (*1*) for details and sample sizes.

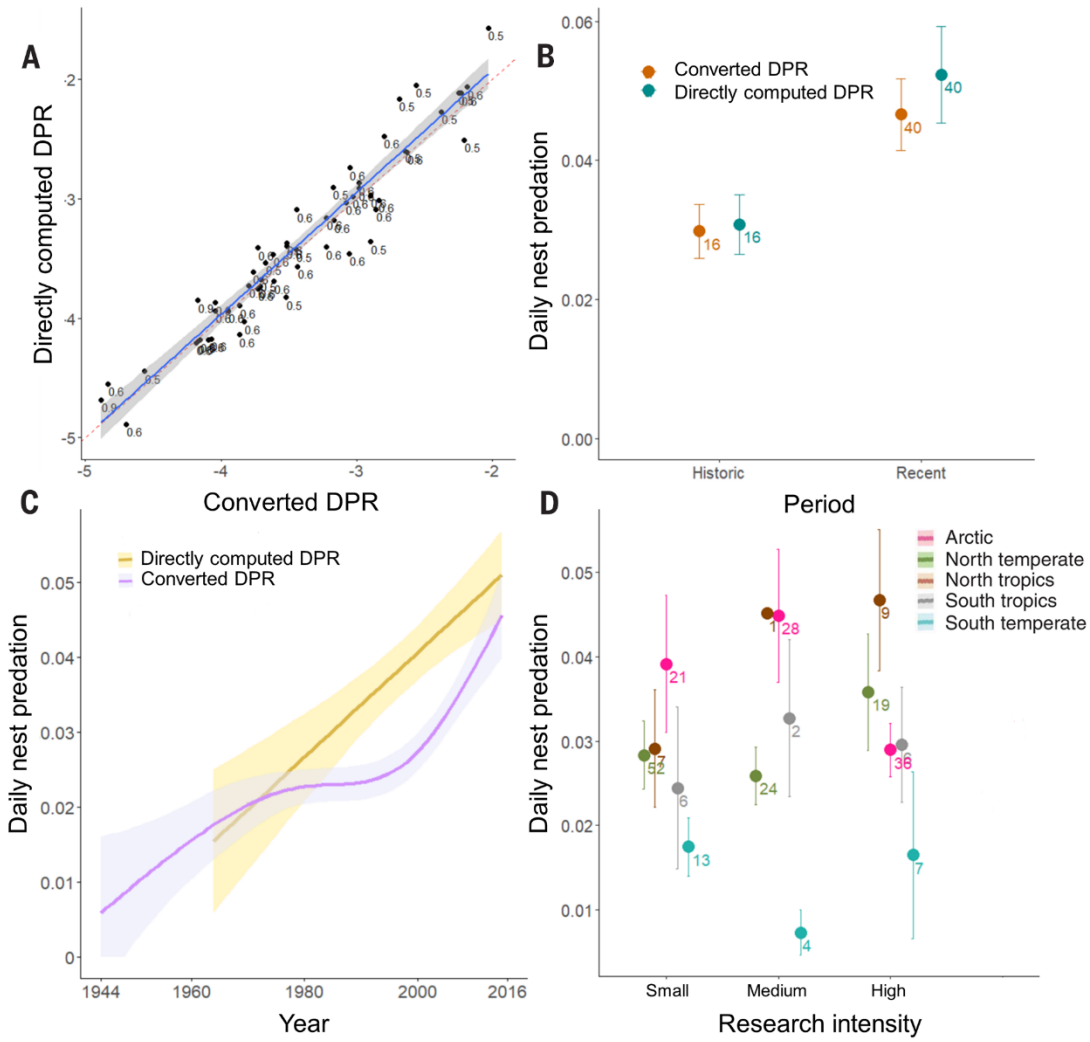


Fig. 2. Variation in DPR in relation to computation method and research intensity. (**A**) Correlation between directly calculated DPR and DPR values using Beintema's conversion (*5*). The relationship (± confidence interval) is in blue, the red dashed line represents parity, and observation time (*B*) is noted next to each data point. (**B**) Directly calculated and converted DPR in historic (before 2000) and recent (after 2000) periods. Data in (A) and (B) represent 56 populations; see (*1*) for details. (**C**) Temporal trend in directly calculated DPR (97 populations) and converted DPR (140 populations). Generalized additive model fits with 95% confidence intervals are shown; see table S3 in (*1*) for model description. (**D**) Daily nest predation variation according to research intensity: Small = infrequent visits to nests (e.g., to measure eggs); Medium = repeated nest visits to deploy nest-monitoring devices (e.g., trail cameras) near nest and/or adult trapping at a few nests; High = trapping and banding of breeding adults on a majority of nests. [Nonsignificant effect of research intensity on DPR: estimate = 0.091, SEM = 0.091, z = 0.09, P = 0.320; the same model structure as in table S3 in (*1*)]. Data are means ± SEM for five latitudinal areas separately; see (*1*, *3*) for details. Number of populations is given next to the relevant data points in (B) and (D).

RESEARCH

Three-way symbiosis protects sea slugs

Zan et al., *p. 1056*

IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

PLANT SCIENCE

How to make almonds palatable

The domesticated almond tree has been feeding humans for millennia. Derivation from the wild, bitter, and toxic almond required loss of the cyanogenic diglucoside amygdalin. Sánchez-Pérez *et al.* sequenced the almond genome and analyzed the genomic region responsible for this shift. The key change turned out to be a point mutation in a transcription factor that regulates production of P450 monooxygenases in the biosynthetic pathway for the toxic compound. —PJH *Science*, this issue p. 1095

Almond orchards in California, USA, produce sweet almonds for human consumption.

CREDITS: (TOP TO BOTTOM) THE PICTURE ART COLLECTION/ALAMY STOCK PHOTO; DESIGN PICS INC/ALAMY STOCK PHOTO

MEMBRANES

Supported graphene-based membranes

Porous graphene sheets have excellent filtration capabilities and are able to block most ions, but their fragility limits their scale-up beyond laboratory demonstrations. Yang *et al.* created a nanoporous graphene membrane reinforced by a network of single-walled carbon nanotubes (SWNTs) to provide mechanical stability (see the Perspective by Mi). The SWNT network also stopped the propagation of cracks in the graphene, effectively localizing the damage to a small area defined by a cell in the carbon nanotube mesh. The membranes showed high water flux rates as well as a high rejection rate for most ions. —MSL

Science, this issue p. 1057; see also p. 1033

STRUCTURAL BIOLOGY

ATP production under lockdown

Cellular processes must respond to change, often by speeding up, slowing down, or stopping altogether. Adenosine triphosphate (ATP) synthases use a transmembrane proton gradient to produce ATP, but this reaction can go in reverse and needs to be halted when conditions are unfavorable. Jinke Gu *et al.* purified a tetrameric ATP synthase complex from pig hearts that contained the endogenous inhibitory protein IF1. Targeted refinement yielded high-resolution views of the mammalian ATP synthase trapped in two different rotation states by IF1. The findings suggest that ATP synthase tetramers can be inhibited by at least three different mechanisms. —MAF

Science, this issue p. 1068

ELECTROCHEMISTRY

Three's a charm for iron and CO_2

Large-scale electrochemical reduction of CO_2 to CO could be a promising first step in sustainable conversion of the greenhouse gas to commodity chemicals. Currently, gold and silver are the most active catalysts for this process, whereas more abundant, less expensive metals tend to require impractically high potentials. Jun Gu *et al.* now report an iron catalyst with activity equaling or exceeding that of the precious metals. The key proved to be stabilization of the dispersed single iron ions in the +3 oxidation state. —JSY

Science, this issue p. 1091

METASURFACES

Dynamic metasurfaces

Nanostructured metasurfaces can function as many passive

optical elements. Now, S.-Q. Li *et al.* demonstrate that metasurfaces can be combined with liquid crystals to provide active control over light beams. With a view toward developing near-eye augmented reality display technology, they combined a dielectric metasurface with a liquid crystal layer to produce a tiny spatial light modulator. The results present a path for the development of dynamic metasurfaces as a platform for miniaturized optical technology with advanced time-dependent functionality. —ISO

Science, this issue p. 1087

NEUROSCIENCE

Longer ripples make better memories

Sharp wave ripples in the hippocampus are thought to play a role in memory formation and action planning. Fernández-Ruiz *et al.* used multisite electrophysiological recordings combined with optogenetic activation of hippocampal pyramidal neurons in rats performing learning tasks. Learning and correct recall in spatial memory tasks were associated with extended sharp wave ripples. Artificially prolonging these ripples improved working memory performance, whereas aborting the late part of ripples decreased performance. —PRS

Science, this issue p. 1082

INNATE LYMPHOID CELLS

Filming the airways

Group 2 innate lymphoid cells (ILC2s) are key drivers of immune responses in the lung. Puttur *et al.* used an interleukin-13 (IL-13) reporter mouse strain to visualize migration of ILC2s in allergic airway inflammation. ILC2s exhibited dynamic ameboid-like movement in response to IL-33. IL-3 induced up-regulation of chemokine receptor 8 on ILC2s, which promoted their homing to deposits of chemokine ligand 8 within the airways. Collagen-I, a component of the extracellular matrix (ECM), played a role in directly regulating ILC2 cell migration. Thus, structural cues from the ECM work in conjunction with immune mediators to influence immune cell functions in tissues. —AB

Sci. Immunol. **4**, eaav7638 (2019).

CANNABIS

Ancient usage of cannabis

The cultivation of cannabis extends back into distant prehistory. During excavations at a cemetery on the Pamir Plateau on the border of China and Tajikistan, Ren *et al.* found cannabinoid oils in wooden braziers. The finding is consistent with psychoactive cannabis use in burial rituals as early as 500 BCE. The cannabis featured high levels of THC (tetrahydrocannabinol)—higher than in wild varieties of the plant. Artifacts recovered from the burials as well as isotopic evidence from human remains suggest a high degree of cultural and economic exchange with neighboring peoples. Thus, people in the region may have been engaged in the hybridization of disparate populations of cannabis plants for the purpose of increasing their potency. —AC and KJP

Sci. Adv. 10.1126/sciadv.aaw1391 (2019).

OBESITY

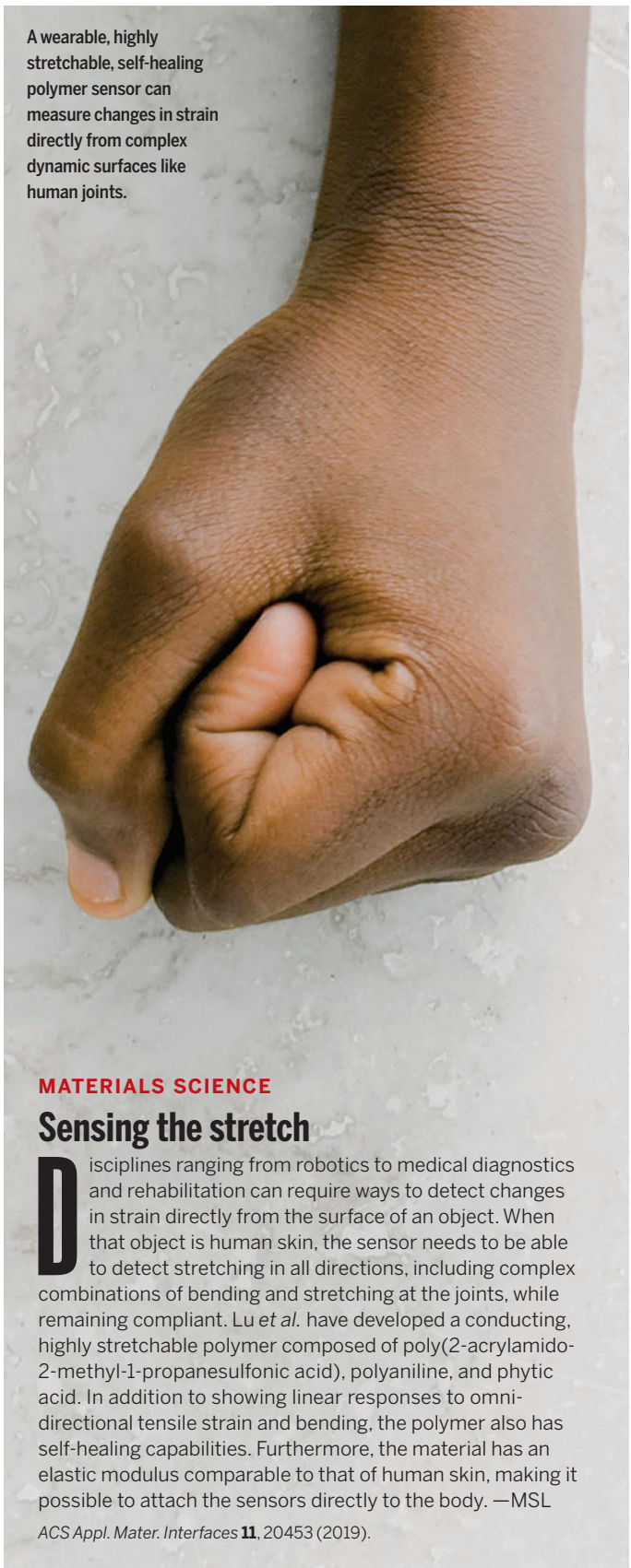
A gut-fat axis

Whether and how the gut microbiome impacts adipose tissue homeostasis is an open question. Virtue *et al.* showed that a high-fat diet in mice led to the activation of *miR-181* in white adipose tissue (WAT). This in turn led to obesity, insulin resistance, and WAT inflammation. The increased expression of this microRNA was linked to a reduction in circulating microbiota-derived metabolites produced by tryptophan metabolism in the gut. In the plasma of obese humans, *miR-181* expression was increased in WAT and indole was reduced, suggesting potential medical relevance for this axis. —CAC

Sci. Transl. Med. **11**, eaav1892 (2019).

IN OTHER JOURNALS

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



A wearable, highly stretchable, self-healing polymer sensor can measure changes in strain directly from complex dynamic surfaces like human joints.

MATERIALS SCIENCE

Sensing the stretch

Disciplines ranging from robotics to medical diagnostics and rehabilitation can require ways to detect changes in strain directly from the surface of an object. When that object is human skin, the sensor needs to be able to detect stretching in all directions, including complex combinations of bending and stretching at the joints, while remaining compliant. Lu *et al.* have developed a conducting, highly stretchable polymer composed of poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyaniline, and phytic acid. In addition to showing linear responses to omnidirectional tensile strain and bending, the polymer also has self-healing capabilities. Furthermore, the material has an elastic modulus comparable to that of human skin, making it possible to attach the sensors directly to the body. —MSL

ACS Appl. Mater. Interfaces **11**, 20453 (2019).

PHOTO: ALTO/ALAMY STOCK PHOTO

PLANT PATHOLOGY

The economic cost of an epidemic

Native tree populations in many parts of the world are threatened by alien pathogens, which are often imported inadvertently via the trade in living plants. Hill *et al.* estimate the likely economic cost of the current epidemic affecting ash (*Fraxinus excelsior*) in Britain. They find that ash dieback, caused by the fungal pathogen *Hymenoscyphus fraxineus* imported from continental Europe in ash saplings, may cost £14.8 billion over the next 100 years, with half of that amount accrued over the coming decade. Most of the cost is in lost economic services such as recreation, avoided runoff, and carbon sequestration, although there are also substantial costs incurred by felling dead trees and replanting. The authors point out that these potential costs dwarf the value of the plant trade. —AMS

Curr. Biol. **29**, R315 (2019).

GENETIC RISK

Gene expression can point to disease risk

Investigations of the genetic mechanisms underlying neuropsychiatric disease are difficult. Such studies involve a large number of variants of relatively small effect and unknown function. Gamazon *et al.* computationally examined expressed genes (the transcriptome) in multiple tissues, including the brain, adrenal glands, blood, and colon of 393 individuals. They identified several genetic variants associated with neuropsychiatric disease. The analysis also revealed novel genetic risk factors not previously linked to neuropsychiatric disorders. Some genes show disease-specific expression in nonbrain tissue, such as the adrenal glands and gut, highlighting the need for multiple tissue analyses to pinpoint the cause of these diseases. —LMZ

Nat. Genet. **55**, 933 (2019).

PLANT SCIENCE

Simple rules generate complex shapes

During development, complex shapes are elaborated from simple origins. Studying plant leaves, Kierzkowski *et al.* use live imaging and computational modeling to deconstruct the growth patterns of a simple leaf compared with those of a complex dissected leaf. Both sorts of leaf use the same tools for growth. The differing outcomes result from shifts in deployment of growth and differentiation programs. Delayed differentiation allows small protrusions to elaborate. Interspersed sites of growth inhibition accentuate protrusions. Thus, diversity in leaf shape is achieved by shifting the regulation of overall differentiation and local growth. The regulatory logic, resembling an incomplete feed-forward loop, might allow fine-tuning of shape without changing overall size. —PJH

Cell **177**, 1405 (2019).

EXOPLANETS

Same masses, different sizes

Exoplanets with masses between those of Earth and Neptune come in two types that display a bimodal distribution of densities: "super Earths," with rocky bodies and small (or no) atmospheres, and "mini Neptunes," with extended atmospheres of hydrogen and helium surrounding rocky cores. That difference might reflect the removal of light gasses from some of their atmospheres by stellar activity. Gandolfi *et al.* have discovered two exoplanets orbiting the star HD15337. Planet c is a mini Neptune with a 50% larger radius than planet b, a super Earth. As both planets have identical masses (within the uncertainties) and orbit the same star, studying them will help explain the atmospheric loss processes that lead to these planet types. —KTS

Astrophys. J. Lett. **876**, L24 (2019).

Complex leaves arise because of local zones of growth inhibition.

PHOTO: PETER HUIJSER AND MILTOS TSIANTIS, MAX PLANCK INSTITUTE FOR PLANT BREEDING RESEARCH

METABOLIC DISEASE

Fatty liver—too much of a bad thing?

Human genetic studies have been instrumental for developing drugs that prevent cardiometabolic disease. Key examples are statins and antibody-based therapies that lower serum levels of low-density lipoprotein (LDL) cholesterol. A similar paradigm may apply to fatty liver disease, an increasingly prevalent disorder that can progress to cirrhosis. Fatty liver is more common in individuals carrying a certain variant of the gene PNPLA3 (patatin-like phospholipase domain-containing protein 3) that leads to accumulation of the protein on lipid droplets in the liver. Whether this aberrant accumulation causes the disease was unknown. BasuRay *et al.* show that removing PNPLA3 from the liver by targeted protein degradation or by RNA interference ameliorates fatty liver in mice and may be a suitable target for therapeutic interventions. —PAK

Proc. Natl. Acad. Sci. U.S.A. **116**, 9521 (2019).

PHYSICS

A charmed violation

Our universe has much more matter than antimatter. This asymmetry requires a violation of combined charge and parity (CP) symmetries. Although the violation of CP symmetry has been observed before in weak interactions, its magnitude is insufficient to explain the size of matter-antimatter asymmetry. Aaij *et al.*, from the LHCb collaboration at the Large Hadron Collider, have now measured a small CP violation present in the decay of D^0 mesons, which consist of one charm and one antiup quark. The researchers found that the CP violation was more than five standard deviations away from a zero value and near the upper bound of the Standard Model prediction. The measurement should provide a valuable theoretical benchmark. —JS

Phys. Rev. Lett. **122**, 211803 (2019).

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

GAS GIANT PLANETS

Cassini's last look at Saturn's rings

During the final stages of the Cassini mission, the spacecraft flew between the planet and its rings, providing a new view on this spectacular system (see the Perspective by Ida). Setting the scene, Spilker reviews the numerous discoveries made using Cassini during the 13 years it spent orbiting Saturn. Iess *et al.* measured the gravitational pull on Cassini, separating the contributions from the planet and the rings. This allowed them to determine the interior structure of Saturn and the mass of its rings. Buratti *et al.* present observations of five small moons located in and around the rings. The moons each have distinctive shapes and compositions, owing to accretion of ring material. Tiscareno *et al.* observed the rings directly at close range, finding complex features sculpted by the gravitational interactions between moons and ring particles. Together, these results show that Saturn's rings are substantially younger than the planet itself and constrain models of their origin. —KTS

Science, this issue p. 1046, p. 1052, p. 1053, p. 1054; see also p. 1028

MICROBIOTA

The dope on L-dopa metabolism

The efficacy of L-dopa treatment for Parkinson's disease is hugely variable between individuals, depending on the composition of their microbiota. L-Dopa is decarboxylated into active dopamine, but if the gut microbiota metabolize L-dopa before it crosses the blood-brain barrier, medication is ineffective. Maini Rekdal *et al.* found that different species of bacterium are involved in L-dopa metabolism (see the Perspective by O'Neill). Tyrosine decarboxylase (TDC) from *Enterococcus faecalis* and dopamine dehydroxylase (Dadh) from *Eggerthella lenta* A2 sequentially metabolized L-dopa into *m*-tyramine. The microbial L-dopa decarboxylase can be inactivated by (*S*)-α-fluoromethyltyrosine (AFMT), which indicates possibilities for developing combinations of Parkinson's drugs to circumvent microbial inactivation. —CA

Science, this issue p. 1055; see also p. 1030

SYMBIOSIS

A little help from a friend

The Hawaiian sea slug *Elysia rufescens* grazes on an alga called *Bryopsis* sp. The alga defends itself from predators using peptide toxins decorated with fatty acids, called kahalalides. Zan *et al.* wondered if a third party was involved in toxin production (see the Perspective by Mascuch and Kubanek). Within the alga, a species of bacterium with a very reduced genome was discovered to be a factory for the nonribosomal assembly of a family of kahalalides. The authors elucidated the pathways for generating this chemical diversity. It seems that the sea slug not only tolerates the toxins but, to protect itself from being eaten by fish, grazes on the alga to accumulate kahalalide. —CA

Science, this issue p. 1056; see also p. 1034

PHASE-CHANGE MEMORY

Structural switch for fast switching

Phase-change materials are important for computer memory. They can quickly switch from glassy to crystalline using a thermal pulse and then lock in that structure for a long time at lower temperature. Zalden *et al.* probed the underlying atomic structure of two phase-change materials during this switching using ultrafast x-rays and simulations (see the Perspective by Rao *et al.*). A liquid-liquid phase transition in both materials allowed fast switching at high temperatures. The lower-temperature glass locks in the structure, allowing for long-term memory storage. —BG

Science, this issue p. 1062; see also p. 1032

PHYSICS

Driving strontium titanate ferroelectric

Hidden phases are metastable collective states of matter that are typically not accessible on equilibrium phase diagrams. Nova *et al.* used infrared pulses to excite higher-frequency lattice modes that drive the crystal into a metastable ferroelectric phase, a phase that can persist for many hours. X. Li *et al.* used terahertz fields to drive the soft mode that moves the ions in the crystal into the positions they occupy in the new phase. The ferroelectric phase in this case was transient, lasting on the order of 10 picoseconds. Because these hidden phases can host exotic properties in otherwise conventional materials, the accessibility to and control of such hidden phases may broaden potential functionality and applications. —ISO

Science, this issue p. 1075, p. 1079

CELL BIOLOGY

Surviving energetic stress with mTORC2

The roles and regulation of the mechanistic target of rapamycin complex 2 (mTORC2), a multiprotein complex that contains the kinase mTOR in association with rictor, have been enigmatic. Kazyken *et al.* identified the energy sensor adenosine monophosphate–activated protein kinase (AMPK) as a kinase that phosphorylated and activated mTORC2 (see the Focus by Jacinto). Activation of AMPK by energetic stress stimulated mTORC2 and its substrate Akt, thereby promoting cell survival. These results may help to explain why AMPK, which is typically thought of as a tumor suppressor, can act as a promoter of tumor growth in some contexts. —WW

Sci. Signal. **12**, eaav3249, eaax5855 (2019).

REVIEW

GAS GIANT PLANETS

Cassini-Huygens' exploration of the Saturn system: 13 years of discovery

Linda Spilker

The Cassini-Huygens mission to Saturn provided a close-up study of the gas giant planet, as well as its rings, moons, and magnetosphere. The Cassini spacecraft arrived at Saturn in 2004, dropped the Huygens probe to study the atmosphere and surface of Saturn's planet-sized moon Titan, and orbited Saturn for the next 13 years. In 2017, when it was running low on fuel, Cassini was intentionally vaporized in Saturn's atmosphere to protect the ocean moons, Enceladus and Titan, where it had discovered habitats potentially suitable for life. Mission findings include Enceladus' south polar geysers, the source of Saturn's E ring; Titan's methane cycle, including rain that creates hydrocarbon lakes; dynamic rings containing ice, silicates, and organics; and Saturn's differential rotation. This Review discusses highlights of Cassini's investigations, including the mission's final year.

Flybys of Saturn by Pioneer 11 in 1979 (*1*), Voyager 1 in 1980 (*2*), and Voyager 2 in 1981 (*3*) provided fleeting glimpses of the Saturn system. Shortly after the Voyager 2 flyby, concepts for an orbiter-and-probe mission like Cassini-Huygens were proposed by the planetary science community to provide answers to important questions raised by these earlier flybys—in particular, to study the unknown surface of Saturn's haze-enshrouded moon, Titan. In 1986, the National Research Council's Committee on Planetary and Lunar Exploration (COMPLEX) stated that the "highest priority for outer planet exploration in the next decade is intensive study of Saturn—the planet, satellites, rings, and magnetosphere—as a system" (*4*). Using the COMPLEX endorsement, NASA initiated joint studies with the European Space Agency (ESA), resulting in the Cassini-Huygens mission.

The motivation for planetary science missions such as Cassini-Huygens includes seeking answers to basic questions such as how planetary systems form, how they evolve through time, and where beyond Earth can requirements for life be found (*5*). The high-level Cassini scientific objectives were to conduct in-depth investigations of the Saturn system, including orbital remote sensing of Saturn's atmosphere, icy satellites, and rings; in situ orbital measurements of charged particles, dust particles, and magnetic fields; and detailed measurements with the Huygens probe during descent through Titan's atmosphere and on the surface. The Cassini-Huygens mission met or exceeded all of these objectives. Details of the mission, orbiter, probe, history, and instruments can be found in (*6*, *7*).

Cassini-Huygens was launched in 1997 and entered orbit around Saturn in 2004 (*8*), beginning its 13-year exploration of the Saturn system, which spanned almost half of a Saturn year, covering late northern winter, spring, and summer. Cassini performed a comprehensive survey of the Saturn system, including detailed mapping of Saturn's rings, moons, and magnetosphere, as well as seasonal studies. NASA's Cassini spacecraft carried ESA's Huygens probe, which landed on the moon Titan in 2005 (*9*). This international mission was a cooperative undertaking between NASA, ESA, and the Italian Space Agency [Agenzia Spaziale Italiana (ASI)].

Cassini's 4-year Prime Mission began its exploration of the Saturn system and raised new questions to address in the following extended missions. The 2-year Equinox Mission continued observations over the equinox crossing in August 2009, when Saturn's rings were edge-on to the Sun. In 2010, the Cassini Solstice Mission began the final 7 years of exploration. It sent the spacecraft on equatorial orbits with many targeted flybys of the icy satellites; inclined orbits with improved views of Saturn's rings and poles; and, finally, highly inclined ring-grazing orbits and the Grand Finale phase. In the Grand Finale, Cassini dove between the innermost D ring and the upper region of Saturn's atmosphere before the mission's end in 2017, a few months after northern summer solstice, with a deliberate crash into Saturn's atmosphere.

Cassini's encounters with Saturn's moons Enceladus and Titan led to findings such as potentially habitable environments in Enceladus' and Titan's subsurface water oceans (*10*–*12*); Enceladus' south polar geysers (*13*), the source of Saturn's E ring (*14*); and Titan's methane-based hydrologic cycle (*15*), with methane rain that creates hydrocarbon lakes and seas (*16*).

Other findings include three-dimensional structures in the planet's dynamic rings (*17*), a giant Saturn storm that completely encircled a northern latitude band for almost a year (*18*), a long-lived hexagonal jet stream encircling the north pole (*19*), and evidence that the captured moon Phoebe may have originated in the outer Solar System's Kuiper Belt (*20*).

The final year of the mission, when Cassini moved closer to Saturn than ever before, provided an array of discoveries regarding the interior of Saturn, its upper atmosphere and rings, and the gap between the rings and the planet.

Enceladus, an ocean world

Before Cassini's arrival, Enceladus, a small moon ~500 km in diameter, was known to reflect almost 90% of the sunlight it receives, making it the brightest moon in the Solar System. During

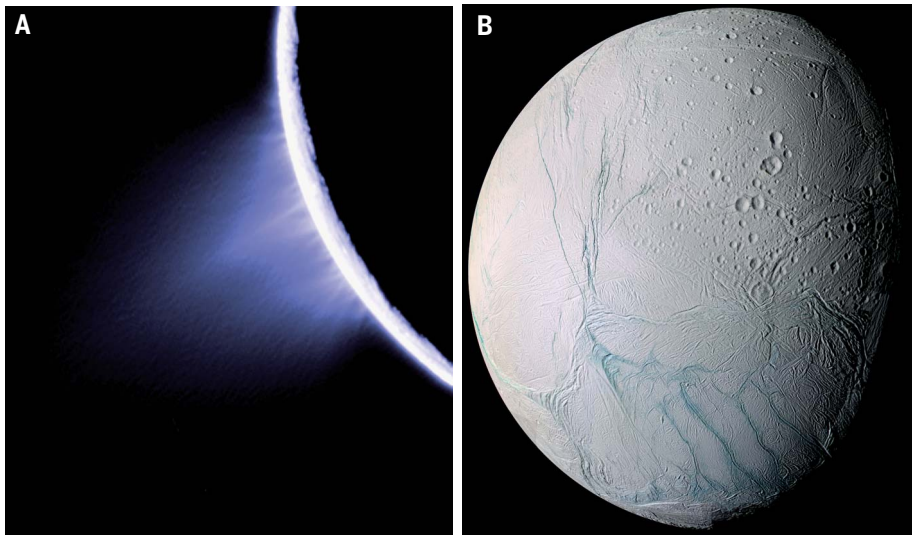


Fig. 1. Icy jets from Enceladus' south pole create the E ring. (**A**) Jets spewing ice particles and water vapor from Enceladus' south pole form a bluish plume in this false-color image, taken by Cassini in 2005 (*128*). (**B**) False-color full-disk view of Enceladus shows softened craters and tectonic fractures. South polar region containing blue-greenish tiger stripes is crater-free and young (*129*).

CREDITS: (A) NASA/JPL/SPACE SCIENCE INSTITUTE; (B) NASA/JPL/SPACE SCIENCE INSTITUTE

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA.
Email: linda.j.spilker@jpl.nasa.gov

Cassini's early flybys, images of the moon revealed a system of extensive cracks in the surface and large regions completely devoid of craters. The south polar region contains large tectonic features, is devoid of craters, and is the youngest surface on Enceladus. It includes a system of four nearly parallel fractures, nicknamed the "tiger stripes," centered near the south pole (*21*). Each tiger stripe is about 130 km long and 2 to 4 km wide and is separated from the others by ~35 km.

Activity on Enceladus was first detected by Cassini's magnetometer as a deflection of Saturn's magnetic field (*22*). During a close targeted flyby in 2005, Cassini's thermal infrared spectrometer discovered a hot spot centered on the south pole (*23*). The four tiger stripe fractures were the warmest regions—more than 100 K warmer than the surrounding areas (*24*)—and were the source of an immense ice plume (Fig. 1A) imaged by Cassini's camera (*13*). The giant plume of water vapor and ice grains is fed by both discrete jets (*25*) and curtains of material (*26*) originating inside each tiger stripe. The localized jets in the tiger stripes are the warmest regions detected on the moon (*27*). Enceladus' extensive water vapor and ice particle plume was unexpected, and later Cassini mission phases were adjusted to further investigate this feature. Cassini ultimately flew through the Enceladus plume seven times, directly sampling its gas (*28*) and icy particles (*29*).

The lack of impact craters in the tiger stripe region demonstrates that it is a geologically fresh surface that is constantly renewed (Fig. 1B). Most of the plume material eventually reimpacts Enceladus, but the smallest grains ejected in the plume are the source of Saturn's tenuous, distended E ring (*14*, *30*), which is densest at the orbit of Enceladus and spreads throughout the system. The E ring interacts with the inner moons, such as Mimas, Tethys, Dione, and Rhea, coating one side of them with E-ring grains (*31*). Cassini scientists discovered that the source of the plume is a global liquid water ocean beneath Enceladus' ice crust (*11*). The ocean has a depth of ~10 km, located beneath an ice shell that is mostly about 26 to 31 km thick but is considerably thinner, as little as 5 km, in the south polar region (*10*).

In addition to water vapor (*32*), Cassini's Ion and Neutral Mass Spectrometer (INMS) showed that the plume contains gaseous carbon dioxide and simple hydrocarbons such as methane, propane, and acetylene (*33*). During Cassini's closest dive through the plume in 2015, INMS also detected molecular hydrogen (H_2) (*34*). Cassini's Cosmic Dust Analyzer (CDA) found salt-rich ice grains containing sodium and potassium (*35*) that are probably frozen droplets from the underground salty ocean (*36*). CDA also detected tiny grains of silica, <10 nm in size, originating from Enceladus' ocean (*37*). These tiny silica grains most likely condensed from hot water spewing from hydrothermal vents on Enceladus' seafloor (*37*). The excess hydrogen discovered by INMS could also be produced by the hydrothermal vents. CDA and INMS also detected evidence for large organic fragments, indicative of complex organic molecules (*38*). The source of energy for the hydrothermal activity on Enceladus is tidal interactions among Enceladus, Dione, and Saturn (*39*). With liquid water, a hydrothermal energy source, and carbon-bearing organic molecules, the subsurface ocean of Enceladus is potentially habitable.

CREDIT: NASA/JPL/ESA/UNIVERSITY OF ARIZONA

Fig. 2. Huygens probe image taken just after landing on Titan's surface in 2005. This image of Titan's surface was generated with spectral reflection data added for color. The largest icy pebbles are 5 to 10 cm across. Flowing methane created their rounded shapes (*130*).

Titan, an Earth-like analog

Titan is the largest moon in the Saturn system, slightly larger than the planet Mercury. It is the only moon in the Solar System with a dense atmosphere, first characterized in detail by Voyager 1 (*40*, *41*). Titan and Earth are the only bodies that have surface liquids: water on Earth, liquid hydrocarbons on Titan. Titan's atmosphere is mostly nitrogen with some methane and a haze layer of organics that give this moon its orange color. Cassini mission results on Titan have been reviewed in two books (*42*, *43*).

The Huygens probe was built and operated by ESA and carried on the Cassini spacecraft. Huygens separated from Cassini in December 2004 and landed on Titan 3 weeks later, on 14 January 2005 (*44*). The probe's 2 hour 27 min parachute descent provided an in situ atmospheric profile of temperature, pressure, density, wind, and composition, as well as detailed images of the surface. The surface pressure was 1.47 times that on Earth (*45*); super-rotating prograde, zonal winds peaked at 430 km $hour^{-1}$, much greater than Titan's equatorial rotation velocity (*46*); and atmospheric composition included the noble gases argon, krypton, and xenon (*47*), whereas nitrogen and methane were confirmed as the primary constituents (*48*). After landing, the Gas Chromatograph Mass Spectrometer (GCMS) measured an increase in abundance of methane gas as the relatively warm GCMS inlet heated Titan's surface (*47*).

Cassini flew overhead to receive and relay the Huygens data, including data from ~72 min on Titan's surface, before Huygens' link to Cassini was lost as Cassini set over the horizon. Huygens landed in a dry lakebed filled with rounded pebbles (Fig. 2), near the Xanadu region, an equatorial area about the size of Australia. Images taken at an altitude of ~10 km captured erosional patterns on a hillside with very steep slopes. Rounded, smoothed pebbles at the landing site are evidence of fluid flow (*49*). For more results from the Huygens probe, see (*44*).

Cassini-Huygens revealed an array of complex hydrocarbons in Titan's atmosphere, created as methane gas is broken apart by sunlight in the upper atmosphere, generating active chemistry (*50*) and a link to haze formation (*51*). The mechanism by which methane is replenished remains a mystery. Over the course of the mission, methane rainfall darkened parts of Titan's surface (*52*, *53*), and methane clouds formed and dissipated (*54*). As the seasons changed, the evolution and breakup of Titan's northern winter polar vortex and the early formation of Titan's southern winter polar vortex were observed (*55*, *56*). The polar vortices appear to be tilted by a few degrees relative to the rotational pole of Titan, and the entire stratosphere is also tilted by several degrees (*57*).

The surface of Titan was observed from orbit using radar and imaging at infrared wavelengths, which showed broad regions of light and dark terrain (*58*) and geologic processes reminiscent of those on Earth. However, with surface temperatures around 93 K, methane plays the role of water on Earth, and solid water ice is an ingredient of Titan's bedrock. Titan's rain is composed of methane, its lakes and seas are filled with liquid methane and ethane (*59*), and its long, linear equatorial dunes are particles of organic solids or ice covered by organics (*60*, *61*). Some seas are about the size of North America's Great Lakes (*62*) and are ~160 m deep (*59*). Titan's hydrocarbon lakes and seas are found near the poles, with most of them located in the north (*16*). As northern summer approached, increasing surface winds (*63*) produced possible surface waves (*64*). Strong evidence exists for a subsurface liquid water ocean (*12*).

The moons Mimas, Iapetus, and Phoebe

The Saturn system contains 62 known moons, many of them small, captured objects that are

distant from Saturn and irregular in shape. Closer to Saturn are 24 regular moons that probably formed from the same gas-and-dust cloud as Saturn. Some of these moons have strong gravitational interactions with the ring system, opening gaps and sculpting the rings. Saturn's icy moons have been reviewed previously (*65*).

Mimas is the innermost and smallest of the intermediate-sized moons, with a diameter of 394 km. During a close flyby in February 2010, Cassini's infrared spectrometer discovered that the leading side of Mimas—the hemisphere that faces forward in the moon's orbit around Saturn—is ~15 K colder than the trailing side (*66*). This thermal anomaly, which has a shape reminiscent of Pac-Man, is the result of differing thermal inertia between the leading and trailing sides of Mimas. The leading face is altered by bombardment of highly energetic electrons, which increases the contact between regolith grains, decreases their porosity, and increases thermal inertia (*67*). Models of Saturn's E ring suggest that the side of Mimas that faces away from the direction of motion should be preferentially coated by particles from the E ring (*31*). Mimas' high reflectivity (it is one of the most reflective moons in the Solar System, after Enceladus) supports this hypothesis. Water ice is the primary compound detected on its surface (*68*). Its heavily cratered appearance demonstrates that there is little current geological activity on the moon (*69*).

Farther out in the Saturn system, one hemisphere of Iapetus is as dark as soot, whereas the other half is nearly as bright as fresh snow. The reason for this light–dark dichotomy was not known before Cassini's arrival at Saturn. Spitzer Space Telescope observations have shown that a very tenuous ring exists at the orbit of Phoebe (*70*), probably due to dust ejected from the moon by tiny meteor impacts. This dust is swept up by Iapetus as it orbits Saturn. The darkest side of the moon is centered in the direction of motion of the satellite. A handful of small impact craters on the dark side punch through to bright material below, suggesting that the dark surface material is only a few meters thick (*71*).

Cassini imaging showed a 20-km-high ridge that circles most of Iapetus' equator (*72*). The ridge breaks up into mountains in some of the dark regions. The ridge must have formed early in the history of Iapetus because it is heavily cratered and eroded. The surface of Iapetus is mainly water ice with small amounts of carbon dioxide, carbon, and complex organic molecules present (*73*).

Early in the mission, Cassini-Huygens flew close to Phoebe, the largest outer irregular moon with a diameter of 220 km. Phoebe moves in an inclined, elliptical, retrograde orbit 13 million kilometers from Saturn. It is covered with impact craters that were probably created by collisions with smaller outer moons. Some of the craters contain icy patches and layered structures; others have unusual conical shapes (*72*). Phoebe reflects only a few percent of the sunlight that falls on it, about the same as that reflected by the dark regions of Earth's Moon. In addition to water ice, Phoebe's surface is composed of carbon and carbon dioxide (*74*). The presence of carbon and its higher density than the other medium-sized moons of Saturn suggest that Phoebe was formed near the edge of the Solar System and migrated inward before being captured by Saturn (*20*). Phoebe might have originated in the Kuiper Belt, the reservoir of ice-and-rock bodies beyond the orbit of Neptune.

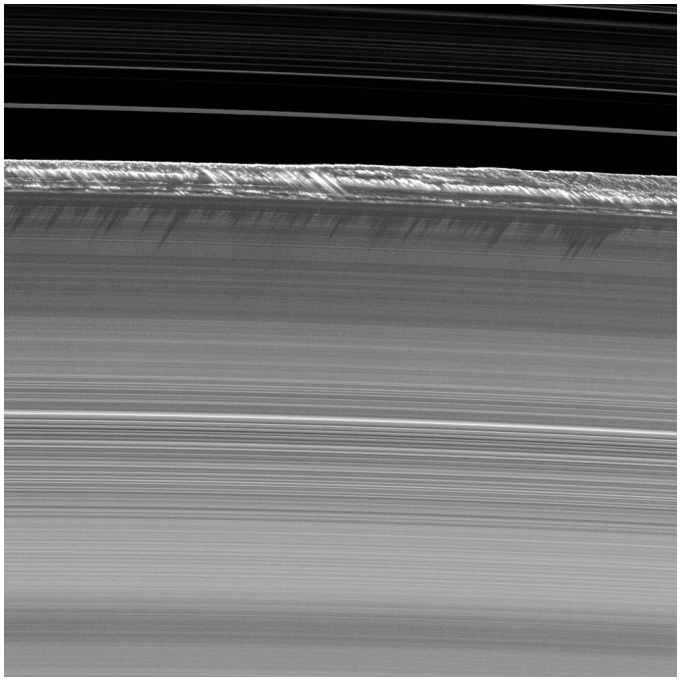


Fig. 3. Vertical structures rise from Saturn's B-ring edge. Among the tallest structures seen in Saturn's main rings, large objects rise abruptly from the edge of the B ring to cast long shadows across the ring. This image was taken by Cassini 2 weeks before the August 2009 equinox; sunlight is from above (*131*).

Discovery of seven small moons

Before the Cassini mission, Saturn was known to host a family of small satellites, including the co-orbital moons Janus and Epimetheus, which switch orbits every 4 years and appear to have been one body until their violent separation (*75*). Other previously known moons include the F-ring shepherds, Prometheus and Pandora, as well as Atlas, orbiting just outside the A-ring edge, and Pan, which clears the Encke gap in the rings (*76*).

Cassini data led to the discovery of seven additional small moons in the Saturn system: Pallene, Aegaeon, Anthe, Methone, Daphnis, S/2009 S 1, and Polydeuces (*77*). Four of these new moons are the sources of particles for co-orbiting rings or ring arcs. Pallene and Aegaeon create diffuse, dusty rings, whereas Anthe and Methone are associated with ring arcs (*78*, *79*). The tiny moon Daphnis was discovered orbiting in the Keeler gap in the A ring, may help to clear the gap, and produces horizontal and vertical waves along both gap edges (*80*). An even smaller moon (S/2009 S 1), only 300 m in diameter, was discovered at Saturn equinox by the shadow it cast inside the B ring (*81*). Aegaeon was discovered inside an arc in Saturn's G ring (*78*), whereas Polydeuces co-orbits with Dione in a gravitationally stable zone known as a Lagrangian point (*76*).

Rings

Saturn's rings are complex, and the processes observed there provide a laboratory for planet formation. Some of Cassini's highest-resolution data on the rings were obtained during Saturn orbit insertion (SOI) in 2004 as well as during the final year of the mission. Tiny propeller shapes and strawlike clumping in the strongest density wave peaks were detected in SOI images taken on the unilluminated side of the rings (*82*). Detailed reviews of the rings are available (*83*, *84*).

Cassini's 13 years in orbit provided an opportunity to observe temporal changes in Saturn's dynamic ring system on both the sunlit and unilluminated sides of the rings. Some regions in Saturn's rings changed on time scales of weeks or months. Dozens of objects (0.1 to 1 km) orbiting in the A ring shifted their locations as they interacted with neighboring material (*85*). Although these objects were too small to be seen directly, they opened propeller-shaped gaps as large as thousands of kilometers in length that were captured by the Cassini cameras and in stellar occultations of the rings (*85*). These migrating propeller-shaped structures (*85*), an analogy for protoplanets forming in a planetary nebula, were followed during the mission. Cassini also witnessed signs of the possible formation of a new moon at the outer edge of Saturn's A ring (*86*).

Throughout the mission, Cassini scientists monitored one of the most active, chaotic rings in our Solar System, Saturn's F ring (*87*). Clumps and dusty jets appeared and disappeared in the F-ring core, generated by embedded moonlets and disturbed by objects with orbits eccentric enough to dive through the F ring (*88*). Cassini observed channels opening and closing in the F ring in response to periodic close approaches by the tiny moon Prometheus (*89*).

Several new ringlets, composed mostly of fine dust grains, appeared during Cassini's 13 years

in orbit (*79*). One ringlet in the outer Cassini Division was barely visible when Cassini arrived but was among the dustiest features in the rings by the end of the mission. Meteoroids impacting Saturn's rings generated ejecta clouds of debris that may be a source of these dusty ringlets (*90*).

Once every half Saturn year the ring plane aligns with the center of the Sun, and for a brief time the northern and southern sides of the rings receive essentially no sunlight. During equinox in August 2009, when the Sun was edge-on to the rings, vertically extended objects cast shadows across the rings (*17*). During this period, Cassini observed shadows up to 2.5 km long (Fig. 3) created by objects larger than the 5-m vertical thickness of the rings, allowing the determination of the heights of structures within the rings, including kilometer-sized objects near the outer edge of the B ring, and the vertical extent of edge waves in the Keeler gap created by the tiny moon Daphnis (*84*). The rings also cooled to their lowest temperatures, as for a few days they were heated only by sunlight reflected from Saturn (*91*).

The particles forming Saturn's A and B rings are almost pure water ice but show a strong ultraviolet absorption that varies in strength across the rings, indicating variation in minor constituents (*92*). A reddish color of varying intensity in the rings is deeper where the ice signature is strongest (*93*). The rings probably darken with time as they are polluted by meteoroid bombardment. The less massive C ring and Cassini Division are redder than the more massive A and B rings (*94*).

Hundreds of stellar and radio occultations of the rings were obtained throughout the mission at a large variety of ring geometries. Detailed horizontal and vertical structure (*95*, *96*) and particle sizes (*97*) were revealed at multiple wavelengths, providing a detailed map of Saturn's rings. Three-dimensional measurements were taken of tendril-like, ephemeral structures in the rings called self-gravity wakes (*98*). These transient clumps form only briefly before being torn apart by Saturn's tides. Similar behavior in a protoplanetary disk might play a role in planet formation. A different kind of microstructure, which behaves like self-gravity but is due to viscous forces in the rings, was seen throughout the densest parts of the rings (*99*).

Modeling the damping behavior of dozens of spiral density and bending waves, generated by resonant interactions between ring particles and Saturn's nearby moons, constrained the mass of most of Saturn's main rings (*100*). However, occultations were not able to probe directly the densest parts of the B ring, so its mass remained uncertain. Gravity measurements during the Grand Finale orbits showed that the total mass of the rings was less than all previous estimates, hinting at a young age for the rings (*101*).

Saturn

Cassini's 13-year mission provided opportunities to observe the formation and dissipation of a giant storm, the characteristics of Saturn's polar vortices, northern hexagonal jet stream, lightning and aurora, and seasonal change. Cassini arrived 2 years after northern winter solstice and the mission ended just after northern summer solstice, providing almost two complete Saturn seasons of observations. Detailed reviews of Saturn are in (*102*). The shadows cast by the rings, originally covering the northern hemisphere, shifted through the equator at equinox and covered the southern hemisphere by the end of the mission. The northern winter hemisphere changed color from blue to golden as sunlight once again fell on that part of Saturn's atmosphere (*103*). Aurorae and lightning were also observed on Saturn. Curtains of auroral emission rose more than 1200 km above the planet (*104*). Lightning was generally associated with Saturn storms (*105*).

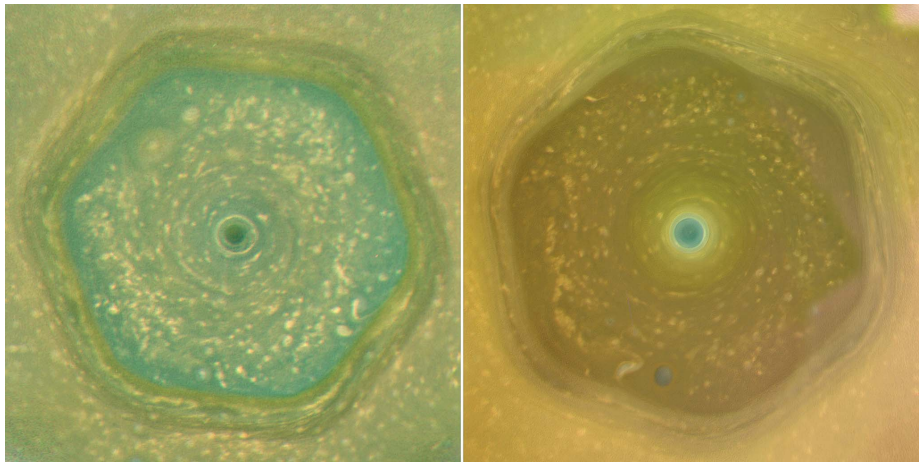
Late in 2010, a giant storm erupted quickly in Saturn's typically bland atmosphere. This type of storm occurs roughly every 30 years, but this one arrived 10 years early. Within months, the storm completely encircled the planet with a swirling band of clouds and vortices (*18*). Large temperature increases were measured in the stratosphere, and previously undetected molecules were observed in Saturn's upper atmosphere (*106*, *107*). The storm began to fade shortly after wrapping itself completely around one latitude band of the planet, about 9 months after it began (*102*).

Saturn's alternating eastward and westward jet streams define the bands of cloud that circle the planet. The bands follow lines of constant latitude up to within 1° of each pole. One of the jet streams, near 75° north latitude, forms a hexagonal pattern that is two Earth diameters across (*19*). Voyager first discovered the hexagon, and it is still there after 35 years. Small clouds move eastward around the corners of the pattern. This hexagonal-shaped jet stream (Fig. 4) is remarkable for its stability and longevity; its source remains a mystery. Using Cassini CIRS (Composite Infrared Spectrometer) thermal data of Saturn's north pole, a thermal hexagonal structure, precisely matching the well-known hexagon, was discovered towering hundreds of kilometers above the cloud tops (*108*). The presence of this thermal hexagon in Saturn's northern summer stratosphere, which is connected to the familiar jet stream hexagon in some way, suggests that there is more to be learned about the dynamics of Saturn's atmosphere.

Cassini scientists observed hurricane-like vortices, 50 times larger than a typical Earth hurricane, centered on both of Saturn's poles (*109*); these phenomena were stable throughout the mission. In the south, Cassini showed a hurricane-like vortex, with eyewall clouds rising 70 km above the clouds in the center (*110*). A warm vortex with well-developed eye walls circles Saturn's north pole as well (*111*).

Length of Saturn's day and probing its interior

Saturn emits long-wavelength radio waves known as Saturn kilometric radiation (SKR). SKR was first observed by Voyager in the early 1980s and interpreted as an indicator of Saturn's internal rotation period (length of day). Cassini results showed that the SKR signals are not coming from the interior of Saturn as originally assumed. The SKR period changed from year to year, incompatible with an interior origin. When Cassini first arrived at Saturn and measured the SKR period, data from the Radio and Plasma Wave Science instrument also showed that the radio waves' frequencies were different in the northern and southern hemispheres (*112*). Planetary period oscillations observed in magnetic field data had a similar variability: The period of these oscillations (although close to the expected planetary period of ~10.7 hours) changed over time, particularly with season, and was different in the northern and southern hemispheres (*113*, *114*).



CREDIT: NASA/JPL/SPACE SCIENCE INSTITUTE

Fig. 4. The hexagon-shaped jet stream encircling Saturn's north pole. This polar hexagon is about two Earth diameters across, more than 35 years old, and very stable. A hurricane-like storm circles at the center of the hexagon in this pair of Cassini images taken in two different seasons: 2013 (left) and 2017 (right) (*132*).

An alternative method for estimating the length of Saturn's day came from the detection of small, Saturn-driven waves in the rings (*115*). Saturn's ring system acts like a seismograph, providing a measure of Saturn's internal oscillations (or normal modes) that directly probe the interior of the planet (*116*) and provide a means for measuring its deep rotation rate. These vibrations, determined by Saturn's nonuniform internal structure, are probably driven by convection inside the planet, which cause oscillations in Saturn's gravity field that manifest themselves as waves in the rings. Preliminary modeling of the propagation behavior of this collection of waves provides an interior rotation rate for Saturn of ~10.6 hours (*117*).

The final year

As Cassini ran low on fuel, it embarked on a series of orbits that took it closer to Saturn. The spacecraft transmitted its final data on 15 September 2017, as it plunged into Saturn's atmosphere, vaporizing to avoid accidental contamination of the potentially habitable moons (*118*). Cassini's final phase covered roughly 10 months.

In late 2016, Cassini transitioned to a series of 20 ring-grazing orbits with closest approach (periapsis) just outside Saturn's F ring, facilitating close flybys of tiny ring moons and high-resolution views of Saturn's A and F rings. A final close Titan flyby in late April 2017 propelled the periapsis across Saturn's main rings to initiate the Grand Finale orbits. During these 22 orbits, Cassini repeatedly dove between Saturn's innermost D ring and the upper atmosphere. The last orbit turned the spacecraft into a Saturn atmospheric entry probe.

Close to Saturn

In the ring-grazing orbits, Cassini performed close flybys of the ring moons Pan, Daphnis, Atlas, Pandora, and Epimetheus. The surface characteristics of these moons are regulated by accretion of both a reddish material from Saturn's main rings and icy grains originating in the Enceladus plume (*119*). The color and brightness of the moons inside or closest to the main rings (Pan, Daphnis, and Atlas) strongly resemble that of the rings. Figure 5 shows a central core surrounded by an equatorial ridge of ring particles for each of these three moons.

The Cassini spacecraft passed close to Saturn's main rings during its final year, during which it obtained high–spatial resolution images, spectral scans, and temperature scans of the rings (*120*). High-resolution images revealed streaky C-ring plateaus and previously unseen bands of particle clumping throughout the rings. Weaker ice bands were identified in the outer A ring, outside the Keeler gap.

The top of the SKR emission region was sampled directly and repeatedly to determine its source (*121*) and elucidate how these planetary radio emissions are generated. SKR was strongly time-variable from orbit to orbit, with a dependence on local time around Saturn. Only three SKR source regions were identified, all on the dawn side of Saturn and controlled by the electron densities in the vicinity (*121*). These regions were embedded in upward currents associated with Saturn's auroral oval.

Saturn's interior

The Cassini Grand Finale at Saturn and the Juno mission at Jupiter provide a detailed view of the magnetic and gravity fields of these giant planets and offer windows into their interiors. Comparing these datasets will aid understanding of planet formation in the early Solar System.

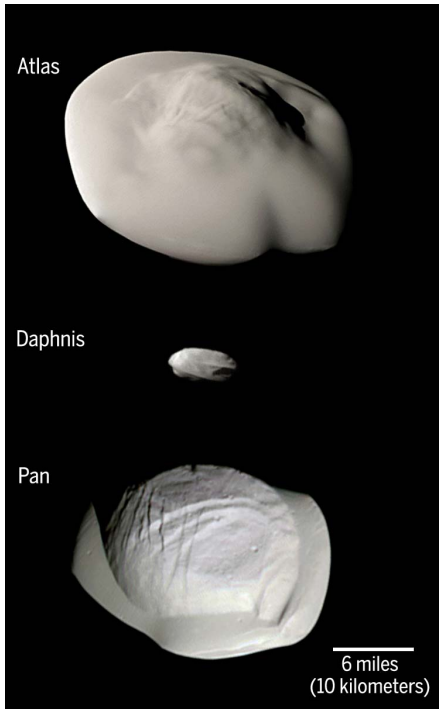


Fig. 5. Montage of Cassini color images of Atlas, Daphnis, and Pan. These moons were individually photographed for this montage during separate flybys (*133*). Their colors are similar to those of the ring particles that form their equatorial ridges and coat their surfaces.

Six Cassini orbits were optimized for gravity measurements to determine the mass distribution in Saturn's interior and the mass of the main rings. The measured values of Saturn's gravitational harmonics require deep differential rotation to a depth of ~9000 km inside the planet (down to ~0.7 Saturn radii), in line with Juno results for Jupiter (*122*). This depth may correspond to the levels of magnetic dissipation. The ring mass is about 1.5×10^{18} kg (0.41 Mimas masses), pointing to a ring age of perhaps 10^7 to 10^8 years (*101*).

Magnetic field measurements of Saturn's internal and external fields showed precise alignment (within 0.008°) between Saturn's spin-axis and its magnetic axis (*114*). This is considerably more axisymmetric than any other measured planetary magnetic field, which is inconsistent with current dynamo field theories. Magnetometer data provide insights into Saturn's conducting interior where zonal flows imply differential rotation, which is consistent with the gravity measurements (*101*). Higher-order magnetic moments suggest secondary dynamo action in the conducting interior of Saturn (*114*).

Magnetometer data also showed a strong, low-latitude, field-aligned current system situated between the inner edge of the D ring and the top of Saturn's atmosphere (*114*). This current is similar in strength to the currents observed in the auroral zone and, as such, may be part of a global current system.

Traversing the gap

While traversing the gap between Saturn and the rings, Cassini's in situ instruments directly measured the composition and mass flux of the infalling ring material at various latitudes. This material is composed of nanograin particles (1 to 20 nm), similar in size to small smoke particles, with the highest concentration within 2° of Saturn's equator (*123*, *124*). Near the equator, collisions with hydrogen atoms provide enough gas drag to decelerate the grains until they plunge into Saturn's atmosphere. The Magnetospheric Imaging Instrument (MIMI) measured at least 5 kg s^{-1} of 1- to 3-nm grains entering Saturn's equatorial atmosphere directly from the inner D ring, possibly from the bright D68 ringlet (*123*). INMS measured the volatile species derived from the equatorial ring grains—which included water, methane, ammonia, carbon monoxide and/or molecular nitrogen, and carbon dioxide—also entering Saturn's atmosphere along the ring plane (*125*). The estimated mass influx rate was 5000 to 40,000 kg s^{-1}, considerably higher than the MIMI estimates, which covered only a portion of the particle size range. This influx of organic-rich nanoparticles from the rings modifies Saturn's equatorial ionosphere and atmosphere (*125*).

Grains at higher latitudes are charged and transported along Saturn's magnetic field lines on each side of the rings (*124*), consistent with previously detected "ring rain" (*126*). These high-latitude grains are primarily impact ejecta from the B and C rings. CDA directly measured the mass and composition of the ~20-nm nanograins to characterize this material falling into Saturn. Two separate nanograin compositions were identified: water ice grains and silicate grains (*124*). Silicate grains accounted for about one-third of the identified nanograins, much greater than the estimated bulk silicate composition of the rings, which is a few percent (*94*). The absence of larger particles, observed in the rings farther from Saturn, indicates that some unidentified process is grinding up particles, gradually transforming the rings to become part of the planet.

During the final orbits, an inner radiation belt was detected by MIMI in the gap between the D ring and the top of Saturn's atmosphere (*127*). Saturn's main rings inhibit the inward passage of trapped charged particles that form

radiation belts. Hence, the radiation belts outside the main rings cannot interact with the inner radiation belt, providing an opportunity to study the inner belt and its interactions with the D ring. Ringlets in the D ring, including D68 and D73, control the structure and outer boundary, respectively, of this radiation belt.

Cassini-Huygens' 13-year exploration of the Saturn system has set the stage for future missions to Saturn, as well as to the ice giants, Uranus and Neptune. This mission leaves a legacy of discoveries that have changed our views of the Saturn system, how solar systems form, and the potential for life beyond Earth.

REFERENCES AND NOTES

1. A. G. Opp, *Science* **207**, 401–403 (1980).
2. E. C. Stone, E. D. Miner, *Science* **212**, 159–163 (1981).
3. E. C. Stone, E. D. Miner, *Science* **215**, 499–504 (1982).
4. National Research Council, *A Strategy for the Exploration of the Outer Planets: 1986-1996* (National Academies Press, 1986).
5. National Research Council, *Vision and Voyages for Planetary Science in the Decade 2013–2022* (National Academies Press, 2011).
6. C. T. Russell, Ed., *The Cassini-Huygens Mission*, volumes 1 to 3 (Kluwer Academic, 2003–2004).
7. M. Meltzer, *The Cassini-Huygens Visit to Saturn: An Historic Mission to the Ringed Planet* (Springer-Praxis, 2015).
8. D. L. Matson, L. J. Spilker, J. P. Lebreton, *Space Sci. Rev.* **104**, 1–58 (2002).
9. J.-P. Lebreton *et al*., *Nature* **438**, 758–764 (2005).
10. L. less *et al*., *Science* **344**, 78–80 (2014).
11. P. Thomas *et al*., *Icarus* **264**, 37–47 (2016).
12. L. less *et al*., *Science* **337**, 457–459 (2012).
13. C. C. Porco *et al*., *Science* **311**, 1393–1401 (2006).
14. C. J. Mitchell, C. C. Porco, J. W. Weiss, *Astron. J.* **149**, 156–172 (2015).
15. A. G. Hayes, R. D. Lorenz, J. I. Lunine, *Nat. Geosci.* **11**, 306–313 (2018).
16. A. G. Hayes, *Annu. Rev. Earth Planet. Sci.* **44**, 57–83 (2016).
17. H. Hoffmann, M. Seiß, H. Salo, F. Spahn, *Icarus* **252**, 400–414 (2015).
18. K. M. Sayanagi *et al*., *Icarus* **223**, 460–478 (2013).
19. K. H. Baines *et al*., *Planet. Space Sci.* **57**, 1671–1681 (2009).
20. T. V. Johnson, J. I. Lunine, *Nature* **435**, 69–71 (2005).
21. P. Helfenstein, *Nat. Geosci.* **3**, 75–76 (2010).
22. M. K. Dougherty *et al*., *Science* **311**, 1406–1409 (2006).
23. J. R. Spencer *et al*., *Science* **311**, 1401–1405 (2006).
24. C. J. A. Howett, J. R. Spencer, J. Pearl, M. Segura, *J. Geophys. Res.* **116**, E03003 (2011).
25. C. Porco, D. DiNino, F. Nimmo, *Astron. J.* **148**, 45–69 (2014).
26. J. N. Spitale, T. A. Hurford, A. R. Rhoden, E. E. Berkson, S. S. Platts, *Nature* **521**, 57–60 (2015).
27. J. R. Spencer, F. Nimmo, *Annu. Rev. Earth Planet. Sci.* **41**, 693–717 (2013).
28. J. H. Waite Jr. *et al*., *Science* **311**, 1419–1422 (2006).
29. F. Postberg, J. Schmidt, J. Hillier, S. Kempf, R. Srama, *Nature* **474**, 620–622 (2011).
30. S. Kempf, U. Beckmann, J. Schmidt, *Icarus* **206**, 446–457 (2010).
31. A. Hendrix, G. Filacchione, C. Paranicas, P. Schenk, F. Scipioni, *Icarus* **300**, 103–114 (2018).
32. C. J. Hansen *et al*., *Science* **311**, 1422–1425 (2006).
33. J. H. Waite Jr. *et al*., *Nature* **460**, 487–490 (2009).
34. J. H. Waite *et al*., *Science* **356**, 155–159 (2017).
35. F. Postberg *et al*., *Nature* **459**, 1098–1101 (2009).
36. C. R. Glein, J. A. Baross, J. H. Waite Jr., *Geochim. Cosmochim. Acta* **162**, 202–219 (2015).
37. H.-W. Hsu *et al*., *Nature* **519**, 207–210 (2015).
38. F. Postberg *et al*., *Nature* **558**, 564–568 (2018).
39. G. Tobie, O. Cadek, C. Sotin, *Icarus* **196**, 642–652 (2008).
40. A. Coustenis, B. Schmitt, R. K. Khanna, F. Trotta, *Planet. Space Sci.* **47**, 1305–1329 (1999).
41. G. Lindal *et al*., *Icarus* **53**, 348–363 (1983).
42. I. Müller-Wodarg, C. A. Griffth, E. Lellouch, T. E. Cravens, Eds., *Titan: Interior, Surface, Atmosphere, and Space Environment* (Cambridge Univ. Press, 2014)
43. R. Brown, J. Lebreton, J. Waite, Eds., *Titan from Cassini-Huygens* (Springer, 2010).
44. J. P. Lebreton *et al*., *Astron. Astrophys. Rev.* **17**, 149–179 (2009).
45. M. Fulchignoni *et al*., *Nature* **438**, 785–791 (2005).
46. M. K. Bird *et al*., *Nature* **438**, 800–802 (2005).
47. H. B. Niemann *et al*., *Nature* **438**, 779–784 (2005).
48. H. B. Niemann *et al*., *J. Geophys. Res.* **115**, E12006 (2010).
49. M. G. Tomasko *et al*., *Nature* **438**, 765–778 (2005).
50. J. H. Waite Jr. *et al*., *Science* **316**, 870–875 (2007).
51. P. Lavvas *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2729–2734 (2013).
52. E. P. Turtle *et al*., *Geophys. Res. Lett.* **36**, L02204 (2009).
53. E. P. Turtle *et al*., *Science* **331**, 1414–1417 (2011).
54. S. Rodrigues *et al*., *Nature* **459**, 678–682 (2009).
55. S. Vinatier *et al*., *Icarus* **250**, 95–115 (2015).
56. R. K. Achterberg, P. J. Gierasch, B. J. Conrath, F. Michael Flasar, C. A. Nixon, *Icarus* **211**, 686–698 (2011).
57. R. K. Achterberg, B. J. Conrath, P. J. Gierasch, R. M. Flasar, C. A. Nixon, *Icarus* **197**, 549–555 (2008).
58. F. Tosi *et al*., *Icarus* **208**, 366–384 (2010).
59. M. Mastrogiuseppe *et al*., *Geophys. Res. Lett.* **41**, 1432–1437 (2014).
60. J. Radebaugh *et al*., *Icarus* **194**, 690–703 (2008).
61. R. D. Lorenz *et al*., *Science* **312**, 724–727 (2006).
62. E. R. Stofan *et al*., *Nature* **445**, 61–64 (2007).
63. R. D. Lorenz, A. G. Hayes, *Icarus* **219**, 468–475 (2012).
64. J. W. Barnes *et al*., *Planet. Sci.* **3**, 3 (2014).
65. P. Schenk, R. Clark, C. J. Howett, A. Verbiscer, J. Waite, Eds., *Enceladus and the Icy Moons of Saturn* (Space Science Series, Univ. of Arizona Press, 2018).
66. C. J. A. Howett *et al*., *Icarus* **216**, 221–226 (2011).
67. P. Schenk *et al*., *Icarus* **211**, 740–757 (2011).
68. G. Filacchione *et al*., *Icarus* **186**, 259–290 (2007).
69. M. R. Kirchoff, P. Schenk, *Icarus* **206**, 485–497 (2010).
70. A. J. Verbiscer, M. F. Skrutskie, D. P. Hamilton, *Nature* **461**, 1098–1100 (2009).
71. T. Denk *et al*., *Science* **327**, 435–439 (2010).
72. C. C. Porco *et al*., *Science* **307**, 1237–1242 (2005).
73. R. N. Clark *et al*., *Icarus* **218**, 831–860 (2012).
74. R. N. Clark *et al*., *Nature* **435**, 66–69 (2005).
75. M. S. Tiscareno, P. C. Thomas, J. A. Burns, *Icarus* **204**, 254–261 (2009).
76. P. C. Thomas, *Icarus* **208**, 395–401 (2010).
77. P. C. Thomas *et al*., *Icarus* **226**, 999–1019 (2013).
78. M. M. Hedman *et al*., *Icarus* **207**, 433–447 (2010).
79. M. M. Hedman *et al*., *Icarus* **199**, 378–386 (2009).
80. C. C. Porco, P. C. Thomas, J. W. Weiss, D. C. Richardson, *Science* **318**, 1602–1607 (2007).
81. C. Porco, Cassini Imaging Team, "S/2009 S1." *IAU Circular* **9091** (2 November 2009).
82. C. C. Porco *et al*., *Science* **307**, 1226–1236 (2005).
83. M. Dougherty, L. Esposito, S. M. Krimigis, Eds., *Saturn from Cassini-Huygens* (Springer, 2009).
84. M. Tiscareno, C. Murray, Eds., *Planetary Ring Systems: Properties, Structure, and Evolution* (Cambridge Univ. Press, 2018).
85. M. S. Tiscareno *et al*., *Astrophys. J.* **718**, L92–L96 (2010).
86. C. D. Murray, N. J. Cooper, G. A. Williams, N. O. Attree, J. S. Boyer, *Icarus* **236**, 165–168 (2014).
87. C. D. Murray *et al*., *Nature* **453**, 739–744 (2008).
88. K. Beurle *et al*., *Astrophys. J.* **718**, L176–L180 (2010).
89. C. D. Murray *et al*., *Nature* **437**, 1326–1329 (2005).
90. M. S. Tiscareno *et al*., *Science* **340**, 460–464 (2013).
91. L. Spilker, C. Ferrari, R. Morishima, *Icarus* **226**, 316–322 (2013).
92. G. Filacchione *et al*., *Icarus* **220**, 1064–1096 (2012).
93. G. Filacchione *et al*., *Icarus* **241**, 45–65 (2014).
94. J. N. Cuzzi *et al*., *Science* **327**, 1470–1475 (2010).
95. M. S. Tiscareno, B. E. Harris, *Icarus* **312**, 157–171 (2018).
96. M. M. Hedman, J. A. Burns, M. R. Showalter, *Icarus* **248**, 137–161 (2015).
97. J. E. Colwell, L. W. Esposito, J. H. Cooney, *Icarus* **300**, 150–166 (2018).
98. J. E. Colwell, L. W. Esposito, M. Sremcevic, *Geophys. Res. Lett.* **33**, L07201–L07204 (2006).
99. J. E. Colwell, L. W. Esposito, M. Sremcevic, G. R. Stewart, W. E. McClintock, *Icarus* **190**, 127–144 (2007).
100. M. S. Tiscareno, J. A. Burns, P. D. Nicholson, M. M. Hedman, C. C. Porco, *Icarus* **189**, 14–34 (2007).
101. L. less *et al*., *Science* eaat2965 (2019).
102. K. Baines, F. Flasar, N. Krupp, T. Stallard, Eds., *Saturn in the 21st Century* (Cambridge Univ. Press, 2018).
103. K. M. Sayangi, *New Astron.* **1**, 584 (2017).
104. U. A. Dyudina, A. P. Ingersoll, S. P. Ewald, D. Wellington, *Icarus* **299**, 117–132 (2018).
105. U. A. Dyudina *et al*., *Icarus* **226**, 1020–1037 (2013).
106. J. I. Moses *et al*., *Icarus* **261**, 149–168 (2015).
107. L. N. Fletcher *et al*., *Science* **332**, 1413–1417 (2011).
108. L. N. Fletcher *et al*., *Nat. Commun.* **9**, 3564 (2018).
109. K. M. Sayanagi, J. J. Blalock, U. A. Dyudina, S. P. Ewald, A. P. Ingersoll, *Icarus* **285**, 68–82 (2017).
110. U. A. Dyudina *et al*., *Science* **319**, 1801 (2008).
111. L. N. Fletcher *et al*., *Science* **319**, 79–81 (2008).
112. D. A. Gurnett *et al*., *Geophys. Res. Lett.* **36**, L16102 (2009).
113. G. Provan, S. W. H. Cowley, J. D. Nichols, *J. Geophy. Res.* **114**, A04205 (2009).
114. M. K. Dougherty *et al*., *Science* **362**, 46 (2018).
115. M. Hedman, P. Nicholson, *Astron. J.* **146**, 12 (2013).
116. J. Fuller, *Icarus* **242**, 283–296 (2014).
117. C. Mankovich, M. S. Marley, J. J. Fortney, N. Movshovitz, *Astrophys. J.* **871**, 1–17 (2019).
118. COSPAR's Planetary Protection Policy, *Space Res. Today* **200**, 12–25 (2017).
119. B. J. Buratti *et al*., *Science* **364**, eaat234 (2019).
120. M. S. Tiscareno *et al*., *Science* **364**, *eaau1017* (2019).
121. L. Lamy *et al*., *Science* **362**, eaat2027 (2018).
122. T. Guillot *et al*., *Nature* **555**, 227–230 (2018).
123. D. G. Mitchell *et al*., *Science* **362**, eaat2236 (2018).
124. H.-W. Hsu *et al*., *Science* **362**, 49 (2018).
125. J. H. Waite Jr. *et al*., *Science* **362**, eaat2382 (2018).
126. J. O'Donoghue *et al*., *Nature* **496**, 193–195 (2013).
127. E. Roussos *et al*., *Science* **362**, eaat1962 (2018).
128. Jet Propulsion Laboratory, PIA08386: Jet Blue (2007); https://photojournal.jpl.nasa.gov/catalog/?IDNumber=PIA08386.
129. Jet Propulsion Laboratory, PIA06254: Zooming in on Enceladus (2005); https://photojournal.jpl.nasa.gov/catalog/?IDNumber=PIA06254.
130. Jet Propulsion Laboratory, PIA07232: First Color View of Titan's Surface (2005); https://photojournal.jpl.nasa.gov/catalog/PIA07232.
131. Jet Propulsion Laboratory, PIA11668: The Tallest Peaks (2010); https://photojournal.jpl.nasa.gov/catalog/PIA11668.
132. Jet Propulsion Laboratory, PIA14945: Spring at the North Pole (2013); https://photojournal.jpl.nasa.gov/catalog/PIA14945.
133. Jet Propulsion Laboratory, PIA21449: Small Wonders, Fig. 1 (2017); https://photojournal.jpl.nasa.gov/catalog/PIA21449.
134. L. Spilker, "The scientific achievements of the Cassini-Huygens mission" in *Serendipities in the Solar System and Beyond*, C.-M. Ko, C.-K. Chang, P.-C. Yu, Eds. (Astronomical Society of the Pacific Conference Series, 2018), vol. 513, p. 131.

ACKNOWLEDGMENTS

I thank S. Edgington and T. Spilker for helpful conversations and feedback on this paper. I also gratefully acknowledge all of the Cassini team members who designed, developed, and operated the Cassini-Huygens mission, which is a joint endeavor of NASA, the European Space Agency (ESA), and the Italian Space Agency (ASI) and is managed by JPL/Caltech under a contract with NASA. This Review is based on (*134*), with permission of the Astronomical Society of the Pacific Conference Series. **Funding:** The research described in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. United States government sponsorship is acknowledged.
Competing interests: There are no competing interests.

10.1126/science.aat3760

RESEARCH ARTICLE SUMMARY

GAS GIANT PLANETS

Measurement and implications of Saturn's gravity field and ring mass

L. Iess*, B. Militzer, Y. Kaspi, P. Nicholson, D. Durante, P. Racioppa, A. Anabtawi, E. Galanti, W. Hubbard, M. J. Mariani, P. Tortora, S. Wahl, M. Zannoni

INTRODUCTION: The interior structure of Saturn, the depth of its winds, and the mass and age of its rings have been open questions in planetary science. The mass distribution inside a fluid and rapidly rotating planet like Saturn is largely driven by the ratio between centrifugal and gravitational forces. In static conditions, the planet should rotate uniformly and its gravity field should be axially and hemispherically symmetric and thus described by even zonal harmonics. However, optical tracking of clouds in the atmospheres of gas giants indicates that their rotation is not uniform. If the velocity field seen at cloud level extends deep into the interior, then there is a redistribution of mass that modifies the gravity field. Gravity measurements therefore constrain both Saturn's deep interior and the depth of its winds. They also provide the mass of its rings, which dynamical and compositional dating methods have shown is related to their age.

RATIONALE: In its final 22 orbits, the Cassini spacecraft passed between the rings and the atmosphere of Saturn. In six of these orbits, while the spacecraft was in free fall under the combined attraction of Saturn and its rings, radio tracking from an antenna on Earth was established to measure the radial velocity of the spacecraft with accuracies between 0.023 and 0.088 mm·s^{-1} at a 30 s integration time. Fitting these data with a dynamical model of the forces acting on the spacecraft allowed us to determine the separate signatures from each zonal harmonic and the ring mass.

RESULTS: The estimated quantities were a zonal gravity model for Saturn and the gravity from a uniformly distributed mass spanning the A, B, and C rings. Additional small accelerations of unknown origin (possibly related to a time-variable gravity field) were required to obtain a good fit of the data. The determination of Saturn's static zonal coefficients and ring mass does not depend on the assumed source of these small effects. We found that the measured values of J_6, J_8, and J_{10} are so large that they cannot be matched with interior models relying on uniform rotation and plausible compositions, but they are in agreement with interior models that assume deep differential rotation, extending from the equator into the interior up to distances of 0.7 to 0.8 Saturn radii from the spin axis. The equatorial outer layers of Saturn must rotate at an angular velocity that is 4% faster than that of the deep interior, whereas regions at higher latitudes must rotate 1 to 2% slower, regardless of the assumed rotation period. A thermal-wind approach shows that flow profiles that are similar in general character to the observed one yield solutions within the uncertainty of the gravity measurements when extended to a depth of ~9000 km, confirming that the flows are very deep and likely extend down to the levels where magnetic dissipation occurs.

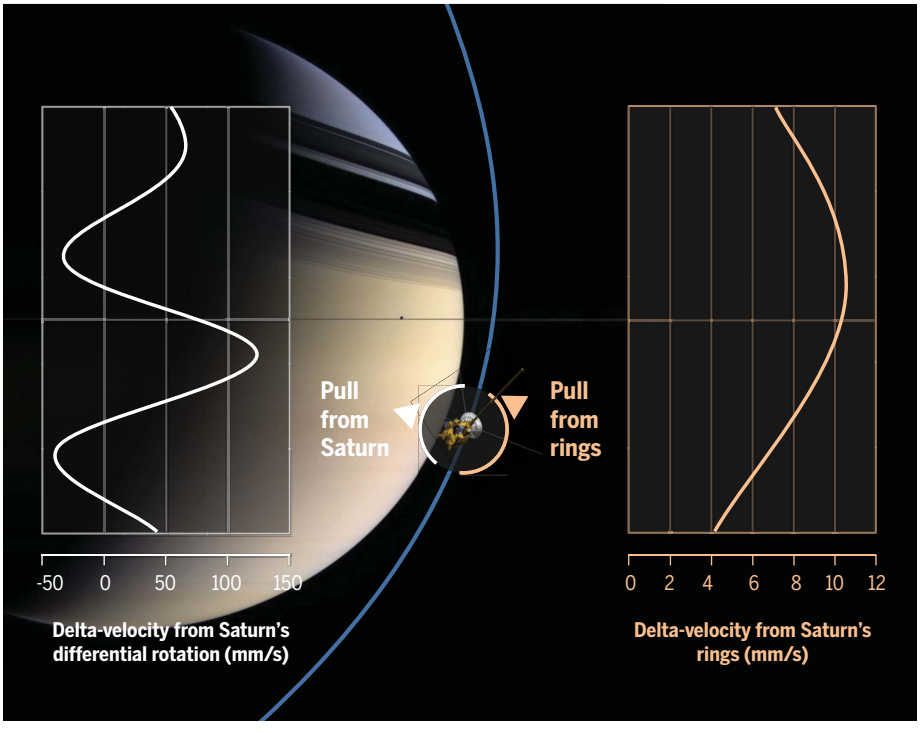
We also found that the mass of the rings is 0.41 times that of the Saturnian moon Mimas, which is at the lower end of the expected values.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat2965

CONCLUSION: Saturn's gravity field measured by Cassini implies a strong and deep differential rotation, extending to a depth of ~9000 km. This differs from Jupiter, where winds are shallower (~3000 km). The gravity measurements are consistent with a mass of Saturn's core of 15 to 18 Earth masses.

The low value of the ring mass suggests a scenario where the present rings of Saturn are young, probably just 10 million to 100 million years old, to be consistent with their pristine icy composition. Nevertheless, the rings may have evolved substantially since their formation and were perhaps once more massive than they are today. Models for a young ring system invoke the chance capture and tidal disruption of a comet or an icy outer Solar System body, suggesting that catastrophic events continued to occur in the Solar System long after its formation 4.6 billion years ago. ■



During the Grand Finale phase of its mission, Cassini passed between the inner edge of Saturn's D ring and the cloud top, enabling the measurement of the ring mass. This orbit allowed gravitational acceleration of the rings to be disentangled from the much larger acceleration resulting from Saturn's oblateness, as they pull the spacecraft in opposite directions. The curves show the spacecraft velocity variation due to the rings and the differential rotation.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: luciano.iess@uniroma1.it
Cite this article as L. Iess *et al.*, *Science* **364**, eaat2965 (2019). DOI: 10.1126/science.aat2965

RESEARCH ARTICLE

GAS GIANT PLANETS

Measurement and implications of Saturn's gravity field and ring mass

L. Iess[1]*, B. Militzer[2,3], Y. Kaspi[4], P. Nicholson[5], D. Durante[1], P. Racioppa[1], A. Anabtawi[6], E. Galanti[4], W. Hubbard[7], M. J. Mariani[1], P. Tortora[8], S. Wahl[3], M. Zannoni[8]

The interior structure of Saturn, the depth of its winds, and the mass and age of its rings constrain its formation and evolution. In the final phase of the Cassini mission, the spacecraft dived between the planet and its innermost ring, at altitudes of 2600 to 3900 kilometers above the cloud tops. During six of these crossings, a radio link with Earth was monitored to determine the gravitational field of the planet and the mass of its rings. We find that Saturn's gravity deviates from theoretical expectations and requires differential rotation of the atmosphere extending to a depth of at least 9000 kilometers. The total mass of the rings is (1.54 ± 0.49) × 10^{19} kilograms (0.41 ± 0.13 times that of the moon Mimas), indicating that the rings may have formed 10^{7} to 10^{8} years ago.

The mass distribution inside a fluid and rapidly rotating planet like Saturn is largely driven by the ratio between centrifugal and gravitational forces. In the absence of internal dynamics, axial and hemispherical symmetry is expected, implying that when the gravitational potential is decomposed into spherical harmonics, only the even zonal harmonics appear (zonal harmonics are longitude-independent). Assuming hydrostatic equilibrium, interior models of gas giant planets indicate that the zonal coefficients J_{2k} can be approximated by $J_{2k} = a_k q^k$, where q is the ratio of the centrifugal and gravitational acceleration at the equator (~0.16 for Saturn), k is a positive integer, and the coefficients a_k depend on the density profile inside the planet (*1*).

Optical tracking of clouds indicates that dynamical phenomena operate on Saturn and Jupiter. The measured zonal (west-east) wind velocity field suggests a state of differential rotation (DR), whereby the angular velocity at any location depends on its distance from the axis of rotation and the depth along this axis (*2*, *3*). If the velocity field seen at the top of the atmosphere (conventionally defined as the level of 1 bar pressure) continues into the interior, then internal dynamics are expected to affect the gravitational field in two ways. First, the equipotential surfaces are perturbed symmetrically, redistributing mass in such a way that the even zonal coefficients deviate from the relation $J_{2k} \sim q^k$ (*2*). Second, any north-south asymmetry in the velocity field leads to nonzero values of the odd zonal harmonics (*4*). These theoretical expectations have been confirmed by the Juno mission at Jupiter (*5*–*7*), where gravity measurements showed that zonal winds are 2000 to 3000 km deep and suggest that the heavy-element core is diffuse (*8*).

Gravity measurements at Saturn can be used to determine the mass of its rings, which is related to their age, by dynamical and compositional dating methods (*9*–*11*). Before the Grand Finale phase of the mission, the pericenter of Cassini's orbit was always outside Saturn's A ring, so that the gravitational effects of the rings could not be separated from those of the oblateness of the planet. During the Grand Finale, Cassini flew between the planet and the rings. This geometry effectively breaks the degeneracy between the even zonal field and the mass of the rings, providing a direct, dynamical estimate of the ring mass.

Cassini gravity measurements

We determined Saturn's gravitational field by reconstruction of Cassini's trajectory during the Grand Finale using a coherent microwave link between Earth tracking stations and the spacecraft. Range-rate measurements were obtained from the Doppler shift of a carrier signal sent from the ground at 7.2 GHz (X band) and retransmitted back to Earth by Cassini's onboard transponder at 8.4 GHz. An auxiliary downlink at Ka band (32.5 GHz) was also recorded.

In April 2017, Cassini was inserted into a series of inclined, highly eccentric orbits, grazing Saturn's cloud tops at each pericenter (table S1). The orbit nodes were chosen such that the angle between the orbit normal and the direction to Earth is close to 90°. This edge-on condition provides the maximum projection of the spacecraft velocity along the line of sight, making it optimal for range-rate measurements and gravity estimations.

Of the 22 Grand Finale orbits (designated Rev 271 through Rev 293), six were selected for gravity measurements. Five orbits (Revs 273, 274, 278, 280, and 284) provided useful data (data from Rev 275 were lost because of a station configuration error). These orbits were selected to minimize neutral particle drag and maximize the spacecraft viewing period around closest approach (C/A).

We produced an orbital solution based on 5 data arcs, using Doppler observations with count times of 30 s, spanning a period of 24 to 36 hours enclosing each closest approach. Tracking data were acquired by the antennas of the three complexes (Goldstone, USA; Madrid, Spain; and Canberra, Australia) of NASA's Deep Space Network (DSN), and two deep space antennas from the European Space Agency's ESTRACK network, located in the southern hemisphere (Malargüe, Argentina, and New Norcia, Australia).

Two-way Doppler measurements at X band make up 93% of the dataset (*12*), with the addition of a few three-way passes to fill gaps during ground station handovers. Available X-band range observables are also included. All data with either uplink or downlink elevation below 15° were discarded to avoid systematic measurement errors due to imperfect calibration of path delays in Earth's troposphere. Calibrations of the tropospheric and ionospheric path delays were provided by the DSN on the basis of pressure, humidity, and Global Positioning System data. Noise from solar plasma, which can strongly affect X-band radio links, was low owing to the large solar elongation angle (>142° on all six gravity orbits) (*13*). The data quality was statistically equivalent in X- and Ka-band data, with a root mean square Doppler noise between 0.020 and 0.088 mm·s^{-1} at a 30-s integration time (fig. S1 and table S1). For comparison, the Doppler signals due to the weakest measurable harmonics (J_3, J_{10}) and Saturn's ring are 40 to 200 times larger than the average Doppler noise (fig. S2).

The X-band Doppler data were favored in our analysis because of the higher signal-to-noise ratio during the unavoidable ring occultation periods. The radio link was maintained during the occultations, except for a blockage of ~10 min on each orbit when the signal crossed the opaque core of the optically thick B ring (region B3). A second, longer blockage period due to ring occultations occurred on the outbound leg of the gravity orbits. Diffraction and near-forward scattering of the radio signal by ring particles caused an increase in the Doppler noise at X band by a factor of 2 to 3 during the occultation periods. Ka band is more sensitive to this effect and suffered repeated signal losses.

Dynamical model

Our orbital fitting is based on the dynamical model previously adopted and tested in the determination of the gravity fields of Titan (*14*) and

[1]Department of Mechanical and Aerospace Engineering, Sapienza Università di Roma, Rome 00184, Italy. [2]Department of Astronomy, University of California, Berkeley, Berkeley, CA 94720, USA. [3]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720, USA. [4]Department of Earth and Planetary Sciences Weizmann Institute of Science, Rehovot 76100, Israel. [5]Department of Astronomy, Cornell University, Ithaca, NY 14853, USA. [6]Jet Propulsion Laboratory–Caltech, Pasadena, CA 91109, USA. [7]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [8]Department of Industrial Engineering, Università di Bologna, Forlì 47100, Italy.
*Corresponding author. Email: luciano.iess@uniroma1.it

Enceladus (*15*), implemented in the Jet Propulsion Laboratory (JPL) navigation code MONTE (*16*). The model was extended to include Saturn's gravitational parameter GM (where G is the gravitational constant and M is its mass), the zonal harmonic coefficients J_2 to J_{20}, the tesseral (longitude-dependent) field of degree 2 (to account for possible nonprincipal axis rotation), and the mass of the rings. The truncation of the zonal field was set to twice the degree of the highest gravity harmonic whose central value is above the uncertainty (degree 10). The estimate of the even zonal harmonics used in our interpretation (degree ≤ 10) is insensitive to changes in the truncation. Although this model was adequate for fitting Juno gravity data at Jupiter (*5*), obtained in an orbital configuration similar to Cassini's Grand Finale, it could not reduce Cassini's Doppler residuals to the noise level. Instead, small stochastic accelerations were added to the model (see below) to produce signature-free residuals (fig. S1).

Although all the rings are included in the dynamical model, only rings A, B, and C can produce an acceleration potentially detectable by Cassini. The rings are assumed to be coplanar with Saturn's equator, and each is assumed to have a constant surface mass density (the data are insensitive to radial variations of the density). Space- and ground-based measurements of ring occultations provide the determination of Saturn's spin axis position and precession rate (*17*). Although Doppler data are sensitive to the orientation of Saturn's equatorial plane, we adopt the prior determination because it is about ten times more accurate than that obtained from our orbital fitting.

Accelerations that are known to be large enough to produce noticeable signatures on range-rate data were accounted for. These include the point-mass gravitational accelerations from Saturn and its satellites (including the ring moons), computed from the JPL planetary and satellite ephemerides DE430, SAT389, and SAT393 (*18*), the acceleration from the Sun, the planets and satellites of the Solar System, and Saturn's tidal response to its satellites (*12*).

The acquired Doppler data were combined in a multi-arc, weighted least-squares estimation filter. In the multi-arc approach, the entire time span of the observations is decomposed into shorter intervals, and a distinction is made between global and local estimated parameters. Global parameters are common to all arcs and estimated using all available observables. These include Saturn's GM, J_2 through J_{20}, C_{21}, C_{22}, S_{21}, S_{22}, and the masses of the A, B, and C rings. Although the estimates of the masses of the three rings are highly correlated, their sum is well determined. The ring masses were initially set to the current best estimates of their values from ring occultation data (*19*–*21*), with an a priori uncertainty of 100% (see table S2). The mass of the B ring had a large uncertainty of 10 Mimas masses (Saturn's moon Mimas has a GM of 2.5026 $km^3 \cdot s^{-2}$). The sum of the masses of Saturn and its rings was constrained to be equal to the value (and uncertainty) estimated from satellite ephemerides (*18*).

Local parameters are those that belong only to a single arc and a different value was estimated for each arc. These included the Cassini position and velocity when the gravity observations began in each of the five orbits, at least 12 hours before transit at pericenter. The a priori uncertainties for position and velocity were set at 100 km and 1 $m \cdot s^{-1}$, respectively.

In the estimation filter, parameters whose uncertainties are large enough to contribute to the final covariance matrix have been included as consider parameters. These include the Love number k_{22} (determining Saturn's tidal response), Saturn's pole direction, accelerations due to thermal emission from Cassini's radioisotope thermoelectric generators (RTGs), and solar radiation pressure. The nominal value and a priori uncertainty of the Love number are taken from previous Cassini constraints (*22*), as were the pole position and precession rate (*17*). The anisotropic thermal acceleration from the RTGs was determined to 5% or better during the Cassini mission by the navigation team. The relative uncertainty associated with solar radiation pressure acceleration was set to 20%.

Gravity determination

Our deterministic model, based on the geophysical expectations for the gravity field of a gas giant like Saturn, can adequately fit the Doppler data if each arc is analyzed separately. However, the same model cannot jointly fit all passes in a combined, multi-arc, gravity and orbital solution. The signatures in range-rate residuals are as large as 0.2 $mm \cdot s^{-1}$ over time scales of 20 to 60 min, corresponding to radial accelerations of the order of 10^{-7} $m \cdot s^{-2}$. This value is an underestimate of the real unmodeled accelerations, as a large fraction of them are aliased in those associated with estimated parameters. The unmodeled accelerations acting on Cassini must be compensated for, to avoid biases in the estimates of the gravity harmonics and ring mass.

The missing accelerations could be due to longitudinally varying density anomalies resulting from wind dynamics or convection in Saturn's deep interior (*23*). For a rocky planet, the corresponding gravitational field would be static and described by tesseral harmonics. However, this approach is not immediately applicable to a fluid planet like Saturn, because vortices move longitudinally with different speeds depending on latitude. The resulting gravity disturbances (caused by the nonzonal wind dynamics) would not be static in any reference frame for the 60-day duration of the Cassini gravity measurements. However, if the density anomalies are deep-seated, below the level of the zonal winds and in the region of uniform rotation, then a static tesseral field associated with convection might be maintained over time scales much longer than the duration of the experiment.

Although a high-degree tesseral field can fit the data, the required degree increases with the number of passes analyzed and depends on the assumed rotation period. Several measurements of Saturn's rotation period have been made using different techniques. In our analysis we use four values—10 hours 32 min 45 s (*24*), 10 hours 39 min 22 s (also called System III) (*25*), 10 hours 45 min 45 s (*26*), and 10 hours 47 min 6 s (*27*)—leading to a required tesseral field of degree and order ranging from 8 to 12 (fig. S4). However, the determination of the tesseral field amplitudes is nonunique, because of the uncertainty in the underlying rotation rate of the deep interior (*12*).

Another potential solution is to assume a time-varying field, generated by acoustic oscillations in the planet (*28*). Fundamental mode oscillations (f-mode, with radial order equal to zero) have been detected by analysis of density waves in the C ring (*29*). These waves were also detected in Cassini ring stellar occultation data and are identified with sectoral modes which have an azimuthal order $2 \le m \le 10$ (sectoral harmonics depend only on longitude). However, their absolute amplitudes are not well determined. In principle, an infinite number of modes can exist within Saturn, only a subset of which can drive waves in the rings. The inclusion of such normal modes in the dynamical model is possible, but the choice of the relevant modes is not unique. We found that different combinations of modes can fit the data but that it is impossible to identify unambiguously the dominant ones with the limited data available.

A more general approach, often adopted in orbit determinations in the presence of unknown or poorly modeled dynamics, is to assume empirical, random accelerations unrelated to any specific physical model. In the absence of geophysical evidence to pinpoint the root cause of the residual accelerations, we used this approach to obtain the baseline solution shown in Table 1. Our goal in this analysis was to disentangle the estimates of the zonal harmonics and the ring mass from the unmodeled accelerations. As an additional check on the robustness of the solution, we compared the baseline solution with solutions obtained using a tesseral field and multiple normal modes (*12*).

These random accelerations are assumed to act on Cassini's orbit for a limited time span centered on pericenter and estimated in the orbital frame as local parameters. On each arc, these accelerations can act for equal-duration intervals at a constant value. The duration and the a priori uncertainty of the constant accelerations are adjustable parameters. In principle, it is desirable to optimize the number and duration of the intervals to avoid a degradation of the solution accuracy because of the over-parametrization of the problem. The a priori uncertainty must also be minimized to avoid aliasing part of the zonal gravity signal into the piecewise-constant accelerations. The minimum value of the a priori uncertainty (corresponding approximately to the largest allowed value of the random accelerations) represents the order of magnitude of the dynamical model's incompleteness.

We found that random accelerations with an a priori uncertainty of 4×10^{-7} $m \cdot s^{-2}$, acting for

Table 1. Measured gravity harmonic coefficients of Saturn (unnormalized; reference radius: 60,330 km) and total ring mass (in units of Mimas mass). The J_2 value includes a constant tidal term owing to the average tidal perturbation from the satellites. The associated uncertainties are recommended values to be used for analysis and interpretation. For the zonal harmonics, they correspond to three times the formal uncertainties. The solution for the total ring mass (A + B + C) is stable independently of the adopted dynamical model (table S2) and the uncertainty reported is the 1σ formal uncertainty. See table S2 for our total ring mass estimates for several models of the unknown accelerations.

Parameter	Value	Uncertainty
J_2 (× 10^6)	16,290.573	0.028
J_3 (× 10^6)	0.059	0.023
J_4 (× 10^6)	−935.314	0.037
J_5 (× 10^6)	−0.224	0.054
J_6 (× 10^6)	86.340	0.087
J_7 (× 10^6)	0.108	0.122
J_8 (× 10^6)	−14.624	0.205
J_9 (× 10^6)	0.369	0.260
J_{10} (× 10^6)	4.672	0.420
J_{11} (× 10^6)	−0.317	0.458
J_{12} (× 10^6)	−0.997	0.672
Ring mass (M_M)	0.41	0.13

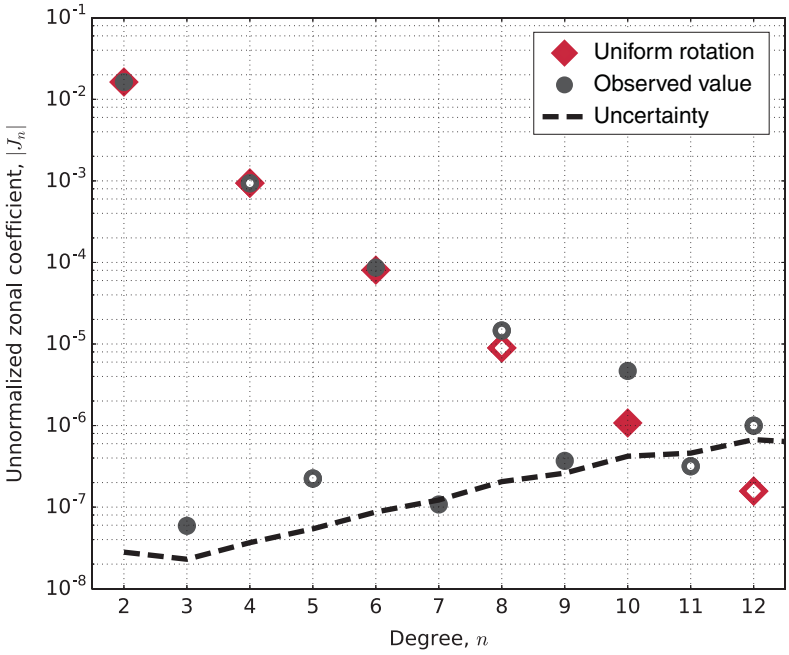


Fig. 1. Zonal gravity harmonic coefficients J_2 to J_{12}. The dashed line shows the uncertainties from Table 1. Positive values are marked as solid circles, negative values as empty circles. Theoretical predictions for uniform rotation (Table 2) are plotted with red diamonds (solid for positive values, empty for negative values) and diverge from the measurements at degrees > 8 (see also fig. S6 for a comparison with Jupiter).

10-min intervals within ±1 hour from pericenter, are able to remove the excess signatures in the Doppler residuals. The profile of the estimated random accelerations is reported in fig. S3.

The reference multi-arc solution is reported in Table 1 and Fig. 1. Before Cassini's Grand Finale, the determination of Saturn's even zonal coefficients J_2, J_4, and J_6 was obtained from perturbations of the orbits of the moons, as well as from direct perturbations on Cassini itself (*18*). The data acquired during the Grand Finale orbits are consistent with these previous estimates; they also provide determination of the higher-degree gravity harmonics (J_8 and J_{10}) and the odd harmonic coefficients J_3 and J_5, the only odd harmonics whose values are larger than the associated uncertainties.

The data effectively constrain the sum of the masses of the A, B, and C rings, but the individual masses are poorly determined because of their large mutual correlations. However, the masses of the more-transparent A and C rings have been estimated from density waves seen in stellar occultations (*20*, *21*). We therefore estimated the masses of the A and C rings subject to those constraints (a priori values and uncertainties). Our final estimates of the A and C ring masses are essentially the same as the a priori values (see table S2). The uncertainty in the total ring mass (Table 1) is obtained from the correlation submatrix of the A, B, and C ring masses. We find a total ring GM value of 1.02 ± 0.41 $km^3 \cdot s^{-2}$, equivalent to 0.41 ± 0.13 Mimas masses (1σ uncertainties) (table S2).

The alternative tesseral field and normal modes approaches were compared with the piecewise-constant acceleration case to check the stability of our solution (*12*). The three solutions are in agreement with each other, as shown by the error ellipses in fig. S4 for pairs of gravity harmonics and in table S2 for the ring masses. We conclude that the estimate of the zonal harmonics is robust and largely independent of the dynamical model. This is especially true for the ring mass, which is consistent in all models and insensitive to the assumptions.

Saturn interior models with uniform rotation

Although gravity measurements provide constraints on the interior of gas giants, every model using gravity data unavoidably suffers from uncertainties. We tested whether interior models based on uniform rotation can explain the measured gravity harmonics. We developed a suite of interior models based on the common assumption that the fluid in Saturn's interior rotates uniformly like a solid body. All our models have a molecular and a metallic layer, each represented by an adiabat (a constant entropy curve in the temperature-pressure diagram) consistent with an equation of state (EOS) for hydrogen-helium mixtures determined from first-principles simulations (*30*, *31*) and characterized by an entropy, S, a helium mass fraction, Y, and a mass fraction of heavy elements, Z. We assume that helium rain occurs in Saturn's interior wherever hydrogen and helium become immiscible, because hydrogen becomes metallic but helium does not (*32*). We thus introduce a helium rain layer that starts and ends at pressure-temperature conditions that are compatible with the hydrogen-helium immiscibility zone. We treat the helium rain layer as a smooth transition of the parameters across a range of pressures P_1 to P_2, defined by the intersections of the adiabat with the immiscibility curve (*32*). These adiabatic profiles are thus described by six parameters (S_{mol}, Y_{mol}, Z_{mol}, S_{met}, Y_{met}, Z_{met}), where the subscript "mol" denotes the outer molecular envelope and "met" the inner metallic envelope. The models match

the planet's total mass and include dense cores of various sizes. Initially, we assumed a fractional radius of the core as r_c = 0.2 (*33*) and later refined the core radii by assuming a terrestrial iron-silicate composition (0.325:0.675) as well as a solar iron-silicate-water ice composition (0.1625:0.3375:0.5), which is consistent with Callisto's interior (*34*). For these two compositions, we derived fractional core radii of r_c = 0.188 and 0.231, respectively, assuming hydrostatic equilibrium inside the core and previously published equations of state (*35*).

For each set of parameters, we construct a model for Saturn's interior by performing a calculation with the concentric MacLaurin spheroid (CMS) method (*36*, *37*) to find a self-consistent shape and gravitational field for the planet. The simulation suite includes models with the four uniform rotation periods for Saturn's interior observations (*24*–*27*). A range of interior distributions of helium were adopted, with a gradual gradient at a depth consistent with the phase diagram (*32*), and the planet-wide mass fraction of helium matching the solar fraction (*38*). A wide range of Y_{mol} was considered, bounded by the solar helium fraction Y = 0.274 and the lowest prediction Y = 0.055 from infrared measurements of Saturn's atmosphere by Voyager (*39*). The range in entropy of the deep interior (S_{met}) is between the value in the outer envelope S_{mol} = 6.84 and the maximum value consistent with the predicted hydrogen-helium phase separation conditions, S ~ 7.20. Values of Z_{mol} up to five times solar fraction and fractional core radii up to 0.3 were considered. In each model, the core mass and deep heavy-element fraction (Z_{met}) were tuned to match the observed J_2, and an iterative approach was used to identify the narrower range of model parameters matching the observed J_4.

The resulting gravity harmonics are compared in Table 2 with the Cassini observations. Although a subset of the uniform rotation models is found to simultaneously match the observed J_2 and J_4, the magnitudes of the observed values for J_6, J_8, and J_{10} are much larger than those predicted by the uniform rotation models (Fig. 1 and section 6 of the supplementary materials). Despite the wide range of physical parameters and interior structures considered, the tight correlation between J_4 and the higher-order harmonics J_6 to J_{10} precludes any possibility of finding a model that simultaneously matches all the observed harmonics. This result is unlike the gravity measurements of Jupiter from the Juno spacecraft. In this case, static models can be constructed, for which all even harmonics differ by less than 0.1×10^{-6} from the measurements (*8*), whereas Table 2 shows that the deviations for Saturn are much larger.

To test the robustness of this result, we added additional flexibility to the interior density distribution. Starting from a reasonable physical model, we introduced arbitrary density jumps of up to ±4% at four points in pressure that were chosen at random. We fit the models to the observed J_2 value but report the full range of the higher J_{2k} value without minimizing the discrepancy from the observations. These additional density modifications increased the range of all J_{2k} values in our ensemble of models (Table 2), but a sizable discrepancy from the observed J_8 and J_{10} values remained. Figure 1, fig. S6, and Table 2 show the large and systematic deviations between the observations and models that assume uniform rotation. This demonstrates that specific assumptions regarding the EOS and conditions of helium rain cannot explain the inability of the models to match the higher-order harmonics.

The inability to reproduce the unusually large values of J_6, J_8, and J_{10} leads us to conclude that models with uniform rotation are ruled out by the gravity data. Instead, the observations require us to introduce DR (*2*) into our models for Saturn's interior. We study its effects with two different techniques. First, we introduce DR on cylinders into the CMS method, which allows us to construct consistent interior models but requires constant rotation rate on cylinders that penetrate all the way through the planet. Second, we use the thermal-wind method (*3*) to derive the difference in the gravity signature between a uniformly and a differentially rotating body. The latter approach has the advantage that the DR profiles can have a finite depth and are not required to be north-south symmetric.

Differential rotation on cylinders with the CMS method

In the CMS with DR approach, we simultaneously optimize interior parameters and DR profiles, $\omega(l)$, where ω is angular frequency and l is the cylindrical radius from the axis of rotation. We derive a centrifugal potential, $Q(l) = \int_0^l dl'\ l'\ \omega(l')^2$, that we introduce into the CMS method (*2*, *40*, *41*). Because this approach is based on potential theory, $\omega(l)$ can only depend on l and may not decay with depth along the cylinders. The benefit of the combined CMS + DR approach is that it represents a fully self-consistent velocity profile, interior density distribution, and gravitational field, rather than treating the winds as a correction to a model with uniform rotation.

We find that the assumption of DR can account for the unusually large magnitude of coefficients J_6 to J_{10}. By assuming Saturn's equatorial region rotates ~4% faster than the deep interior, agreement between models and data for all coefficients J_2 to J_{10} can be achieved (Table 2). This assumption is also compatible with the equatorial part of the wind profile obtained from optical observations (*42*), showing strong eastward flow near the equator. We thus favored models that were in general agreement with the observed winds when we subsequently constructed an ensemble of interior models. For the planet's deep interior, we assumed the same four rotation periods as above (*24*–*27*).

Starting from a parameter set chosen at random, we performed ~10^4 independent model optimizations with the simplex algorithm. The 11 best models were used as starting points for subsequent Markov chain Monte Carlo (MCMC) calculations to explore the allowed parameter space more carefully. Models that matched the observed gravitational moments J_2 to J_{10} are compared in Figs. 2 and 3. Figure 2 shows the rotational velocity profiles as a function of l and compares these to the corresponding north-south averaged surface winds derived from optical tracking of clouds (*42*, *43*).

From the best-fitting MCMC models, we selected quantities to elucidate the allowed region of parameter space (Fig. 3). Models that assume

Table 2. Comparison of observed and calculated gravitational harmonics (unnormalized; reference radius: 60,330 km). Where two values are given, they denote the minimum and maximum values from the suite of models. The physical models in column 3 match the observed J_2 and J_4 in Table 1, over a parameter space considering ranges of S_{met}, Y_{mol}, Z_{mol}, and r_c and rotation periods from 10 hours 32 min 44 s to 10 hours 47 min 6 s. For the same span of rotation periods, column 4 reports a wider range from models that match only J_2 and allow for density modifications assuming r_c = 0.2. For J_6 to J_{10}, the discrepancy between measurements and uniform rotation models is large for all models that assume uniform rotation. Column 5 shows a representative model with DR on cylinders and a deep rotation period of 10 hours 39 min 22 s that matches measurements from J_2 to J_{10}.

Parameter	Measurements	Physical models with uniform rotation		Uniform rotation model with modified density profiles		Physical model with differential rotation
J_2	16,290.573 ± 0.028	16,290.57		16,290.57		16,290.573
J_4	−935.314 ± 0.037	−935.31		−990.12	−902.93	−935.312
J_6	86.340 ± 0.087	80.74	81.76	75.69	90.42	86.343
J_8	−14.624 ± 0.205	−8.96	−8.70	−10.26	−7.97	−14.616
J_{10}	4.672 ± 0.420	1.08	1.13	0.97	1.33	4.677

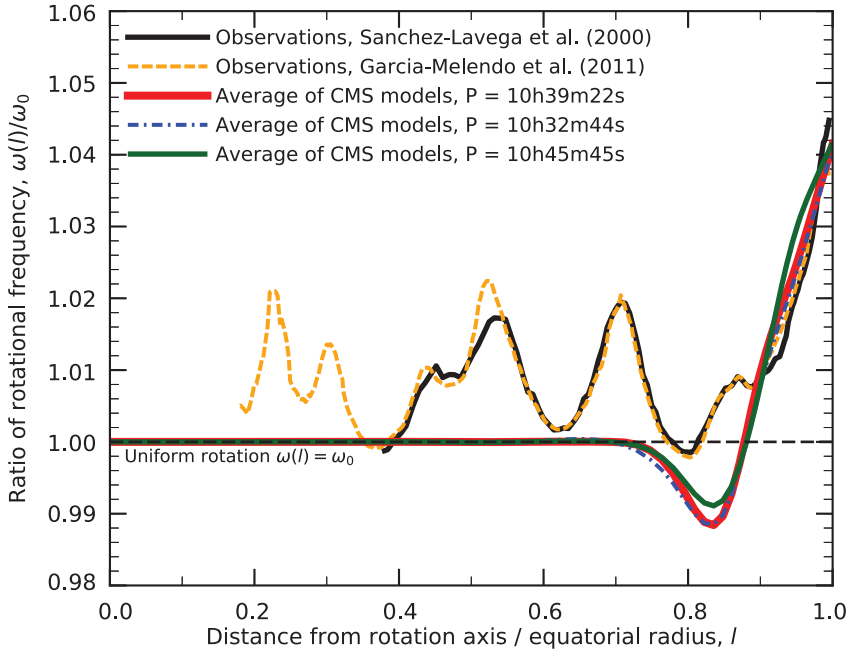
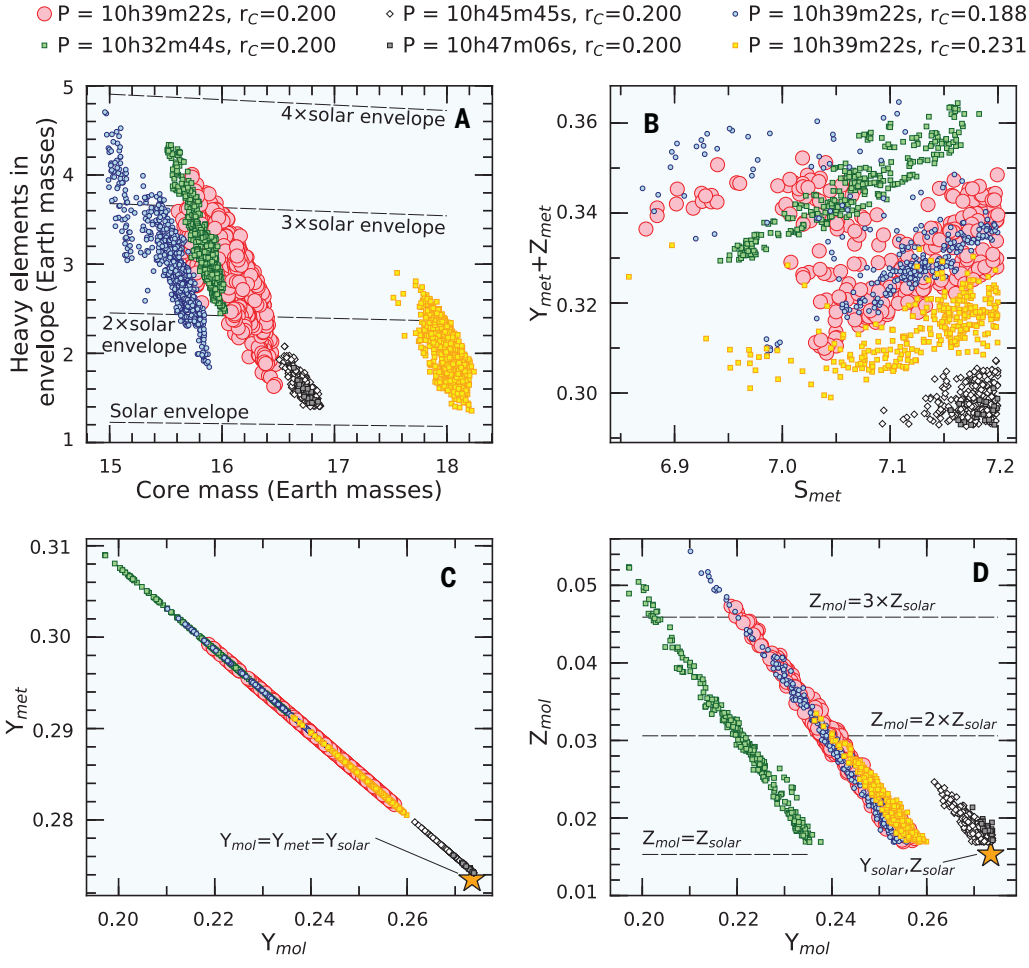


Fig. 2. Differential rotation profiles for CMS models. CMS profiles are compared with north-south symmetrized wind profiles from optical observations (*42*) (black line) and (*43*) (dashed orange line). Our Monte Carlo averages for models with three different rotation periods, *P*, and core radius of r_c = 0.2 are also plotted.

core masses between 15.0 and 18.2 Earth masses best match the gravity measurements (Fig. 3A). This range of core masses is broadly consistent with predictions of earlier models (*44*–*47*) built on previous gravity measurements and different assumptions for the EOS or structure of the deep interior. The predicted core masses increase if a smaller core radius is assumed, because the H-He mixture surrounding the core is exposed to higher pressures and becomes denser. The core masses also increase when a longer rotation period is assumed. As the total planet mass is constrained, the core mass anticorrelates with the fraction of heavy elements in the envelope (Z_{mol}). In comparison to the core (which we assume here consists entirely of elements heavier than helium), the mass of heavy elements in the envelope is relatively small, with a permitted range of 1.3 to 4.8 Earth masses. By contrast, analogous interior models of Jupiter predict a core mass slightly lower than Saturn's, but with greater heavy-element enrichment in the deep envelope (*8*). Figure 3B shows the correlation between $Y_{met} + Z_{met}$ and entropy of the metallic region S_{met}. An increase in entropy implies a higher temperature and a lower density. This density reduction can then be compensated by higher fractions of helium and heavy elements.

Fig. 3. Composition of parameters from six sets of interior models including differential rotation. The assumed rotation periods and core radii are indicated by the color and symbol, as specified in the legend. (**A**) The distribution of heavy-element mass between the core and envelope. (**B**) The variation of the mass fraction of helium and hydrogen plus heavy elements with entropy in the metallic envelope. (**C**) The variation of helium mass fraction in the molecular and metallic envelopes. (**D**) The trade-off between heavy-element and helium mass fractions in the molecular envelope. In panels (B) and (C), solar values (*69*) are shown with a yellow star.



The helium fraction in the metallic layer and in the molecular layer are linearly related (Fig. 3C), as we require all models to have an overall solar helium abundance in the envelope. All models assumed that some helium rain has occurred, depleting the molecular layer and enriching the metallic layer in helium. We find that almost no helium sequestration is predicted for models that assume the longer rotation periods of 10 hours 45 min 45 s or 10 hours 47 min 6 s. This effectively removes one degree of freedom for the interior models and further constrains the range of other model parameters. For example, when we performed MCMC simulations with variable core radii for a period of 10 hours 47 min 6 s, no models with core radius larger than 0.21 and core masses larger than 17.2 Earth masses were obtained. Figure 3D shows that the heavy-element fraction in the molecular envelope may vary between 1 and 3.5 times the solar value. The heavy Z fraction is anticorrelated with the helium fraction, which implies that there is only limited capacity for elements heavier than hydrogen. Longer rotation periods imply that there is a greater capacity for such elements.

All models that match the even gravity harmonics (J_2 to J_{10}) show a rapid decrease in angular velocity from $l = 1.0$ to $l \sim 0.9$ (Fig. 2). For all rotation periods under consideration, our CMS + DR models also require a minimum region near $l \sim 0.83$ where the fluid rotates slower than in the deep interior. This radius corresponds to a depth of ~10,000 km from the surface at the equator. By construction, the DR profiles converge to the assumed rotation rate for small l. The CMS + DR models are unable to reproduce the wind profile for small l (high latitudes) because we assume DR on cylinders.

Differential rotation with finite depth flows

An alternative approach, the thermal-wind method (*3*), can incorporate wind profiles with a finite depth. In this approach, the flow is not limited to full cylindrical symmetry and the flows are not required to be equal in the Northern and Southern hemispheres. This allows the model to account for both the even and odd gravity harmonics; the latter reflect a north-south asymmetry in the gravity field (*4*). It has been shown that for the barotropic case (flows limited to cylindrical symmetry), the thermal-wind and the CMS method with DR produce consistent results (*41*, *48*). We apply the thermal-wind method using an adjoint inverse model (*49*), where the measured gravity harmonics in Table 1 are used to identify the flow structure that best fits the data. Although it has these advantages, the thermal-wind method can only account for the dynamical part of the gravity spectrum (attributable to the flows), and therefore relies on having prior knowledge of a reference interior model for the even gravity harmonics. The predicted gravity values thus depend on the reference model (Table 2).

For Jupiter, the extension of the cloud-level flow into the interior results in a gravity spectrum that matches the observations (*6*) and the background density profile. The case of Saturn is more complicated. We find that regardless of the vertical structure of the zonal flow, with the background density profiles corresponding to the 11 models discussed in the previous section, extending the observed cloud-level flow (*43*) into the interior results in dynamical gravity harmonics (ΔJ_n) that are too small by at least a factor of two. For example, the gravity signature of the observed surface winds extended along cylinders to large depths gives roughly $\Delta J_8 = -1.5 \times 10^{-6}$ and $\Delta J_{10} = 1 \times 10^{-6}$ (*4*, *50*). The measured J_8 and J_{10} are -14.6×10^{-6} and 4.7×10^{-6}, respectively, whereas the uniform rotation contribution is, at most, -9×10^{-6} and 1×10^{-6} (Table 2). This implies that the observed cloud-level flow cannot account for the difference between the measurements and the uniform rotation contribution regardless of how it extends into the interior. As seen in Fig. 2, the only way to match the data is to alter the latitudinal profile of the zonal wind, by assuming that, because of other processes, the bulk of the sub-cloud-level flow does not resemble that observed at cloud level.

Therefore, to match the Cassini gravity measurements, another degree of freedom is added to the thermal-wind gravity inversion, whereby the meridional profile of the flow is allowed to vary from the observed winds in addition to the flow depth. We then optimize for the depth and meridional structure of the zonal flow so that the calculated gravity harmonics match the measured values. To compare the results to those of the hemispherically symmetric CMS method, and because, unlike Jupiter (*6*), the odd harmonics are small and contribute little to the flow structure, we use only the even gravity harmonics, focusing on those which are most strongly affected by the flows, namely J_8 and J_{10}. Because of the uncertainty resulting from the uniform rotation reference model, and the nonuniqueness of the problem, there is little merit in optimizing for the vertical profile of the flow, as can be done for Jupiter (*6*), so the vertical decay is approximated by a radial hyperbolic tangent profile with a width of 500 km and is then optimized for its depth.

The reconstructed meridional profile of the zonal flow at the planet's surface is shown in Fig. 4A, achieved with a vertical profile depth of 9363 ± 357 km. With these parameters, we are able to match the measured values of J_8 and J_{10}, once the uniform rotation contribution in Table 2 is subtracted. Both gravity harmonics are within the uncertainties of the measured values (Table 3). The ΔJ we optimize for are calculated from the difference between the measurements and the 11 preferred models run without DR and using the Voyager rotation period. Although the reconstructed zonal wind profile varies from the observed one, it still retains its main characteristics (Fig. 4A). The uncertainty of the wind depth is obtained from the variance in the 11 solutions of the internal models, the variance in the internal models and thermal-wind calculation for different rotation rates between 10 hours 32 min and 10 hours 47 min, and using different internal models, by statistically varying the depth, and finally searching for solutions that satisfy the condition that both harmonics are within the uncertainties. In all cases, the reported depth is the inflection point of the hyperbolic tangent vertical profile. The exact meridional structure of the zonal winds is not uniquely determined, meaning that other profiles can give results within the measurement uncertainties.

Regardless of the exact meridional profile, all vertical flow profiles are constrained to contain a very deep flow of ~9000 km. This depth, corresponding to 15% of Saturn's radius, matches the predictions from magnetohydrodynamic theories, suggesting that the flow should extend down to the levels of magnetic dissipation (*51*, *52*). It has been suggested that a higher-order expansion to the vorticity balance, beyond a thermal wind, is necessary to reproduce the dynamical gravity harmonics (*53*). To compare those solutions [known as the thermal-gravity equation, (*53*)], we calculate the gravity harmonics using the thermal-gravity equation for a similar wind profile. The comparison (Table 3) shows that the solutions from both methods are within the observational uncertainties, confirming that the thermal-wind solution is the leading-order vorticity balance (*48*). The thermal-gravity solution was obtained by solving the full integro-differential equation on a sphere (*54*).

The CMS solutions with DR on full cylinders show a flow signature only for distances from the axis of rotation $l \geq 0.6$ (Fig. 2). Using the thermal-wind inversion model, we can examine the nature of the optimized flow obtained under a similar restriction. In Fig. 4B, we show that taking the same approach with the thermal-wind model, restricting the flow to be equatorward only for $l \geq 0.6$ (latitudes $\leq 60°$), gives similar results. The reconstructed flow in this case is not as close to the observed flow as in the unrestricted case (Fig. 4A) but still gives results within 1σ of the deviation uncertainty (Table 3). We also find a large flow depth of 8832 ± 295 km in this solution. The deviations between the two flow models in Fig. 4 illustrate the nonuniqueness in the resulting flows when gravity harmonics ΔJ_8 and ΔJ_{10} are modeled. However, all our thermal-wind flow solutions show strong westward flow near latitude 35°, which is in agreement with the westward flowing region calculated with the CMS model in Fig. 2. These results were calculated for a rotation period of 10 hours 39 min 22 s; because the high-degree harmonics are weakly affected by the rotation rate (*55*), we expect consistent results if a different period is assumed.

We find therefore that with the thermal-wind approach, extending the exact observed flows into the interior in a simple manner does not lead to an exact match of measured gravity values. However, flow profiles that are similar in general character to the observed ones (Fig. 4) yield solutions within the deviation uncertainty when extended to a depth of ~9000 km, indicating that the flows are very deep and likely extend down to the levels of magnetic dissipation.

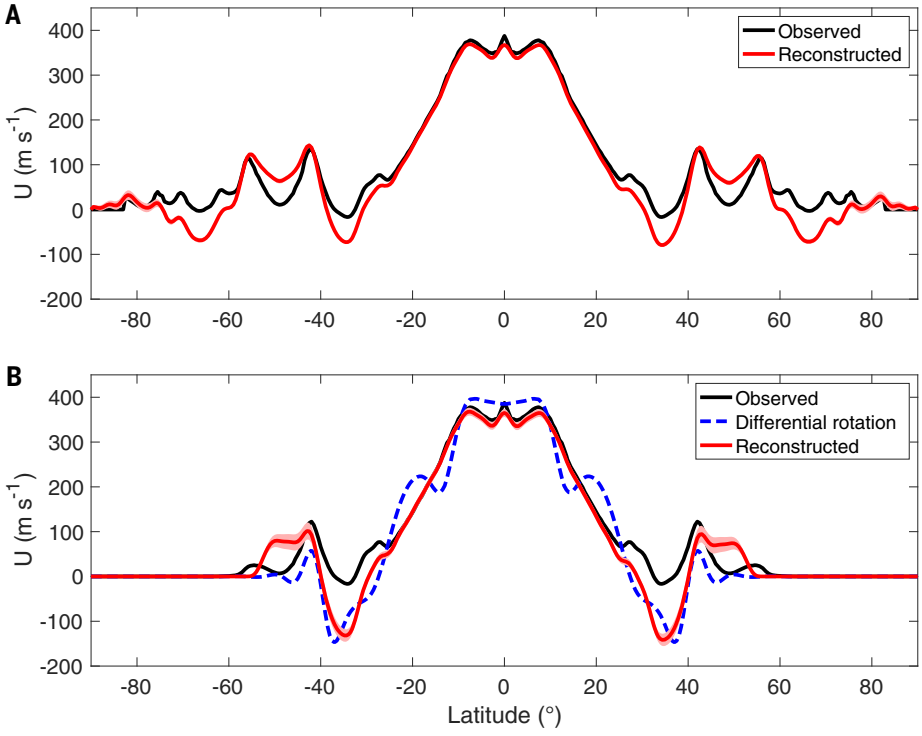


Fig. 4. Comparison of observed and reconstructed wind profiles. (**A** and **B**) Observed zonal winds relative to the Voyager rotation rate 10 hours 39 min 32 s (*43*) and meridional profiles of the zonal wind, optimized with the thermal-wind method, for a full pole-to-pole wind profile (A) and a wind profile truncated at latitude 60° (B). In both cases, the flows must to extend to a depth of ~9000 km for the computed gravity signal to be consistent with observations for J_8 and J_{10}. The dashed blue line is one of the CMS models shown in Fig. 2 projected into geocentric latitude at the planet's surface, and the red shading is the calculation uncertainty on the optimized meridional profile of the zonal wind.

Table 3. Contribution to the higher gravity harmonics ΔJ_8 and ΔJ_{10} resulting from differential rotation and thermal-wind optimization. The deviation (column 1) is the difference between the measured J_8 and J_{10} (Table 1) and the average of the computed values from the 11 CMS models with uniform rotation (Table 2). Two optimizations are shown: one without latitudinal truncation of the zonal flow, resulting in the reconstructed zonal wind profile shown in Fig. 4A and with a flow depth of 9363 km (column 2), and the second with the flows truncated at latitude 60° (Fig. 4B) and a flow depth of 8832 km (column 3). Columns 4 and 5 show the deviations calculated with the thermal-gravity equation (*48*) for similar wind profiles. The solutions from thermal wind are closer to the measurement because the optimization was done using the thermal-wind method, but the thermal-gravity solutions also match the observations within 10%.

Parameter	Deviation	Thermal-wind solution	Thermal-wind solution truncated at latitude 60°	Thermal-gravity solution	Thermal-gravity solution truncated at latitude 60°
ΔJ_8	−5.600 ± 0.205	−5.624	−5.533	−5.758	−5.759
ΔJ_{10}	3.528 ± 0.659	3.570	3.660	3.974	4.037

The agreement between the CMS and thermal-wind solutions, based on two substantially different approaches, indicate that the interpretation of the Cassini gravity data is robust. Regardless of the exact interior model or vertical decay profile chosen, the flows must extend deep below the surface, to ~15% of the planet's radius.

Mass and age of Saturn's rings

We now consider the mass of the main rings (i.e., the A, B, and C rings) and how this quantity can constrain theoretical models for the age and origin of the rings. (The masses of the diffuse D, E, F, and G rings are expected to be negligible.) Before the Voyager flybys in 1980 and 1981, it was known that the cross section–weighted average ring particle radius was at least 10 cm, as required to account for the rings' high radar reflectivity (*56*). Voyager results provided estimates of the rings' mass, on the basis of local surface mass densities determined from the wavelengths of 13 density waves in the A and B rings and an optical depth profile derived from a stellar occultation (*57*). The estimated total ring mass was 2.8×10^{19} kg, or 0.75 Mimas masses. However, it was subsequently argued that substantially more mass might be hidden in the opaque parts of the B ring (*58*).

Cassini observations of density waves in the rings have led to additional local surface density estimates of the A (*20*), C (*29*), and B rings (*19*). The B-ring density is higher than the others, with a mean value of ~600 kg·m^{-2}, but not higher by as much as its large optical depth would suggest. Combining these estimates, we find a likely total ring mass of GM = 1.01 km^3·s^{-2}, or 0.40 Mimas masses. However, the presence of self-gravity wakes in the A and B rings has led to suggestions that substantial amounts of material may be present but does not contribute to density wave estimates. Numerical simulations of these structures (*58*) suggest wake surface densities as high as 5000 kg·m^{-2} and an upper limit to the total ring mass of 9.7×10^{19} kg, or ~2.5 Mimas masses.

Our estimate of $GM = 0.58 \pm 0.48$ km^3·s^{-2} for the B ring, when combined with the previously determined GM products of the A (0.38 km^3·s^{-2}) and C (0.06 km^3·s^{-2}) rings, leads to a total ring GM of 1.02 ± 0.43 km^3·s^{-2}, equivalent to 0.41 ± 0.13 Mimas masses (table S2). This estimate is somewhat smaller than the Voyager value but consistent with those inferred from Cassini observations of density waves.

The ring mass has implications for the age of the rings (*59*). Traditional age estimates of Saturn's ring-satellite system fall into three categories: dynamical estimates based on the rate of recession of the small satellites from the rings owing to gravitational torques and the back-reaction on the A ring (*10*, *60*); structural evolution time scales based on the evolution of unconfined edges, such as the inner edges of the A and B rings (*11*); and compositional time scales

based on the assumption that the rings were formed as pure water ice and have subsequently been steadily darkened by the infall of interplanetary debris (*9*). Most of these time scale estimates depend, directly or indirectly, on the mass of the rings; both dynamical and compositional considerations suggest that low-mass rings are likely to be young, and both approaches yield evolutionary ages ~10^8 years for the A and B rings, assuming the Voyager measurements of ring masses and interplanetary impact fluxes (*59*). Low-mass rings therefore pose a challenge for models in which the ring system is assumed to have formed in the early Solar System.

The Cassini results have sharpened these arguments. Analysis of the main rings' non-icy fraction derived from their passive microwave emission, using the pre-Cassini interplanetary flux estimates and a minimum-mass B ring, led to estimates of the ages of the A and B rings of 80 to 150 million years (Ma) and 30 to 100 Ma, respectively (*61*). Our revised mass estimate for the B ring would increase the latter estimate by ~25%. Cassini dust measurements have also provided a refined estimate of the interplanetary dust flux at the rings (*62*), indicating that the interplanetary dust impact flux on the rings is higher by almost a factor of 10 compared to the Voyager estimates, because of an improved understanding of the gravitational focusing by Saturn. These new dust fluxes would reduce the age estimates for the A and B rings.

As the rings evolve viscously, they may spread radially beyond the Roche limit and therefore gradually lose mass to form new generations of small satellites, which would then move rapidly away from the rings (*63*). In this scenario, the rings would have been more massive in the past, with proportionally longer evolutionary time scales. Current models of viscous evolution (*64*) predict that such a ring would approach an asymptotic mass of ~1.5×10^{19} kg, or 0.40 Mimas masses, after 5 billion years, close to our estimate of 0.41 ± 0.13 Mimas masses. Taken at face value, and subject to the limitations of the simulations (which ignore the gravitational torques from the newly formed moons), this suggests that the rings may be older and formed with more mass than they have today. However, the simulations also show that a massive, primordial ring initially loses mass very quickly before settling into a long period of progressively slower evolution (*64*). An old high-mass ring would thus quickly evolve to a mass not much greater than that measured today and would therefore still be subjected to rapid ballistic and compositional evolution.

On balance, we favor a scenario whereby the present rings of Saturn are relatively young, at least compared with the planet itself, although they may have evolved substantially in the past 10^7 to 10^8 years and were perhaps once more massive than they are today. Our data do not indicate how the ring system formed within such a recent period. Models of that process invoke the chance capture and tidal disruption of a comet or Centaur (*65*, *66*). Alternatively, rings may arise from the catastrophic disruption of an earlier population of midsized icy satellites ~100 Ma ago (*67*), although whether the resulting debris can migrate into the ring region before it reaccretes into new satellites is uncertain (*68*). Regardless of how the rings formed, the ring mass we derived from Cassini gravity data indicates a recent origin for Saturn's ring system, which we consider a fitting way to end Cassini's mission.

REFERENCES AND NOTES

1. W. B. Hubbard, *Planetary Interiors* (Van Nostrand Reinhold, 1984).
2. W. B. Hubbard, Gravitational signature of Jupiter's deep zonal flows. *Icarus* **137**, 357–359 (1999). doi: 10.1006/icar.1998.6064
3. Y. Kaspi, W. B. Hubbard, A. P. Showman, G. R. Flierl, Gravitational signature of Jupiter's internal dynamics. *Geophys. Res. Lett.* **37**, L01204 (2010). doi: 10.1029/2009GL041385
4. Y. Kaspi, Inferring the depth of the zonal jets on Jupiter and Saturn from odd gravity harmonics. *Geophys. Res. Lett.* **40**, 676–680 (2013). doi: 10.1029/2012GL053873
5. L. less *et al*., Measurement of Jupiter's asymmetric gravity field. *Nature* **555**, 220–222 (2018). doi: 10.1038/nature25776; pmid: 29517001
6. Y. Kaspi *et al*., Jupiter's atmospheric jet streams extend thousands of kilometres deep. *Nature* **555**, 223–226 (2018). doi: 10.1038/nature25793; pmid: 29516995
7. T. Guillot *et al*., A suppression of differential rotation in Jupiter's deep interior. *Nature* **555**, 227–230 (2018). doi: 10.1038/nature25775; pmid: 29517000
8. S. M. Wahl *et al*., Comparing Jupiter interior structure models to Juno gravity measurements and the role of a dilute core. *Geophys. Res. Lett.* **44**, 4649–4659 (2017). doi: 10.1002/2017GL073160
9. J. N. Cuzzi, P. R. Estrada, Compositional evolution of Saturn's rings due to meteoroid bombardment. *Icarus* **132**, 1–35 (1998). doi: 10.1006/icar.1997.5863
10. N. Borderies, P. Goldreich, S. Tremaine, "Unsolved problems in planetary ring dynamics," in *Planetary Rings* (Univ. of Arizona Press, 1984), pp. 713–734.
11. P. R. Estrada, R. H. Durisen, J. N. Cuzzi, D. A. Morgan, Combined structural and compositional evolution of planetary rings due to micrometeroid impacts and ballistic transport. *Icarus* **252**, 415–439 (2015). doi: 10.1016/j.icarus.2015.02.005
12. Materials and methods are provided in the supplementary materials.
13. L. less *et al*., Astra: Interdisciplinary study on enhancement of the end-to-end accuracy for spacecraft tracking techniques. *Acta Astronaut.* **94**, 699–707 (2014). doi: 10.1016/j.actaastro.2013.06.011
14. L. less *et al*., The tides of Titan. *Science* **337**, 457–459 (2012). doi: 10.1126/science.1219631; pmid: 22745254
15. L. less *et al*., The gravity field and interior structure of Enceladus. *Science* **344**, 78–80 (2014). doi: 10.1126/science.1250551; pmid: 24700854
16. S. Evans *et al*., MONTE: The next generation of mission design and navigation software. *CEAS Space J.* **10**, 79–86 (2018). doi: 10.1007/s12567-017-0171-7
17. R. G. French *et al*., Noncircular features in Saturn's rings IV: Absolute radius scale and Saturn's pole direction. *Icarus* **290**, 14–45 (2017). doi: 10.1016/j.icarus.2017.02.007
18. Jet Propulsion Laboratory, California Institute of Technology, NAIF: Planetary Data System Navigation Node; https://naif.jpl.nasa.gov/pub/naif/generic_kernels/spk/.
19. M. M. Hedman, P. D. Nicholson, The B-ring's surface mass density from hidden density waves: Less than meets the eye? *Icarus* **279**, 109–124 (2016). doi: 10.1016/j.icarus.2016.01.007
20. M. S. Tiscareno, J. A. Burns, P. D. Nicholson, M. M. Hedman, C. C. Porco, Cassini imaging of Saturn's rings: II. A wavelet technique for analysis of density waves and other radial structure in the rings. *Icarus* **189**, 14–34 (2007). doi: 10.1016/j.icarus.2006.12.025
21. K. Baillié, J. E. Colwell, J. J. Lissauer, L. W. Esposito, M. Sremčević, Waves in Cassini UVIS stellar occultations: 2. The C ring. *Icarus* **216**, 292–308 (2011). doi: 10.1016/j.icarus.2011.05.019
22. V. Lainey *et al*., New constraints on Saturn's interior from Cassini astrometric data. *Icarus* **281**, 286–296 (2017). doi: 10.1016/j.icarus.2016.07.014
23. D. Kong, K. Zhang, G. Schubert, Using Jupiter's gravitational field to probe the Jovian convective dynamo. *Sci. Rep.* **6**, 23497 (2016). pmid: 27005472
24. R. Helled, E. Galanti, Y. Kaspi, Saturn's fast spin determined from its gravitational field and oblateness. *Nature* **520**, 202–204 (2015). doi: 10.1038/nature14278; pmid: 25807487
25. M. D. Desch, M. L. Kaiser, Voyager measurement of the rotation period of Saturn's magnetic field. *Geophys. Res. Lett.* **8**, 253–256 (1981). doi: 10.1029/GL008i003p00253
26. D. A. Gurnett *et al*., Radio and plasma wave observations at Saturn from Cassini's approach and first orbit. *Science* **307**, 1255–1259 (2005). doi: 10.1126/science.1105356; pmid: 15604362
27. G. Giampieri, M. K. Dougherty, E. J. Smith, C. T. Russell, A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* **441**, 62–64 (2006). doi: 10.1038/nature04750; pmid: 16672964
28. D. Durante, T. Guillot, L. less, The effect of Jupiter oscillations on Juno gravity measurements. *Icarus* **282**, 174–182 (2017). doi: 10.1016/j.icarus.2016.09.040
29. M. M. Hedman, P. D. Nicholson, More Kronoseismology with Saturn's rings. *Mon. Not. R. Astron. Soc.* **444**, 1369–1388 (2014). doi: 10.1093/mnras/stu1503
30. B. Militzer, Equation of state calculations of hydrogen-helium mixtures in solar and extrasolar giant planets. *Phys. Rev. B* **87**, 014202 (2013). doi: 10.1103/PhysRevB.87.014202
31. B. Militzer, W. B. Hubbard, Ab initio equation of state for hydrogen-helium mixtures with recalibration of the giant-planet mass-radius relation. *Astrophys. J.* **774**, 148 (2013). doi: 10.1088/0004-637X/774/2/148
32. M. A. Morales, S. Hamel, K. Caspersen, E. Schwegler, Hydrogen-helium demixing from first principles: From diamond anvil cells to planetary interiors. *Phys. Rev. B* **87**, 174105 (2013). doi: 10.1103/PhysRevB.87.174105
33. S. M. Wahl, W. B. Hubbard, B. Militzer, The Concentric Maclaurin Spheroid method with tides and a rotational enhancement of Saturn's tidal response. *Icarus* **282**, 183–194 (2017). doi: 10.1016/j.icarus.2016.09.011
34. O. L. Kuskov, V. A. Kronrod, Internal structure of Europa and Callisto. *Icarus* **177**, 550–569 (2005). doi: 10.1016/j.icarus.2005.04.014
35. S. Seager, M. Kuchner, C. A. Hier-Majumder, B. Militzer, Mass-radius relationship of solid exoplanets. *Astrophys. J.* **669**, 1279–1297 (2007). doi: 10.1086/521346
36. W. B. Hubbard, High-precision Maclaurin-based models of rotating liquid planets. *Astrophys. J.* **756**, L15 (2012). doi: 10.1088/2041-8205/756/1/L15
37. W. B. Hubbard, Concentric Maclaurin spheroid models of rotating liquid planets. *Astrophys. J.* **768**, 43 (2013). doi: 10.1088/0004-637X/768/1/43
38. N. Grevesse, M. Asplund, A. J. Sauval, P. Scott, The chemical composition of the Sun. *Astrophys. Space Sci.* **328**, 179–183 (2010). doi: 10.1007/s10509-010-0288-z
39. B. J. Conrath, D. Gautier, R. A. Hanel, J. S. Hornstein, The helium abundance of Saturn from Voyager measurements. *Astrophys. J.* **282**, 807–815 (1984). doi: 10.1086/162267
40. J. Wisdom, W. B. Hubbard, Differential rotation in Jupiter: A comparison of methods. *Icarus* **267**, 315–322 (2016). doi: 10.1016/j.icarus.2015.12.030
41. Y. Kaspi, J. E. Davighi, E. Galanti, W. B. Hubbard, The gravitational signature of internal flows in giant planets: Comparing the thermal wind approach with barotropic potential-surface methods. *Icarus* **276**, 170–181 (2016). doi: 10.1016/j.icarus.2016.04.001
42. A. Sánchez-Lavega, J. F. Rojas, P. V. Sada, Saturn's zonal winds at cloud level. *Icarus* **147**, 405–420 (2000). doi: 10.1006/icar.2000.6449
43. E. García-Melendo, S. Pérez-Hoyos, A. Sánchez-Lavega, R. Hueso, Saturn's zonal wind profile in 2004–2009 from Cassini ISS images and its long-term variability. *Icarus* **215**, 62–74 (2011). doi: 10.1016/j.icarus.2011.07.005
44. W. B. Hubbard, M. S. Marley, Optimized Jupiter, Saturn, and Uranus interior models. *Icarus* **78**, 102–118 (1989). doi: 10.1016/0019-1035(89)90072-9
45. N. Nettelmann, R. Püstow, R. Redmer, Saturn layered structure and homogeneous evolution models with different EOSs. *Icarus* **225**, 548–557 (2013). doi: 10.1016/j.icarus.2013.04.018
46. D. Saumon, T. Guillot, Shock compression of deuterium and the interiors of Jupiter and Saturn. *Astrophys. J.* **609**, 1170–1180 (2004). doi: 10.1086/421257
47. R. Helled, T. Guillot, Interior models of Saturn: Including the uncertainties in shape and rotation. *Astrophys. J.* **767**, 113 (2013). doi: 10.1088/0004-637X/767/2/113

48. E. Galanti, Y. Kaspi, E. Tziperman, A full, self-consistent treatment of thermal wind balance on oblate fluid planets. *J. Fluid Mech.* **810**, 175–195 (2017). doi: 10.1017/jfm.2016.687
49. E. Galanti, Y. Kaspi, An adjoint based method for the inversion of the Juno and Cassini gravity measurements into wind fields. *Astrophys. J.* **820**, 91 (2016). doi: 10.3847/0004-637X/820/2/91
50. E. Galanti, Y. Kaspi, Decoupling Jupiter's deep and atmospheric flows using the upcoming Juno gravity measurements and a dynamical inverse model. *Icarus* **286**, 46–55 (2017). doi: 10.1016/j.icarus.2017.01.004
51. J. Liu, P. M. Goldreich, D. J. Stevenson, Constraints on deep-seated zonal winds inside Jupiter and Saturn. *Icarus* **196**, 653–664 (2008). doi: 10.1016/j.icarus.2007.11.036
52. H. Cao, D. Stevenson, Zonal flow magnetic field interaction in the semi-conducting region of giant planets. *Icarus* **296**, 59–72 (2017). doi: 10.1016/j.icarus.2017.05.015
53. D. Kong, K. Zhang, G. Schubert, Odd gravitational harmonics of Jupiter: Effects of spherical versus nonspherical geometry and mathematical smoothing of the equatorially antisymmetric zonal winds across the equatorial plane. *Icarus* **277**, 416–423 (2016). doi: 10.1016/j.icarus.2016.05.037
54. K. Zhang, D. Kong, G. Schubert, Thermal-gravitational wind equation for the wind-induced gravitational signature of giant gaseous planets: Mathematical derivation, numerical method, and illustrative solutions. *Astrophys. J.* **806**, 270 (2015). doi: 10.1088/0004-637X/806/2/270
55. E. Galanti, Y. Kaspi, Prediction for the flow-induced gravity field of Saturn: Implications for Cassini's Grand Finale. *Astrophys. J.* **843**, L25 (2017). doi: 10.3847/2041-8213/aa7aec
56. J. N. Cuzzi, J. J. Lissauer, L. W. Esposito, J. B. Holberg, E. A. Marouf, G. L. Tyler, A. Boischot, "Saturn's rings—Properties and processes," in *Planetary Rings* (Univ. of Arizona Press, 1984), pp. 73–199.
57. L. W. Esposito, M. O'Callaghan, R. A. West, The structure of Saturn's rings—Implications from the Voyager stellar occultation. *Icarus* **56**, 439–452 (1983). doi: 10.1016/0019-1035(83)90165-3
58. S. J. Robbins, G. R. Stewart, M. C. Lewis, J. E. Colwell, M. Sremčević, Estimating the masses of Saturn's A and B rings from high-optical depth *N*-body simulations and stellar occultations. *Icarus* **206**, 431–445 (2010). doi: 10.1016/j.icarus.2009.09.012
59. S. Charnoz, "Origin and evolution of Saturn's ring system" in *Saturn From Cassini-Huygens*, M. Dougherty, L. Esposito, S. Krimigis, Eds. (Springer, 2009), chap. 17, pp. 537–575.
60. R. Tajeddine, P. D. Nicholson, P.-Y. Longaretti, M. El Moutamid, J. A. Burns, What confines the rings of Saturn? *Astrophys. J. Suppl. Ser.* **232**, 28 (2017). doi: 10.3847/1538-4365/aa8c09
61. Z. Zhang *et al*., Exposure age of Saturn's A and B rings, and the Cassini division as suggested by their non-icy material content. *Icarus* **294**, 14–42 (2017). doi: 10.1016/j.icarus.2017.04.008
62. S. Kempf, "The age of Saturn's rings constrained by the meteroid flux into the system," paper presented at the American Geophysical Union Fall Meeting, New Orleans, LA, 13 December 2017.
63. S. Charnoz, J. Salmon, A. Crida, The recent formation of Saturn's moonlets from viscous spreading of the main rings. *Nature* **465**, 752–754 (2010). doi: 10.1038/nature09096; pmid: 20535205
64. J. Salmon, S. Charnoz, A. Crida, A. Brahic, Long-term and large-scale viscous evolution of dense planetary rings. *Icarus* **209**, 771–785 (2010). doi: 10.1016/j.icarus.2010.05.030
65. L. Dones, A recent cometary origin for Saturn's rings? *Icarus* **92**, 194–203 (1991). doi: 10.1016/0019-1035(91)90045-U
66. H. C. Dones, C. B. Agnor, E. Asphaug, Formation of Saturn's rings by tidal disruption of a centaur. *Bulletin of the American Astronomical Society* **39**, 7.07 (2007).
67. M. Ćuk, L. Dones, D. Nesvorný, Dynamical evidence for a late formation of Saturn's moons. *Astrophys. J.* **820**, 97 (2016). doi: 10.3847/0004-637X/820/2/97
68. R. Hyodo, S. Charnoz, Dynamical evolution of the debris disk after a satellite catastrophic disruption around Saturn. *Astron. J.* **154**, 34 (2017). doi: 10.3847/1538-3881/aa74c9
69. K. Lodders, H. Palme, H. P. Gail, *Solar System* (Springer, 2009), vol. 4B.

ACKNOWLEDGMENTS

We thank J. W. Armstrong, J. Cuzzi, M. Di Benedetto, S. Finocchiaro, J. Fortney, R. French, L. Gomez-Casajus, M. Gregnanin, M. Hedman, R. A. Jacobson, E. Marouf, N. Movshovitz, and L. Spilker for their suggestions and contributions. We are grateful to the Cassini Radio Science operations team members (E. Barbinis, D. Kahan, J. Gao, C. Lee) for supporting the experiments described in this paper. We also thank the DSN and ESA personnel involved in the acquisition of the data. L.I. thanks the late Bruno Bertotti for his mentorship and contributions to the Cassini mission. **Funding:** Support for this work was provided by the Italian Space Agency (L.I., D.D., P.R., M.J.M., P.T., and M.Z.), NASA's CDAP program (B.M., W.H., and S.W.), the Cassini Project (P.N. and A.A.), and the Israeli Space Agency (Y.K. and E.G.). All authors acknowledge support from the Cassini Project. **Author contributions:** L.I. led the experiment and supervised the data analysis. D.D. and P.R. carried out the gravity data analysis. B.M., W.H., and S.W. contributed interior models and the differential rotation with CMS method. Y.K. and E.G. contributed the models on differential rotation with finite depth flows. P.N. contributed the ring mass interpretation. A.A. coordinated the experiment operations and data acquisition. M.J.M., P.T., and M.Z. helped with the data analysis. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The Cassini data used in this paper is available through the NASA Planetary Data System at https://atmos.nmsu.edu/pdsd/archive/data/co-s-rss-1-sagr18-v10/ and https://atmos.nmsu.edu/pdsd/archive/data/co-s-rss-1-sagr19-v10/. The MONTE space navigation code was obtained through a license agreement between NASA and the Italian Space Agency; the terms do not permit redistribution. MONTE licenses may be requested at https://montepy.jpl.nasa.gov/. Our output gravitational field model is listed in Table 1. Executable code for running the interior models is available at https://github.com/smwahl/CMS_Saturn (CMS method) and https://github.com/egalanti/TW_Saturn (thermal-wind method, requires MATLAB).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/eaat2965/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References (*70*–*72*)

12 February 2018; accepted 19 December 2018
Published online 17 January 2019
10.1126/science.aat2965

RESEARCH ARTICLE SUMMARY

GAS GIANT PLANETS

Close Cassini flybys of Saturn's ring moons Pan, Daphnis, Atlas, Pandora, and Epimetheus

B. J. Buratti*, P. C. Thomas, E. Roussos, C. Howett, M. Seiß, A. R. Hendrix, P. Helfenstein, R. H. Brown, R. N. Clark, T. Denk, G. Filacchione, H. Hoffmann, G. H. Jones, N. Khawaja, P. Kollmann, N. Krupp, J. Lunine, T. W. Momary, C. Paranicas, F. Postberg, M. Sachse, F. Spahn, J. Spencer, R. Srama, T. Albin, K. H. Baines, M. Ciarniello, T. Economou, H.-W. Hsu, S. Kempf, S. M. Krimigis, D. Mitchell, G. Moragas-Klostermeyer, P. D. Nicholson, C. C. Porco, H. Rosenberg, J. Simolka, L. A. Soderblom

INTRODUCTION: Saturn's main ring system is associated with a family of small moons. Pan and Daphnis orbit within the A-ring's Encke Gap and Keeler Gap, respectively, whereas Pandora and Prometheus orbit just outside the F-ring and Atlas just outside the A-ring. The latter three moons help to confine ring particles.

The moons Janus and Epimetheus are in closely spaced orbits that they exchange approximately every 4 years; these two objects may be collisional fragments of a larger body. All these moons have densities much less than 1000 kg/m^3, indicating that they formed from ring debris that accumulated around a preexisting core.

RATIONALE: During the final stages of the Cassini mission, the spacecraft made a series of close observations of Saturn's rings. Flybys of Pan, Daphnis, Pandora, Atlas, and Epimetheus were performed to investigate the geologic processes shaping their surfaces, their composition, their thermal and ultraviolet properties, their relationship to Saturn's ring system, and their interactions with particles in Saturn's magnetosphere.

RESULTS: The moons that orbit in ring gaps or are adjacent to the main rings have equatorial ridges of material consisting of accreted particles that are distinct from their rounded central cores. The cores are more structurally sound than ridges, with rougher surfaces and more impact craters. Complex patterns of grooves formed by tidal stresses crisscross the moons.

A visible-infrared reflectance spectrum of Pan, which is embedded in the rings, shows that it is redder than any of the other ring moons. The color of the moons becomes more red as the distance to Enceladus increases. This suggests that the optical properties of the moons are determined by the balance of two external effects: addition of a red coloring agent from the main rings, and accretion of neutral-colored icy particles or water vapor possibly from the E-ring, which is formed by Enceladus's plume. The exact composition of the red material from the main ring system is unknown, although a mixture containing organic silicates and iron is likely. Differences in particle size also affect the moons' spectra.

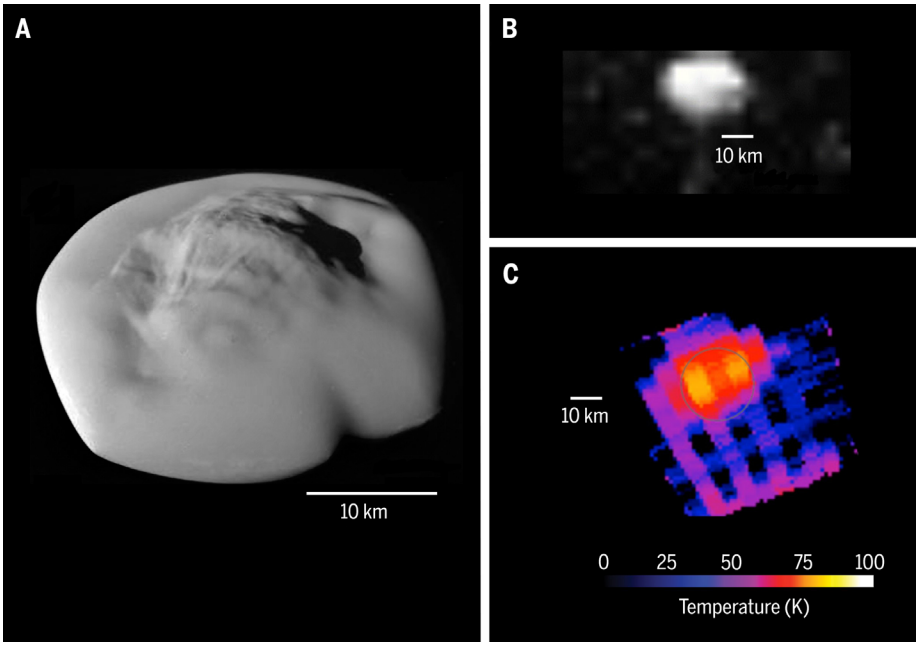
ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat2349

Measurements of the spectral slope of Epimetheus in the ultraviolet suggest that it is less affected by particles from the E-ring than are the mid-sized main moons farther from Saturn. Temperature maps were derived for both Atlas and Epimetheus, whose blackbody temperatures were 82 ± 5 K and 90 ± 3 K, respectively.

Carbon dioxide, which is present on the eight mid-sized saturnian moons, was not detected on the ring moons, nor were any volatiles other than water ice. Measurements showed a scarcity of high-energy ions in the vicinity of the ring moons and only transient energetic electron populations; in the ring gaps, no trapped electron or proton radiation was detected. Although particle bombardment alters both the albedo and color of the main moons' surfaces, for the ring moons it appears to be unimportant.

CONCLUSION: Saturn's ring moons record a complex geologic history with groove formation caused by tidal stresses and accretion of ring particles. The moons embedded within the rings or near their edges have solid cores with equatorial ridges of more weakly consolidated material. The finding of a porous surface further supports substantial accretion. High-resolution images strongly suggest exposures of a solid substrate distinct from the mobile regolith that frequently covers essentially all small Solar System objects. These exposures may eventually help to reveal systematic trends of the evolution of moons and their geologic structures for the whole of Saturn's satellite system. ■



The A-ring "shepherd" moon Atlas showing ridges of accumulated ring particles. (**A**) Visible-light image of Atlas showing the core and equatorial ridge. (**B**) Infrared image of the moon at 2.0 μm. (**C**) Thermal infrared scan of Atlas consistent with a mean temperature of 82 ± 5 K.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: bonnie.buratti@jpl.nasa.gov
Cite this article as B. J. Buratti *et al.*, *Science* **364**, eaat2349 (2019). DOI: 10.1126/science.aat2349

RESEARCH ARTICLE

GAS GIANT PLANETS

Close Cassini flybys of Saturn's ring moons Pan, Daphnis, Atlas, Pandora, and Epimetheus

B. J. Buratti[1]*, P. C. Thomas[2], E. Roussos[3], C. Howett[4], M. Seiß[5], A. R. Hendrix[6], P. Helfenstein[2], R. H. Brown[7], R. N. Clark[6], T. Denk[8], G. Filacchione[9], H. Hoffmann[5], G. H. Jones[10], N. Khawaja[8,11], P. Kollmann[11], N. Krupp[3], J. Lunine[2], T. W. Momary[1], C. Paranicas[12], F. Postberg[8,11], M. Sachse[5], F. Spahn[5], J. Spencer[4], R. Srama[13], T. Albin[13], K. H. Baines[1], M. Ciarniello[9], T. Economou[14], H.-W. Hsu[15], S. Kempf[15], S. M. Krimigis[12], D. Mitchell[12], G. Moragas-Klostermeyer[13], P. D. Nicholson[2], C. C. Porco[16], H. Rosenberg[8], J. Simolka[13], L. A. Soderblom[17]

Saturn's main ring system is associated with a set of small moons that either are embedded within it or interact with the rings to alter their shape and composition. Five close flybys of the moons Pan, Daphnis, Atlas, Pandora, and Epimetheus were performed between December 2016 and April 2017 during the ring-grazing orbits of the Cassini mission. Data on the moons' morphology, structure, particle environment, and composition were returned, along with images in the ultraviolet and thermal infrared. We find that the optical properties of the moons' surfaces are determined by two competing processes: contamination by a red material formed in Saturn's main ring system and accretion of bright icy particles or water vapor from volcanic plumes originating on the moon Enceladus.

Saturn possesses a family of small inner irregular moons that orbit close to its rings. Two moons orbit in gaps within Saturn's main ring system: Daphnis, which dwells in the A-ring's Keeler Gap (*1*), and Pan, which is found in the Encke Gap, also in the A-ring (*2*). Three others, called shepherd moons, orbit at the edges of the A-ring (Atlas) or the F-ring (Pandora and Prometheus) (*3*) (fig. S2). The co-orbital moons Janus and Epimetheus share horseshoe orbits outside the F-ring and swap their positions every 4 years (fig. S2). Saturn's rings are almost certainly tied to the origin and continued existence of these moons (*1*). It remains unclear whether the rings formed from the breakup of an inner moon, or whether the present ring moons formed from the consolidation of existing ring material, either primordial or impact-created. The alteration processes acting on these moons and the rings, past and present, are also unknown.

Prior to Saturn's exploration by spacecraft, the main rings were thought to be unconsolidated primordial debris, unable to form a moon because of tidal forces (*4*, *5*). Evidence from the two Voyager spacecraft suggested that the rings and inner moons constituted debris from the breakup of a single parent body, or perhaps several parent bodies, with the moons being the largest fragments (*4*). Measurement of the rings' and moons' bulk densities using Cassini data (*5*), along with dynamical studies and the existence of ridges around the equators of Atlas and Pan (*5*, *6*), suggested a more complicated, multistage formation. The ring moons—from Pan out to Pandora, but possibly also Janus and Epimetheus—likely formed from the very early accretion of low-density debris around a denser seed, presumably a collisional shard from the breakup of a preexisting moon (*5*). In the cases of Atlas and Pan, this was followed by a second stage of accretion of material onto the equator, after the rings had settled into their present very thin disk (*6*, *7*). In this scenario, the surfaces of these moons should be similar in composition to the rings.

Analysis of the optical properties of the moons, including color, albedo, and spectral properties in the visible and infrared between 0.35 and 5.2 μm, has shown that they resemble the ring systems in which they are embedded or which they abut (*8*–*11*). An unidentified low-albedo reddish material that could be organic molecules, silicates, or iron particles (*9*–*12*) appears to be abundant in the rings and has also tinged the moons (*8*–*12*), further supporting a common origin and implying continuing accretion of particles onto the moons' surfaces. The interactions of the ring system with the inner moons may form two distinct zones: an inner region in the vicinity of the main ring system that is dominated by the red chromophore, and an outer region that is dominated by fresh, high-albedo icy particles from the E-ring. Complicating the picture, however, is the possible influence of interactions with magnetospheric particles, which have been shown to alter the color and albedo of the main moon system of Saturn (*13*, *14*). It is unclear whether any volatiles other than water ice exist on the ring moons. The presence of molecules with higher volatility than water ice would indicate material originating in a colder region outside the saturnian system; for example, the discovery of CO_2 ice on the irregular outer moon Phoebe suggested that it originated in the Kuiper Belt (*15*).

The last phase of Cassini's mission began on 30 November 2016 and ended on 15 September 2017, with two distinct periods: the ring-grazing (or F-ring) orbits, in which 20 close passes to the F-ring were performed, and the proximal orbits (or "Grand Finale"), which executed 23 dives between the planet and the main ring system.

Table 1. Summary of the close flybys of Saturn's ring moons during the ring-grazing orbits.

Moon	Semimajor axis (R_S)	Rotation rate (days)	Date of flyby	Closest approach (km)	Best image resolution (m/pixel)
Pan	2.22	0.575	7 Mar 2017	22,247	147
Daphnis	2.26	0.594	16 Jan 2017	22,336	170
Atlas	2.29	0.602	12 Apr 2017	10,848	76
Pandora	2.35	0.629	18 Dec 2016	22,157	132
Epimetheus	2.51	0.695	30 Jan 2017	3,625	36
			21 Feb 2017	8,266	82

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. [2]Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA. [3]Max Planck Institute for Solar System Research, 37077 Göttingen, Germany. [4]Southwest Research Institute, Boulder, CO 80302, USA. [5]Department of Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany. [6]Planetary Sciences Institute, Tucson, AZ 85719, USA. [7]Lunar and Planetary Lab, University of Arizona, Tucson, AZ 85721, USA. [8]Institute of Geological Sciences, Freie Universität Berlin, 12249 Berlin, Germany. [9]Istituto Nazionale di Astrofisica, Rome, Italy. [10]University College London, London, UK. [11]Institute of Earth Sciences, University of Heidelberg, 69120 Heidelberg, Germany. [12]Applied Physics Laboratory, Johns Hopkins University, Laurel, MD 20723, USA. [13]University of Stuttgart, 70569 Stuttgart, Germany. [14]Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA. [15]Physics Department, University of Colorado, Boulder, CO 80303, USA. [16]Space Sciences Institute, Boulder, CO 80301, USA, and Department of Astronomy, University of California, Berkeley, CA 94720, USA. [17]U.S. Geological Survey, Flagstaff, AZ 86001, USA.
*Corresponding author. Email: bonnie.buratti@jpl.nasa.gov

During the ring-grazing orbits, Cassini performed its closest flybys of Pan, Daphnis, Atlas, Pandora, and Epimetheus (Table 1). A second flyby of Epimetheus was performed at a slightly greater distance. Data were obtained using several instruments on Cassini: the Imaging Science Subsystem (ISS) (*16*); the Visual Infrared Mapping Spectrometer (VIMS), taking medium-resolution spectra between 0.35 and 5.1 μm (*17*); the Cassini Composite Infrared Spectrometer (CIRS) (*18*); the Ultraviolet Imaging Spectrometer (UVIS) (*19*); the Cosmic Dust Analyzer (CDA) (*20*); and the Magnetospheric Imaging Instrument (MIMI) (*21*). The dust and plasma environment in the vicinity of the small inner moons was observed by CDA and MIMI during the subsequent proximal orbits.

Geology and morphology

Previous images of the ring moons showed distinctive equatorial ridges on Pan and Atlas (*5*, *6*) that were interpreted as likely formed by accretion of ring particles, whereas those of Daphnis were ambiguous. The small satellites are all in synchronous rotation, tidally locked to the planet (*7*). Prometheus's and Pandora's orbits straddle the F-ring, and although they exhibit different surface morphology, their densities are nearly identical (*3*) (table S1). The small (mean radius <5 km) satellites Aegaeon, Methone, and Pallene, which orbit in diffuse rings or ring arcs (*22*, *23*), have smooth ellipsoidal shapes indicative of hydrostatic equilibrium (*7*). The co-orbital satellites Epimetheus and Janus, by far the largest of the inner small moons, have nearly identical mean densities (table S1), which are also the highest among the inner small moons. Grooves had been observed on Epimetheus (*24*), and there were suggestions of discrete crater-filling sediments on both Janus and Epimetheus (*7*). Epimetheus experiences a ~7° forced wobble (libration) around a purely synchronous rotation (*25*). Table S1 summarizes the shapes, volumes, and calculated mean densities of the small satellites of Saturn based on the images taken during the flybys (*26*, *27*). Epimetheus and Janus have densities substantially above 500 kg m^{-3}; the lowest density (and highest uncertainty) is that of Daphnis, at 274 ± 142 kg m^{-3}. Surface accelerations vary substantially across each object because of their irregular shapes and tidal accelerations (table S1).

Main ring moons and ridges

The flyby images in Fig. 1 show that the equatorial ridges on Pan and Atlas are morphologically distinct from the more rounded central component of each moon. The ridges are different sizes on each moon: The fractional volumes of the ridges are Pan ~10%, Daphnis ~1%, and Atlas ~25%. Atlas's ridge appears smooth in the highest-resolution image (76 m/pixel), with some elongate brighter albedo markings. The ridge contacts the central component that has distinct ridge and groove topography (Fig. 1C); it has a previously known slight polygonal equatorial profile (*7*). Pan's ridge has a distinct boundary with the central component, a somewhat polygonal equatorial shape, some grooves, small ridges, and several small impact craters. The profile of Pan's ridge varies considerably with longitude. Figure 2 shows Pan's northern hemisphere, with calculated relative gravitational topography and surface slopes using existing techniques (*3*, *5*, *7*) (data S1). Unlike some equatorial ridges on small asteroids (*28*, *29*), Pan's ridge was not formed by material sliding toward lower gravitational potential areas generated by rotation and tides, because the slope directions are not latitudinally directed. The distinct boundary between

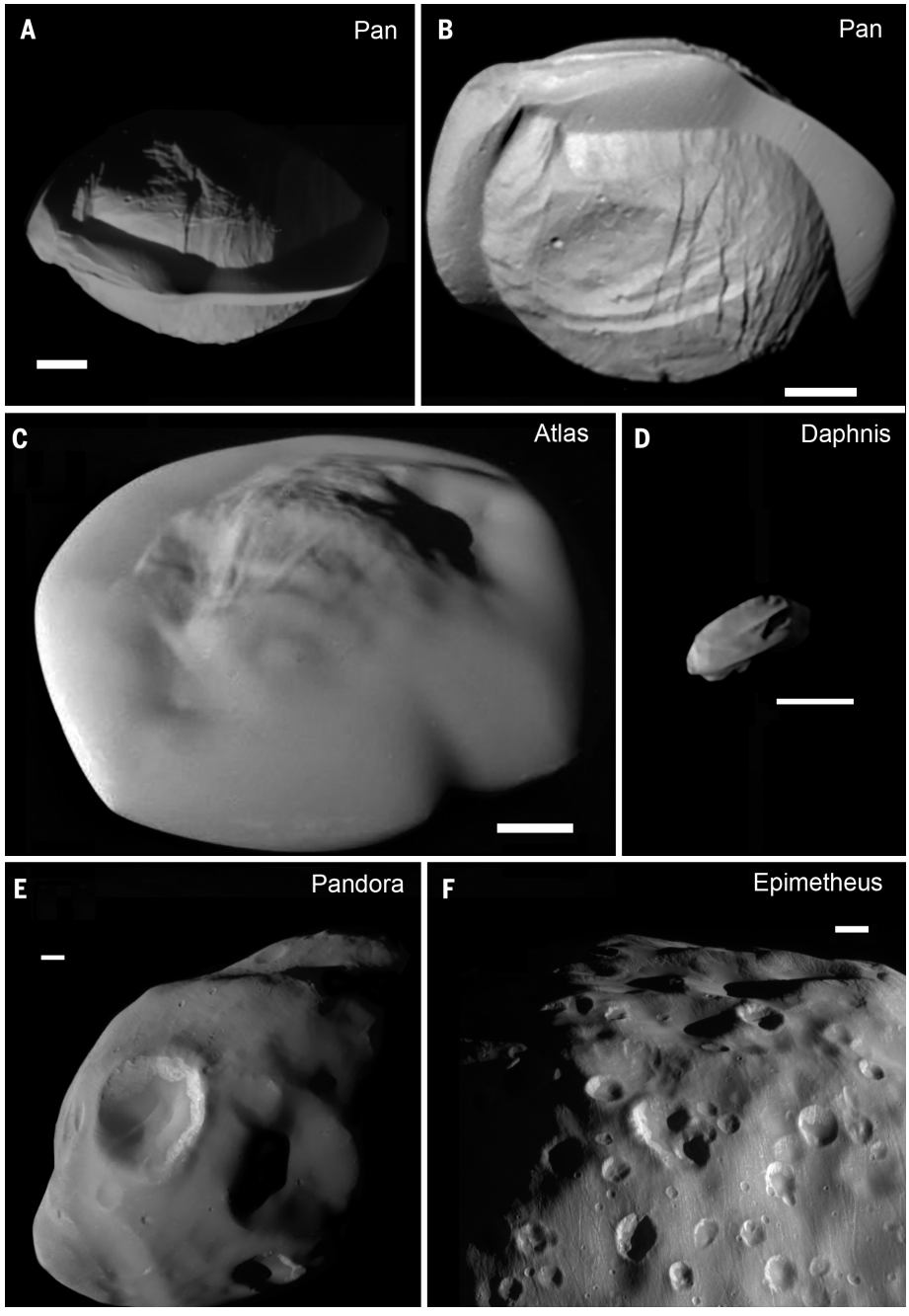


Fig. 1. Grayscale images of the ring moons obtained with ISS during the Cassini flybys. (**A**) Pan, image number N1867606181, Clear/Clear filters, from 26°S, at a scale of 182 m/pixel. (**B**) Pan, N186704669, Clear/Clear filters, from 39°N; 147 m/pixel. (**C**) Atlas, N1870699087, Clear/IR3 filters, from 40°N; anti-Saturn point at lower left; 108 m/pixel. (**D**) Daphnis, N1863267232, Clear/Green filters, from 14°N; anti-Saturn point to left; 170 m/pixel. (**E**) Pandora, N1860790629, Clear/Green filters, 240 m/pixel. The sub-spacecraft point is 35°N, 98°W; Pandora's north pole is close to two small craters above the large, bright-walled crater. (**F**) Epimetheus, N1866365809, Clear/UV3 filters, 99 m/pixel. Grooves and craters dominate the surface. All scale bars, 5 km. Images were chosen for scale and viewing geometry; different filters have little effect on visibility of morphologic detail.

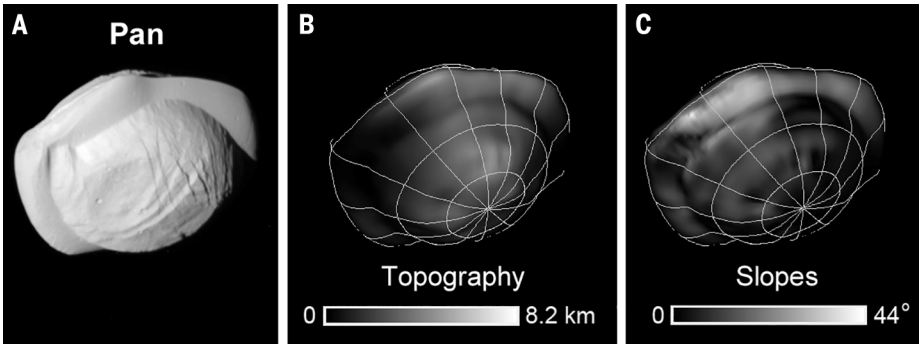


Fig. 2. Relative topography and slopes on Pan. (**A**) Grayscale image N1867604669 from 39°N, 217°W (rotated from view in Fig. 1). (**B**) Dynamic topography is the relative potential energy at the surface (due to mass, rotation, and tides) divided by an average surface acceleration (*5*, *7*). A homogeneous interior density is assumed. (**C**) Slopes are the angles between the surface normals and the (negative) net acceleration vectors.

the ridge and the central component core, the differing surface morphology on each, and the large differences in relative heights along the ridge require the ridge formation to be unrelated to surface, gravity-driven processes. These observations are consistent with formation of the ridge by the accretion of particles, with a distribution dictated by the relative orbital and rotational dynamics of the moon and ring particles (*6*).

The calculated mean densities of Pan, Atlas, and Daphnis result in calculated surface accelerations near zero at the sub- and anti-Saturn points, which suggests that those points cannot accrete additional material. The rest of the surfaces have inward-directed net accelerations. The surfaces of these three moons may be crudely divided into three units on the basis of morphology, geography, and surface texture visible at the available resolutions (Fig. 3). The equatorial ridges generally are the smoothest terrain on each moon.

The central components have more impact craters than do the ridges on Pan and Atlas, which display a few sub-kilometer impact craters. Pan and Atlas's central components show lineated topography indicative of structures such as faults or fractures. Pan has two distinct global sets of quasi-parallel faults. The first is roughly concentric to the long axis and exhibits conspicuous scarps and terracing, likely formed by equatorward displacements. The axial symmetry of this system suggests that tidal forces were involved in its development. The second system is oriented obliquely to the first and is visible in both the north and south hemispheres (Figs. 1A and 3C). In contrast, Atlas's central component core exhibits patterns of elongated ridge and groove topography that do not have fault scarp morphologies; the core appears to be covered by at least tens of meters of loose material (regolith).

Pan's equatorial ridge is thickest north-south at longitudes of approximately 220°, 310°, 135°, and 50°W, yet its radial extent peaks at longitudes of about 5°, 55°, 100°, 180°, 235°, and 310°W. The ridge supports grooves and small craters; their presence suggests some cohesion in this low-gravity environment (<2 mm s^{-2}). Atlas's equatorial profile is also somewhat polygonal but not as pronounced as Pan's.

The classification of some material units on Pan's southern hemisphere is ambiguous, in part because these are not directly illuminated by the Sun, only by light reflected off Saturn. These unclassified units in Fig. 3C include knobby streaks of hummocked material oriented approximately parallel to the equator, as well as hummocky deposits outlining a curvilinear depression on the Saturn-facing side.

The spatial resolution of the Daphnis imagery is 170 m/pixel, poorer than that of Pan (147 m/pixel) and Atlas (76 m/pixel). Daphnis is only about one-quarter the dimensions of the other ring moons. As a result, it is not clear whether its near-equatorial ridge is smoother or otherwise different from the rest of the surface. The equatorial ridge extends at least from 75°W to 185°W. An additional ridge at 22°N runs from ~60°W to 120°W. Both ridges are 300 to 400 m north-south and perhaps radially 300 m in extent. The core has an elongated (2.5 km) depression that is roughly aligned east-west.

F-ring moons

Prometheus and Pandora orbit inside and outside the F-ring, respectively. The images taken during the Pandora flyby show grooves and debris on the surface of this shepherd moon (Fig. 1E). Although many of the grooves form a pattern concentric to the major axis of the body, there is a slight offset between them, especially noticeable on the sub-Saturn side, which reflects orientations seen in previous observations (*22*).

Part of Pandora's leading hemisphere is smooth in comparison to other regions on this moon (Figs. 1E and 3D). The smooth deposits are most continuous near the equator but become patchy at high latitudes, where they appear to be too thin to mute the coarse surface relief along protruding crater rims. The smooth deposits extend approximately ±60° in latitude, slightly more than the maximum latitude of the ridge on Atlas. This arrangement might indicate the accretion of material, as with the main ring moons. If so, the accretion efficacy on Pandora is at least two orders of magnitude smaller than on Pan and Atlas, and much broader latitudinally. However, variations in resolution, illumination, and viewing geometry make mapping of textural variations on Pandora ambiguous.

Co-orbital moons

The highest-resolution images of the flybys were of Epimetheus, the smaller of the co-orbital moons, reaching scales of 36 and 49 m/pixel. These data enabled enhanced mapping of grooves and sediment coverings seen in previous observations (*24*). The grooves are global in occurrence, largely beaded to straight, elongated depressions that appear to be formed in loose regolith. There are some exposures of brighter material apparently devoid of regolith cover (Fig. 1F) that also show elongated lineations, generally slight depressions. These align with the grooves nearby that appear to be regolith features and largely align with the regolith groove global patterns. This association appears to support a previously proposed relation of at least some regolith grooves with fractures or other structures in a more rigid underlying bedrock, although the variety of groove morphologies on many objects suggests a multiplicity of groove origins (*24*, *30*–*32*).

Colors of the small ring satellites and Pandora

The whole-disk colors of the ring satellites as measured in ISS broadband filters (*33*) follow trends with distance from Saturn similar to those found by the VIMS instrument (*8*–*11*). The ISS Narrow Angle Camera (NAC) uses paired broadband filters. The CL1:UV3 pair (0.341 μm) and CL1:IR3 pair (0.930 μm) span the spectral range of the camera, and IR3/UV3 ratios represent the observed brightness value in each CL1:UV3 broadband filter relative to the corresponding value in the CL1:IR3 filter (*7*). For reference, Enceladus, the presumed source of ice particles that mute colors on other satellites, has an effectively neutral IR3/UV3 ratio of 1.03 ± 0.02 (*34*).

The spatially resolved colors of Pan, Daphnis, and Atlas can be used to show the effects of material deposited from the rings (*3*) (table S2). Closest to Saturn, Pan's average IR3/UV3 ratio is red at 2.5 ± 0.2 but is significantly smaller than the value of 3.3 ± 0.2 for the adjacent A-ring (i.e., Pan is less red than the rings). Farther out, the A-ring IR3/UV3 ratio decreases from 2.7 ± 0.2 on the inside of the Keeler Gap (which contains Daphnis) to 2.2 ± 0.3 on the outside. The mean value is not statistically different from that of Daphnis itself, 2.3 ± 0.3. The equatorial ridges on the ring satellites may be very old (*5*), but the colors most likely reflect a patina of material deposited from geologically recent and ongoing processes. Atlas, which falls just outside the A-ring, has an IR3/UV3 ratio of 2.4 ± 0.1. Pandora, which is near the F-ring and farther from Saturn, has a lower IR3/UV3 ratio of 1.9 ± 0.1. It lacks an equatorial ridge but possesses smooth deposits, which on the leading side extend from the equator to mid-latitudes.

Among the terrains shown in Fig. 3, color differences can be identified in the high-resolution images of all moons except Daphnis, for which the CL1:UV3 images were badly blurred by spacecraft motion. The IR3/UV3 ratio for cratered materials on Pan is about 19% higher than for its equatorial ridge and reaches approximately the average global value. Similarly, the ratio for cratered materials on Atlas is about 16% higher than for its ridge, but in this case, the global average value closely matches that of Atlas's larger equatorial ridge. For Pandora, the cratered materials have an IR3/UV3 ratio that is 15% lower than for the smooth materials toward the equator. The global average ratio falls between that of the cratered material and the smooth deposits. Exposed substrate is visible as a scarp on Pan and a bright exposed crater wall on Pandora. On Pan, the IR3/UV3 ratio of exposed substrate is intermediate between the ridge materials and crater materials. However, on Pandora, the corresponding ratio for the exposed crater wall is not statistically distinguishable from that of the cratered material.

Composition

Compositional information on the surfaces of the moons was acquired using VIMS (*17*). Prior to the close flybys of the ring moons, spectra taken by VIMS from greater distances were obtained (*8*–*11*). Water ice was the only volatile identified, but the moons' visible colors varied, especially in the 0.35- to 0.55-μm spectral region, which suggested contamination by a reddish chromophore that perhaps came from the ring system itself. This coloring agent is distinct from the low-albedo red material from the Phoebe ring that is deposited on the leading hemisphere of Iapetus and on Hyperion (*8*, *9*).

The close flybys of the embedded moons Daphnis and Pan enabled the acquisition of spectra of these moons, although only an IR spectrum (1.0 to 5.0 μm) for Daphnis was successfully obtained. These data provide a test for the origin of the red chromophore in the inner saturnian system. They also provide rudimentary information on spatial variations in composition on the moon's surfaces, although the spatial resolution is only about 1 to 2% (depending on the instrument mode) of ISS's resolution (*3*) (fig. S1).

Figure 4 shows the spectrum of each moon from 0.35 to 4.2 or 3.6 μm (1 to 3.6 μm for Daphnis). The only absorption bands detectable are those of water ice at 1.25, 1.6, 2.0, and 3.0 μm. No other volatiles are detectable, including CO_2, although its strongest absorption band in this spectral region is at 4.26 μm, in the noisy region of the spectrum beyond about 3.5 μm. There is a deep absorption band for crystalline water ice at 1.65 μm. This spectral band is sensitive to radiation damage (*35*); its unusual depth relative to previous Cassini spectra of icy moons (*8*–*11*) implies a lack of radiation damage in the ring environment, which is expected given the dearth of high-energy particles in the rings (see below). Water ice spectral bands are also sensitive to grain size, with deeper bands signifying larger grains (*36*). A larger particle size could signify larger regolith grains in the main ring system than in the E-ring, or it could simply be due to gravitational escape of the smaller particles, some of which could be formed by continual impacts.

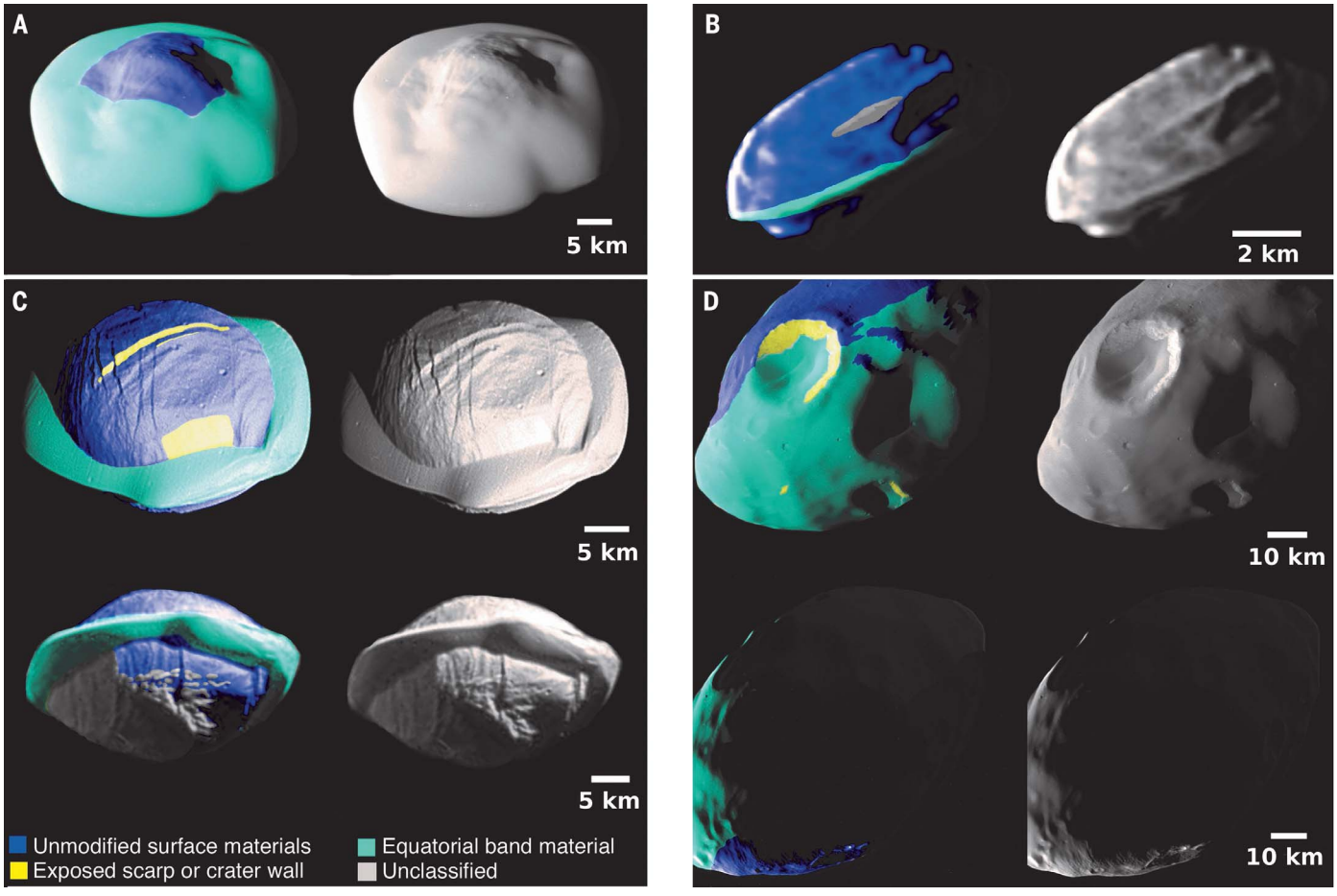


Fig. 3. Distribution of geological units on Pan, Atlas, Daphnis, and Pandora. In each panel, the three main units are highlighted in color, with the uninterpreted grayscale image alongside for comparison. Cratered surfaces (blue) have numerous craters, relatively crisp surface relief, and regolith typical of other small bodies in the saturnian system. Smooth terrains (cyan) are distinctly smooth relative to typical small-body cratered surfaces; some is material collected in crater floors. Exposed substrates (yellow) are relatively bright with lineations, more typical of rigid materials than of loose regolith. Unclassified areas (gray) are those for which insufficient data are available to resolve ambiguities between terrain types. (**A**) Atlas, resolution 94 m/pixel. (**B**) Daphnis, 167 m/pixel. (**C**) Pan, 144 m/pixel (top), 279 m/pixel (bottom). (**D**) Pandora, 137 m/pixel (top), 200 m/pixel (bottom).

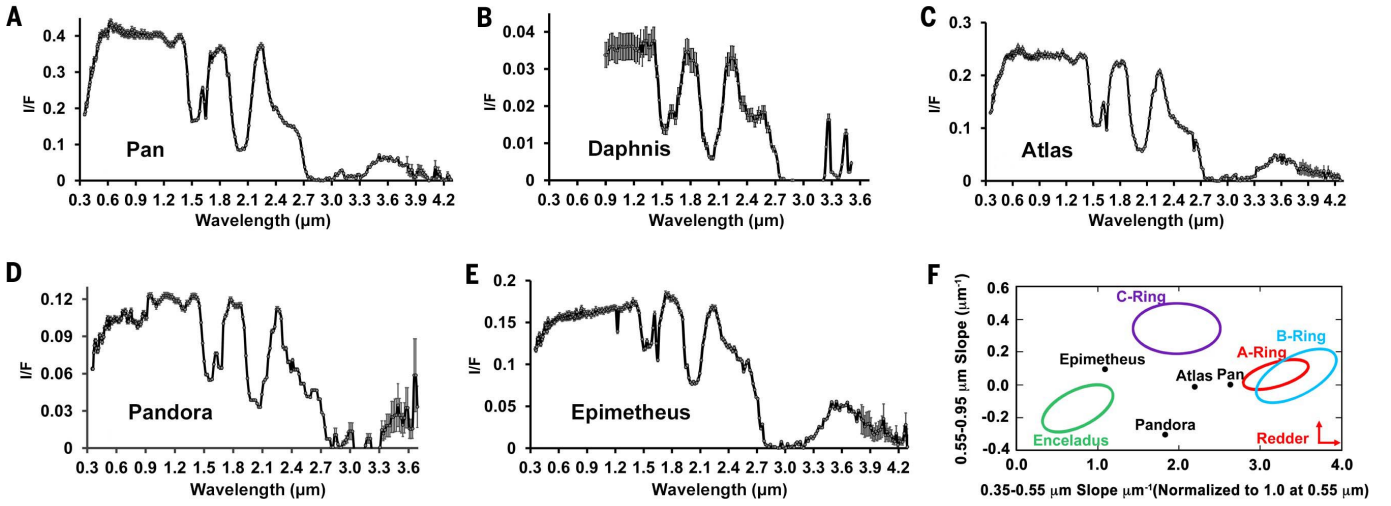


Fig. 4. VIMS spectra and colors of the five moons and the A- to C-rings. (**A** to **E**) Spectra of Pan (A), Daphnis (B), Atlas (C), Pandora (D), and Epimetheus (E). Noisy data at the long wavelengths are not shown. *I/F* is the reflected intensity compared with the incident solar flux. (**F**) Color-color plot of Saturn's main ring system and Enceladus (*8*, *9*) compared with Epimetheus, Atlas, Pandora, and Pan.

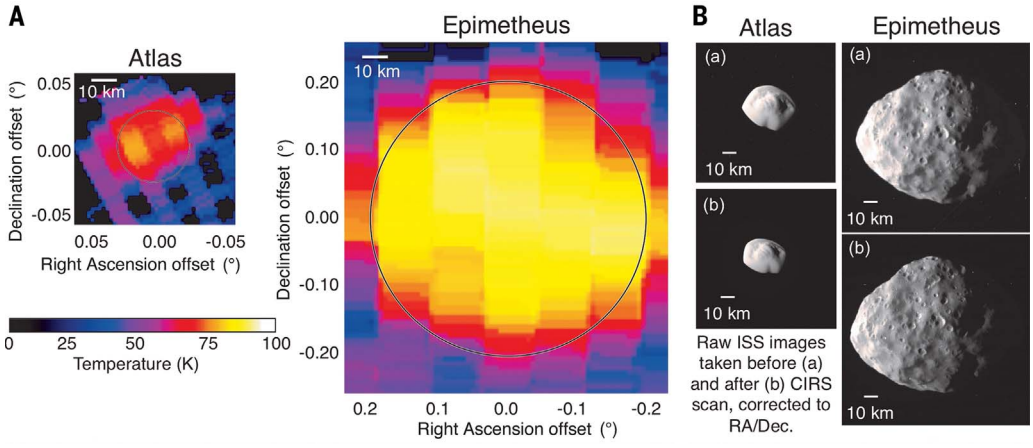


Fig. 5. CIRS thermal infrared and ISS visible-light observations of Atlas and Epimetheus. (**A**) Blackbody temperature distributions of the two moons, determined by fitting a blackbody curve to the full CIRS radiance spectrum at each location. The axes are offsets in right ascension and declination, with the origin at the center of the target (note that the IR images do not fall on this location because only the hotter side is detected). (**B**) ISS observations of both targets taken immediately before and after the CIRS scan, on the same orientation (*3*).

In general, there is a gradient depending on the position of the moon with respect to the rings, with Pan, which is embedded in the Encke Gap, being the reddest and Epimetheus, which is farthest from the rings and closest to Enceladus, being the bluest. (The one exception to this pattern is Pandora's bluer color in the 0.55- to 0.95-μm region.) This effect results from the countervailing processes of contamination by a red chromophore from the main rings and ice particles or water vapor from the E-ring, which originates from Enceladus's plume.

The VIMS colors agree with those derived by ISS. The VIMS equivalent values at the same wavelengths as the effective wavelengths of the ISS filters yield IR3/UV3 ratios of 2.7 ± 0.3 for Pan, 2.2 ± 0.2 for Atlas, 1.7 ± 0.2 for Pandora, and 1.5 ± 0.1 for Epimetheus. (The VIMS spectrum extends to only 0.35 μm; the visible slopes of the spectra were linearly extrapolated to 0.34 μm to match the wavelength of the ISS UV3 filter.) The moons embedded in the rings show spectral differences with the surrounding rings; in general they are less red (Fig. 4F). The VIMS ratio image of Atlas (fig. S1) shows uniformity between the main body and its equatorial ridge, at least in water ice abundance, which implies the accumulation of particles away from the equator to provide a globally homogeneous surface. Color differences below the spatial resolution of VIMS exist, as detected by ISS in the visible spectrum.

The spectrum of Pan is redder in the 0.35- and 0.55-μm region than other saturnian moons. Atlas, the shepherd moon just outside the A-ring, is also red but less so than Pan, and Pandora, which is associated with the F-ring, even less. The color of Epimetheus is more like that of the medium-sized moons Enceladus and Mimas (*8*–*10*). Thus, there is a gradient in color with distance from Saturn's ring system, with the embedded Pan being the reddest. Figure 4, A to E, shows that the slope of the visible spectrum increases as the distance to Saturn increases, and this is quantified in Fig. 4F, which shows the visible colors derived from the flybys with the colors of the main ring system of Saturn (*9*). These results imply that the red chromophore comes from the rings themselves. However, the differences in color between the moons and their adjacent rings—the small moons are consistently bluer than their surrounding rings—could be due to another contaminant: particles of almost pure water ice or vapor from the E-ring. This ring is a diffuse torus that is fed from the plume of Enceladus. The particles have a wide range of orbital elements and predominantly impact the leading sides of the main moons (and the trailing side of Mimas), altering their albedos and colors (*37*–*39*). The ring moons' leading hemispheres would tend to accrete more fresh grains of water ice than the surrounding ring particles.

The depth of the water ice band at 2.0 μm compared to the continuum at 1.8 μm (i.e., the 1.8/2.0 μm ratio) is 5.2 ± 0.1 for Pan, 5.0 ± 0.2 for Daphnis, 4.4 ± 0.1 for Atlas, 3.4 ± 0.1 for

Pandora, and 2.4 ± 0.1 for Epimetheus. The band depths increase closer to Saturn, most likely as a result of the increasing particle sizes (*36*). This is consistent with the moons embedded in the ring (Pan and Daphnis) being coated with main ring particles rather than with smaller particles from the E-ring. (The absorption band at 1.6 μm shows a similar but weaker trend.) Because the main ring system provides a shield against the E-ring, particle size may be a factor in determining the color and reflectivity of these moons.

Interactions between moons and magnetospheric particles can also alter the moons' colors and albedos (*13*, *14*). However, there is a dearth of high-energy particles in the vicinity of these moons (see below). Another factor that may alter spectral slopes and band depths is the particle size of the accreted ring particles (*36*), which may not be the same as that of the native particles.

Ultraviolet and thermal infrared observations of the moons

During the ring-grazing orbits, the spacecraft was in a high-radiation, high-dust environment that produced high background levels in ultraviolet observations with UVIS. The only moon detected was Epimetheus, during the encounter on 21 February 2017 (*3*) (fig. S4), in which the signal is above the background for only the longest far-UV wavelengths, ~0.170 to 0.19 μm. However, this single UV measurement of reflectance places some constraints on surface composition and external effects on Epimetheus. At 72° solar phase angle (the angle between the spacecraft, Epimetheus, and the Sun), the derived normal reflectance (average over 0.17 to 0.19 μm) is 0.09 ± 0.02. For comparison, this is lower than the reflectance measured at Tethys under similar viewing geometry by a factor of roughly 1.5 to 2 (*40*); however, Tethys has a higher visible geometric albedo [~1.2, versus ~0.73 for Epimetheus (*37*)], which indicates that Epimetheus may have a roughly uniformly lower reflectance than Tethys in the UV-visible range. The UV-visible spectral slope and albedo are strongly driven by external effects, because this spectral range senses the uppermost layer of the regolith affected by processes including plasma and E-ring grain bombardment. The UVIS measurement combined with the visible albedo suggests that Epimetheus is not as affected by the brightening effects of the E-ring grains as Tethys is (*37*), or that there is some other darkening agent or process important at Epimetheus's location. Thus, the UV-visible albedo of Epimetheus may simply reflect the relative importance of the alteration by the reddish lower-albedo chromophore and the icy E-ring particles at this moon's distance.

Thermal infrared observations with CIRS detected two moons: Epimetheus and Atlas (Fig. 5) (*3*). The results were modeled with a blackbody fitted to the observed radiance over all wavelengths. Both Epimetheus and Atlas are visible above the background sky. The mean surface temperatures are 90.1 ± 2.7 K on Epimetheus and 82.4 ± 4.7 K on Atlas.

Particle observations

Throughout the ring-grazing orbits, the particle and electromagnetic field instruments CDA and MIMI measured Saturn's plasma and dust environment, including the regions around the small inner moons. During this period, Cassini passed close to the orbits of the co-orbital moons Janus and Epimetheus. During 11 of the 20 ring plane crossings, the CDA's High Rate Detector (HRD) detected a total of about 2000 dust grains with radii larger than 0.8 μm (Fig. 6). The vertically integrated number density of grains smaller than 1.6 μm does not depend on the radial distance to Saturn, whereas the density of larger grains drops by about 50% over a radial distance of approximately 3500 km. Because the larger particles are less susceptible to nongravitational forces, large particles ejected from the moons stay closer to their parent bodies and form a more confined ring, which was previously detected by the Cassini camera (*41*). Fitting a Gaussian distribution to the HRD data, and accounting for the dust background from the F- and G-rings, constrains the radial full width at half maximum (FWHM) of the Janus-Epimetheus ring to about 4300 km. This implies a total number of 2 (±1) × 10^{19} ring particles larger than 1.6 μm.

Many dust rings are formed by ejecta from high-velocity impacts of interplanetary micrometeoroids eroding the surfaces of satellites without atmospheres. The measured particle number in the Janus-Epimetheus ring constrains the poorly known parameters of the impact-ejection dust creation model (*42*, *43*) at Saturn. Using an unfocused flux of >3.6 × 10^{-16} kg m^{-2} s^{-1} with a mean impact speed of 4.3 km s^{-1} (*44*), the dust production rate from both moons totals about 0.81 kg s^{-1} (0.57 kg s^{-1} from Janus and 0.24 kg s^{-1} from Epimetheus). This corresponds to 9.1 × 10^{11} particles larger than 1.6 μm per second (6.4 × 10^{11} s^{-1} from Janus and 2.7 × 10^{11} s^{-1} from Epimetheus), assuming a cumulative power-law size distribution for a dust diameter $d \propto d^{-\alpha}$ with α = 2.4 and a maximal ejecta mass of 1 × 10^{-8} kg, consistent with observations of impact-generated dust clouds around the Galilean moons (*42*, *45*).

To explain the measured number of ring particles, this comparably high production rate requires a shallow slope of the cumulative ejecta velocity v distribution $\propto v^{-\gamma}$ (γ = 1) and a kinetic energy dissipation at the higher end of the values predicted by laboratory experiments (*46*, *47*). The kinetic energy ratio of ejecta to impactor is 5%. This points to a highly dissipative and porous (snow or regolith) surface. We find that most ejecta are gravitationally bound to the moons and fall back to their surface, whereas only about 5% of them for Janus and 7% for Epimetheus escape to the surrounding ring. Numerical simulations (*3*) show that most of the ring particles are recaptured by the source moons after an average lifetime of 60 years, resulting in an estimate of 9.8 × 10^{19} ring particles larger than 1.6 μm. This value is in rough agreement with the observed value of 2 (±1) × 10^{19} particles, which in turn constrains the poorly known parameters of the impact-ejection model, which can vary by orders of magnitude.

The CDA Chemical Analyzer (*9*) recorded mass spectra of sub–micrometer-sized dust particles (0.1 to 0.4 μm). The compositional analysis of these spectra recorded near the ring plane shows mostly ice grains but also about 3% pure silicate grains or ice-silicate mixtures (*3*) (fig. S5). The source of the icy particles could be either the inner edge of the E-ring or surface ejecta of the nearby small ice moons. Because silicate-rich grains of this size have not been detected in the E-ring (*48*), these must originate from

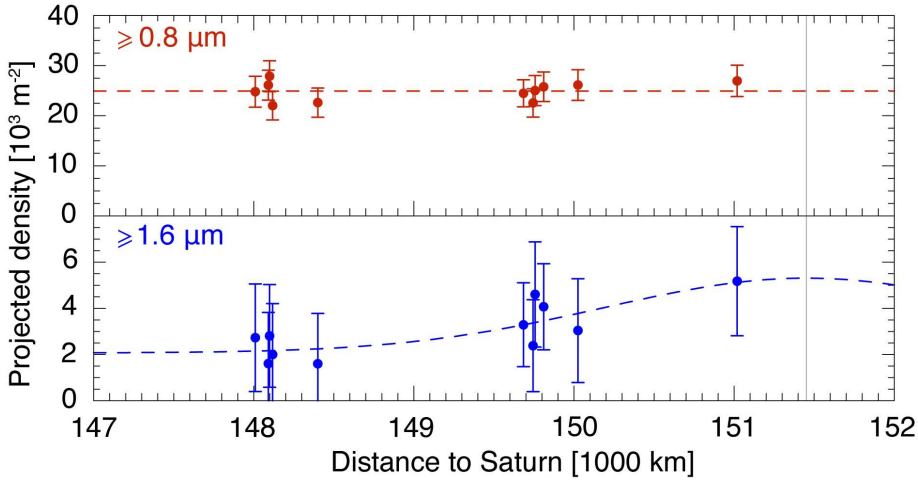


Fig. 6. Radial dust density distribution obtained from CDA-HRD measurements. The density of the ≥0.8-μm particles indicates a constant profile (red dashed line), whereas the density of the ≥1.6-μm particles decreases inward from the orbit of Janus and Epimetheus (vertical gray line). The dust distribution of the larger particles is modeled by a Gaussian distribution (blue dashed line) with a maximum at the mean radial position of Janus and Epimetheus, plus a constant background density. Error bars are derived from Poisson statistics.

a different source, possibly the nearby moons Janus and Epimetheus or the F- and G-rings.

The Low Energy Magnetospheric Measurements System (LEMMS) of the MIMI energetic charged particle detector surveyed the planet's radiation belts inward of Saturn's G-ring and monitored the energetic particle environment of the five small moons. LEMMS measures energetic electrons and ions above 18 and 27 keV, respectively, reaching into the MeV energy range. The region inward of Saturn's G-ring has been sampled in the past on several occasions with Pioneer 11 and Cassini (*49*–*51*). It contains the location where both Saturn's proton and electron radiation belts have their highest intensities, which lies between the G-ring and Janus and Epimetheus's orbits. Inward of that maximum, intensities drop gradually up to the outer edge of Saturn's A-ring, which absorbs all energetic particles. Superimposed on the radial profile of radiation belt fluxes are localized dropouts originating from Saturn's moons and rings (*52*). Although several of these features can be attributed to specific moons [e.g., Janus and Epimetheus (*53*)], any influences by Pandora, Prometheus, and Atlas (orbiting within the radiation belt boundaries) are less clear. These moons orbit close to Saturn's A- and F-rings, complicating the separation of the different contributions. Understanding how effectively these moons sweep out particle radiation also determines the radiation environment to which their surfaces are exposed.

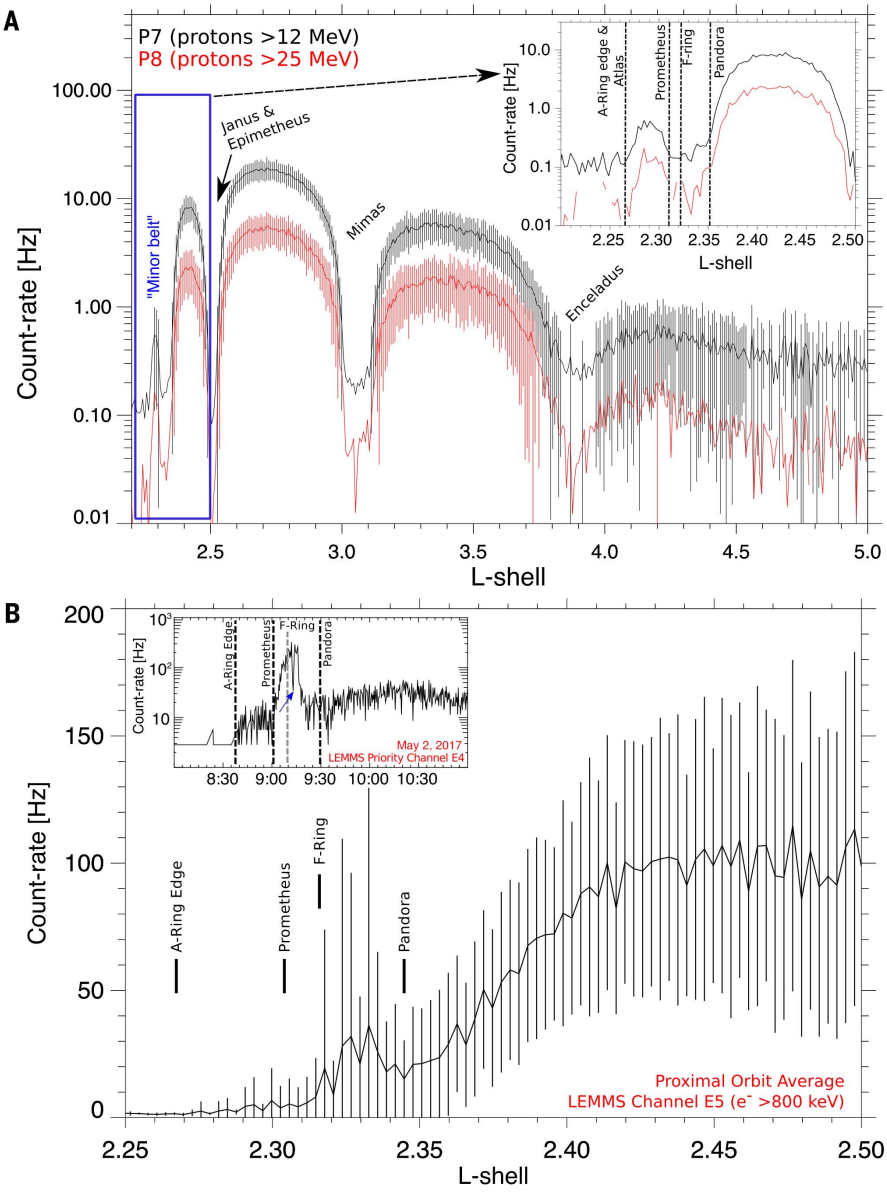


Fig. 7. Average count rates for protons and electrons, as measured by MIMI/LEMMS. The channels are >12 and >25 MeV for protons (**A**) and >800 keV for electrons (**B**); both sets of data are shown as a function of L-shell, with 1σ error bars. Absence of error bars indicates an uncertainty larger than the corresponding mean value. The orbits of several of Saturn's large icy moons are also marked. The inset in (A) zooms into the region of the Minor Belt, highlighting the absorbing effects of Atlas, Pandora, Prometheus, and the A- and F-rings. The inset in (B) shows a high–time resolution series of observations (1 sample per 0.3125 s) from LEMMS obtained during the second proximal orbit on 2 May 2017. The blue arrow marks an electron microsignature within one of the MeV electron spikes seen consistently during Cassini's outbound crossings near the L-shell of the A-ring's outer edge.

Figure 7A shows count rates of >12-MeV and >25-MeV protons as a function of the L-shell [expressed as multiples (L) of Saturn's radius of 60,268 km] averaged over the proximal orbits. The L-shell is defined as the distance from Saturn that a magnetic field line intersects the magnetic equator. The spacecraft L-shell at the time of data collection is determined by mapping along Saturn's magnetic field using a third-order multipole model for Saturn's internal magnetic field (*54*). Figure 7 shows the previously established sectorization of the MeV proton radiation belts, due to the moons and rings that absorb any protons diffusing across their orbits (*55*, *56*). Among these different sectors, the least well characterized by previous observations is the "Minor Belt," centered at approximately L = 2.29. The gap immediately outside the Minor Belt is centered near the F-ring (L ~ 2.32); we find that the gap boundaries coincide with the L-shells of Prometheus and Pandora (Fig. 7A). Pandora and Prometheus are therefore absorbing protons at a rate that is high enough to counter the diffusive influx of protons from the surrounding belt sectors. Effectively, the two moons and the F-ring form an extended obstacle to proton radiation. The net result is that the weathering of Pandora's and Prometheus's surfaces by energetic protons is negligible because they orbit within the proton radiation gaps they create. Atlas's effects cannot be distinguished from those of the A-ring, but that moon is also exposed to very low proton fluxes. Overall, almost all of Saturn's inner moons (except Dione, Rhea, or minor moons such as Anthe or Pallene) orbit in regions free from energetic ions (*57*–*59*). This is unlike Jupiter's satellites, whose surface chemistry and thin atmospheric properties are strongly affected by irradiation from high fluxes of keV and MeV particles (*60*, *61*).

Figure 7B shows the proximal orbit averages of electron count rates from LEMMS channel E5 (>0.8 MeV) as a function of the L-shell. Electron radiation levels are more variable than those of protons, as the large error bars indicate, because moons and rings are not effective in sweeping out electrons from their orbits (*52*, *62*). Inside L = 2.4 (inward of the Janus and Epimetheus orbits), electron rates fall slowly toward the outer edge of the A-ring (L = 2.27). This drop is interrupted by an enhancement of the mean electron rates near the L-shells of the F-ring, Pandora, and Prometheus. In the absence of a MeV electron source, such an enhancement,

which was absent from past observations at the same L-shell (*53*, *63*), is unexpected. The 1σ error bars in that location span more than two orders of magnitude in amplitude, indicating much higher variability than in the surrounding regions. This large scatter is attributed to spikes of enhanced MeV electron flux observed in 18 of the 22 outbound crossings outward of the A-ring's edge and between L = 2.31 and L = 2.35. The radial extent of an individual spike is less than 1800 km along the equatorial plane, and the electron intensity within them can be enhanced by as much as a factor of 300 relative to their surroundings. The inset of Fig. 7B shows one such resolved spike, captured by the high–time resolution measurements of LEMMS priority channel E4 (0.8 to 4.2 MeV) on 2 May 2017. Because most measurements in the inbound portion of Cassini's orbit showed no evidence of similar spikes in the same L-shell range, we deduce that these features are usually located a few hours after local noon, and their longitudinal extent ranges between 22° and 37° in the clockwise direction, starting from a magnetospheric local time of 13:20. The longitudinal extent cannot be constrained in the anticlockwise direction. Most of these enhancements were seen around the L-shells of the F-ring, Prometheus, and Pandora. This electron belt component is therefore limited in local-time range. As a result, energetic electron bombardment of the three moons is variable in intensity, is episodic, and occurs only for a fraction of their orbit around Saturn. Material interaction signatures of energetic electrons are seen as localized depletions (microsignatures) within the electron spikes. These may be due to Atlas, Prometheus, Pandora, or F-ring clumps (*63*); an example is shown in the inset of Fig. 7B and could have formed only after the electron enhancement developed.

There is no discernible signal of trapped electron or proton radiation at the orbits of the Keeler and Encke Gaps, where Daphnis and Pan are orbiting (*54*).

Summary and conclusions

The low densities of the small moons of Saturn, measured during the flybys, are consistent with a multistage formation scenario involving accretion of ring material (*5*, *6*). The colors of the moons embedded in the A-ring are more consistent with the rings the closer the moons are to Saturn. This suggests the ongoing accretion of a reddish chromophore, possibly a mixture of organics and iron (*9*–*12*), onto the surfaces of the moons. The difference in color between the moons and their adjacent ring may be explained by the accretion of bright, icy particles or, more likely, water vapor from the E-ring. Each moon's surface is subjected to a balance between these two ongoing processes, with their distance from Saturn and Enceladus determining the resulting color, as illustrated in Fig. 4F. The detection of abundant ice grains by CDA supports this view. The bluer core of Atlas is also explained by the accretion of E-ring particles, which have a wider range of inclinations than main ring particles. If the ring moons formed from the same material as the rings, they would have been the same color, and the color gradient may be solely due to contamination by the E-ring. The size of particles on the moons' surfaces also plays a role, especially for the moons embedded in the main ring system, which would shield these moons from the E-ring.

The dearth of high-energy ions close to the moons lessens the alteration processes caused by bombardment with magnetospheric particles. The strong crystalline water ice band at 1.65 μm also suggests low radiation damage. This low-energy plasma environment is unlike the main moons of Saturn, especially Dione and Rhea, as they dwell in a region where alteration by ions is substantial. Particle radiation would tend to darken and redden the surfaces, so the red chromophore on the trailing hemispheres of the main moons may be unrelated to the red material contributing to the colors of the ring moons (*64*). Contamination of Saturn's rings by bright icy particles or water vapor offers counterevidence to previous arguments that the observed brightness of the rings indicates recent formation (*65*).

The moons' geology records a complex history, including groove formation caused by tidal stresses and accretion of ring particles. The CDA finding of a porous surface further supports substantial accretion. Although the topography and surface slopes strongly suggest that the equatorial ridges of Pan and Atlas are accreted from the rings and are not formed by normal surface transport, the ridges on these objects appear in a variety of forms. The flyby images strongly suggest exposures of a solid substrate distinct from the mobile regolith that covers many small Solar System objects.

REFERENCES AND NOTES

1. C. C. Porco *et al.*, Cassini Imaging Science: Initial results on Saturn's rings and small satellites. *Science* **307**, 1226–1236 (2005). doi: 10.1126/science.1108056; pmid: 15731439
2. M. R. Showalter, Visual detection of 1981S13, Saturn's eighteenth satellite, and its role in the Encke gap. *Nature* **351**, 709–713 (1991). doi: 10.1038/351709a0
3. See supplementary materials.
4. B. A. Smith *et al.*, Encounter with Saturn: Voyager 1 imaging science results. *Science* **212**, 163–191 (1981). doi: 10.1126/science.212.4491.163; pmid: 17783827
5. C. C. Porco, P. C. Thomas, J. W. Weiss, D. C. Richardson, Saturn's small inner satellites: Clues to their origins. *Science* **318**, 1602–1607 (2007). doi: 10.1126/science.1143977; pmid: 18063794
6. S. Charnoz, A. Brahic, P. C. Thomas, C. C. Porco, The equatorial ridges of Pan and Atlas: Terminal accretionary ornaments? *Science* **318**, 1622–1624 (2007). doi: 10.1126/science.1148631; pmid: 18063797
7. P. C. Thomas *et al.*, The inner small satellites of Saturn: A variety of worlds. *Icarus* **226**, 999–1019 (2013). doi: 10.1016/j.icarus.2013.07.022
8. G. Filacchione *et al.*, Saturn's icy satellites investigated by *Cassini*-VIMS. II. Results at the end of nominal mission. *Icarus* **206**, 507–523 (2010). doi: 10.1016/j.icarus.2009.11.006
9. G. Filacchione *et al.*, Saturn's icy satellites and rings investigated by *Cassini*-VIMS: III - Radial compositional variability. *Icarus* **220**, 1064–1096 (2012). doi: 10.1016/j.icarus.2012.06.040
10. G. Filacchione *et al.*, The radial distribution of water ice and chromophores across Saturn's system. *Astrophys. J.* **766**, 76–80 (2013). doi: 10.1088/0004-637X/766/2/76
11. B. J. Buratti *et al.*, *Cassini* spectra and photometry 0.25-5.1 μm of the small inner satellites of Saturn. *Icarus* **206**, 524–536 (2010). doi: 10.1016/j.icarus.2009.08.015
12. R. N. Clark *et al.*, The surface composition of Iapetus: Mapping results from *Cassini* VIMS. *Icarus* **218**, 831–860 (2012). doi: 10.1016/j.icarus.2012.01.008
13. P. Schenk *et al.*, Plasma, plumes and rings: Saturn system dynamics as recorded in global color patterns on its midsize icy satellites. *Icarus* **211**, 740–757 (2011). doi: 10.1016/j.icarus.2010.08.016
14. A. R. Hendrix, G. Filacchione, C. Paranicas, P. Schenk, F. Scipioni, Icy Saturnian satellites: Disk-integrated UV-IR characteristics and links to exogenic processes. *Icarus* **300**, 103–114 (2018). doi: 10.1016/j.icarus.2017.08.037
15. T. V. Johnson, J. I. Lunine, Saturn's moon Phoebe as a captured body from the outer Solar System. *Nature* **435**, 69–71 (2005). doi: 10.1038/nature03384; pmid: 15875015
16. C. C. Porco *et al.*, *Cassini* maging science: Instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115**, 363–497 (2004). doi: 10.1007/s11214-004-1456-7
17. R. H. Brown *et al.*, The *Cassini* Visual and Infrared Mapping Spectrometer (VIMS) investigation. *Space Sci. Rev.* **115**, 111–118 (2004). doi: 10.1007/s11214-004-1453-x
18. F. M. Flasar *et al.*, Exploring the Saturn system in the thermal infrared: The composite infrared spectrometer. *Space Sci. Rev.* **115**, 169–297 (2004). doi: 10.1007/s11214-004-1454-9
19. L. W. Esposito *et al.*, The *Cassini* ultraviolet imaging spectrograph investigation. *Space Sci. Rev.* **115**, 299–361 (2004). doi: 10.1007/s11214-004-1455-8
20. R. Srama *et al.*, The *Cassini* cosmic dust analyzer. *Space Sci. Rev.* **114**, 465–518 (2004). doi: 10.1007/s11214-004-1435-z
21. S. M. Krimigis *et al.*, Magnetosphere IMaging Instrument (MIMI) on the *Cassini* mission to Saturn/Titan. *Space Sci. Rev.* **114**, 233–329 (2004). doi: 10.1007/s11214-004-1410-8
22. M. M. Hedman *et al.*, Three tenuous rings/arcs for three tiny moons. *Icarus* **199**, 378–386 (2009). doi: 10.1016/j.icarus.2008.11.001
23. M. M. Hedman *et al.*, Aegaeon (Saturn LIII), a G-ring object. *Icarus* **207**, 433–446 (2010). doi: 10.1016/j.icarus.2009.10.024
24. S. J. Morrison, P. C. Thomas, M. S. Tiscareno, J. A. Burns, J. Veverka, Grooves on small Saturnian satellites and other objects: Characteristics and significance. *Icarus* **204**, 262–270 (2009). doi: 10.1016/j.icarus.2009.06.003
25. M. S. Tiscareno, P. C. Thomas, J. A. Burns, The rotation of Janus and Epimetheus. *Icarus* **204**, 254–261 (2009). doi: 10.1016/j.icarus.2009.06.023
26. N. J. Cooper, S. Renner, C. D. Murray, M. W. Evans, Saturn's inner satellites: Orbits, masses, and the chaotic motion of Atlas from new *Cassini* imaging observations. *Astron. J.* **149**, 27–45 (2015). doi: 10.1088/0004-6256/149/1/27
27. J. W. Weiss, C. C. Porco, M. S. Tiscareno, Ring edge waves and the masses of nearby satellites. *Astron. J.* **138**, 272–286 (2009). doi: 10.1088/0004-6256/138/1/272
28. S. J. Ostro *et al.*, Radar imaging of binary near-Earth asteroid (66391) 1999 KW4. *Science* **314**, 1276–1280 (2006). doi: 10.1126/science.1133622; pmid: 17038586
29. A. W. Harris, E. G. Fahnestock, P. Pravec, On the shapes and spins of "rubble pile" asteroids. *Icarus* **199**, 310–318 (2009). doi: 10.1016/j.icarus.2008.09.012
30. S. J. Weidenschilling, A possible origin for the grooves of Phobos. *Nature* **282**, 697–698 (1979). doi: 10.1038/282697a0
31. D. L. Buczkowski, O. S. Barnouin-Jha, L. M. Prockter, 433 Eros lineaments: Global mapping and analysis. *Icarus* **193**, 39–52 (2008). doi: 10.1016/j.icarus.2007.06.028
32. M. Nayak, E. Asphaug, Sesquinary catenae on the Martian satellite Phobos from reaccretion of escaping ejecta. *Nat. Commun.* **7**, 12591 (2016). doi: 10.1038/ncomms12591; pmid: 27575002
33. R. A. West *et al.*, In-flight calibration of the *Cassini* imaging science sub-system cameras. *Planet. Space Sci.* **58**, 1475–1488 (2010). doi: 10.1016/j.pss.2010.07.006
34. P. C. Thomas, M. Tiscareno, P. Helfenstein, in *Enceladus and the Icy Moons of Saturn*, P. Schenk, R. Clark, C. Howett, A. Verbiscer, H. Waite, Eds. (University of Arizona, 2018).
35. R. M. E. Mastrapa, R. H. Brown, Ion irradiation of crystalline H_2O-ice: Effect on the 1.65-μm band. *Icarus* **183**, 207–214 (2006). doi: 10.1016/j.icarus.2006.02.006
36. R. N. Clark, Water frost and ice - The near-infrared spectral reflectance 0.65-2.5 microns. *J. Geophys. Res.* **86**, 3087–3096 (1981). doi: 10.1029/JB086iB04p03087
37. D. P. Hamilton, J. A. Burns, Origin of Saturn's E Ring: Self-Sustained, Naturally. *Science* **264**, 550–553 (1994). doi: 10.1126/science.264.5158.550; pmid: 17732738

38. B. J. Buratti, J. A. Mosher, T. V. Johnson, Albedo and color maps of the Saturnian satellites. *Icarus* **87**, 339–357 (1990). doi: 10.1016/0019-1035(90)90138-Y
39. A. Verbiscer, R. French, M. Showalter, P. Helfenstein, Enceladus: Cosmic graffiti artist caught in the act. *Science* **315**, 815 (2007). doi: 10.1126/science.1134681; pmid: 17289992
40. E. M. Royer, A. Hendrix, First far-ultraviolet disk-integrated phase curve analysis of Mimas, Tethys and Dione from the Cassini-UVIS data sets. *Icarus* **242**, 158–171 (2014). doi: 10.1016/j.icarus.2014.07.026
41. C. C. Porco, *IAU Circular 8759* (2006).
42. A. V. Krivov, M. Sremčević, F. Spahn, V. V. Dikarev, K. V. Kholshevnikov, Impact-generated dust clouds around planetary satellites: Spherically symmetric case. *Planet. Space Sci.* **51**, 251–269 (2003). doi: 10.1016/S0032-0633(02)00147-2
43. F. Spahn *et al*., E-ring dust sources: Implications from *Cassini*'s dust measurements. *Planet. Space Sci.* **54**, 1024–1032 (2006). doi: 10.1016/j.pss.2006.05.022
44. N. Altobelli, S. Kempf, F. Postberg, A. Poppe, C. Fischer, T. Albin, R. Srama, Exogenous Dust at Saturn. Cassini Science Symposium (Boulder, CO, 2018); http://lasp.colorado.edu/media/projects/cassini/meetings/css2018/2-TUESDAY/3-Rings_3_(R3-)/1_Altobelli.ppt.
45. H. Krüger, A. V. Krivov, M. Sremčević, E. Grün, Impact-generated dust clouds surrounding the Galilean moons. *Icarus* **164**, 170–187 (2003). doi: 10.1016/S0019-1035(03)00127-1
46. N. Asada, Fine fragments in high-velocity impact experiments. *J. Geophys. Res.* **90**, 12,445–12,453 (1985). doi: 10.1029/JB090iB14p12445
47. W. K. Hartmann, Impact experiments. I - Ejecta velocity distributions and related results from regolith targets. *Icarus* **63**, 69–98 (1985). doi: 10.1016/0019-1035(85)90021-1
48. J. K. Hillier *et al*., Interplanetary dust detected by the Cassini CDA Chemical Analyser. *Icarus* **190**, 643–654 (2007). doi: 10.1016/j.icarus.2007.03.024
49. J. A. Van Allen *et al*., Pioneer 11 observations of energetic particles in the jovian magnetosphere. *Science* **188**, 459–462 (1975). doi: 10.1126/science.188.4187.459; pmid: 17734361
50. F. B. McDonald, A. W. Schardt, J. H. Trainor, If you've seen one magnetosphere, you haven't seen them all - Energetic particle observations in the Saturn magnetosphere. *J. Geophys. Res.* **85**, 5813–5830 (1980). doi: 10.1029/JA085iA11p05813
51. S. M. Krimigis *et al*., Dynamics of Saturn's magnetosphere from MIMI during Cassini's orbital insertion. *Science* **307**, 1270–1273 (2005). doi: 10.1126/science.1105978; pmid: 15731445
52. E. Roussos *et al*., Evidence for dust-driven, radial plasma transport in Saturn's inner radiation belts. *Icarus* **274**, 272–283 (2016). doi: 10.1016/j.icarus.2016.02.054
53. J. A. Van Allen, Findings on rings and inner satellites of Saturn of Pioneer 11. *Icarus* **51**, 509–527 (1982). doi: 10.1016/0019-1035(82)90143-9
54. E. Roussos *et al*., Sources, sinks and transport of energetic electrons near Saturn's main rings. *Geophys. Res. Lett.* **45**, 10.1029/2018gl078097 (2018). doi: doi:10.1029/2018GL078097
55. E. Roussos *et al*., Discovery of a transient radiation belt at Saturn. *Geophys. Res. Lett.* **35**, L22106 (2008). doi: 10.1029/2008GL035767
56. P. Kollmann, E. Roussos, A. Kotova, C. Paranicas, N. Krupp, The evolution of Saturn's radiation belts modulated by changes in radial diffusion. *Nat. Astron.* **1**, 872–877 (2017). doi: 10.1038/s41550-017-0287-x
57. B. D. Teolis *et al*., Cassini finds an oxygen-carbon dioxide atmosphere at Saturn's icy moon Rhea. *Science* **330**, 1813–1815 (2010). doi: 10.1126/science.1198366; pmid: 21109635
58. P. Kollmann *et al*., Energetic particle phase space densities at Saturn: *Cassini* observations and interpretations. *J. Geophys. Res.* **116**, A05222 (2011). doi: 10.1029/2010JA016221
59. C. Paranicas *et al*., Energetic charged particle weathering of Saturn's inner satellites. *Planet. Space Sci.* **61**, 60–65 (2012). doi: 10.1016/j.pss.2011.02.012
60. B. H. Mauk *et al*., Energetic ion characteristics and neutral gas interactions in Jupiter's magnetosphere. *J. Geophys. Res.* **109**, A09S12 (2004). doi: 10.1029/2003JA010270
61. J. F. Cooper, R. E. Johnson, B. H. Mauk, H. B. Garrett, N. Gehrels, Energetic Ion and Electron Irradiation of the Icy Galilean Satellites. *Icarus* **149**, 133–159 (2001). doi: 10.1006/icar.2000.6498
62. E. Roussos *et al*., The variable extension of Saturn's electron radiation belts. *Planet. Space Sci.* **104**, 3–17 (2014). doi: 10.1016/j.pss.2014.03.021
63. J. N. Cuzzi, J. A. Burns, Charged particle depletion surrounding Saturn's F ring - Evidence for a moonlet belt? *Icarus* **74**, 284–324 (1988). doi: 10.1016/0019-1035(88)90043-7
64. R. E. Johnson, T. I. Quickenden, Photolysis and radiolysis of water ice on outer solar system bodies. *J. Geophys. Res.* **102**, 10985–10996 (1997). doi: 10.1029/97JE00068
65. Z. Zhang *et al*., Exposure age of Saturn's A and B rings, and the *Cassini* Division as suggested by their non-icy material content. *Icarus* **294**, 14–42 (2017). doi: 10.1016/j.icarus.2017.04.008
66. H. Hoffmann *et al*., hohoff/ddx: DDX (2019). doi: 10.5281/zenodo.2572758

ACKNOWLEDGMENTS

The authors are grateful to the Cassini project engineers and staff for their dedicated service that led to the success of the final stages of the mission. **Funding:** This paper was funded by the Cassini Project. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract to NASA. Also supported by Deutsches Zentrum für Luft- und Raumfahrt grants OH 1401 and 1503 and by Deutsche Forschungsgemeinschaft (DFG) grants Ho5720/1-1 and OH 1401 (M.Se., R.S., H.H., M.Sa., F.S., T.A., S.K., N.Kh., G.M.-K., F.P., and J.Si.), 50OH1501 (J.Si., G.M.-K., and T.A.), and 50OH1503 (T.D. and H.R.); the Italian Space Agency (G.F. and M.C.); and DFG projects PO 1015/2-1, /3-1, /4-1 and ERC Consolidator Grant 724908-Habitat OASIS (N.Kh. and F.P.). **Author contributions:** B.J.B., P.C.T., E.R., C.H., M.Se., A.R.H., P.H., H.H., N.Kh., H.-W.H., T.W.M.: planning observations, data analysis, and writing; R.H.B., R.S., T.A., K.H.B., S.K., S.M.K., D.M., G.M.-K., P.D.N., C.C.P., H.R., J.Si., L.A.S.: instrument development and planning observations; R.N.C., T.D., M.Sa., F.S., J.Sp., N.Kr., F.P., C.P., G.H.J., P.K., J.L., H.-W.H.: data analysis and planning observations; G.F., M.C., T.E.: data analysis. **Competing interests:** None. **Data and materials availability:** All data used in this paper are archived in NASA's Planetary Data System (PDS). The ISS, VIMS, CIRS, and UVIS data can be found at https://pds-rings.seti.org/cassini/. Relevant periods of data acquisition (in Universal Time): Pan, 7 March 2017, 16:35 to 19:05; Daphnis, 16 January 2017, 11:33 to 14:03; Atlas, 12 April 2017, 11:30 to 14:10; Pandora, 18 December 2016, 19:59 to 21:54; Epimetheus, 30 January 2017, 19:22 to 21:12, and 21 February 2017, 09:33 to 10:43. The CDA and MIMI data were acquired continuously throughout the F-ring orbital period, lasting from 30 November 2016 to 22 April 2017 and during the proximal orbits, which lasted from the end of the F-ring orbits until the end of the mission on 15 September 2017. CDA observations can be found at https://pds.nasa.gov/ds-view/pds/viewDataset.jsp?dsid=CO-D-CDA-3/4/5-DUST-V1.0, and MIMI data at https://pds-ppi.igpp.ucla.edu/search/view/?f=yes&id=pds://PPI/CO-S-MIMI-4-LEMMS-CALIB-V1.0. The Pan model slope and topography shown in Fig. 2 are provided in data S1. The software for the CDA modeling in the supplementary materials can be found at https://github.com/hohoff/ddx (*66*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/eaat2349/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S4
Data S1
References (*67*–*86*)

6 February 2018; accepted 12 March 2019
Published online 28 March 2019
10.1126/science.aat2349

RESEARCH ARTICLE SUMMARY

GAS GIANT PLANETS

Close-range remote sensing of Saturn's rings during Cassini's ring-grazing orbits and Grand Finale

Matthew S. Tiscareno*, Philip D. Nicholson, Jeffrey N. Cuzzi, Linda J. Spilker, Carl D. Murray, Matthew M. Hedman, Joshua E. Colwell, Joseph A. Burns, Shawn M. Brooks, Roger N. Clark, Nicholas J. Cooper, Estelle Deau, Cecile Ferrari, Gianrico Filacchione, Richard G. Jerousek, Stéphane Le Mouélic, Ryuji Morishima, Stu Pilorz, Sébastien Rodriguez, Mark R. Showalter, Sarah V. Badman, Emily J. Baker, Bonnie J. Buratti, Kevin H. Baines, Christophe Sotin

INTRODUCTION: Saturn's rings are an accessible exemplar of astrophysical disk processes and a delicate tracer of the Saturn system's dynamical processes and history.

RATIONALE: During its ring grazing orbits and Grand Finale, the Cassini spacecraft passed very close to Saturn's main rings and obtained very high–spatial-resolution images, spectral scans, and temperature scans.

RESULTS: We find structures related to the detailed sculpting of rings by embedded masses, including structures near the moon Daphnis that have apparently experienced markedly different perturbations compared to the surrounding ring material, and complex structure elements within the largest propeller-shaped disturbances. Interpreting certain such elements in terms of the Hill radius yields diameters of 1.0 to 1.6 km for the largest propeller-causing moons.

Several classes of subkilometer structure in the ring, which we call textures, are found in well-defined radial bands, which in many cases are difficult to correlate with other ring properties. The plateaux in the C ring exhibit a characteristic streaky texture. We hypothesize that these textures indicate variation in properties that affect the results of particle-particle collisions.

Medium-strength density waves neither alter the spectral characteristics of the region surrounding them nor exclude swarms of propellers from their vicinity, as the largest-density waves are known to do. We also confirm that even the strongest bending waves (such as Mimas 5:3) do not exhibit any signs of spectral halos. However, medium-strength density waves do exhibit clumpy texture in their troughs, and they also alter the propeller size distribution.

"Mini-jets" in the F ring are found in clusters, whose members evolve in lockstep with each other. This provides the strongest evidence yet that impacts onto the rings are commonly due to (Saturn-orbiting) streams of material, rather than lone impacting objects.

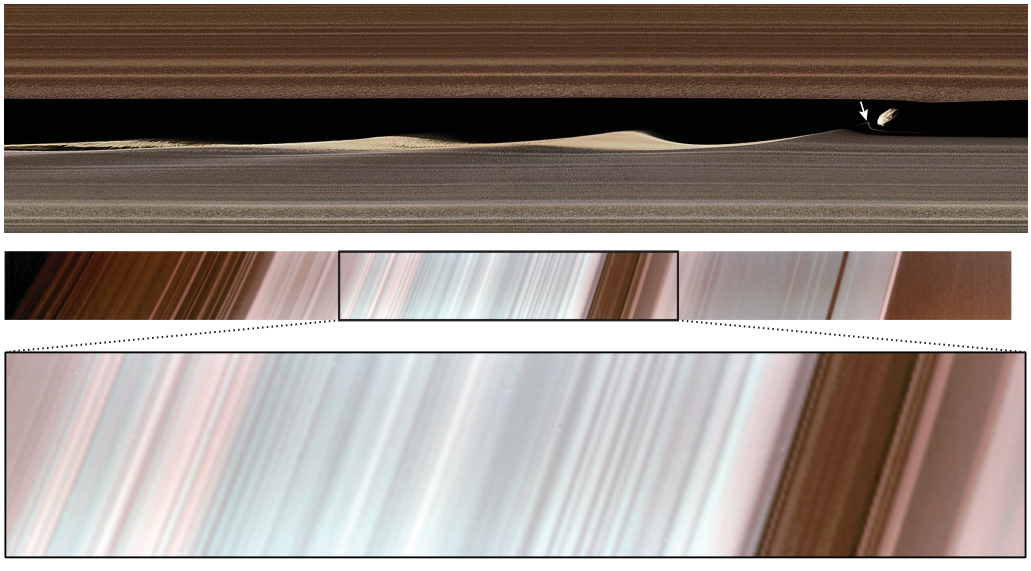
ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau1017

The distinct light-scattering characteristics of the narrow region outward of the Keeler gap—weaker water ice absorption bands, higher reflectivity, grayish rather than reddish color—transition abruptly from the rest of the A ring, although different degrees of abruptness are seen in the visible and the near-infrared. The combination of weaker water-ice band depths with higher reflectivity is difficult to understand.

Water-ice band depth and general color slope are closely correlated with optical depth, and temperature is anticorrelated (that is, denser regions are colder), even in sharply banded regions, down to the spatial resolution limit of ~3 km px^{-1}. However, the narrow bright bands in the C ring, called plateaux, have similar color slopes and water-ice band depths to those of the surrounding C ring, despite their marked difference in brightness. Furthermore, denser regions are warmer in some fine-scaled structures, including C-ring plateaux and structure in the B ring, both on the lit side only, and strong waves in the A ring on both the lit and unlit sides.

CONCLUSION: The rings are sculpted by embedded masses, producing structure visible down to our resolution limit. Correlations of spectral properties and temperature with optical depth are tight at many locations, although exceptions are found that deepen puzzles in certain regions. Many of these results are likely related to radial stratification in particle properties, rather than in chemical composition or surface mass density. ■



False-color images of Saturn's rings. (Top) A mosaic showing Daphnis in the Keeler gap on the lit side of the rings, with three wave crests of the structure raised by Daphnis in the gap's outer edge. (Middle and bottom) Visual and Infrared Mapping Spectrometer radial scans across the lit side of the main rings, displayed as false-color images. Reddish colors signify a higher fraction of components other than water ice. The boxed region in the middle panel indicates the location of the bottom panel.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: matt@seti.org
Cite this article as M. S. Tiscareno *et al.*, *Science* **364**, eaau1017 (2019). DOI: 10.1126/science.aau1017

RESEARCH ARTICLE

GAS GIANT PLANETS

Close-range remote sensing of Saturn's rings during Cassini's ring-grazing orbits and Grand Finale

Matthew S. Tiscareno[1*], Philip D. Nicholson[2], Jeffrey N. Cuzzi[3], Linda J. Spilker[4], Carl D. Murray[5], Matthew M. Hedman[6], Joshua E. Colwell[7], Joseph A. Burns[2,8], Shawn M. Brooks[4], Roger N. Clark[9], Nicholas J. Cooper[5], Estelle Deau[4,10], Cecile Ferrari[11], Gianrico Filacchione[12], Richard G. Jerousek[7], Stéphane Le Mouélic[13], Ryuji Morishima[4,10], Stu Pilorz[1], Sébastien Rodriguez[11], Mark R. Showalter[1], Sarah V. Badman[14], Emily J. Baker[15], Bonnie J. Buratti[4], Kevin H. Baines[4], Christophe Sotin[4]

Saturn's rings are an accessible exemplar of an astrophysical disk, tracing the Saturn system's dynamical processes and history. We present close-range remote-sensing observations of the main rings from the Cassini spacecraft. We find detailed sculpting of the rings by embedded masses, and banded texture belts throughout the rings. Saturn-orbiting streams of material impact the F ring. There are fine-scaled correlations among optical depth, spectral properties, and temperature in the B ring, but anticorrelations within strong density waves in the A ring. There is no spectral distinction between plateaux and the rest of the C ring, whereas the region outward of the Keeler gap is spectrally distinct from nearby regions. These results likely indicate that radial stratification of particle physical properties, rather than compositional differences, is responsible for producing these ring structures.

Saturn's main rings (*1*) are the regions known as the A, B, C, and F rings and the Cassini Division, which together constitute a broad, dense disk of orbiting particles lying in the equatorial plane between 74,600 and 140,200 km (that is, 1.2 to 2.3 Saturn radii) from Saturn's center. The normal optical depth τ_n (*2*) in the main rings is at least $\tau_n \gtrsim 0.1$, and in some locations $\tau_n \gg 1$. Particle composition is dominated by water ice, with trace contaminants, and particle size ranges from millimeters to meters (although a few localities, including the F ring, are rich in dust). By contrast, the dusty rings are inward (the D ring) and outward (the E and G rings) of the main rings, have $\tau_n \ll 0.1$, and are dominated by micrometer-sized particles (*3*).

The 13-year mission of the Cassini spacecraft at Saturn culminated in two stages. During the ring-grazing orbits (RGOs), from December 2016 to April 2017, Cassini executed near-polar orbits that crossed the ring plane just outward of the F ring, providing close-range viewing of the outer portion of the main rings. During the Grand Finale (GF), from April to September 2017, the orbit shifted so that closest approach to Saturn was inward of the D ring, passing a few thousand km above Saturn's cloud tops, providing close-range flybys of the inner portion of the main rings. We present and analyze optical remote-sensing data taken during these flybys (*4*). We use data from Cassini's Imaging Science Subsystem (ISS) (*5*), Visual and Infrared Mapping Spectrometer (VIMS) (*6*), Composite Infrared Spectrometer (CIRS) (*7*), and Ultraviolet Imaging Spectrograph (UVIS) (*8*).

Ring structure: Embedded moons and propellers

Each ring particle follows its own orbit around the planet. However, collective effects, including collisions and mutual self-gravity, affect the structure of the rings, as do interactions with larger orbiting objects. Moons ranging in size from Titan to Pan raise spiral density and bending waves in the ring (*9*–*11*). Moderate-sized moons Pan (radius $R \sim 14$ km) and Daphnis ($R \sim 4$ km) orbit within the outer part of the A ring and clear sharp-edged gaps (the Encke and Keeler gaps, respectively). Smaller moons, but ones that are still substantially larger than the largest continuum ring particles, create local disturbances that have been dubbed "propellers," because of their shape (*12*).

Propellers are similar in principle to a circumferential gap like the Encke or Keeler gap, except that the embedded moon's influence is overcome some distance downstream as the gap is filled in by the viscosity of the disk. Swarms of relatively small propellers, likely due to embedded moons ~100 m in size, are present in the Propeller Belts of the mid-A ring (*13*–*15*). There are probably more than 10^6 propellers over the several-thousand-km annulus of the Propeller Belts (*15*). Solitary larger propellers, likely due to embedded moons ~1 km in size, orbit beyond the Encke gap (*16*). At least a half-dozen of the latter have had their orbits tracked for the duration of the Cassini mission, and their shifting orbital attributes have been analyzed theoretically (*12*).

Among the science objectives of the RGO and GF (presented in Fig. 1) were close-range flybys of Pan and Daphnis (*17*) and of three large propellers (namely, those nicknamed Blériot, Earhart, and Santos-Dumont), as well as high-resolution imaging of the Propeller Belts.

Daphnis and its vicinity

A false-color mosaic of Daphnis and its vicinity is shown in Fig. 1A. Daphnis itself is discussed in a companion paper (*17*); here we focus on the ring material perturbed by Daphnis. The image conveys the dichotomy in brightness and color that occurs at the Keeler gap (see below). Immediately to the lower left of Daphnis in Fig. 1A, within the gap, a strand of material is seen to follow a looping streamline. This material was likely pulled out of the sharp edge of the gap by a recent passage of Daphnis and is now spreading along its streamline as a result of Keplerian shear (*18*). Cassini was 13.5° above the ring plane when this image was taken, and the wavy edges of the Keeler gap are known to have a vertical component due to the orbital inclination of Daphnis (*19*), so the streamline seen here may be vertical and/or in-plane, which cannot be disentangled, owing to projection effects.

Downstream of Daphnis (which is trailing on the outer edge of the gap, but leading on the inner edge, due to Keplerian shear), the edge is scalloped with a characteristic wavelength equal to $3\pi\Delta a$, where Δa is the difference in semimajor axis between Daphnis and the gap edge (*12*). It was previously known that the wavy edges due to Daphnis die away only a few wave crests downstream, whereas those due to Pan in the edges of the Encke gap persist for the entire circumferential extent of the gap (*20*–*22*). Figure 1A shows structure in the third wave crest downstream of Daphnis. The trailing structure in the outer edge is seen on the left-hand side of Fig. 1A and

[1]Carl Sagan Center for the Study of Life in the Universe, SETI Institute, Mountain View, CA 94043, USA. [2]Department of Astronomy, Cornell University, Ithaca, NY 14853, USA. [3]NASA Ames Research Center, Moffett Field, CA 94035, USA. [4]NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. [5]Astronomy Unit, Queen Mary University of London, London E1 4NS, UK. [6]Department of Physics, University of Idaho, Moscow, ID 83844, USA. [7]Department of Physics, University of Central Florida, Orlando, FL 32816, USA. [8]College of Engineering, Cornell University, Ithaca, NY 14853, USA. [9]Planetary Science Institute, Tucson, AZ 85719, USA. [10]Department of Earth, Planetary, and Space Sciences, University of California at Los Angeles, Los Angeles, CA 90095, USA. [11]Institut de Physique du Globe de Paris, Centre National de la Recherche Scientifique (CNRS)–Unité Mixte de Recherche (UMR) 7154, Université Paris-Diderot, Université Sorbonne-Paris-Cité (USPC), Paris, France. [12]INAF-IAPS (Istituto Nazionale di AstroFisica-Istituto di Astrofisica e Planetologia Spaziali), Rome, Italy. [13]Laboratoire de Planetologie et Geodynamique, CNRS-UMR 6112, Université de Nantes, 44322 Nantes, France. [14]Physics Department, Lancaster University, Lancaster LA1 4YB, UK. [15]Space Science Institute, Boulder, CO 80301, USA.
*Corresponding author. Email: matt@seti.org

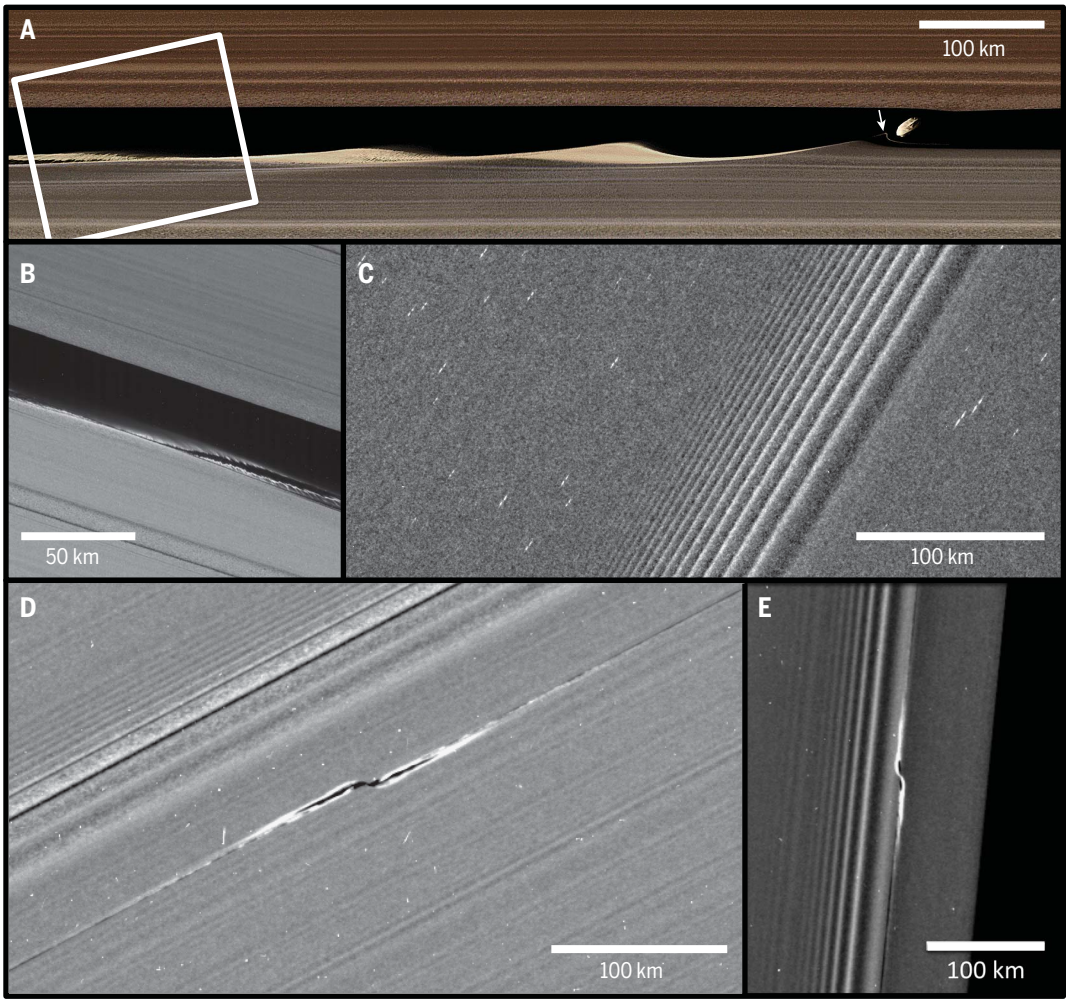


Fig. 1. Embedded moons cause structure in the main rings. (**A**) False-color image mosaic (*4*) showing Daphnis in the Keeler gap on the lit side of the rings, with three wave crests of the structure raised by Daphnis in the gap's outer edge. Each successive wave crest from right to left is older, having elapsed an additional orbital period (~13 hours) since its close approach to Daphnis. The direction to Saturn is up, and the orbital direction is to the right. A thin strand of material (indicated by the arrow) is present in the gap to the lower left of Daphnis, and there is intricate structure in the third wave crest downstream. The box on the left-hand side indicates the approximate footprint of the image in (B). (**B**) A further close-up of the third wave crest, on the unlit side of the rings. (**C**) Propellers in the Propeller Belts of the mid-A ring on the unlit side of the rings. The large multibanded structure across the center of the panel is the Prometheus 9:8 spiral density wave. (**D**) The propeller Blériot on the unlit side of the rings. (**E**) The propeller Earhart on the lit side of the rings.

magnified in Fig. 1B. Macroscopic clumps of ring particles, which are not visible in the first or second wave crest, dominate the third wave crest, as the effects of Daphnis diminish. Furthermore, a gap or rift appears between the clump-studded wave crest and the rest of the ring. We interpret this as a vertical feature: The ribbon of material may be rising above the ring plane, having been launched by the inclination of Daphnis, with the line of sight (due to Cassini's oblique viewing angle) passing through the vertical gap. As evidence for this, further downstream (in the left-hand portion of Fig. 1B), the same ribbon of material can be seen to pass behind the main portion of the ring, from Cassini's point of view. This hypothesis does not exclude some degree of simultaneous radial structure in the third wave crest. The difference in appearance between the third wave crest and the first two may be partly due to a shroud of smaller particles (whose structure perhaps is primarily radial) that smooths the appearance of the first two wave crests, which may have been absorbed into larger clumps in the third wave crest.

The propeller belts

The propellers observed (*13*) in close-range images taken during Saturn orbit insertion (SOI) were substantially smaller than those seen in the Propeller Belts during the remainder of the Cassini mission (*14*, *15*). Figure 1C shows propellers of a size similar to those seen during SOI. We identified and analyzed 41 propellers in this image (*4*). We find a substantial range in propeller sizes, with the largest propellers being more than four times as large as the smallest (table S1). We also find that the population density, average size, and size distribution may vary substantially with ring radius on finer scales than previously identified (fig. S1), particularly in the vicinity of a moderately strong density wave. This implies that some combination of the parent bodies, subsequent evolution, and/or visibility of propellers is highly radially stratified.

Giant propellers

Close-range flyby images of propellers Blériot and Earhart are shown in Fig. 1, D and E, respectively. Blériot was captured on the unlit side of the rings, where regions can be dark because of low optical depth (insufficient material to scatter sunlight toward the camera) or because optical depths are so high that the region is opaque. Earhart was captured on the lit side, where brightness usually increases monotonically as a function of optical depth. Additionally, close-range flybys captured images of propeller Santos-Dumont on both the lit and unlit sides of the rings, in images taken on both sides of ring-plane crossing during a single RGO pass (fig. S2).

Complex structure is seen in all these images. The dark band through the center of Blériot, which is due to the disturbed region being opaque, extends further downstream than the similar dark band in the unlit-side image of Santos-Dumont, likely indicating higher surface mass densities.

Scalloping is present on the inner edges of the perturbed regions, analogous to patterns seen on the edges of the Encke and Keeler gaps. Equating the wavelength of the scalloping with $3\pi\Delta a$, and taking the full width of the gap (i.e., twice Δa) to be four times the Hill radius r_H (*23*), where $r_H \equiv a(m/3M)^{1/3}$ for Saturn's mass M and the central moon's mass m and semimajor axis a, we calculate $\Delta a \simeq 2r_H$ to be 1.0 km for Santos-Dumont and 1.6 km for Blériot.

The Hill radii imply masses of 7×10^{14} g for the central moon of Santos-Dumont and 3×10^{15} g for that of Blériot. Most likely, the actual size of each central moon is similar to its Hill radius (*16*), so that its bulk density is equal to the critical density for this region of the A ring, which is ~0.4 g cm^{-3} (*24*), so that the above-calculated Δa approximates the triaxial moon's longest axis.

The central propeller moon for each of these close-range propellers should be two to three pixels across in these images. However, we do not detect them, likely because they are obscured by the disturbed ring material swirling around them.

Ring structure: Texture belts

The main rings appear smooth in most images, although a clumpy straw-like texture has been identified in the troughs of strong density waves (*25*, *26*). Other textures have been dimly perceptible in images throughout the Cassini mission but have not been described or analyzed in any detail.

The RGO and GF images show several classes of ring texture, in well-defined radial bands, which do not correlate with other ring properties. Examples of these textures are shown in Fig. 2.

There is growing evidence that some of the sharply bounded features in the main rings are not due to changes in surface mass density and thus must be due to variations in particle properties (*27*, *28*). Composition, particle size, and regolith character (*29*) are candidate causes, although it is not clear how ring particles are sorted according to these properties. We find that these ring textures are localized to sharply defined radial bands (Fig. 2). We speculate that this may be due to differences in how ring particles bounce off each other when they collide and thus might be correlated with regolith character (*30*).

Clumpy straw-like texture is seen not only in the troughs of strong density waves (Fig. 2a) but also in smaller waves such as those due to Prometheus (Fig. 2B) and at other types of locations. At 124,230 km from Saturn's center (in the inner A ring), a region characterized by fine-scale radial structure that is not well understood, a 100-km radial band exhibits clumpy structure, whereas the regions around it do not (Fig. 2C). The clumpy texture at this location is correlated with a relatively bright band, but other relative brightness maxima nearby do not exhibit clumpy texture. The clumpy texture at this location is not associated with any spiral density wave, although such waves do occur at adjacent locations.

Many locations in the B ring (e.g., Fig. 2D) exhibit a more elongated texture. This streaky or feathery texture is often correlated with local minima in optical depths (although all optical depths are quite high in this region), which appear dark in lit-side images such as Fig. 2D. The landmark triple-humped belt in the outer Cassini Division also exhibits feathery texture (fig. S3). Here the dark (gap-like) streaks are more discrete and may be related to features termed "ghosts" identified in UVIS occultations (*31*).

Another streaky texture appears in all 10 of the narrow and sharp-edged increases in brightness and optical depth in the C ring known as "plateaux." For example, Plateau P1 (Fig. 2E) exhibits streaky texture, whereas the nonplateau C-ring material around it exhibits either clumpy texture or no texture at all. Plateau P5 also exhibits streaky texture (fig. S4). Their surface mass densities do not appear to be markedly different from those of the background C ring (*32*), so their defining characteristic must be related to particle properties (*33*).

An occultation of the star κ Orionis by the rings was observed using UVIS (*4*). Gaps ≳20 m across were seen in the plateaux, with an average separation of 0.5 to 1 km. We modeled the plateaux with a simplified bimodal transparency model using 19 UVIS stellar occultations (*4*). This model yields gap widths of 10 to 100 m and vertical thicknesses of the plateaux of 10 to 40 m. These average gap widths may represent the narrowest in a range of gap sizes, observable only with the high spatial resolution of the stellar occultations, with the dark streaks seen in the images being gaps on the wider end of the distribution.

Ring structure: F-ring objects

Images of the F-ring core obtained during the RGO and GF (Fig. 3) confirm the existence of a population of smaller objects by the detection of their associated mini-jets (*34*). Although the core has an irregular, jagged shape [figure 4 in (*35*)]), the population of mini-jets visible in Fig. 3, D and E, are all in phase, implying that they formed from a set of objects that collided with the core almost simultaneously. This observation suggests that rather than representing a random population of small, colliding objects, the mini-jets in Fig. 3, D and E, had a common origin, probably as the result of a breakup of a single object. This supports previous inferences (*36*, *37*) that impacts onto the rings may sometimes be due to Saturn-orbiting streams of material rather than to lone impacting objects. Additional evidence for this is shown in the inset to Fig. 3B, where a collection of objects extending over

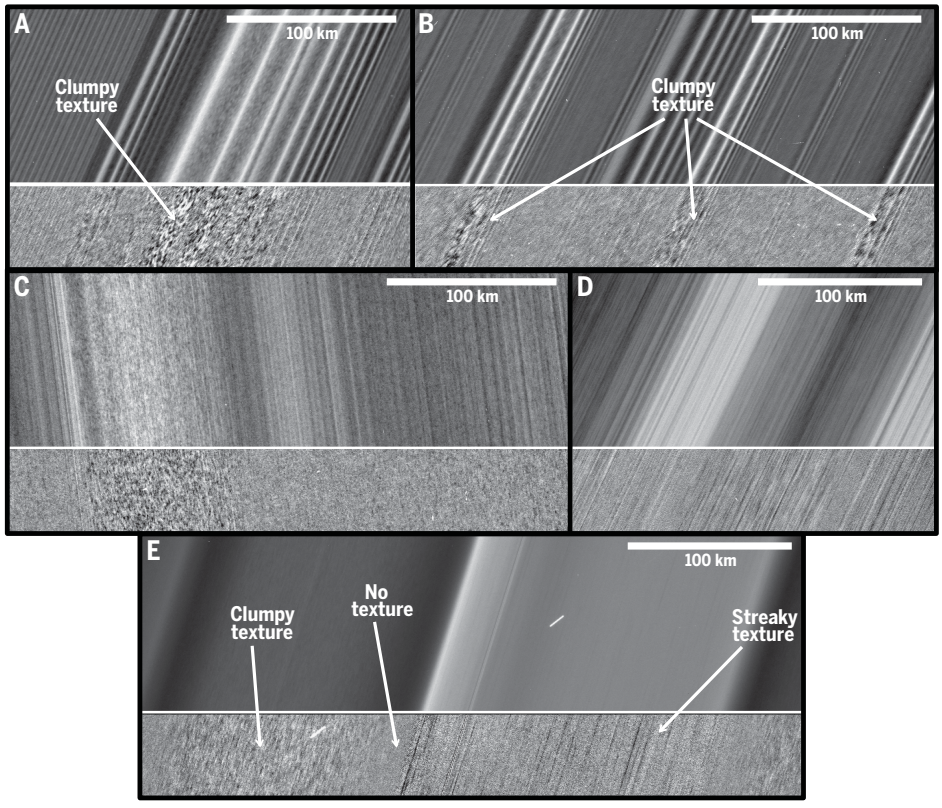


Fig. 2. Belts of textures in the main rings. A strip along the bottom of each panel has been filtered by subtracting the average radial profile of the image, so that local structures and textures are more visible. All images in this figure show the lit side of the rings (*4*). (**A**) Straw-like clumps in the troughs of the strong Janus 6:5 density wave in the outer A ring. (**B**) Straw in the troughs of the Prometheus 26:25, 27:26, and 28:27 density waves (left to right) in the outer A ring. (**C**) Straw-like texture in one radial band of the inner A ring but not in surrounding regions. (**D**) Feathery texture in some radial bands of the outer B ring (especially those with lower brightness in the main part of the image) but not in surrounding regions. (**E**) Plateau P1 and its environs in the C ring, with three different textures in close proximity to each other.

~0.05° (~100 km) appears to have recently emerged from the core. We interpret this feature as a more compact version of the multiple objects seen to cause the mini-jets in Fig. 3, D and E. The calculated radial velocities (*4*) of ~1.4 m s^{-1} are consistent with those of previously observed mini-jets (*34*).

We observe a narrow component in the F ring, 1 to 2 km in radial extent but perhaps unresolved, located ~10 km radially inward of the center of the bright core and visible in ~50% of Fig. 3A (i.e., extending over ~2°). Although evidence for such a component has been seen before [figure 4h in (*35*)]), we infer from Fig. 3D that the original impact events that produced the mini-jet features in this part of the F ring originated when the source objects collided with the narrow component, implying the existence of substantial mass at this location. It has been suggested (*38*) that the true core of the F ring consists of discontinuous ring arcs, <1 km in radial width, each extending over ~2° in longitude, wherein most of the F ring's mass is contained. The observations in Fig. 3 do not constrain the mass of the F ring, but they do provide supporting evidence for this interpretation and imply that the true core of the F ring may be radially offset from the brighter components that are more commonly seen.

Ring composition, particle size, and grain size

The reflectance spectrum of Saturn's main rings in the ultraviolet (UV), visible, and near-infrared (IR) regions is dominated by fine-grained crystalline water ice with relatively small amounts of non-icy material (*39*, *40*). The latter seems to consist of two distinct components. The first is responsible for the strong absorption at UV and blue wavelengths that gives the rings their pale tan or reddish color, and is now generally thought to be either organic in nature (*41*–*43*) or nanophase particles of metallic iron or iron oxides (*44*), whereas the second component is a spectrally neutral absorber generally assumed to be silicates and/or carbon (*45*). Together, this icy mixture makes up the regolith coating the surfaces of the individual ring particles, whose sizes range from a few millimeters up to ~5 m (*39*, *46*–*50*).

The UV absorber appears to be present as inclusions within the individual grains of water ice (sometimes referred to as an intramix), and its abundance increases monotonically with proximity to Saturn (*51*–*53*). Its nature has been debated (*54*), although it is suspected to be a primordial component of the rings. Recent observations and modeling (*43*) favor an organic material, perhaps tholins or mixtures of polycyclic aromatic hydrocarbons (PAHs). Large organic molecules have been detected falling into Saturn's upper atmosphere (*55*), which may have come from the D ring. The neutral absorber, by contrast, is inferred to be intimately mixed with the icy grains (i.e., as separate particles in close proximity, like a mixture of salt and pepper) and is most abundant in the optically thin C ring and Cassini Division (*52*). It may be produced by the exposure of the originally pure icy rings to bombardment by interplanetary debris (*45*).

Previous observations by Cassini have shown small-scale spatial variations in the main rings' visible and near-IR spectrum (*51*, *56*). VIMS spectra have revealed that regolith grains are larger in the A and B rings; grain size and/or composition varies locally in the vicinity of strong density waves in the A ring and at several locations in the B and C rings (*52*, *53*, *57*). UVIS optical depth variance measurements have yielded similar conclusions about small-scale variations in the ring particle-size distribution (*33*).

Color and spectral variations

During the final orbits of the Cassini mission, high-resolution multicolor and spectral scans were taken across the entire ring system, covering both the lit and unlit sides of the rings (Figs. 4 and 5).

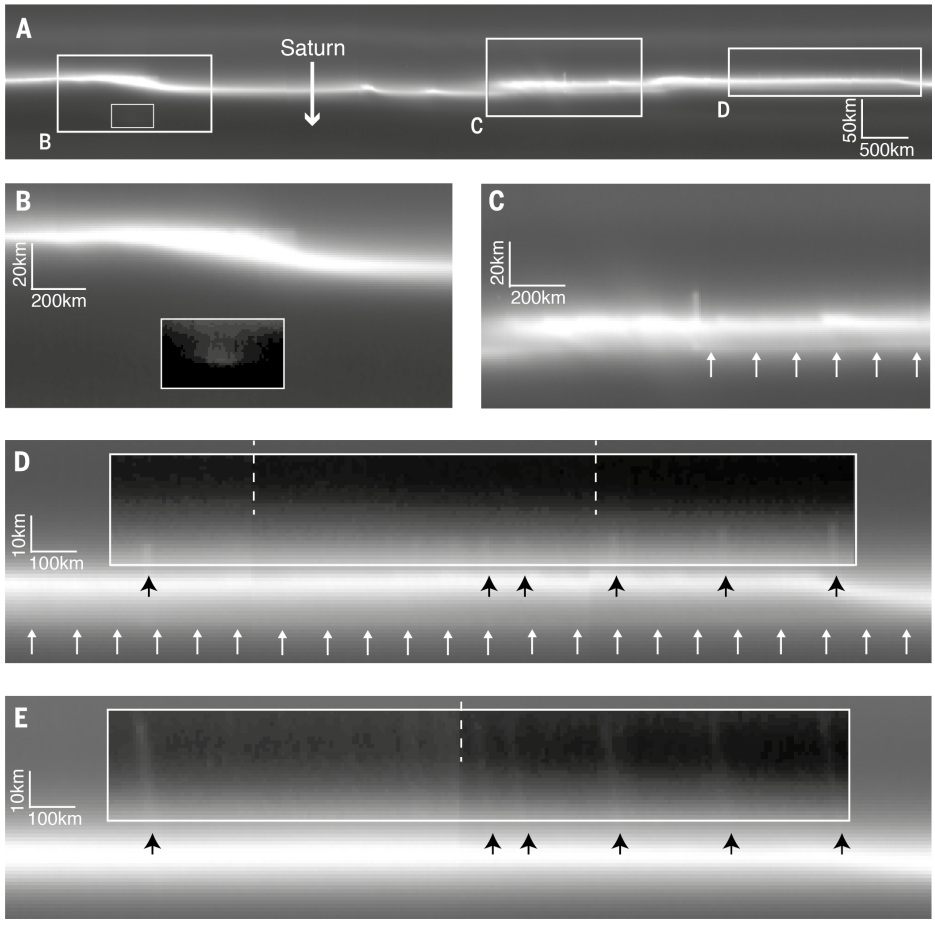


Fig. 3. Mini-jets and other structures in the F ring. A high-resolution mosaic (*4*) of the F-ring core region showing evidence for a population of small objects, their evolution, and a possible narrow component. (**A**) Context mosaic. White boxes indicate the regions shown in the other panels. (**B**) Enlarged portion of (A), showing a linear feature near the core with a contrast-enhanced inset showing a faint, linear feature below it. (**C**) Enlarged portion of (A), showing a mini-jet emerging ~10 km from the core with evidence of a narrow ring component parallel to and below the core. There is also evidence of a fan-shaped structure (*4*) to the left, characteristic of an embedded object on an eccentric orbit. (**D**) Enlarged portion of (A), with a region immediately above the core (indicated by the white rectangle) selectively contrast enhanced to bring out faint features. Several mini-jets (black arrows) emerge from the core, leaving trails that extend between ~7 km (left) and ~14 km (right) from the core. There is also evidence that the mini-jets originate in a narrow ring component parallel to the core but ~10 km below it (white arrows). (**E**) Same as (D), except portion of a different image mosaic of the same region, taken 85 min later. Although the resolution is poorer, the evolution of the mini-jets is evident, having doubled in length and undergone Keplerian shear. The dashed lines in (D) and (E) indicate longitudes where adjacent images overlap, leading to the production of slight offsets in the mosaics. The white arrows in (C) and (D) indicate the location of the narrow component.

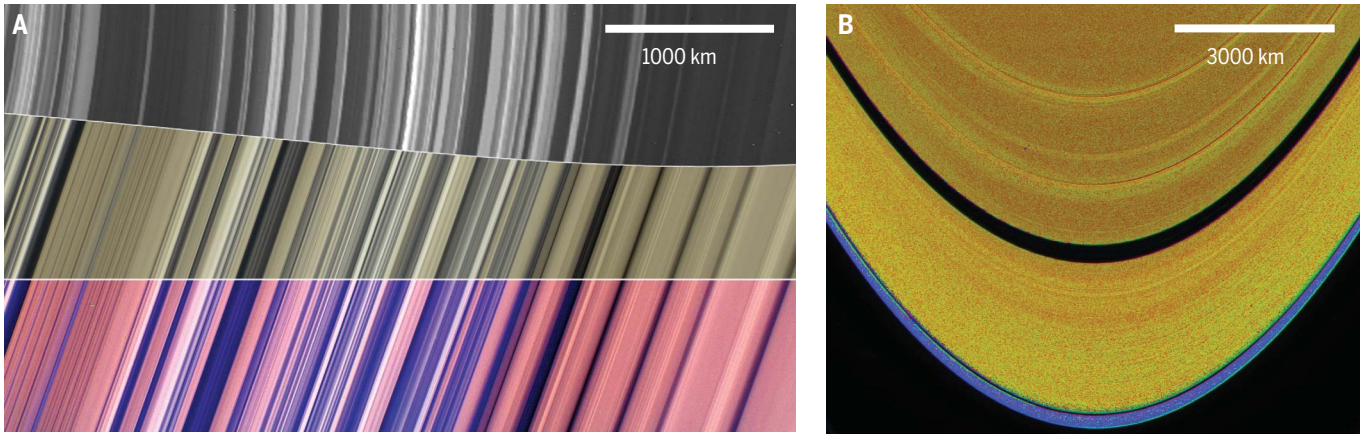


Fig. 4. Color images of the B and A rings. (**A**) The inner-central B ring, covering the region from 98,600 to 105,500 km from Saturn's center at about 3 km px^{-1}. The top panel is a context image in black and white, while the middle and bottom panels are, respectively, true-color and enhanced-color versions. (**B**) The mid- and outer A ring (*4*). Figure S10 shows a VIMS image of the same region.

Color imaging with ISS was conducted at a resolution of ~3 km px^{-1}, although this is a lower resolution than for the black-and-white images (e.g., Figs. 1 and 2). Figure 4A displays a composite of true and enhanced color images (bottom two panels), with context given by a single unlit-face image (upper panel) in which darker bands are more optically thick. The pale tan color seen in the middle panel is generally not perceptible in a telescope, especially because Saturn is a yellowish tan itself (see below). Figure 4B is an enhanced color composite from early in the Cassini mission, illustrating the different color outward of the Keeler gap.

The observing mode for VIMS (*4*) resulted in a single, continuous image of the A, B, and C rings (referred to as a noodle) 64 pixels wide by ~1000 pixels long. The spatial resolution ranges from ~30 to 60 km px^{-1}, the gradient in resolution from one end of a noodle to the other being due to the changing spacecraft range. In the VIMS false-color images shown in Fig. 5, the red channel corresponds to the center of a strong water-ice absorption band, so reddish colors in the images signify less ice absorption, while the blue channel corresponds to a much weaker water-ice band. The green channel is a continuum wavelength. The red color is most evident in the C ring and Cassini Division but is also visible in the extreme outer part of the A ring. The inner B ring also has slightly weaker ice bands than do the outer B or A rings (*4*).

Figure 6 plots a set of standard spectral parameters for the main rings, derived from the VIMS lit-side scan on Rev 287 (*58*) and rebinned to a sampling interval of 20 km. Figure 6A shows two representative brightness profiles of the rings at continuum wavelengths. Figure 6C shows an optical depth profile of the rings for context, derived from a VIMS stellar occultation. In Fig. 6B are plotted water-ice band depths at 1.5 and 2.0 μm, calculated from the average spectrum for each radial bin (*4*). Also shown are spectral slopes in the visible portion of the spectrum, between 0.35 and 0.55 μm and between 0.55 and 0.85 μm (these are sometimes called the blue slope and the red slope, respectively, after the spectral regions where they are centered, rather than their shapes) Also plotted is the height of the 3.6-μm peak, which is defined by strong absorptions at 3 and ~4.5 μm, using the continuum level at 1.8 μm as a reference (*4*, *51*, *52*, *57*).

As seen in previous work (*51*, *52*), ice-band depths are greatest in the outer half of the B ring and the outer two-thirds of the A ring and lowest in the inner C ring and the Cassini Division. Band depths decrease smoothly from the middle A ring into the outer Cassini Division, and also from the innermost part of the B ring into the outer C ring, and then decrease further across the C ring. We confirm that regional transitions in band depths are gradual (*51*, *52*), except that we identify a more abrupt transition at the Keeler gap (see below). More detailed views of portions of Fig. 6 are shown in figs. S9 to S11.

Consistent with previous observations (*51*, *52*, *57*), we find that the fine-scale variations in the 0.35- to 0.55-μm slope and the IR water-ice band depths in the A and B rings are well correlated. These curves exhibit very similar shapes down to the spatial resolution limit of the scans, and these parameters are more closely correlated with each other than either is with the ring brightness I/F (*59*). The 0.35- to 0.55-μm slope and ice bands show different behaviors within the lower optical depth C ring and Cassini Division, as well as across the transitions between those rings and the innermost B and A rings. Detailed spectrophotometric modeling of VIMS observations of the A, B, and C rings (*40*, *52*) reveals that these varying trends can be attributed to a combination of variable particle regolith properties and compositions across the rings. For example, most of the fine-scale variations in the 0.35- to 0.55-μm slope and ice bands across the A and B rings can be attributed to shifts in grain sizes, whereas the deviations between these parameters in the lower optical depth C ring and Cassini Division are due to higher concentrations of darkening materials (neutral absorber) in those regions (*40*). After accounting for these variations, the differences between the overall trends across the inner B ring between the 0.35- to 0.55-μm slope and the ice-band strengths could be due to an increase in the fractional abundances of the UV absorber closer to the planet (*52*).

Although the 0.55- to 0.85-μm slope shows some of the same features in the A and B rings as the other spectral parameters, its overall behavior is quite different and so traces a different aspect of the rings' texture or composition. In particular, the 0.55- to 0.85-μm slope is elevated in the middle part of the C ring, which is consistent with evidence for enhanced silicate or carbon-rich material at that location inferred from radio wavelengths (*60*).

A ring

In the A ring, halos of reduced water-ice band depth around the strong density waves are seen in the VIMS observations. In fig. S9, we plot the band depths in the A ring at a larger scale, together with profiles of the 0.35- to 0.55-μm and 0.55- 0.85-μm slopes (although the latter is fairly uniform in this region). Apart from the general inward decrease in band depths noted above, the most prominent features in the A ring profile are the peaks and dips seen at each of the strong density waves, especially those due to Janus. Centered on each of the Janus 4:3, 5:4, and 6:5 resonances is a peak in all ice-band depths—and in the 0.35- to 0.55-μm spectral slope—symmetrically flanked by a region of reduced band depths. These features have been dubbed "halos" (*51*), from VIMS SOI data. The overall radial width of each halo is ~1000 km, which is much larger than the extent of the density waves associated with these resonances, and the halos are centered on the waves. The halo associated with the Mimas 5:3 resonance is less evident, even though it is one of the strongest density waves in the A ring. (The nearby Mimas 5:3 bending wave at 131,800 km shows neither peak nor halo in the band depth profiles.) Although the reason for the peaks and halos remains unclear,

they may reflect an increase in the interparticle collision rate within the density wave itself (exposing fresh, coarser-grained ice from deeper layers in the ring particles), accompanied by the ejection of debris that falls back onto the surrounding regions (which may reduce their mean grain size by spreading a cloud of small icy particles). UVIS optical depth variance results support this conclusion (*33*).

There are several ripples in the band depths and 0.35- to 0.55-μm slope in the innermost part of the A ring, inwards of 124,500 km. There are two moderately strong density waves in this region, due to the Pandora 5:4 and Prometheus 6:5 resonances, and the optical depth profile shows several sharp peaks in this neighborhood, but there is no correlation between the band depths and either the resonances or optical depth structure (*61*).

The region around and outward of the Keeler gap is less red, shows less backscattering, and has weaker ice bands than the rest of the A ring, as noted first in Voyager data (*39*, *62*), and pronounced color variations are seen in early Cassini images such as Fig. 4B, where the color transition is very sharp and coincides with the location of the gap. In our observations, the spectral variations appear to start near 136,200 km and extend across the gap to the ring's outer edge at 136,770 km. They include decreases in both ice-band depths and the 0.35- to 0.55-μm slope (figs. 6 and S9), accompanied by an increase in the continuum brightness of the rings beyond the Keeler gap. This region is also known to have a lower surface mass density (*63*) and different particle-size properties (*49*, *50*, *64*) from the rest of the A ring.

It remains unknown whether the spectral difference of the trans-Keeler region implies a different composition and origin of the underlying material, or whether some other factor is responsible. Figures 6B and S9 suggest weaker water-ice bands in the trans-Keeler region. This may imply less water ice and less of the accompanying organic reddening material. However, this hypothesis is difficult to reconcile with the relatively high reflectivity (*I*/*F*) in the same region, as seen in Fig. 6A and fig. S9. A finer regolith grain size outside the Keeler gap could explain all three observations: flatter (less-red) visual wavelength spectra, higher *I*/*F*, and weaker ice bands, whether or not the material has an intrinsically different composition.

B ring

The B ring, the most opaque and massive of Saturn's main rings (*65*), is conventionally divided into four or five subregions, based on varying structure and mean optical depth (*66*). The innermost or B1 region, between radii of 92,000 and 99,000 km, has a normal optical depth τ_n of 1 to 2 and is characterized by smooth, quasi-sinusoidal variations in τ_n and brightness with typical length scales of ~100 km (*67*). B1 is uniform in its spectral parameters (*51*) and in its optical colors (*68*, *69*). Our VIMS data in Fig. 6 show very little variation in this region, except for a modest inward increase in both the 0.35 - to 0.55-μm and 0.55- to 0.85-μm slopes and a slight drop in the ice-band depths toward the C ring ramp over the innermost 500 km.

The B2 region is a zone of transition between the relatively low–optical depth B1 region and the opaque B3 region (outward of 104,000 km) where $\tau_n > 3$ almost everywhere and frequently exceeds 5. The B2 region is characterized by a series of abrupt local transitions in τ_n, between a minimum of ~2 and a maximum of 5 or greater, at irregular intervals of 40 to 200 km. The VIMS data for this region (shown at a larger scale in fig. S10) show a positive correlation between the IR ice-band depths, the 0.35- to 0.55-μm slope, and τ_n, with every peak (or dip) in band depth being associated with a maximum (or minimum) in τ_n. At somewhat lower spatial resolution, a similar relationship between optical depth and ice-band depths was previously demonstrated for the whole B ring (*52*); our data show that this correlation also holds down to the 50- to 100-km scale in the B2 region. The correlation of optical depth with *I*/*F* is much weaker, with peaks in τ_n being associated with both maxima and minima in *I*/*F* in this region. This suggests that many of the *I*/*F* variations in the central B ring are due to variations in albedo and/or phase function rather than particle density—a conclusion previously drawn from Voyager imaging data (*68*). Figure 4A shows a high-resolution ISS color image, also of the B2 region. There are sharp boundaries between the discrete ~100-km-wide bands in this region, which Cassini radio occultation data have shown to be even sharper than the ISS resolution (*66*). The narrow ringlets in the middle of this figure are about 40 km wide, whereas the broader bands near the right edge are 300 to 500 km across. It remains unclear what causes the variable brightness of these ringlets and bands; the inherent reflectivity of the ring particle material, shadowing on their surfaces, their absolute abundance, and/or their packing density may all play a role. The bottom panel of Fig. 4A is a color-enhanced version of the middle panel, in which blue colors indicate a "less red" or flatter spectrum at visible wavelengths than the redder colors, which mean a steeper-than-average spectrum. Observations by Voyager showed these color variations at lower resolution (*68*, *69*); we find that such well-defined color contrasts are sharply defined even down to the 3-km radial scale.

Across the very opaque B3 and B4 regions, between radii of 99,000 and 115,000 km (*66*), we find only small variations in the VIMS ice band depths or the 0.35- to 0.55-μm slope. A series of six low-amplitude oscillations in the band depths between 112,500 and 114,200 km are positively correlated with similar variations in both τ_n and *I*/*F* (*52*). Only within ~1300 km of the outer edge of the B ring is there a small but distinct decrease in all four of these spectral indicators, followed by an abrupt upturn in the final 500 km.

An exception to the generally flat band depths in this central part of the B ring is a broad but shallow dip in the ice band depths and 0.35- to 0.55-μm slope centered at 109,100 km and an even shallower dip centered at 107,400 km. These two features coincide with prominent peaks in *I*/*F* but with no obvious features in the optical depth profile. They may represent halos associated with the otherwise-invisible density waves that are expected to be driven by the Prometheus and Pandora 3:2 resonances (*52*), as seen in the A ring (fig. S9). Both dips are located 500 to 600 km exterior to the actual resonance locations (*52*). Two similar shallow dips in the band depths are seen in the outermost part of the B ring, at 115,700 and 116,500 km, and bear a

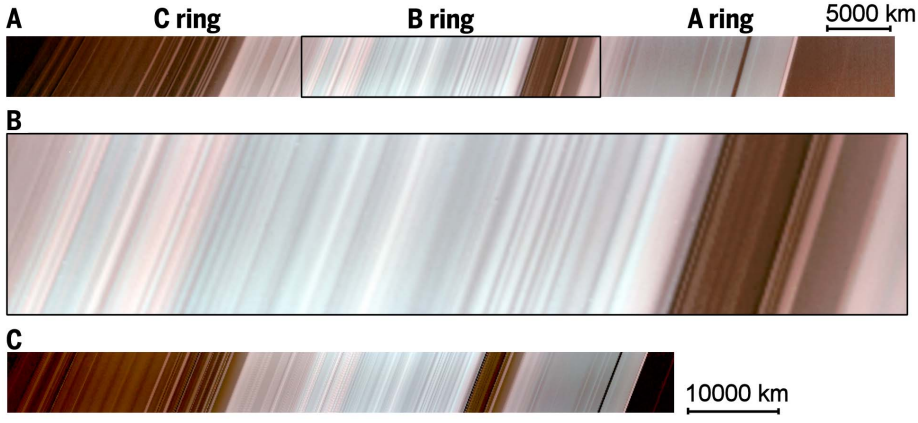


Fig. 5. Near-IR spectral scans. VIMS radial scans across the lit side of the main rings, displayed as false-color images. (**A**) The scan obtained on Rev 262, on 21 February 2017. (**B**) An enlarged view of the outer two-thirds of the B ring, and Cassini Division. (**C**) The scan obtained on Rev 287, on 7 August 2017. The scale bars indicate the average radial scale for each scan, although in both cases, the true scale varies across the scan because of the varying projected radial velocity of the spacecraft. On Rev 262, the velocity was a maximum in the B ring and the radial scale is compressed there by ~15%. On Rev 287, the velocity decreased from right to left by ~30% and the scale in the C ring is correspondingly stretched out.

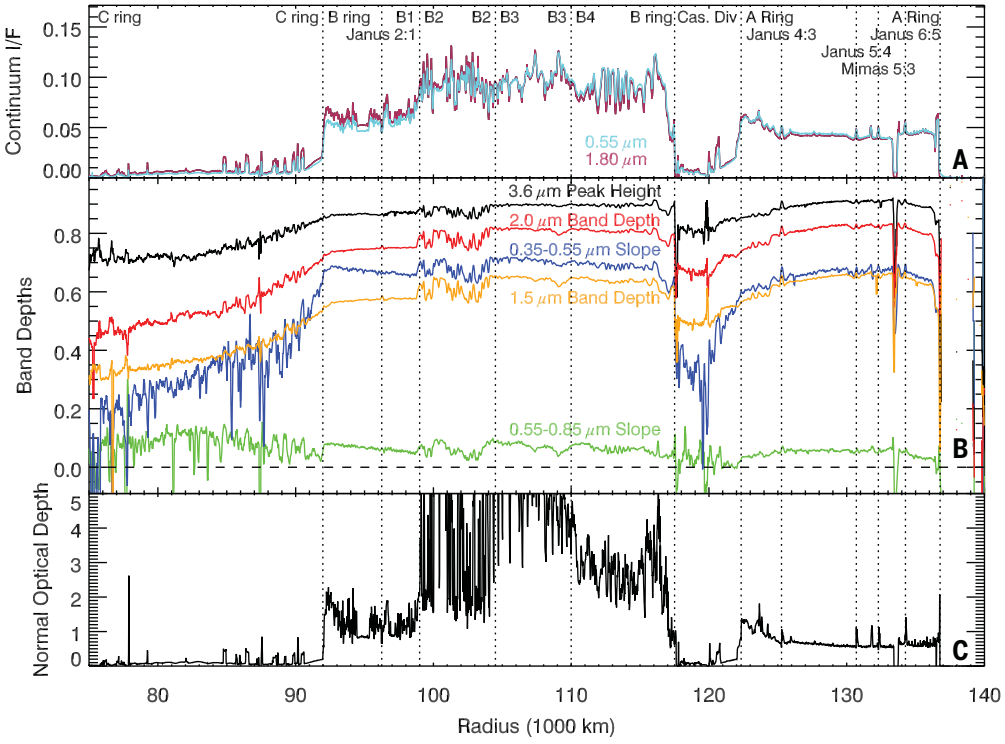


Fig. 6. Radial profiles of spectral parameters. Data shown are from the VIMS lit-side scan on Rev 287. Vertical dotted lines indicate (as labeled) the locations of density waves or of boundaries between ring regions, including the A, B, and C rings and the B-ring subregions B1 through B4. (**A**) Reflectivity of the rings at continuum wavelengths of 0.55 and 1.8 μm, with the locations of major ring boundaries and the strongest density waves identified by vertical dotted lines. (**B**) Fractional depths of the water-ice bands at 1.55, 2.0, and 3.6 μm in orange, red, and black, respectively, as well as the 0.35- to 0.55-μm and 0.55- to 0.85-μm slopes in blue and green (see text for definitions). (**C**) Optical depth profile of the rings obtained from a VIMS occultation of the star γ Crucis on Rev 082 (*52*, *70*). Versions of this figure that zoom on the A, mid-B, and outer C rings, respectively, are shown in figs. S9, S10, and S11.

similar spatial relationship to the Enceladus 3:1 and Janus 3:2 resonances. This region is quite complex, with several other potentially interfering resonances due to Mimas and Prometheus (*70*).

C ring

Large-scale variations across the C ring visible in Fig. 6 include a steady inward decrease in IR-band depths and the 0.35- to 0.55-μm slope parameter. On a finer scale, the most obvious structural features in the C ring are the plateaux. A typical plateau is ~150 km wide, with fairly abrupt edges and a peak optical depth of ~0.4. They were detected in Voyager images (*1*) but are of unknown origin. Measurements of surface mass densities in the C ring (*32*) indicate that the plateau mass densities are similar to that of the adjacent background C ring ($\tau_n \simeq 0.1$), suggesting that their particle size distribution may be quite different (fewer particles in the meter-size range). VIMS observations at SOI yielded a marginal detection of variation in ice-band depths between the plateaux and background regions (*51*). There is no discernible signature of the plateaux in the lit-side profiles of ice band depths, but there is a weak enhancement in the 0.35- to 0.55-μm slope parameter, which may be due to a local increase in the abundance of the UV absorber (*52*).

In the outer C ring (fig. S11), the VIMS spectral data show that the 0.35- to 0.55-μm slope again tracks the depth of the ice band at 1.55 μm, although the profiles for the C ring are noisier at visible wavelengths. These data show very little difference, if any, in spectral characteristics between the plateaux and the background. In several cases, a very small increase in band depths within the plateaux may be detectable, at the level of 0.01 or 0.02, or 1.5 to 3%. These signatures are more apparent in the 0.35- to 0.55-μm slope than in any of the IR ice bands.

Correlation with optical depth

Color variations and spectral variations often imply compositional variations. However, this is not necessarily correct in Saturn's rings. Figure 4A shows that ring color correlates with optical depth down to the smallest scales discernible: Optically thicker bands (darker in the top panel) are redder. A similar correlation between optical depth, ice-band depths, and spectral slope is visible in fig. S9 and discussed above. The intrinsic composition is unlikely to vary in such a correlated way and on such small radial scales. Meteoroid bombardment spreads ejecta around the rings on 100-km radial scales, smoothing out any such variations (*71*). However, variable optical depth might lead to other local changes in ring particle and/or ring layer properties. Collisional dynamics in regions of varying optical depth can lead to varying volume or packing density of particles, which might affect the degree of multiple interparticle scattering. Multiple scattering is known to enhance spectral contrast and the ring material is red, so more multiple scattering can make a region redder and also increase the relative depths of absorption bands (*72*). The local packing density can also lead to nonclassical scattering effects (e.g., shadowing), which can affect the ring brightness as a function of particle albedo, also affecting spectral contrast (*73*). The frequency and/or speed of collisions may also vary between regions of different optical depth, and this constant jostling can affect the roughness, porosity, or grain size in the surfaces of the local particles (*30*, *52*, *74*, *75*). Shadows on rough particle surfaces can change the particle phase function (*76*) in such a way as to exaggerate spectral contrasts at moderate to high phase angles. Monte Carlo models (*4*) of scattering in closely packed rings verified against similar models (*40*, *73*, *77*) suggest that the simplest explanation of the optical depth–dependent reddening is increased multiple scattering in the optically thicker regions by the already reddish ring particles, given particle phase functions and surface reflectances like those of the ring particles (*43*), in the geometry of the observation (the ISS image in Fig. 4A was taken at a relatively high phase angle, defined as Sun-target-observer angle, of $\alpha - 110°$). At shorter wavelengths, particle albedos are low enough that multiple scattering is minimal, leading to the variable reddening. Therefore, the fine-scale color variations are consistent with a generally invariant composition, at least on 10- to 100-km length scales.

Ring temperature

CIRS measured the thermal emission from Saturn's main rings—the A, B, and C rings and the Cassini Division (*7*, *78*, *79*). During the RGOs, CIRS obtained radial scans at spatial resolutions ranging from 260 to 470 km for the A ring to 660 to 930 km for the C ring for both their lit (northern) and unlit (southern) sides. Temperature T of the ring material was derived from spectra measured with focal plane 1 (FP1) between 10 and 600 cm^{-1} (17 μm to 1 mm) at a spectral resolution of 15 cm^{-1}. Rings were viewed nearly face-on for both the lit and unlit sides at absolute spacecraft elevation angles B

(defined with respect to the ring plane) between 70° and 83°, and at phase angles of α ~ 60° for the lit side and α ~ 120° to 130° for the unlit side (table S2). The representative ring temperatures and associated scaling factors were derived from Planck functions fitted to FP1 spectral data (*80*).

The lit-side temperatures varied from 80 to 100 K, and the unlit-side temperatures varied from 70 to 90 K (Fig. 7). The radially averaged ring temperatures were [C, 93 K; B, 85 K; CD (Cassini Division), 88 K; A, 84 K] on the lit face and (C, 87 K; B, 73 K; CD, 80 K; A, 78 K) on the unlit face. Ring temperatures are driven by direct solar flux, solar flux reflected from Saturn, and Saturn thermal flux and depend on the transport of heat vertically through the rings via radiation, conduction, and particle transport. The flux due to Saturn-shine falls off approximately as the square of the distance from Saturn, and the solar flux varies with the sine of the solar elevation angle above the ring plane. The optically thin C ring and Cassini Division are warmer than the optically thick A and B rings because incident flux from the Sun and Saturn penetrates through the ring, resulting in more efficient heating (*81*). The magnitudes of the thermal radiation emitted from the lit and unlit sides of the C ring and Cassini Division are consistent with the unlit side being in thermal equilibrium with the radiation field. That is not true of the B and A rings, where the unlit sides radiate at a much higher temperature than would be expected if they were in equilibrium with the small amount of radiation that penetrates through the ring.

The heat transport within the C ring and Cassini Division is thought to be similar, although the C ring is always warmer than the Cassini Division because of its proximity to Saturn; similarly, the A- and B-ring temperatures are comparable, and at most epochs, the lit B ring is warmer than the lit A ring because of their relative distances from Saturn. This Saturn-shine effect is at a minimum for the current data because at solstice the solar flux is large compared to the Saturn-shine, so that the lit A-ring temperatures are comparable to those in the B ring. Details of the effect of Saturn-shine on the heating have been studied by using the radial dependence of temperature at Saturn Equinox when solar flux is nearly zero (*81*), where the effect of reduced mutual shadowing between particles in the optically thinner rings could be modeled directly, given details of the incident flux.

Within each ring, the temperature difference between the lit and unlit face correlates to some extent with optical depth, although the actual magnitudes of the temperatures differ between rings. The correlations are shown graphically in fig. S8 and discussed below with reference to Fig. 7. In the thick rings, this effect has been noted before and depends at least partially on efficiency of heat transport across the rings (*82*, *83*). Detailed models indicate contributions to these broad trends from a positive correlation between albedo and optical depth across the rings (*84*), of a nature consistent with preferential darkening of optically thinner rings due to meteoritic bombardment (*45*). Our observations show that, at finer scale, this correlation (of thermal gradient with optical depth) is not everywhere consistent, possibly because of local variations in ring structure and particle properties.

The temperature difference between the lit and unlit faces of the B ring correlates with optical depth, owing mostly to variation on the unlit face. This could be caused by mutual collisions between particles in high–optical depth regions hampering heat transport. Observational studies have shown that the thermal transmittance of the B ring correlates inversely with optical depth (*82*). Transport models of the B ring based on effective medium theory indicate that if heat transport by the vertical motion of particles is negligible, then the observed correlation of the ring thermal gradient with optical depth can be explained only if the increase in optical depth is due to a vertically thicker ring and not to a larger filling factor (*83*).

Figure 7 indicates that in the A ring, the correlation of temperature differential with optical depth is consistent from the inner A ring to the flat outer regions, but there is a reverse in this trend in several locations where unlit-side

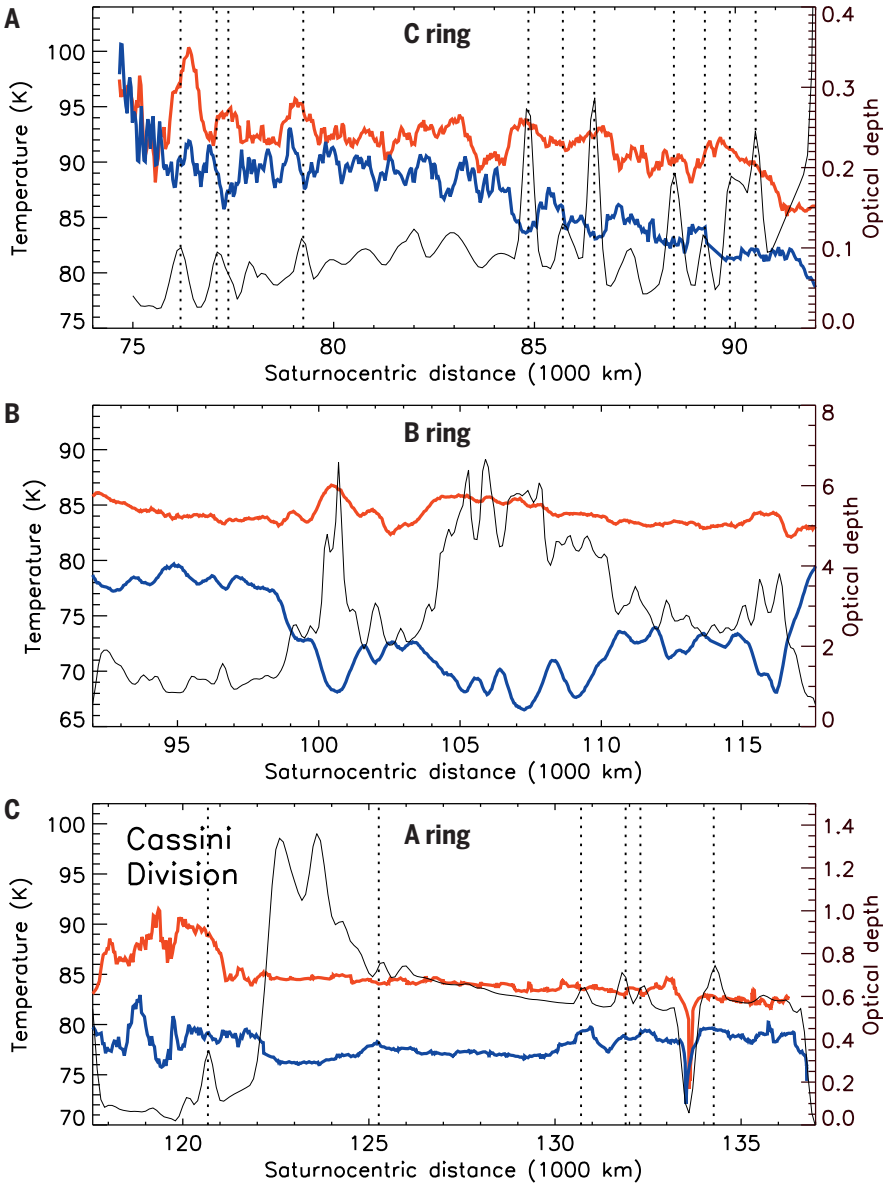


Fig. 7. Radial temperature profiles. Observations by CIRS of the lit (red) and unlit (blue) faces of (**A**) the C ring, (**B**) the B ring, and (**C**) the Cassini Division and the A ring. The black curve shows, for context, the optical depth measured by Cassini UVIS (*8*) and smoothed over a radial extent comparable to the typical CIRS footprint size. The vertical dotted lines show locations of prominent features: the plateaux in the C ring and the Cassini Division and the strongest density and bending waves in the A ring (table S3).

temperatures of the A ring have local maxima correlating with the halos of the strongest density waves (Janus 4:3, 5:4, and 6:5 and Mimas 5:3) and bending waves (Mimas 5:3). This decreases the temperature difference, even though the optical depths of these resonant locations are higher than those of the surrounding regions. This effect may be due to enhanced vertical heat transport or particle spin rates as a result of increased stirring of the ring particles in these regions. On the lit side of the rings, no temperature differences between the resonant locations and their surrounding regions are seen.

Correlations between temperature gradient and optical depth are not obvious in the C ring and Cassini Division, partially because of the lower signal-to-noise ratio in those regions. There appear to be dips in the unlit-side temperature within at least the largest plateaux. In contrast to the B and A rings, where the lit-side temperature varies little, there are features in the lit-side temperature of the C ring, but these do not correlate well with the optical depth. The net result is that there is either a weak correlation or none between optical depth and temperature difference in the optically thin regions and that other effects seem to be important in this regard. Temperature maxima near the largest plateaux could be associated with mutual heating between particles (*84*) or could be associated with radial variations in particle properties or ring structure, the presence of which is hinted at by the streaky texture seen in images of the C-ring plateaux.

Measurements by VIMS of the wavelength of the broad continuum peak around ~3.6 m can be used to infer the surface temperature of water ice. This method is applicable to ring particles, for which water ice is the dominant endmember (*53*). Spatial and temporal variations of the temperature as derived by this method are seen in the RGO and GF lit-side data. For most of the A ring, this peak was found at about 3.58 μm, corresponding to T ~ 88 K, with lower temperatures in the outermost parts of the ring. The same temperature prevails inward to the mid-B ring, but it increases to T ~ 107 K (peak at 3.60 μm) in the inner B ring. In the C ring and Cassini Division, the 3.6-μm reflectance peak is faint and distorted by Saturn-shine, making temperature retrieval uncertain. However, data collected before the GF (*53*) have shown systematically higher temperatures; the inner C ring is 30 K warmer than the B ring, and the Cassini Division is 20 K warmer than the outer A ring. These VIMS results mirror the large-scale variations in mean albedo and optical depth, in the sense that the particles are colder where these quantities are higher. Furthermore, with the exception of the Cassini Division, the ring temperature decreases with distance from Saturn (*53*). However, the temperatures inferred by VIMS are generally higher than those measured by CIRS. Different depths to which ring particles are penetrated by the radiation measured by VIMS (~1 mm) and CIRS (≳10 mm) may play a role, as may the larger-pixel footprint for CIRS, which could average warmer and cooler temperatures together.

Conclusions

We have reported imaging and spectral observations of Saturn's rings taken during the final year of the Cassini mission. The rings are sculpted by embedded masses, producing structure visible down to our resolution limit. Correlations of spectral properties and temperature with optical depth are tight at many locations, although exceptions are found in certain regions already known to be enigmatic. Some spectral variations may be due to local variations in optical depth, rather than being solely due to particle composition or regolith grain size. Sharply stratified variations in particle properties, possibly including regolith character and bulk porosity as well as the distribution of particle sizes, likely play a role in producing many of the structures described here, and in some cases they may supplant variations in surface mass density as the dominant effect.

REFERENCES AND NOTES

1. J. N. Cuzzi, G. Filacchione, E. A. Marouf, *Planetary Ring Systems: Properties, Structure, and Evolution*, M. S. Tiscareno, C. D. Murray, Eds. (Cambridge Univ. Press, 2018), pp. 51–92.
2. Optical depth τ quantifies the attenuation of a beam of light passing through the disk, measured in e-folding terms. That is, $\tau \equiv -\ln\mathcal{T}$ for the fractional transparency $\mathcal{T}$. The normal optical depth τ_n corrects for the effects of an inclined line of sight, thus approximating what the optical depth would be if the line of sight were normal to the ring plane (that is, if observed face-on to the rings).
3. M. M. Hedman, F. Postberg, D. P. Hamilton, S. Renner, H.-W. Hsu, *Planetary Ring Systems: Properties, Structure, and Evolution*, M. S. Tiscareno, C. D. Murray, Eds. (Cambridge Univ. Press, 2018), pp. 308–337.
4. Materials and methods are available as supplementary materials.
5. C. C. Porco *et al*., Cassini imaging science: Instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115**, 363–497 (2004). doi: 10.1007/s11214-004-1456-7
6. R. H. Brown *et al*., The Cassini Visual and Infrared Mapping Spectrometer (Vims) Investigation. *Space Sci. Rev.* **115**, 111–168 (2004). doi: 10.1007/s11214-004-1453-x
7. F. M. Flasar *et al*., Exploring The Saturn System In The Thermal Infrared: The Composite Infrared Spectrometer. *Space Sci. Rev.* **115**, 169–297 (2004). doi: 10.1007/s11214-004-1454-9
8. J. E. Colwell *et al*., Cassini UVIS stellar occultation observations of Saturn's rings. *Astron. J.* **140**, 1569–1578 (2010). doi: 10.1088/0004-6256/140/6/1569
9. P. Goldreich, S. Tremaine, The Dynamics of Planetary Rings. *Annu. Rev. Astron. Astrophys.* **20**, 249–283 (1982). doi: 10.1146/annurev.aa.20.090182.001341
10. F. H. Shu, *Planetary Rings*, R. Greenberg, A. Brahic, Eds. (Univ. of Arizona Press, Tucson, 1984), pp. 513–561.
11. M. S. Tiscareno, J. A. Burns, P. D. Nicholson, M. M. Hedman, C. C. Porco, Cassini imaging of Saturn's rings. *Icarus* **189**, 14–34 (2007). doi: 10.1016/j.icarus.2006.12.025
12. F. Spahn *et al*., *Planetary Ring Systems: Properties, Structure, and Evolution*, M. S. Tiscareno, C. D. Murray, Eds. (Cambridge Univ. Press, 2018), pp. 157–197.
13. M. S. Tiscareno *et al*., 100-metre-diameter moonlets in Saturn's A ring from observations of 'propeller' structures. *Nature* **440**, 648–650 (2006). doi: 10.1038/nature04581; pmid: 16572165
14. M. Sremcević *et al*., A belt of moonlets in Saturn's A ring. *Nature* **449**, 1019–1021 (2007). doi: 10.1038/nature06224; pmid: 17960236
15. M. S. Tiscareno, J. A. Burns, M. M. Hedman, C. C. Porco, The population of propellers in Saturn's A ring. *Astron. J.* **135**, 1083–1091 (2008). doi: 10.1088/0004-6256/135/3/1083
16. M. S. Tiscareno *et al*., Physical characteristics and non-Keplerian orbital motion of "propeller" moons embedded in Saturn's rings. *Astrophys. J.* **718**, L92–L96 (2010). doi: 10.1088/2041-8205/718/2/L92
17. B. J. Buratti *et al*., Close Cassini flybys of Saturn's ring moons Pan, Daphnis, Atlas, Pandora, and Epimetheus. *Science* **364**, eaat2349 (2019).
18. Keplerian shear is the difference in orbital velocity between adjacent streamlines of ring material. For angular velocity n and radial distance a from Saturn's center, it is straightforward to differentiate Kepler's third law and obtain $dn/da = (3/2)(n/a)$.
19. J. W. Weiss, C. C. Porco, M. S. Tiscareno, Ring edge waves and the masses of nearby satellites. *Astron. J.* **138**, 272–286 (2009). doi: 10.1088/0004-6256/138/1/272
20. M. S. Tiscareno *et al*., *AAS Division for Planetary Sciences Meeting Abstracts* **37**, 64.02 (2005).
21. P. A. Torrey, M. S. Tiscareno, J. A. Burns, C. C. Porco, *AAS Division on Dynamical Astronomy Meeting Abstracts* **39**, 15.19 (2008).
22. R. Tajeddine *et al*., Dynamical phenomena at the inner edge of the Keeler gap. *Icarus* **289**, 80–93 (2017). doi: 10.1016/j.icarus.2017.02.002
23. M. Seiß, F. Spahn, M. Sremčević, H. Salo, Structures induced by small moonlets in Saturn's rings: Implications for the Cassini Mission. *Geophys. Res. Lett.* **32**, L11205 (2005). doi: 10.1029/2005GL022506
24. M. S. Tiscareno, M. M. Hedman, J. A. Burns, J. C. Castillo-Rogez, Compositions and origins of outer planet systems: insights from the Roche critical density. *Astrophys. J.* **765**, L28 (2013). doi: 10.1088/2041-8205/765/2/L28
25. C. C. Porco *et al*., Cassini Imaging Science: Initial results on Saturn's rings and small satellites. *Science* **307**, 1226–1236 (2005). doi: 10.1126/science.1108056; pmid: 15731439
26. M. C. Lewis, G. R. Stewart, Expectations for Cassini observations of ring material with nearby moons. *Icarus* **178**, 124–143 (2005). doi: 10.1016/j.icarus.2005.04.009
27. M. S. Tiscareno, M. M. Hedman, J. A. Burns, J. W. Weiss, C. C. Porco, Probing the inner boundaries of Saturn's A ring with the Iapetus −1:0 nodal bending wave. *Icarus* **224**, 201–208 (2013). doi: 10.1016/j.icarus.2013.02.026
28. M. M. Hedman, P. D. Nicholson, Kronoseismology: using density waves in Saturn's C ring to probe the planet's interior. *Astron. J.* **146**, 12 (2013). doi: 10.1088/0004-6256/146/1/12
29. Regolith is any layer of loose surface material.
30. E. Déau *et al*., The opposition effect in Saturn's main rings as seen by Cassini ISS: 4. Correlations of the surge morphology with surface albedos and VIMS spectral properties. *Icarus* **305**, 324–349 (2018). doi: 10.1016/j.icarus.2017.12.025
31. K. Baillié, J. E. Colwell, L. W. Esposito, M. C. Lewis, Meter-sized moonlet population in Saturn's C ring and Cassini Dy. *Astron. J.* **145**, 171 (2013). doi: 10.1088/0004-6256/145/6/171
32. M. M. Hedman, P. D. Nicholson, More Kronoseismology with Saturn's rings. *Mon. Not. R. Astron. Soc.* **444**, 1369–1388 (2014). doi: 10.1093/mnras/stu1503
33. J. E. Colwell, L. W. Esposito, J. H. Cooney, Particle sizes in Saturn's rings from UVIS stellar occultations 1. Variations with ring region. *Icarus* **300**, 150–166 (2018). doi: 10.1016/j.icarus.2017.08.036
34. N. O. Attree, C. D. Murray, G. A. Williams, N. J. Cooper, A survey of low-velocity collisional features in Saturn's F ring. *Icarus* **227**, 56–66 (2014). doi: 10.1016/j.icarus.2013.09.008
35. C. D. Murray *et al*., The determination of the structure of Saturn's F ring by nearby moonlets. *Nature* **453**, 739–744 (2008). doi: 10.1038/nature06999; pmid: 18528389
36. M. S. Tiscareno *et al*., Observations of ejecta clouds produced by impacts onto Saturn's rings. *Science* **340**, 460–464 (2013). doi: 10.1126/science.1233524; pmid: 23620048
37. M. M. Hedman, J. A. Burns, M. R. Showalter, Corrugations and eccentric spirals in Saturn's D ring: New insights into what happened at Saturn in 1983. *Icarus* **248**, 137–161 (2015). doi: 10.1016/j.icarus.2014.10.021
38. J. N. Cuzzi *et al*., Saturn's F Ring core: Calm in the midst of chaos. *Icarus* **232**, 157–175 (2014). doi: 10.1016/j.icarus.2013.12.027
39. J. Cuzzi *et al*., *Saturn from Cassini-Huygens*, M. Dougherty, L. Esposito, S. M. Krimigis, Eds. (Springer, 2009), pp. 459–509.
40. M. Ciarniello *et al*., Cassini-VIMS observations of Saturn's main rings: II. A spectrophotometric study by means of Monte Carlo ray-tracing and Hapke's theory. *Icarus* **317**, 242–265 (2019). doi: 10.1016/j.icarus.2018.07.010
41. F. Poulet, J. N. Cuzzi, The Composition of Saturn's Rings. *Icarus* **160**, 350–358 (2002). doi: 10.1006/icar.2002.6967
42. F. Poulet, D. P. Cruikshank, J. N. Cuzzi, T. L. Roush, R. G. French, Compositions of Saturn's rings A, B, and C from high resolution near-infrared spectroscopic observations. *Astron. Astrophys.* **412**, 305–316 (2003). doi: 10.1051/0004-6361:20031123

43. J. N. Cuzzi *et al*., HST-STIS spectra and the redness of Saturn's rings. *Icarus* **309**, 363–388 (2018). doi: 10.1016/j.icarus.2018.02.025
44. R. N. Clark *et al*., The surface composition of Iapetus: Mapping results from Cassini VIMS. *Icarus* **218**, 831–860 (2012). doi: 10.1016/j.icarus.2012.01.008
45. J. N. Cuzzi, P. R. Estrada, Compositional Evolution of Saturn's Rings Due to Meteoroid Bombardment. *Icarus* **132**, 1–35 (1998). doi: 10.1006/icar.1997.5863
46. E. A. Marouf, G. L. Tyler, H. A. Zebker, R. A. Simpson, V. R. Eshleman, Particle size distributions in Saturn's rings from voyager 1 radio occultation. *Icarus* **54**, 189–211 (1983). doi: 10.1016/0019-1035(83)90192-6
47. H. A. Zebker, E. A. Marouf, G. L. Tyler, Saturn's rings: Particle size distributions for thin layer models. *Icarus* **64**, 531–548 (1985). doi: 10.1016/0019-1035(85)90074-0
48. R. G. French, P. D. Nicholson, Saturn's Rings II Particle Sizes Inferred from Stellar Occultation Data. *Icarus* **145**, 502–523 (2000). doi: 10.1006/icar.2000.6357
49. R. A. Harbison, P. D. Nicholson, M. M. Hedman, The smallest particles in Saturn's A and C Rings. *Icarus* **226**, 1225–1240 (2013). doi: 10.1016/j.icarus.2013.08.015
50. T. M. Becker, J. E. Colwell, L. W. Esposito, A. D. Bratcher, Characterizing the particle size distribution of Saturn's A ring with Cassini UVIS occultation data. *Icarus* **279**, 20–35 (2016). doi: 10.1016/j.icarus.2015.11.001
51. P. D. Nicholson *et al*., A close look at Saturn's rings with Cassini VIMS. *Icarus* **193**, 182–212 (2008). doi: 10.1016/j.icarus.2007.08.036
52. M. M. Hedman *et al*., Connections between spectra and structure in Saturn's main rings based on Cassini VIMS data. *Icarus* **223**, 105–130 (2013). doi: 10.1016/j.icarus.2012.10.014
53. G. Filacchione *et al*., Cassini–VIMS observations of Saturn's main rings: I. Spectral properties and temperature radial profiles variability with phase angle and elevation. *Icarus* **241**, 45–65 (2014). doi: 10.1016/j.icarus.2014.06.001
54. J. N. Cuzzi *et al*., An evolving view of Saturn's dynamic rings. *Science* **327**, 1470–1475 (2010). doi: 10.1126/science.1179118; pmid: 20299586
55. J. H. Waite Jr. *et al*., Chemical interactions between Saturn's atmosphere and its rings. *Science* **362**, eaat2382 (2018). doi: 10.1126/science.aat2382; pmid: 30287634
56. R. H. Brown *et al*., Observations in the Saturn system during approach and orbital insertion, with Cassini's visual and infrared mapping spectrometer (VIMS). *Astron. Astrophys.* **446**, 707–716 (2006). doi: 10.1051/0004-6361:20053054
57. G. Filacchione *et al*., Saturn's icy satellites and rings investigated by Cassini–VIMS: III – Radial compositional variability. *Icarus* **220**, 1064–1096 (2012). doi: 10.1016/j.icarus.2012.06.040
58. The Rev code is a three-digit label for Cassini's orbits around Saturn. For operational reasons, Saturn orbit insertion (SOI) was followed by Revs 00A, 00B, and 00C, and then by Revs 003 through 293.
59. *I*/*F* is a measure of the observed brightness *I* as normalized by a perfect Lambert surface's reflection of the incident solar flux πF.
60. Z. Zhang *et al*., Cassini microwave observations provide clues to the origin of Saturn's C ring. *Icarus* **281**, 297–321 (2017). doi: 10.1016/j.icarus.2016.07.020
61. The sharp peak at 123,570 km in both τ_n and *I*/*F* is probably associated with the Prometheus resonance.
62. L. Dones, J. N. Cuzzi, M. R. Showalter, Voyager Photometry of Saturn's A Ring. *Icarus* **105**, 184–215 (1993). doi: 10.1006/icar.1993.1118
63. M. S. Tiscareno, B. E. Harris, Mapping spiral waves and other radial features in Saturn's rings. *Icarus* **312**, 157–171 (2018). doi: 10.1016/j.icarus.2018.04.023
64. R. G. Jerousek, J. E. Colwell, L. W. Esposito, P. D. Nicholson, M. M. Hedman, Small particles and self-gravity wakes in Saturn's rings from UVIS and VIMS stellar occultations. *Icarus* **279**, 36–50 (2016). doi: 10.1016/j.icarus.2016.04.039
65. L. less *et al*., Measurement and implications of Saturn's gravity field and ring mass. *Science* **364**, eaat2965 (2019).
66. J. E. Colwell *et al*., *Saturn from Cassini-Huygens*, M. Dougherty, L. Esposito, S. M. Krimigis, Eds. (Springer, 2009), pp. 375–412.
67. L. J. Horn, J. N. Cuzzi, Characteristic Wavelengths of Irregular Structure in Saturn's B Ring. *Icarus* **119**, 285–310 (1996). doi: 10.1006/icar.1996.0021
68. P. R. Estrada, J. N. Cuzzi, Voyager Observations of the Color of Saturn's Rings. *Icarus* **122**, 251–272 (1996). doi: 10.1006/icar.1996.0124
69. P. R. Estrada, J. N. Cuzzi, M. R. Showalter, Voyager color photometry of Saturn's main rings: A correction. *Icarus* **166**, 212–222 (2003). doi: 10.1016/j.icarus.2003.06.001
70. M. M. Hedman, P. D. Nicholson, The B-ring's surface mass density from hidden density waves: Less than meets the eye? *Icarus* **279**, 109–124 (2016). doi: 10.1016/j.icarus.2016.01.007
71. P. R. Estrada, R. H. Durisen, J. N. Cuzzi, D. A. Morgan, Combined structural and compositional evolution of planetary rings due to micrometeoroid impacts and ballistic transport. *Icarus* **252**, 415–439 (2015). doi: 10.1016/j.icarus.2015.02.005
72. F. Poulet, J. N. Cuzzi, D. P. Cruikshank, T. Roush, C. M. Dalle Ore, Comparison between the Shkuratov and Hapke Scattering Theories for Solid Planetary Surfaces: Application to the Surface Composition of Two Centaurs. *Icarus* **160**, 313–324 (2002). doi: 10.1006/icar.2002.6970
73. H. Salo, R. G. French, The opposition and tilt effects of Saturn's rings from HST observations. *Icarus* **210**, 785–816 (2010). doi: 10.1016/j.icarus.2010.07.002
74. E. Déau *et al*., The opposition effect in Saturn's main rings as seen by Cassini ISS: 1. Morphology of phase functions and dependence on the local optical depth. *Icarus* **226**, 591–603 (2013). doi: 10.1016/j.icarus.2013.01.015
75. E. Déau, The opposition effect in Saturn's main rings as seen by Cassini ISS: 2. Constraints on the ring particles and their regolith with analytical radiative transfer models. *Icarus* **253**, 311–345 (2015). doi: 10.1016/j.icarus.2013.08.031
76. J. N. Cuzzi, L. B. Chambers, A. R. Hendrix, Rough surfaces: Is the dark stuff just shadow? *Icarus* **289**, 281–294 (2017). doi: 10.1016/j.icarus.2016.10.018
77. C. C. Porco *et al*., Simulations of the dynamical and light-scattering behavior of Saturn's rings and the derivation of ring particle and disk properties. *Astron. J.* **136**, 2172–2200 (2008). doi: 10.1088/0004-6256/136/5/2172
78. V. G. Kunde *et al*., in *Proc. SPIE Vol. 2803, Cassini/Huygens: A Mission to the Saturnian Systems*, L. Horn, Ed. (SPIE, 1996), pp. 162–177.
79. D. E. Jennings *et al*., Composite infrared spectrometer (CIRS) on Cassini. *Appl. Opt.* **56**, 5274–5294 (2017). doi: 10.1364/AO.56.005274; pmid: 29047582
80. L. J. Spilker *et al*., Cassini thermal observations of Saturn's main rings: Implications for particle rotation and vertical mixing. *Planet. Space Sci.* **54**, 1167–1176 (2006). doi: 10.1016/j.pss.2006.05.033
81. L. Spilker, C. Ferrari, R. Morishima, Saturn's ring temperatures at equinox. *Icarus* **226**, 316–322 (2013). doi: 10.1016/j.icarus.2013.06.002
82. S. Pilorz, N. Altobelli, J. Colwell, M. Showalter, Thermal transport in Saturn's B ring inferred from Cassini CIRS. *Icarus* **254**, 157–177 (2015). doi: 10.1016/j.icarus.2015.01.002
83. E. Reffet, M. Verdier, C. Ferrari, Thickness of Saturn's B ring as derived from seasonal temperature variations measured by Cassini CIRS. *Icarus* **254**, 276–286 (2015). doi: 10.1016/j.icarus.2015.04.006
84. R. Morishima *et al*., A multilayer model for thermal infrared emission of Saturn's rings II: Albedo, spins, and vertical mixing of ring particles inferred from Cassini CIRS. *Icarus* **210**, 330–345 (2010). doi: 10.1016/j.icarus.2010.06.032

ACKNOWLEDGMENTS

We thank the Cassini project, the science planners, and the instrument teams for making these observations possible. We thank T. Denk for the color mosaic image processing used in Fig. 1A. **Funding:** M.S.T., J.E.C., and R.G.J. acknowledge funding from the NASA Cassini Data Analysis program (grants NNX16AI33G and NNX15AH22G) and the Cassini project. P.D.N. acknowledges funding from NASA via the Cassini Project under JPL contract no. 1403282. The work of L.J.S., S.M.B., E.D., and R.M. was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA; U.S. government sponsorship is acknowledged. C.D.M., N.J.C., and S.V.B. acknowledge UK Science and Technology Facilities Council grants ST/P000622/1, ST/R000816/1, and ST/M005534/1. G.F. was supported by the Italian Space Agency and Italian National Institute for Astrophysics. C.F. acknowledges funding from the French Space Agency CNES. **Author contributions:** M.S.T., P.D.N., J.N.C., L.J.S., C.D.M., M.M.H., and J.E.C. led the execution and analysis of the observations, with contributions from J.A.B., S.M.B., R.N.C., N.J.C., E.D., C.F., G.F., R.G.J., S.L.M., R.M., S.P., S.R., and M.R.S. S.V.B., E.J.B., B.J.B., K.H.B., and C.S. contributed to executing the observations. **Competing interests:** J.N.C. is also affiliated with the University of California Santa Cruz. **Data and materials availability:** The Cassini data used in this paper are available from the Ring-Moon Systems Node of NASA's Planetary Data System (https://pds-rings.seti.org). The Product IDs of the images that we used are specified in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/eaau1017/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S4
References (*85*–*100*)

7 May 2018; accepted 7 May 2019
10.1126/science.aau1017

RESEARCH ARTICLE SUMMARY

MICROBIOTA

Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism

Vayu Maini Rekdal, Elizabeth N. Bess, Jordan E. Bisanz, Peter J. Turnbaugh*, Emily P. Balskus*

INTRODUCTION: Parkinson's disease is a debilitating neurological condition affecting more than 1% of the global population aged 60 and above. The primary medication used to treat Parkinson's disease is levodopa (L-dopa). To be effective, L-dopa must enter the brain and be converted to the neurotransmitter dopamine by the human enzyme aromatic amino acid decarboxylase (AADC). However, the gastrointestinal tract is also a major site for L-dopa decarboxylation, and this metabolism is problematic because dopamine generated in the periphery cannot cross the blood-brain barrier and causes unwanted side effects. Thus, L-dopa is coadministered with drugs that block peripheral metabolism, including the AADC inhibitor carbidopa. Even with these drugs, up to 56% of L-dopa fails to reach the brain. Moreover, the efficacy and side effects of L-dopa treatment are extremely heterogeneous across Parkinson's patients, and this variability cannot be completely explained by differences in host metabolism.

RATIONALE: Previous studies in humans and animal models have demonstrated that the gut microbiota can metabolize L-dopa. The major proposed pathway involves an initial decarboxylation of L-dopa to dopamine, followed by conversion of dopamine to *m*-tyramine by means of a distinctly microbial dehydroxylation reaction. Although these metabolic activities were shown to occur in complex gut microbiota samples, the specific organisms, gene, and enzymes responsible were unknown. The effects of host-targeted inhibitors such as carbidopa on gut microbial L-dopa metabolism were also unclear. As a first step toward understanding the gut microbiota's effect on Parkinson's disease therapy, we sought to elucidate the molecular basis for gut microbial L-dopa and dopamine metabolism.

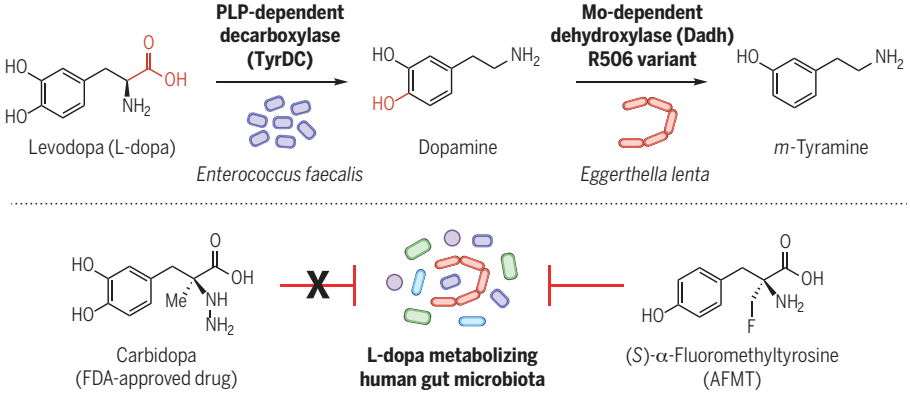
RESULTS: Hypothesizing that L-dopa decarboxylation would require a pyridoxal phosphate (PLP)–dependent enzyme, we searched gut bacterial genomes for candidates and identified a conserved tyrosine decarboxylase (TyrDC) in *Enterococcus faecalis*. Genetic and biochemical experiments revealed that TyrDC simultaneously decarboxylates both L-dopa and its preferred substrate, tyrosine. Next, we used enrichment culturing to isolate a dopamine-dehydroxylating strain of *Eggerthella lenta*, a species previously implicated in drug metabolism. Transcriptomics linked this activity to a molybdenum cofactor–dependent dopamine dehydroxylase (Dadh) enzyme. Unexpectedly, the presence of this enzyme in gut bacterial genomes did not correlate with dopamine metabolism; instead, we identified a single-nucleotide polymorphism (SNP) in the *dadh* gene that predicts activity. The abundance of *E. faecalis*, *tyrDC*, and the individual SNPs of *dadh* correlated with L-dopa and dopamine metabolism in complex gut microbiotas from Parkinson's patients, indicating that these organisms, genes, enzymes, and even nucleotides are relevant in this setting.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aau6323

We then tested whether the host-targeted AADC inhibitor carbidopa affected L-dopa decarboxylation by *E. faecalis* TyrDC. Carbidopa displayed greatly reduced potency toward bacteria and was completely ineffective in complex gut microbiotas from Parkinson's patients, suggesting that this drug likely does not prevent microbial L-dopa metabolism in vivo. To identify a selective inhibitor of gut bacterial L-dopa decarboxylation, we leveraged our molecular understanding of gut microbial L-dopa metabolism. Given TyrDC's preference for tyrosine, we examined tyrosine mimics and found that (*S*)-α-fluoromethyltyrosine (AFMT) prevented L-dopa decarboxylation by TyrDC and *E. faecalis* as well as complex gut microbiota samples from Parkinson's patients. Coadministering AFMT with L-dopa and carbidopa to mice colonized with *E. faecalis* also increased the peak serum concentration of L-dopa. This observation is consistent with inhibition of gut microbial L-dopa metabolism in vivo.

CONCLUSION: We have characterized an interspecies pathway for gut bacterial L-dopa metabolism and demonstrated its relevance in human gut microbiotas. Variations in these microbial activities could possibly contribute to the heterogeneous responses to L-dopa observed among patients, including decreased efficacy and harmful side effects. Our findings will enable efforts to elucidate the gut microbiota's contribution to treatment outcomes and highlight the promise of developing therapies that target both host and gut microbial drug metabolism. ■



Gut microbes metabolize the Parkinson's drug L-dopa. Decarboxylation of L-dopa by *E. faecalis* TyrDC and human AADC likely limits drug availability and contributes to side effects. *E. lenta* dehydroxylates dopamine produced from L-dopa using a molybdenum-dependent enzyme. Although the host-targeted drug carbidopa did not affect gut bacterial L-dopa decarboxylation, AFMT inhibited this activity in complex human gut microbiotas.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: balskus@chemistry.harvard.edu (E.P.B.); peter.turnbaugh@ucsf.edu (P.J.T.)
Cite this article as V. Maini Rekdal *et al.*, *Science* **364**, eaau6323 (2019). DOI: 10.1126/science.aau6323

RESEARCH ARTICLE

MICROBIOTA

Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism

Vayu Maini Rekdal[1], Elizabeth N. Bess[2,3,4], Jordan E. Bisanz[2], Peter J. Turnbaugh[2,5]*, Emily P. Balskus[1]*

The human gut microbiota metabolizes the Parkinson's disease medication Levodopa (L-dopa), potentially reducing drug availability and causing side effects. However, the organisms, genes, and enzymes responsible for this activity in patients and their susceptibility to inhibition by host-targeted drugs are unknown. Here, we describe an interspecies pathway for gut bacterial L-dopa metabolism. Conversion of L-dopa to dopamine by a pyridoxal phosphate-dependent tyrosine decarboxylase from *Enterococcus faecalis* is followed by transformation of dopamine to *m*-tyramine by a molybdenum-dependent dehydroxylase from *Eggerthella lenta*. These enzymes predict drug metabolism in complex human gut microbiotas. Although a drug that targets host aromatic amino acid decarboxylase does not prevent gut microbial L-dopa decarboxylation, we identified a compound that inhibits this activity in Parkinson's patient microbiotas and increases L-dopa bioavailability in mice.

A growing body of evidence links the trillions of microbes that inhabit the human gastrointestinal tract (the human gut microbiota) to neurological conditions, including the debilitating neurodegenerative disorder Parkinson's disease (*1*, *2*). Gut microbes from Parkinson's patients exacerbate motor deficits when transplanted into germ-free mouse models of disease (*2*). This effect is reversed with antibiotic treatment, suggesting a causal role for gut microbes in neurodegeneration. Multiple studies have revealed differences in gut microbiota composition in Parkinson's disease patients compared with healthy controls that may correlate with disease severity (*3*–*9*). However, the influence of the human gut microbiota on the treatment of Parkinson's and other neurodegenerative diseases remains poorly understood.

The primary treatment for Parkinson's disease is Levodopa (L-dopa) (*10*), which is prescribed to manage motor symptoms that result from dopaminergic neuron loss in the substantia nigra. After crossing the blood-brain barrier, L-dopa is decarboxylated by aromatic amino acid decarboxylase (AADC) to give dopamine, the active therapeutic agent. However, dopamine generated in the periphery by AADC cannot cross the blood-brain barrier, and only 1 to 5% of L-dopa reaches the brain, owing to extensive presystemic metabolism in the gut by enzymes such as AADC (*11*–*13*). Peripheral production of dopamine also causes gastrointestinal side effects, can lead to orthostatic hypotension through activation of vascular dopamine receptors, and may induce cardiac arrhythmias (*14*, *15*). To decrease peripheral metabolism, L-dopa is coadministered with AADC inhibitors such as carbidopa. Despite this, 56% of L-dopa is metabolized peripherally (*16*), and patients display highly variable responses to the drug, including loss of efficacy over time (*17*).

Multiple lines of evidence suggest that gut microbial interactions with L-dopa influence treatment outcomes (*18*). Administering broad-spectrum antibiotics improves L-dopa therapy, suggesting that gut bacteria interfere with drug efficacy (*19*, *20*). The gut microbiota can also metabolize L-dopa, potentially reducing its bioavailability and leading to side effects (*21*–*24*). The major proposed pathway involves an initial decarboxylation of L-dopa to dopamine followed by a distinctly microbial dehydroxylation reaction that converts this neurotransmitter to *m*-tyramine by selectively removing the para hydroxyl group of the catechol ring (Fig. 1A) (*25*, *26*). When we began our work, the gut microbial species, genes, and enzymes involved in these transformations were unknown because previous studies examined undefined and uncharacterized consortia. The clinical relevance of this pathway was also unclear given the potential effects of coadministered inhibitors of host peripheral L-dopa metabolism on these gut microbial activities.

[1]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA. [2]Department of Microbiology and Immunology, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA. [3]Department of Chemistry, University of California, Irvine, 1102 Natural Sciences 2, Irvine, CA 92617, USA. [4]Department of Molecular Biology and Biochemistry, University of California, Irvine, 1102 Natural Sciences 2, Irvine, CA 92617, USA. [5]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.
*Corresponding author. Email: balskus@chemistry.harvard.edu (E.P.B.); peter.turnbaugh@ucsf.edu (P.J.T.)

The human gut bacterium *Enterococcus faecalis* decarboxylates L-dopa

We sought to elucidate the genetic and biochemical bases for gut microbial L-dopa metabolism and understand how coadministered AADC inhibitors affect this pathway. Using a genome-mining approach, we first identified strains that encode candidate L-dopa decarboxylating enzymes. Aromatic amino acid decarboxylation is typically performed by enzymes using pyridoxal-5′-phosphate (PLP), an organic cofactor that provides an electron sink (*27*). Recently, a PLP-dependent tyrosine decarboxylase (TyrDC) from the food-associated strain *Lactobacillus brevis* CGMCC 1.2028 was shown to have promiscuous activity toward L-dopa in vitro (*28*). To locate TyrDC homologs in human gut bacteria, we performed a BLASTP (Protein Basic Local Alignment Search Tool) search against the complete set of Human Microbiome Project (HMP) reference genomes available through the National Center for Biotechnology Information (NCBI). The majority of hits were found in the neighboring genus *Enterococcus*, with some hits within lactobacilli and Proteobacteria (Fig. 1B, fig. S1, and data file S1). We selected 10 representative gut strains that contain TyrDC homologs (29 to 100% amino acid ID) and examined their ability to decarboxylate L-dopa in anaerobic culture. Although both *Enterococcus faecalis* and *Enterococcus faecium* displayed activity, only *E. faecalis* showed complete decarboxylation across all strains tested (Fig. 1C). All *E. faecalis* strains tested share the highly conserved four-gene *tyrDC* operon (fig. S2), and we found *tyrDC* in 98.4% of the *E. faecalis* assemblies deposited in NCBI with a median amino acid identity of 99.8 (range 97.0 to 100). This high degree of sequence conservation and prevalence is consistent with tyrosine decarboxylation being a common phenotypic trait of *E. faecalis* (*29*). We therefore chose this prevalent, genetically tractable gut organism as a model for characterizing L-dopa decarboxylation (*30*).

Although lyophilized *E. faecalis* cells decarboxylate L-dopa (*31*) and the *tyrDC* operon's role in tyrosine decarboxylation in *E. faecalis* is well-characterized (*32*), the connection between *tyrDC* and L-dopa decarboxylation was unknown. We used genetics and in vitro biochemistry experiments to confirm that TyrDC is necessary and sufficient for L-dopa decarboxylation by *E. faecalis*. *E. faecalis* MMH594 mutants carrying a 2-kb Tet-cassette disrupting *tyrDC* could not decarboxylate L-dopa (Fig. 1D and fig. S3) and displayed no growth defects compared with wild type (fig. S4). In vitro characterization of TyrDC revealed a five-fold higher catalytic efficiency toward L-tyrosine compared with L-dopa, suggesting that drug metabolism arises from promiscuous enzyme activity (Fig. 1E, fig. S5, and table S1). This selectivity contrasts sharply with that of AADC, which displays very low activity toward L-tyrosine (*33*). Although TyrDC from *E. faecalis* was previously shown to decarboxylate tyrosine and phenylalanine

(*34*–*37*), its ability to accept L-dopa had not been demonstrated. A recent, independent report also corroborates this finding (*38*).

We next tested whether tyrosine, which is the preferred substrate for TyrDC and is present in the small intestine, could interfere with L-dopa decarboxylation by *E. faecalis* (*39*, *40*). In competition experiments, purified TyrDC (fig. S6) and anaerobic *E. faecalis* cultures decarboxylated L-dopa and tyrosine simultaneously (500 μM tyrosine, approximating the resting small intestinal concentration) (Fig. 1F and fig. S7) (*40*). This observation sharply contrasts with previous investigations of phenylalanine, which is metabolized by *E. faecalis* only when tyrosine is completely consumed (*36*). Simultaneous decarboxylation of L-dopa and tyrosine also occurred in *E. faecalis* MMH594 cultures that contained higher tyrosine concentrations (1.5 mM, approximating the small intestinal post-meal concentration) (fig. S8) and in three human fecal suspensions (fig. S9). As observed previously for tyrosine, L-dopa decarboxylation occurred more rapidly at lower pH across all strains tested (figs S7 and S8), suggesting that this metabolism is likely accelerated at the lower pH of the upper small intestine (*41*, *42*). Because the Michaelis constant (K_m) of TyrDC for L-dopa (1.5 mM) is below the estimated maximum in vivo small intestinal L-dopa concentration even at its lowest clinically administered dose (5 mM), these data strongly suggest that peripheral decarboxylation is performed by both host and gut bacterial enzymes.

Eggerthella lenta dehydroxylates dopamine using a molybdenum-dependent enzyme

Having identified a gut bacterial L-dopa decarboxylase, we next examined the conversion of dopamine to *m*-tyramine because this activity may influence the side effects associated with peripheral L-dopa decarboxylation. *E. faecalis* did not further metabolize dopamine, indicating that this step was performed by a different microorganism. Dehydroxylation of dopamine has not been reported for any bacterial isolate, and a screen of 18 human gut strains failed to uncover metabolizers. Therefore, we used enrichment culturing to obtain a dopamine-dehydroxylating organism. Recognizing the chemical parallels between this reductive dehydroxylation and reductive dehalogenation of chlorinated aromatics, which enables anaerobic respiration in certain bacteria (*43*), we inoculated a stool sample from a human donor into a minimal growth medium containing 0.5 mM dopamine as the sole electron acceptor (figs. S10 and S11). Passaging over multiple generations enriched for active strains, as assessed by means of a colorimetric assay for catechol dehydroxylation (fig. S11). This effort identified a strain of the gut Actinobacterium *Eggerthella lenta* (referred to herein as strain A2) that is capable of selectively removing the para hydroxyl group of dopamine to give *m*-tyramine (fig. S12). Because *E. lenta* also inactivates the cardiac drug digoxin, our finding suggests a wider role for this gut organism in drug metabolism (*44*, *45*).

Catechol dehydroxylation is a chemically challenging reaction that has no equivalent in synthetic chemistry and likely involves unusual enzymology. To identify the dopamine-dehydroxylating enzyme, we first searched the *E. lenta* A2 genome for genes that encode homologs of the only characterized aromatic *para*-dehydroxylase, 4-hydroxybenzoyl-CoA reductase (*46*), but found no hits. Assays with *E. lenta* A2 cell lysates showed dopamine dehydroxylation required anaerobic conditions and was induced by dopamine (fig. S13). We therefore used RNA-sequencing of *E. lenta* A2 to identify the dehydroxylase. This experiment revealed >2500-fold up-regulation of three colocalized genes in response to dopamine (Fig. 2A and table S2). These genes encode a predicted bis-molybdopterin guanine dinucleotide cofactor (moco)–containing enzyme belonging to the dimethyl sulfoxide reductase family. Moco-dependent enzymes catalyze a wide variety of oxygen-transfer reactions but have not been demonstrated to catalyze catechol dehydroxylation in vitro (*47*). We therefore hypothesized that this enzyme was a dopamine dehydroxylase (Dadh).

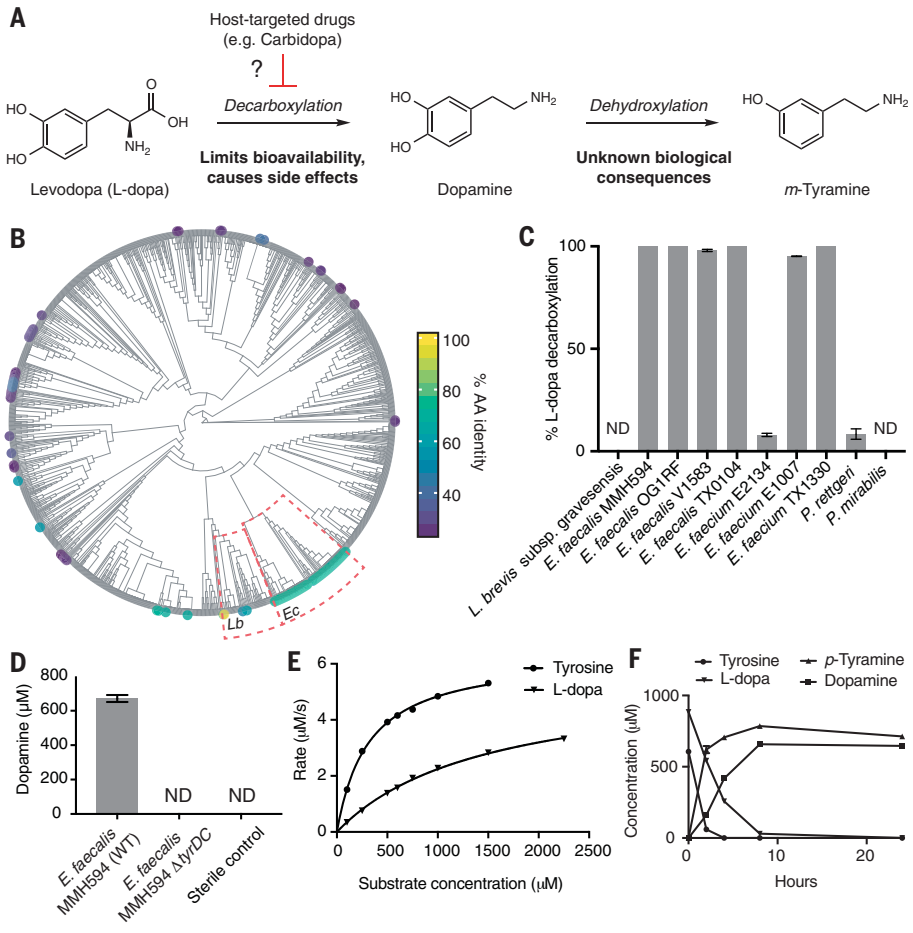


Fig. 1. *E. faecalis* metabolizes L-dopa using a PLP-dependent tyrosine decarboxylase. (**A**) Proposed major pathway for L-dopa metabolism by the human gut microbiota and potential for interaction with host-targeted drugs. (**B**) Phylogenetic distribution of TyrDC in the human microbiota. Human Microbiome Project reference genomes were queried by means of BLASTP for homologs of the *L. brevis* TyrDC, and the results are visualized on a cladogram of phylogeny [based on 16S ribosomal RNA (rRNA) alignment]. TyrDC homologs found sporadically within *Lactobacillus* spp. (*Lb*) are widely distributed among *Enterococcus* (*Ec*; average amino acid identity 67.8% over 97.6% query length). (**C**) Testing representative gut microbial strains encoding TyrDC reveals that *E. faecalis* strains reproducibly convert L-dopa to dopamine. Strains were cultured for 48 hours anaerobically. Bar graphs represent the mean ± SEM of three biological replicates. (**D**) Deletion of *tyrDC* abolishes L-dopa decarboxylation by *E. faecalis*. Dopamine was detected in culture supernatants after 48 hours of anaerobic growth with 0.5 mM L-dopa. Bar graphs represent the mean ± SEM of three biological replicates. (**E**) Kinetic analysis of *E. faecalis* TyrDC reveals a preference for tyrosine. Error bars represent the mean ± SEM of three biological replicates. ND, not detected. (**F**) L-dopa and tyrosine are simultaneously decarboxylated in anaerobic cultures of *E. faecalis* MMH594 grown at pH 5 with 1 mM L-dopa and 0.5 mM tyrosine. Bar graphs represent the mean ± SEM of three biological replicates.

To assess Dadh's role in dopamine dehydroxylation, we first explored whether this activity was molybdenum-dependent by culturing *E. lenta* A2 in the presence of tungstate. Substitution of molybdate with tungstate during moco biosynthesis generates an inactive metallocofactor (fig. S14) (*48*). Treating cultures of *E. lenta* A2 with tungstate inhibited dopamine dehydroxylation without affecting growth (Fig. 2B and fig. S15), whereas incubating cell lysates with tungstate had no effect, which is consistent with inhibition requiring active moco biosynthesis (fig. S16). We next confirmed the activity of Dadh in vitro. Heterologous expression of >20 constructs in multiple hosts failed to provide active enzyme, prompting us to pursue native purification. Anaerobic activity–guided fractionation of *E. lenta* A2 cell lysates yielded a dopamine-dehydroxylating fraction containing four proteins as assessed by means of SDS–polyacrylamide gel electrophoresis (Fig. 2C, fig. S17, and table S3). Dehydroxylation activity correlated with a 115-kDa band that was confirmed with mass spectrometry (MS) to be Dadh. Dadh was the only isolated protein up-regulated in the presence of dopamine (tables S2 and S3). Together, these data strongly support the assignment of this enzyme.

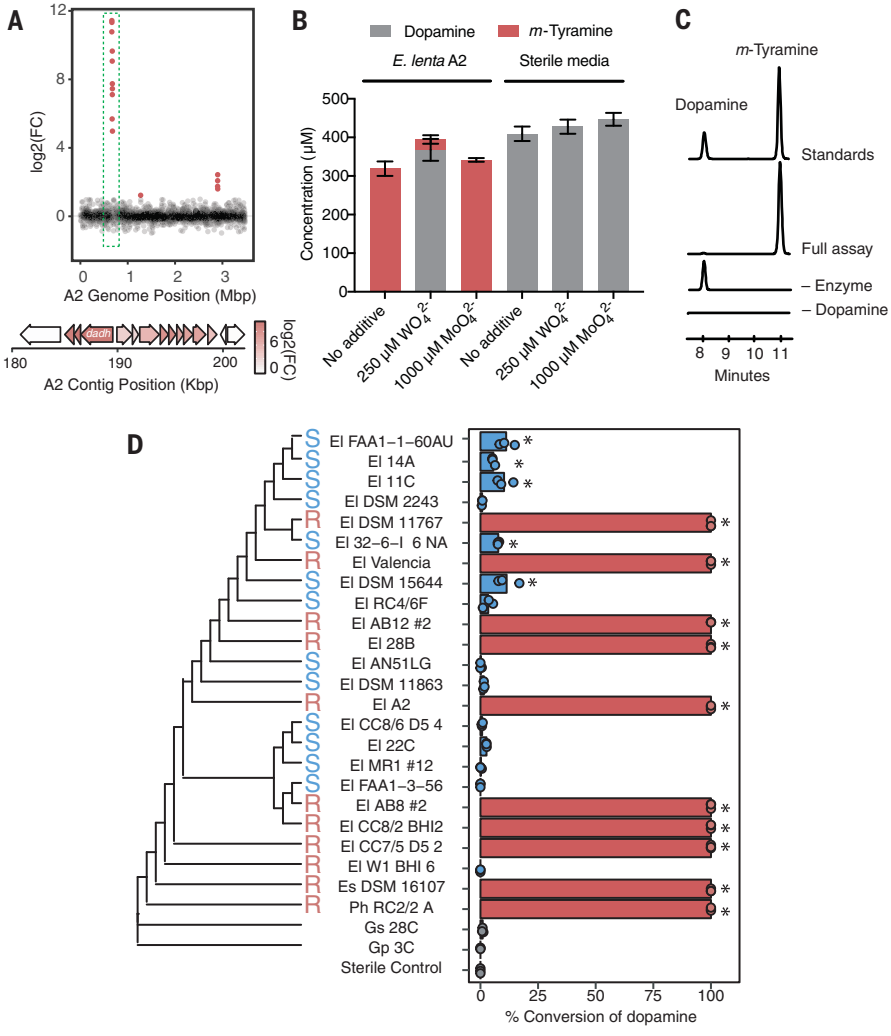


Fig. 2. *E. lenta* dehydroxylates dopamine using a molybdenum-dependent enzyme. (**A**) RNA-sequencing identifies a putative molybdenum (moco)–dependent dopamine dehydroxylase (Dadh) in *E. lenta* A2. Differentially expressed candidate genes (false discovery rate < 0.1 and fold change > |2|) are plotted as a function of genome position, revealing three discrete loci of differentially expressed genes. (Inset) Analysis of the largest cluster of differentially expressed genes at 0.665 Mbp in the scaffolded assembly (190 kg base pairs in the reference contig) revealed that a putative *dadh* was up-regulated by 2568-fold in response to dopamine. (**B**) Tungstate treatment inhibits dehydroxylation of dopamine by *E. lenta* A2. Cultures were grown anaerobically with tungstate (WO_4^{2-}) or molybdate (MoO_4^{2-}) for 48 hours with 0.5 mM dopamine. Bar graphs represent the mean ± SEM of three biological replicates. (**C**) In vitro activity of Dadh-containing fractions purified from *E. lenta* A2. Extracted LC-MS/MS ion chromatograms for simultaneous detection of dopamine and *m*-tyramine after 12 hours of anaerobic incubation of enzyme preparation with 500 μM dopamine and artificial electron donors at room temperature. Peak heights show the relative intensity of each mass, and all chromatograms are shown on the same scale. (**D**) A single amino acid variant predicts dopamine metabolism by *E. lenta* and related strains ($P = 0.013$ Fisher's exact test) and does not correlate with phylogeny. Strains were cultured anaerobically with 500 μM dopamine for 48 hours (*El*, *E. lenta*; *Es*, *Eggerthella sinensis*; Gs, *Gordonibacter* sp.; and Gp, *Gordonibacter pamelaeae*; Ph, *Paraeggerthella hongkongensis*). High (100% conversion) and low (<11% conversion) metabolizers are denoted in red and blue. For each strain, data points represent biological replicates (*$P < 0.05$ analysis of variance with Dunnett's test versus sterile controls).

We next assessed whether the presence of *dadh* in microbial genomes correlated with dopamine dehydroxylation. A BLASTP search revealed that this enzyme is restricted to *E. lenta* and its close Actinobacterial relatives (table S4), prompting us to screen a collection of 26 gut Actinobacterial isolates (*49*) for their ability to dehydroxylate dopamine in anaerobic culture. Although Dadh appeared to be encoded by 24 of the 26 strains (92 to 100% amino acid ID) (fig. S18 and table S5), only 10 *Eggerthella* strains quantitatively converted dopamine to *m*-tyramine, with low (<11%) or no metabolism in the others (Fig. 2D). This strain-level variability in dopamine metabolism reinforces that gut microbial species identity is often not predictive of metabolic functions (*49*, *50*).

To better understand this variation, we first performed RNA-sequencing experiments with metabolizing (*E. lenta* 28B) and nonmetabolizing (*E. lenta* DSM2243) strains in the presence and absence of dopamine. Surprisingly, *dadh* was up-regulated in response to dopamine in both strains, indicating that lack of activity in *E. lenta* DSM2243 did not arise from differences in transcription (tables S6 and S7). Aligning the Dadh protein sequences, we instead found a single amino acid substitution that almost perfectly predicted metabolizer status: Position 506 is an arginine in metabolizing strains and a serine in inactive strains (Fig. 2D and fig. S19). This change arises from a single-nucleotide polymorphism (SNP) in *dadh*. The only exception, *E. lenta* W1BHI6, has the Arg^{506} variant and an additional substitution nearby (Cys^{500}) (fig. S19). Thus, specific amino acid residues in the Dadh enzyme, rather than presence or transcription of *dadh*, predict dopamine dehydroxylation among gut bacterial strains. The Dadh variants do not correlate with *E. lenta* phylogeny (Fig. 2D), suggesting that this activity has been gained and/or lost multiple times.

E. faecalis and *E. lenta* metabolize L-dopa in human gut microbiotas

Having identified organisms and enzymes that perform the individual steps in the L-dopa pathway, we next tested whether *E. faecalis* and *E. lenta* generated *m*-tyramine in coculture. Wild-type *E. faecalis* grown with *E. lenta* A2

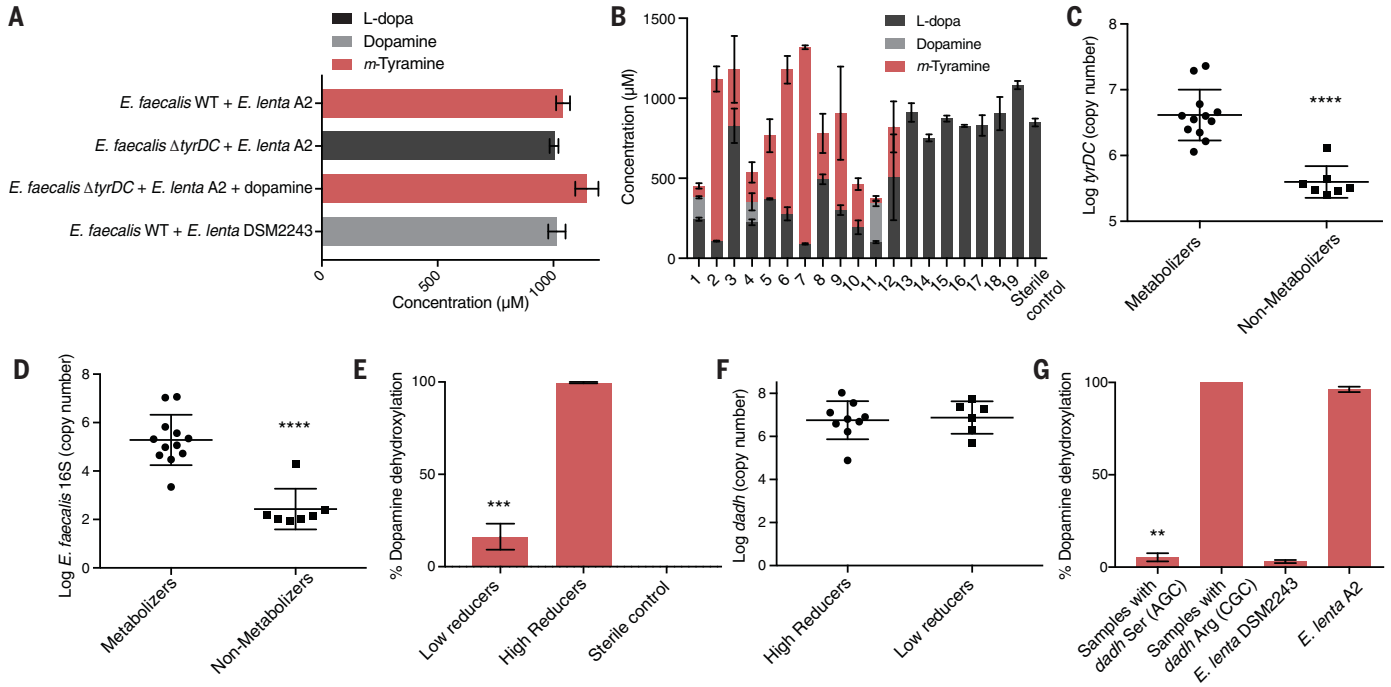


Fig. 3. *E. faecalis* and *E. lenta* Dadh predict L-dopa metabolism in complex human gut microbiotas. (**A**) Metabolism of L-dopa by cocultures of *E. faecalis* and *E. lenta* strains cocultured for 48 hours with 1 mM d_3-phenyl-L-dopa or 1 mM dopamine. Results are mean ± SEM (n = 3 replicates). (**B**) Metabolism of d_3-phenyl-L-dopa by 19 unrelated human gut microbiota samples ex vivo. Samples were cultured anaerobically with d_3-phenyl-L-dopa (1 mM) for 72 hours. Results are mean concentration ± SEM (n = 3 replicates). (**C**) The abundance of *tyrDC* predicts L-dopa decarboxylation in human gut microbiota samples. Data represent the average *tyrDC* abundance (as assessed with qPCR) across the three replicates for samples in (B). Results are mean ± SEM (****P < 0.0001, one tailed Mann-Whitney test). (**D**) The abundance of *E. faecalis* (as assessed with qPCR) predicts L-dopa decarboxylation in human gut microbiota samples. Each data point is the average abundance across three biological replicates for each sample shown in (B). Results are mean ± SEM (****P < 0.0001, one-tailed Mann-Whitney test). (**E**) Dopamine dehydroxylation by gut microbiota samples of 15 unrelated individuals. Samples were cultured for 48 hours with 0.5 mM dopamine. Bars are mean ± SEM of n = 6 for low reducers (<50%) and n = 9 for high reducers (>50%) (*** P = 0.0002, one tailed Mann-Whitney test). (**F**) *Dadh* abundance does not correlate with dehydroxylation by human gut microbiotas. Data represent qPCR with *Dadh*-specific primers. Each data point is the *dadh* abundance in each sample shown in (E). Bars represent the mean and SE. (**G**) Dadh sequence variants predict dopamine dehydroxylation ex vivo. Full-length *dadh* from each culture in (E) was sequenced by using primers specific for the region containing position 506. Samples in which a mix of variants were present (n = 5) were removed. Bars represent the mean and SEM [n = 3 for samples encoding the Arg506 Dadh variant, n = 7 for samples encoding the Ser506 Dadh variant., n = 3 for DSM2243, and n = 3 for A2] (** P = 0.0083, one-tailed Mann-Whitney test, CGC samples versus AGC samples).

(Arg^{506}) fully converted L-dopa to *m*-tyramine (Fig. 3A). Although a coculture containing the *E. faecalis tyrDC* mutant could not consume L-dopa, *m*-tyramine was produced when exogenous dopamine was added to this culture, revealing that *E. lenta* A2 was still metabolically active. Incubating wild-type *E. feacalis* with the nonmetabolizing *E. lenta* DSM2243 (Ser506) strain produced only dopamine, indicating that this Dadh variant is also inactive in a coculture setting (Fig. 3A).

To investigate whether *E. faecalis* and *E. lenta* transform L-dopa in the human gut microbiota, we assessed the metabolism of deuterated L-dopa by fecal suspensions ex vivo. Whereas 7 of 19 samples did not show detectable depletion of L-dopa, the remaining samples displayed substantial variability in metabolism, ranging from partial (25%) to almost full conversion (98%) of L-dopa to *m*-tyramine (Fig. 3B). We next asked whether the abundance of *tyrDC* predicted metabolism in these samples. Quantitative polymerase chain reaction (qPCR) enumeration of *tyrDC* (*51*) and *E. faecalis* discriminated metabolizing and nonmetabolizing samples (P < 0.0001, one-tailed Mann-Whitney test) (Fig. 3, C and D). By contrast, *E. lenta* abundance showed no association with L-dopa metabolism (fig. S20). We found a strong linear correlation between *tyrDC* abundance and *E. faecalis* abundance [coefficient of determination (R^2) = 0.99, P < 0.0001] (fig. S21), which likely reflects the high conservation of *tyrDC* in *E. faecalis* genomes. These data also suggest that *E. faecalis* is the dominant microorganism responsible for L-dopa decarboxylation in these complex human gut microbial communities. Consistent with this, *E. faecalis* abundance significantly correlated with *tyrDC* abundance in 1870 human gut microbiomes (R^2 > 0.812, P < 2.2×10^{-16}, Pearson's correlation) (fig. S22).

To confirm that *E. faecalis* could decarboxylate L-dopa in complex gut microbiotas, we added this organism to nonmetabolizing samples. Although introducing the *tyrDC*-deficient strain did not change L-dopa levels, including the wild-type strain led to complete depletion of L-dopa (fig. S23, B to E). In some samples, addition of wild-type *E. faecalis* was sufficient to yield quantitative production of *m*-tyramine, indicating the presence of dopamine-dehydroxylating organisms in these communities (fig. S23, B and D). Last, addition of both the wild-type *E. faecalis* and the metabolizing strain *E. lenta* A2 to nonmetabolizing samples or the addition of *E. lenta* A2 alone to a decarboxylating sample generated *m*-tyramine (fig. S23, A and C to E). Taken together, these data indicate that the abundance of *E. faecalis* and its encoded *tyrDC* predicts the considerable interindividual variation in L-dopa metabolism observed in complex human gut microbiota samples.

As expected from our previous experiments, neither the abundance of *E. lenta* nor *dadh* predicted dopamine dehydroxylation in complex gut microbial communities (Fig. 3, E and F, and fig. S24). However, when we amplified *dadh* from these cultures and determined the SNP status at position 506, we found samples that contained the Arg^{506} variant quantitatively

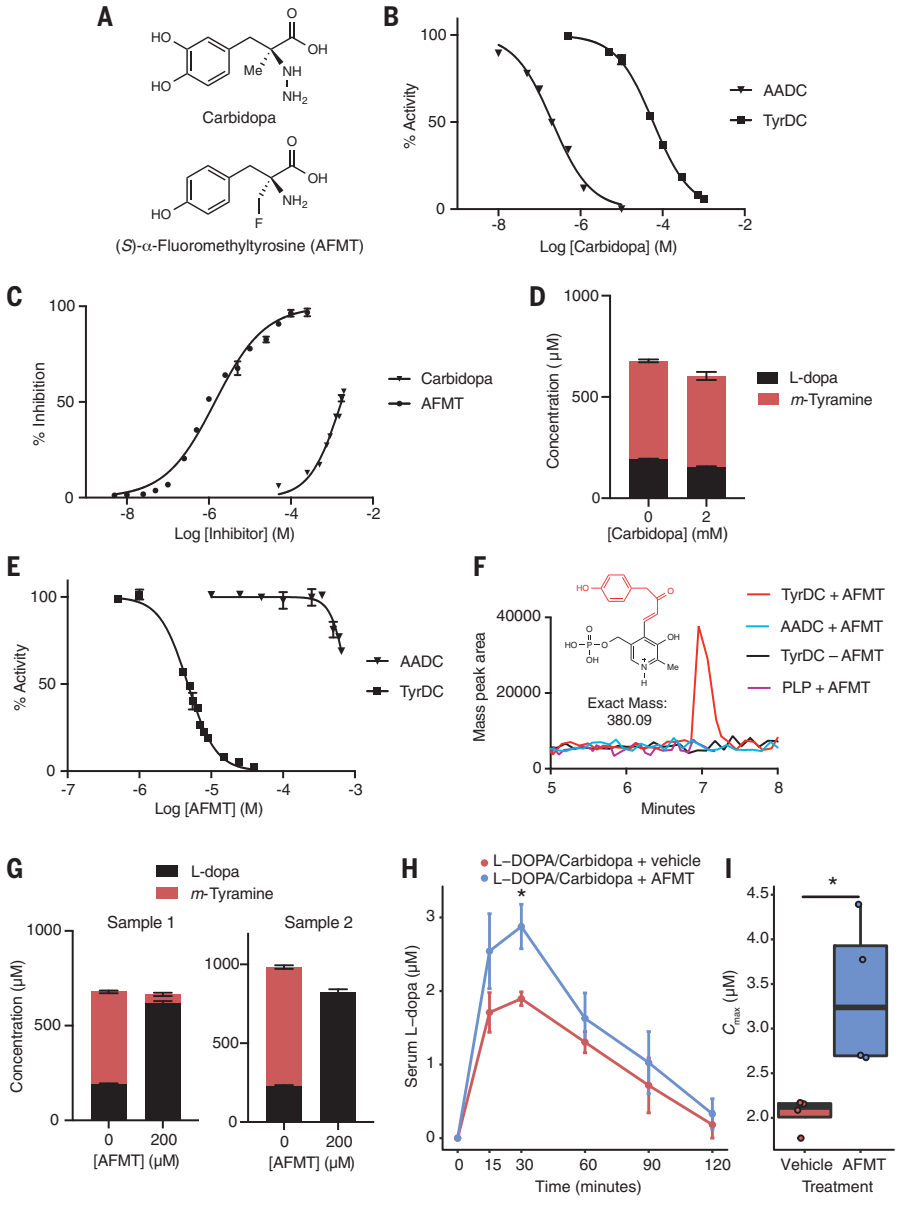


Fig. 4. L-dopa decarboxylation by *E. faecalis* is inhibited by AFMT but not the host-targeted drug carbidopa. (**A**) Carbidopa and AFMT. (**B**) Carbidopa preferentially inhibits human AADC over TyrDC. AADC or TyrDC were incubated with inhibitor, and reaction rates were measured with LC-MS/MS. "% Activity" represents the rate relative to a no inhibitor (vehicle) control. Results are mean ± SEM (n = 3 replicates). (**C**) Activity of carbidopa and AFMT in cultures of *E. faecalis* grown for 16 hours anaerobically with 0.5 mM L-dopa. Error bars represent the mean ± SEM for three biological replicates. (**D**) Activity of carbidopa in a human fecal microbiota from a Parkinson's patient. The sample was cultured anaerobically with carbidopa and 1 mM d_3-phenyl-L-dopa for 72 hours. Error bars represent the mean ± SEM for three biological replicates. (**E**) AFMT preferentially inhibits TyrDC over AADC in vitro. AADC or TyrDC were incubated with inhibitor, and reaction rates were measured with LC-MS/MS. "% Activity" represents the rate relative to a no inhibitor (vehicle) control. Error bars represent the mean ± SEM for three biological replicates. (**F**) Detection of an AFMT-PLP covalent adduct after incubation of TyrDC or AADC with AFMT for 1 hour. The data shown is the extracted ion chromatogram of the mass of the predicted covalent adduct. (**G**) Action of AFMT in human fecal microbiotas from Parkinson's patients incubated anaerobically with AFMT and 1 mM d_3-phenyl-L-dopa for 72 hours. Error bars represent the mean ± SEM for three biological replicates. (**H**) Pharmacokinetic analysis in gnotobiotic mice colonized with *E. faecalis* and given L-dopa + carbidopa + AFMT demonstrates higher serum L-dopa relative to vehicle controls. Error bars represent the mean ± SEM. (**I**) The maximum serum concentration (C_{max}) of L-dopa is significantly higher with AFMT relative to vehicle controls. In (H) and (I), *P < 0.05, Mann-Whitney U test; n = 4 to 5 mice per group.

metabolized dopamine, whereas the activity of samples that carried the Ser^{506} variant was indistinguishable from the nonmetabolizing *E. lenta* DSM2243 strain (Fig. 3G). These findings indicate that a single amino acid residue in a gut microbial enzyme predicts dopamine metabolism in complex communities. Given that *dadh* is highly prevalent (>70%) in gut microbiomes from human subjects and the two *dadh* variants are present among this population (figs. S22 and S25), we speculate that SNPs may influence xenobiotic metabolism in the context of both the host genome (*52*) and the human gut microbiome (*53*).

To further explore the clinical relevance of our findings, we assessed the metabolism of dopamine and L-dopa by fecal suspensions from Parkinson's disease patients ex vivo. Similar to control subjects, these individuals displayed substantial variability in metabolism of L-dopa (fig. S26A). qPCR assays revealed that *tyrDC* abundance and *E. faecalis* abundance discriminated between L-dopa decarboxylating and nondecarboxylating samples (P < 0.005, one-tailed Mann-Whitney test) (fig. S26, C and D). We also observed depletion of L-dopa without corresponding production of dopamine or *m*-tyramine in three samples (fig. S26A). Instead, L-dopa was converted to hydroxyphenylpropionic acid (fig. S26B), a pathway thought to make a minor contribution to drug metabolism in vivo (*22*, *25*, *26*). Last, we found that the *dadh* SNP predicted dopamine dehydroxylation in these samples (fig. S27). Overall, these data support a role for gut bacteria in the extensive interindividual variability in L-dopa decarboxylation observed in Parkinson's patients (*13*). A recent study reported that stool *tyrDC* abundance is positively correlated with L-dopa dosage in patients (*38*) but did not demonstrate a connection between *tyrDC* and L-dopa decarboxylation in these samples. Our findings indicate this metabolic activity may indeed affect L-dopa therapeutic efficacy.

(*S*)-α-Fluoromethyltyrosine (AFMT) inhibits gut microbial L-dopa metabolism

Having shown that *E. faecalis* and *E. lenta* enzymes predict L-dopa metabolism by complex patient gut microbiotas, we next investigated whether this interspecies pathway was susceptible to inhibition by drugs that target peripheral L-dopa decarboxylation. In the United States, Parkinson's patients are coprescribed carbidopa (Fig. 4A), an L-dopa mimic that inhibits AADC by forming a stable, covalent hydrazone linkage with its PLP cofactor (*54*). We found carbidopa was 200 times less active toward purified *E. faecalis* TyrDC [half maximal inhibitory concentration (IC_{50}) = 57 μM] relative to *H. sapiens* AADC (IC_{50} = 0.21 μM) and showed only ~50% inhibition of L-dopa decarboxylation by *E. faecalis* cultures at the solubility limit of 2 mM (Fig. 4, B and C, and table S8), which is consistent with recently reported findings (*38*). Additionally, carbidopa did not affect growth of *E. faecalis* or metabolism or growth of *E. lenta* (figs. S28 to S30). Given the maximum

predicted gastrointestinal concentration of carbidopa (0.4 to 9 mM), these data suggest that this drug does not fully inhibit gut bacterial L-dopa decarboxylation in Parkinson's patients. We found that 2 mM carbidopa did not alter the kinetics of L-dopa degradation (fig. S31) or endpoint *m*-tyramine production in stool samples from both Parkinson's patients and neurologically healthy controls (Fig. 4D and fig. S32). These observations support previous findings that carbidopa administration does not affect *m*-tyramine production in patients (*55*).

Our results also highlight the possibility of therapeutically targeting gut microbial L-dopa decarboxylation to increase L-dopa efficacy. To selectively manipulate gut bacterial TyrDC in complex microbiotas, we turned to α-fluoromethyl amino acids, which are known mechanism-based inhibitors of PLP-dependent decarboxylases (*33*). A survey of potential amino acid substrates revealed that TyrDC requires a *p*-hydroxyl group for robust activity, whereas AADC prefers a *m*-hydroxyl substituent (fig. S33), leading us to hypothesize that the L-tyrosine analog (*S*)-α-fluoromethyltyrosine (AFMT) (Fig. 4A) might selectively inhibit the microbial enzyme. In vitro, AFMT strongly inhibited L-dopa decarboxylation by TyrDC (IC_{50} = 4.7 μM) but not AADC (~20% inhibition at solubility limit of 650 μM) (Fig. 4E and table S8). Consistent with this selectivity, AFMT formed a covalent PLP adduct only in the presence of TyrDC (Fig. 4F). AFMT was also effective in *E. faecalis* cultures (EC_{50} = 1.4 μM) (Fig. 4C), outperforming carbidopa by 1000-fold without affecting growth (table S8 and fig. S29). It also reduced L-dopa decarboxylation by co-cultures of *E. faecalis* and *E. lenta* without affecting growth or metabolism of *E. lenta* (figs. S29, S30, and S34). Last, AFMT completely inhibited L-dopa decarboxylation in gut microbiota samples from Parkinson's disease patients and neurologically healthy control subjects (Fig. 4G and fig. S35) and was nontoxic to eukaryotic cells (fig. S36).

To investigate AFMT activity in vivo, we administered either AFMT (25 mg/kg) or a vehicle control in combination with L-dopa (10 mg/kg) and carbidopa (30 mg/kg) to gnotobiotic mice colonized with *E. faecalis* MMH594 (Fig. 4H). We found that AFMT significantly increased the peak serum concentration (C_{max}) of L-dopa compared with vehicle ($P < 0.05$, two-tailed Mann Whitney test) (Fig. 4I), which is consistent with inhibition of first-pass gut microbial metabolism in the intestine. Although we cannot rule out the possibility that AFMT modulates additional, uncharacterized targets, this observation is consistent with our in vitro inhibition data. This result also aligns with a recent report that small intestinal *tyrDC* abundance negatively correlates with plasma L-dopa levels in conventional rats receiving L-dopa and carbidopa (*38*). Overall, these data suggest that AFMT could be a promising tool compound for the study of bacterial L-dopa metabolism (*56*) and highlight the promise of developing L-dopa–based combination therapies containing drugs that target both host and gut microbial decarboxylation.

Conclusions

We have used chemical knowledge and interdisciplinary tools to decipher the molecular mechanisms by which gut bacteria interfere with the treatment of Parkinson's disease. The decarboxylation of L-dopa by *E. faecalis* mirrors host drug metabolism and, together with human AADC, likely limits drug availability and contributes to interindividual variation in efficacy. Together with recent work dissecting host and gut microbial contributions to the antiviral drug brivudine (*57*), our findings show that gut bacterial metabolism need not be chemically distinct from host activities to alter drug efficacy and suggest that such interactions may be underappreciated. Moreover, carbidopa's failure to prevent L-dopa decarboxylation by *E. faecalis* implies that additional host-targeted drugs may lack efficacy toward activities also present in the gut microbiota. Although a recent, independent study also characterized *E. faecalis* TyrDC's role in L-dopa decarboxylation and its lack of susceptibility to carbidopa (*38*), it did not show that this activity occurs in human gut microbiotas or identify strategies for inhibiting the bacterial enzyme. By contrast, we demonstrate that TyrDC predicts drug metabolism in Parkinson's patient microbiotas and use an understanding of its substrate specificity to identify a small molecule that prevents L-dopa decarboxylation in patient samples and increases L-dopa bioavailability in vivo. Through discovery of predictive biomarkers for L-dopa metabolism and identification of an inhibitor of this activity, this work will enable efforts to elucidate the contribution of the gut microbiota to drug availability, patient drug response, and treatment outcomes.

We also show that *E. lenta* further metabolizes the dopamine produced by L-dopa decarboxylation using a distinctly microbial reaction, catechol dehydroxylation. It is possible that this transformation influences the multiple side effects of L-dopa administration linked to dopamine production. This discovery also raises questions about the biological consequences of gut microbial metabolism of endogenous dopamine, which is present in the gastrointestinal tract and has been linked to phenotypes ranging from gut motility to pathogen colonization (*58*–*60*). The biological activity of the gut microbial metabolite *m*-tyramine in the host and the benefits of this metabolism for *E. lenta* are also poorly understood. Our findings will enable further study of these phenomena. Given that gut microbes dehydroxylate catechol groups found in numerous aromatic drugs and dietary compounds (*18*, *61*–*63*), the discovery of Dadh will enable identification of additional catechol dehydroxylases and help to elucidate the biological role of this enigmatic transformation. Uncovering the unexpected effect of SNPs on gut microbial dopamine metabolism suggests that simply detecting functional genes may not accurately predict the activities encoded by the human gut microbiome and underscores the importance of studying enzymes from this community.

Materials and methods summary

Our methods for the identification and biochemical characterization of *E. faecalis* TyrDC; characterization of anaerobic L-dopa metabolism by *E. faecalis* and gut microbiota samples; enrichment culturing for dopamine dehydroxylating organisms; RNA-sequencing; culture-based assays; purification of Dadh; assays of anaerobic dopamine metabolism by Actinobacteria and complex gut microbiota samples; PCR and qPCR assays; liquid chromatography–MS (LC-MS) methods; and assays for evaluating inhibitors in vitro, ex vivo, and in vivo are provided in the supplementary materials. Additional information about our protocols, including references to the supplementary materials, can be found throughout the main text.

REFERENCES AND NOTES

1. G. Sharon, T. R. Sampson, D. H. Geschwind, S. K. Mazmanian, The central nervous system and the gut microbiome. *Cell* **167**, 915–932 (2016). doi: 10.1016/j.cell.2016.10.027; pmid: 27814521
2. T. R. Sampson *et al*., Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. *Cell* **167**, 1469–1480.e12 (2016). doi: 10.1016/j.cell.2016.11.018; pmid: 27912057
3. F. Scheperjans *et al*., Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov. Disord.* **30**, 350–358 (2015). doi: 10.1002/mds.26069; pmid: 25476529
4. M. M. Unger *et al*., Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism Relat. Disord.* **32**, 66–72 (2016). doi: 10.1016/j.parkreldis.2016.08.019; pmid: 27591074
5. J. R. Bedarf *et al*., Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Med.* **9**, 39 (2017). doi: 10.1186/s13073-017-0428-y; pmid: 28449715
6. E. M. Hill-Burns *et al*., Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov. Disord.* **32**, 739–749 (2017). doi: 10.1002/mds.26942; pmid: 28195358
7. T. Minato *et al*., Progression of Parkinson's disease is associated with gut dysbiosis: Two-year follow-up study. *PLOS ONE* **12**, e0187307 (2017). doi: 10.1371/journal.pone.0187307; pmid: 29091972
8. V. A. Petrov *et al*., Analysis of gut microbiota in patients with Parkinson's disease. *Bull. Exp. Biol. Med.* **162**, 734–737 (2017). doi: 10.1007/s10517-017-3700-7; pmid: 28429209
9. A. Heintz-Buschart *et al*., The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Mov. Disord.* **33**, 88–98 (2018). doi: 10.1002/mds.27105; pmid: 28843021
10. O. Hornykiewicz, L-DOPA. *J. Parkinsons Dis.* **7** (s1), S3–S10 (2017). doi: 10.3233/JPD-179004; pmid: 28282813
11. P. S. Papavasiliou *et al*., Levodopa in Parkinsonism: Potentiation of central effects with a peripheral inhibitor. *N. Engl. J. Med.* **286**, 8–14 (1972). doi: 10.1056/NEJM197201062860102; pmid: 4550085
12. K. F. Gey, A. Pletscher, Distribution and metabolism of DL-3,4-dihydroxy[2-14C]-phenylalanine in rat tissues. *Biochem. J.* **92**, 300–308 (1964). doi: 10.1042/bj0920300; pmid: 5838073
13. S. Bergmann *et al*., The absorption and metabolism of a standard oral dose of levodopa in patients with Parkinsonism. *Br. J. Clin. Pharmacol.* **1**, 417–424 (1974). doi: 10.1111/j.1365-2125.1974.tb00280.x; pmid: 22454921
14. M. M. Goldenberg, Medical management of Parkinson's disease. *P&T* **33**, 590–606 (2008). pmid: 19750042
15. A. C. Whitfield, B. T. Moore, R. N. Daniels, Classics in chemical neuroscience: Levodopa. *ACS Chem. Neurosci.* **5**, 1192–1197 (2014). doi: 10.1021/cn5001759; pmid: 25270271
16. SINEMET CR, (carbidopa-levodopa) [package insert]. Princeton: Bristol-Myers Squibb Company, NJ; 2008.
17. R. Gray *et al*., Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): A large, open-label, pragmatic randomised trial. *Lancet* **384**, 1196–1205 (2014). doi: 10.1016/S0140-6736(14)60683-8; pmid: 24928805
18. P. Spanogiannopoulos, E. N. Bess, R. N. Carmody, P. J. Turnbaugh, The microbial pharmacists within us: A metagenomic view of

xenobiotic metabolism. *Nat. Rev. Microbiol.* **14**, 273–287 (2016). doi: 10.1038/nrmicro.2016.17; pmid: 26972811
19. H. Hashim *et al*., Eradication of *Helicobacter pylori* infection improves levodopa action, clinical symptoms and quality of life in patients with Parkinson's disease. *PLOS ONE* **9**, e112330 (2014). doi: 10.1371/journal.pone.0112330; pmid: 25411976
20. A. Fasano *et al*., The role of small intestinal bacterial overgrowth in Parkinson's disease. *Mov. Disord.* **28**, 1241–1249 (2013). doi: 10.1002/mds.25522; pmid: 23712625
21. F. Deeds, A. N. Booth, F. T. Jones, Methylation and dehydroxylation of phenolic compounds by rats and rabbits. *J. Biol. Chem.* **225**, 615–621 (1957). pmid: 13416264
22. B. R. Goldin, M. A. Peppercorn, P. Goldman, Contributions of host and intestinal microflora in the metabolism of L-dopa by the rat. *J. Pharmacol. Exp. Ther.* **186**, 160–166 (1973). pmid: 4723308
23. M. Sandler, B. L. Goodwin, C. R. Ruthven, D. B. Calne, Therapeutic implications in Parkinsonism of *m*-tyramine formation from L-dopa in man. *Nature* **229**, 414–415 (1971). doi: 10.1038/229414a0; pmid: 4926994
24. M. Sandler, C. R. Ruthven, B. L. Goodwin, K. R. Hunter, G. M. Stern, Variation of levodopa metabolism with gastrointestinal absorption site. *Lancet* **1**, 238–240 (1974). doi: 10.1016/S0140-6736(74)92547-1; pmid: 4130247
25. P. Goldman, M. A. Peppercorn, B. R. Goldin, Metabolism of drugs by microorganisms in the intestine. *Am. J. Clin. Nutr.* **27**, 1348–1355 (1974). doi: 10.1093/ajcn/27.11.1348; pmid: 4155602
26. B. L. Goodwin, C. R. J. Ruthven, G. S. King, M. Sandler, B. G. S. Leask, Metabolism of 3, 4-dihydroxyphenylalanine, its metabolites and analogues in vivo in the rat: Urinary excretion pattern. *Xenobiotica* **8**, 629–651 (1978). doi: 10.3109/00498257809069575; pmid: 716472
27. A. C. Eliot, J. F. Kirsch, Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **73**, 383–415 (2004). doi: 10.1146/annurev.biochem.73.011303.074021; pmid: 15189147
28. H. Zhu *et al*., Crystal structure of tyrosine decarboxylase and identification of key residues involved in conformational swing and substrate binding. *Sci. Rep.* **6**, 27779 (2016). doi: 10.1038/srep27779; pmid: 27292129
29. V. Ladero *et al*., Is the production of the biogenic amines tyramine and putrescine a species-level trait in enterococci? *Food Microbiol.* **30**, 132–138 (2012). doi: 10.1016/j.fm.2011.12.016; pmid: 22265293
30. F. Lebreton *et al*., Tracing the Enterococci from Paleozoic origins to the hospital. *Cell* **169**, 849–861.e13 (2017). doi: 10.1016/j.cell.2017.04.027; pmid: 28502769
31. H. Blashcko, Substrate specificity of amino acid decarboxylases. *Biochim. Biophys. Acta* **4**, 130–137 (1950). doi: 10.1016/0006-3002(50)90016-3; pmid: 15403918
32. N. Connil *et al*., Identification of the *Enterococcus faecalis* tyrosine decarboxylase operon involved in tyramine production. *Appl. Environ. Microbiol.* **68**, 3537–3544 (2002). doi: 10.1128/AEM.68.7.3537-3544.2002; pmid: 12089039
33. J. Kollonitsch *et al*., Selective inhibitors of biosynthesis of aminergic neurotransmitters. *Nature* **274**, 906–908 (1978). doi: 10.1038/274906a0; pmid: 683331
34. T. Børresen, N. K. Klausen, L. M. Larsen, H. Sørensen, Purification and characterisation of tyrosine decarboxylase and aromatic-L-amino-acid decarboxylase. *Biochim. Biophys. Acta* **993**, 108–115 (1989). doi: 10.1016/0304-4165(89)90149-9; pmid: 2508758
35. F. Liu *et al*., Heterologous expression and characterization of tyrosine decarboxylase from Enterococcus faecalis R612Z1 and Enterococcus faecium R615Z1. *J. Food Prot.* **77**, 592–598 (2014). doi: 10.4315/0362-028X.JFP-13-326; pmid: 24680070
36. E. Pessione *et al*., First evidence of a membrane-bound, tyramine and beta-phenylethylamine producing, tyrosine decarboxylase in *Enterococcus faecalis*: A two-dimensional electrophoresis proteomic study. *Proteomics* **9**, 2695–2710 (2009). doi: 10.1002/pmic.200800780; pmid: 19405032
37. S. Allenmark, B. Servenius, Characterization of bacterial L-(-)-tyrosine decarboxylase by isoelectric focusing and gel chromatography. *J. Chromatogr.* **153**, 238–245 (1978). doi: 10.1016/S0021-9673(00)89878-7; pmid: 96122
38. S. P. van Kessel *et al*., Gut bacterial tyrosine decarboxylases restrict levels of levodopa in the treatment of Parkinson's disease. *Nat. Commun.* **10**, 310 (2019). doi: 10.1038/s41467-019-08294-y; pmid: 30659181
39. W. B. Abrams, C. B. Coutinho, A. S. Leon, H. E. Spiegel, Absorption and metabolism of levodopa. *JAMA* **218**, 1912–1914 (1971). doi: 10.1001/jama.1971.03190260028007; pmid: 5171067
40. S. A. Adibi, D. W. Mercer, Protein digestion in human intestine as reflected in luminal, mucosal, and plasma amino acid concentrations after meals. *J. Clin. Invest.* **52**, 1586–1594 (1973). doi: 10.1172/JCI107335; pmid: 4718954
41. M. Perez *et al*., Tyramine biosynthesis is transcriptionally induced at low pH and improves the fitness of *Enterococcus faecalis* in acidic environments. *Appl. Microbiol. Biotechnol.* **99**, 3547–3558 (2015). doi: 10.1007/s00253-014-6301-7; pmid: 25529314
42. J. Fallingborg, Intraluminal pH of the human gastrointestinal tract. *Dan. Med. Bull.* **46**, 183–196 (1999). pmid: 10421978
43. C. Holliger, G. Wohlfarth, G. Diekert, Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiol. Rev.* **22**, 383–398 (1998). doi: 10.1111/j.1574-6976.1998.tb00377.x
44. N. Koppel, J. E. Bisanz, M. E. Pandelia, P. J. Turnbaugh, E. P. Balskus, Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *eLife* **7**, e33953 (2018). doi: 10.7554/eLife.33953; pmid: 29761785
45. H. J. Haiser *et al*., Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* **341**, 295–298 (2013). doi: 10.1126/science.1235872; pmid: 23869020
46. M. Boll, B. Schink, A. Messerschmidt, P. M. H. Kroneck, Novel bacterial molybdenum and tungsten enzymes: Three-dimensional structure, spectroscopy, and reaction mechanism. *Biol. Chem.* **386**, 999–1006 (2005). doi: 10.1515/BC.2005.116; pmid: 16218872
47. R. A. Rothery, J. H. Weiner, Shifting the metallocentric molybdoenzyme paradigm: The importance of pyranopterin coordination. *J. Biol. Inorg. Chem.* **20**, 349–372 (2015). doi: 10.1007/s00775-014-1194-6; pmid: 25267303
48. W. Zhu *et al*., Precision editing of the gut microbiota ameliorates colitis. *Nature* **553**, 208–211 (2018). doi: 10.1038/nature25172; pmid: 29323293
49. J. E. Bisanz *et al*., Illuminating the microbiome's dark matter: A functional genomic toolkit for the study of human gut Actinobacteria. *bioRxiv* 10.1101/304840 (2018); accessed 18 April 2019]. doi: 10.1101/304840
50. A. Martínez-del Campo *et al*., Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *mBio* **6**, e00042–e15 (2015). doi: 10.1128/mBio.00042-15; pmid: 25873372
51. S. Torriani *et al*., Rapid detection and quantification of tyrosine decarboxylase gene (tdc) and its expression in gram-positive bacteria associated with fermented foods using PCR-based methods. *J. Food Prot.* **71**, 93–101 (2008). doi: 10.4315/0362-028X-71.1.93; pmid: 18236668
52. S. C. Preissner *et al*., Polymorphic cytochrome P450 enzymes (CYPs) and their role in personalized therapy. *PLOS ONE* **8**, e82562 (2013). doi: 10.1371/journal.pone.0082562; pmid: 24340040
53. J. Collins *et al*., Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* **553**, 291–294 (2018). doi: 10.1038/nature25178; pmid: 29310122
54. P. Burkhard, P. Dominici, C. Borri-Voltattorni, J. N. Jansonius, V. N. Malashkevich, Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase. *Nat. Struct. Biol.* **8**, 963–967 (2001). doi: 10.1038/nsb1101-963; pmid: 11685243
55. M. Sandler, R. D. Johnson, C. R. J. Ruthven, J. L. Reid, D. B. Calne, Transamination is a major pathway of L-dopa metabolism following peripheral decarboxylase inhibition. *Nature* **247**, 364–366 (1974). doi: 10.1038/247364b0; pmid: 4817854
56. B. D. Wallace *et al*., Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010). doi: 10.1126/science.1191175; pmid: 21051639
57. M. Zimmermann, M. Zimmermann-Kogadeeva, R. Wegmann, A. L. Goodman, Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* **363**, eaat9931 (2019). doi: 10.1126/science.aat9931; pmid: 30733391
58. G. Eisenhofer *et al*., Substantial production of dopamine in the human gastrointestinal tract. *J. Clin. Endocrinol. Metab.* **82**, 3864–3871 (1997). doi: 10.1210/jcem.82.11.4339; pmid: 9360553
59. Z. S. Li, C. Schmauss, A. Cuenca, E. Ratcliffe, M. D. Gershon, Physiological modulation of intestinal motility by enteric dopaminergic neurons and the D2 receptor: Analysis of dopamine receptor expression, location, development, and function in wild-type and knock-out mice. *J. Neurosci.* **26**, 2798–2807 (2006). doi: 10.1523/JNEUROSCI.4720-05.2006; pmid: 16525059
60. S. Dichtl *et al*., Dopamine is a siderophore-like iron chelator that promotes *Salmonella enterica* serovar typhimurium virulence in mice. *mBio* **10**, e02624–e18 (2019). doi: 10.1128/mBio.02624-18; pmid: 30723125
61. R. R. Scheline, R. T. Williams, J. G. Wit, Biological dehydroxylation. *Nature* **188**, 849–850 (1960). doi: 10.1038/188849a0; pmid: 13747445
62. D. J. Sweeny *et al*., Metabolism of fostamatinib, the oral methylene phosphate prodrug of the spleen tyrosine kinase inhibitor R406 in humans: Contribution of hepatic and gut bacterial processes to the overall biotransformation. *Drug Metab. Dispos.* **38**, 1166–1176 (2010). doi: 10.1124/dmd.110.032151; pmid: 20371637
63. N. Koppel, V. Maini Rekdal, E. P. Balskus, Chemical transformation of xenobiotics by the human gut microbiota. *Science* **356**, eaag2770 (2017). doi: 10.1126/science.aag2770; pmid: 28642381

ACKNOWLEDGMENTS

We acknowledge the Broad Institute Microbial Omics Core (MOC) for assistance with RNA and 16*S* rRNA gene sequencing analysis and experimental design, the Harvard Bauer Core Proteomics facility for assistance with proteomics, M. Wilson (Harvard University) for helpful discussions and input, M. Gilmore and E. Selleck (Massachusetts Eye and Ear, Harvard Medical School, and Broad Institute) for supplying the *E. faecalis tyrDC* mutant and helpful discussions, F. Lebreton (Massachusetts Eye and Ear and Harvard Medical School) for supplying *E. faecalis* and *E. faecium* strains, R. Nayak [University of California, San Francisco (UCSF)] and M. Krueger (UCSF) for help with sample collection from healthy control subjects, and the Biocollective for collection and provision of stool samples from Parkinson's patients. We thank Merck for the gift of AFMT. **Funding:** This work was supported by the Packard Fellowship for Science and Engineering (2013-39267) (E.P.B), the Howard Hughes Medical Institute (HHMI)–Gates Faculty Scholars Program (OPP1158186) (E.P.B), the National Institutes of Health (R01HL122593) (P.J.T.), the Searle Scholars Program (SSP-2016-1352) (P.J.T), the UCSF-Stanford Arthritis Center of Excellence (supported in part by the Arthritis Foundation) (P.J.T), and the Rheumatology Research Foundation (P.J.T.). V.M.R. is the recipient of a National Science Foundation Graduate Research Fellowship, a Gilliam Fellowship from HHMI, and the Ardis and Robert James Graduate Research Fellowship from Harvard University and acknowledges support from a National Institutes of Health Training Grant (5T32GM007598-38). E.N.B. is a Howard Hughes Medical Institute fellow of the Life Sciences Research Foundation. J.E.B. has fellowship support from the Natural Sciences and Engineering Research Council of Canada. P.J.T. is a Chan Zuckerberg Biohub investigator and a Nadia's Gift Foundation Innovator supported, in part, by the Damon Runyon Cancer Research Foundation (DRR-42-16). **Author contributions:** V.M.R. and E.P.B. conceived of the project. V.M.R. purified and characterized all proteins biochemically with and without inhibitors, performed assays for L-dopa and dopamine metabolism in pure cultures and complex microbiotas in the presence and absence of inhibitors, performed qPCR experiments and analysis, and performed RNA-sequencing experiments in *E. lenta* A2 and *E. lenta* 28B. E.N.B. performed RNA-sequencing experiments in *E. lenta* DSM2243, performed assays for dopamine metabolism across the Actinobacterial library, and contributed to the design of the ex vivo experiments and other culture-based assays. J.E.B. performed RNA sequencing analysis, comparative genomics, and metagenomic analysis and contributed to the design of and conducted in vivo AFMT experiments. V.M.R., E.P.B., E.N.B., J.E.B., and P.J.T. provided critical feedback on experiments. V.M.R., J.E.B., P.J.T., and E.P.B. wrote the manuscript. **Competing interests:** E.P.B. has consulted for Merck, Novartis, and Kintai Therapeutics; is on the Scientific Advisory Boards of Kintai Therapeutics and Caribou Biosciences; and is an Associate Member of the Broad Institute of Harvard and MIT. P.J.T. is on the scientific advisory board for Kaleido, Seres, uBiome, and WholeBiome. **Data and materials availability:** The *E. lenta* A2 genome has been deposited into GenBank (PRJNA412637). RNA-sequencing data has been deposited into the Sequence Read Archive available by way of BioProject PRJNA507796. The small-molecule AFMT was obtained under a materials transfer agreement with Merck.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/eaau6323/suppl/DC1
Materials and Methods
Figs. S1 to S36
Tables S1 to S8
References (*64*–*84*)
Data File S1

29 June 2018; resubmitted 18 April 2019
Accepted 2 May 2019
10.1126/science.aau6323

RESEARCH ARTICLE SUMMARY

SYMBIOSIS

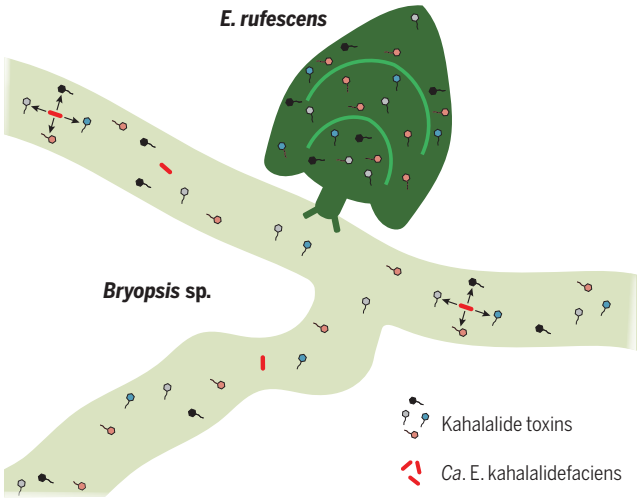
A microbial factory for defensive kahalalides in a tripartite marine symbiosis

Jindong Zan*, Zhiyuan Li*, Ma. Diarey Tianero, Jeanette Davis, Russell T. Hill, Mohamed S. Donia†

INTRODUCTION: Chemical defense strategies, in which organisms use toxic molecules for protection against pathogens or predators, are widespread in the marine environment. In some cases, the same defensive molecules are shared by taxonomically distant organisms, raising questions about their molecular origin. The actual source of these molecules may be the organism itself, as observed in marine algae; a microbial symbiont, as commonly seen in marine sponges and tunicates; or diet, as in several marine mollusks. Elucidating the molecular basis of toxin production in chemically defended organisms is important for a complete understanding of their ecological interactions.

RATIONALE: In this work, we studied toxin production in the Hawaiian marine alga *Bryopsis* sp. and its predator, the mollusk *Elysia rufescens*. Both organisms are chemically defended against predators by a diverse library of lipopeptide toxins, the kahalalides, but the details of kahalalide production and diversification are unknown (see the figure). One of these molecules, kahalalide F, is a potent cytotoxin and has been evaluated clinically as an anticancer agent. The molecular structures of the kahalalides show several features of microbial biosynthesis: They are fatty acid–cyclic peptide hybrids with several D- and non-proteinogenic amino acids, thus motivating us to hypothesize that the kahalalides are produced by a bacterial or fungal symbiont of *Bryopsis* sp. or of both *Bryopsis* sp. and *E. rufescens*. We combined metagenomic, metatranscriptomic, and chemical analyses with microbial cultivation, fluorescence microscopy, and evolutionary genomics to determine the molecular bases of kahalalide production and evolution in this tripartite marine symbiosis.

RESULTS: Using metagenomic analyses, we discovered a bacterium—termed "*Candidatus* Endobryopsis kahalalidefaciens"—that has no cultured close relatives, and we show that it lives in symbiosis with the alga *Bryopsis* sp. (see the figure). Using fluorescence microscopy, bacterial cultivation, and comparative genomics, we show that "*Ca.* E. kahalalidefaciens" is an intracellular, obligate, genome-reduced symbiont that has lost essential functions for free living (e.g., amino acid biosynthesis). Despite this reduced metabolic capacity, 20% of the "*Ca.* E. kahalalidefaciens" genome encodes a diverse set of 20 nonribosomal peptide synthetase pathways. We link nine of these pathways to nine structurally diverse kahalalides (including kahalalide F, the main defensive toxin of both *Bryopsis* sp. and *E. rufescens*), which we then chemically identify in the same sample of *Bryopsis* sp.



Chemical defense in a tripartite marine symbiosis. The bacterial symbiont "*Ca.* E. kahalalidefaciens," which lives intracellularly in the marine alga *Bryopsis* sp., produces a diverse library of toxins (the kahalalides) that protect the host from predation. The mollusk *E. rufescens* sequesters the same toxins from its algal diet and employs them for its own defense.

None of the amino acid substrates that make up the kahalalides can be produced by "*Ca.* E. kahalalidefaciens" itself; therefore, these substances are mostly provided by the autotrophic *Bryopsis* sp., highlighting an unusual strategy of collaborative biosynthesis between a symbiotic bacterium and its host. Moreover, using metagenomic analysis and fluorescence microscopy, we show that "*Ca.* E. kahalalidefaciens" is not a symbiont of the mollusk *E. rufescens*, establishing chemical sequestration as the means by which this animal indirectly acquires bacterially produced kahalalides from its algal diet.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw6732

Detailed analysis of the "*Ca.* E. kahalalidefaciens" genome reveals a high level of plasticity and a distinctive model of diversifying evolution that is independent of horizontal gene transfer, consistent with the intracellular lifestyle of this symbiont. In this model, new nonribosomal peptide synthetase pathways arise through duplication and divergence, accompanied by extensive interpathway recombination events. Finally, metatranscriptomic analysis reveals that 26% of the transcriptional activity of "*Ca.* E. kahalalidefaciens" is dedicated to kahalalide biosynthesis and that biosynthetic pathways for different kahalalides vary widely in their expression levels, further emphasizing the importance of these molecules in maintaining a successful symbiosis.

CONCLUSION: In a chemically defended, tripartite marine symbiosis, we show that an obligate bacterial symbiont of a marine alga produces a library of defensive molecules that protect the host from predation, and that the same molecules are in turn hijacked by a predatory mollusk and used for its own defense. Living intracellularly in algal cells, the symbiont acts as a microbial factory for the biosynthesis of complex defensive molecules from simple host-derived substrates. Symbiont-derived production of defensive molecules in marine algae and indirect acquisition of microbial products by predatory mollusks may thus be important yet rarely studied phenomena in marine ecosystems. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: donia@princeton.edu
Cite this article as J. Zan *et al.*, *Science* **364**, eaaw6732 (2019). DOI: 10.1126/science.aaw6732

RESEARCH ARTICLE

SYMBIOSIS

A microbial factory for defensive kahalalides in a tripartite marine symbiosis

Jindong Zan[1]*, Zhiyuan Li[2,3]*, Ma. Diarey Tianero[1], Jeanette Davis[4]†, Russell T. Hill[4], Mohamed S. Donia[1]‡

Chemical defense against predators is widespread in natural ecosystems. Occasionally, taxonomically distant organisms share the same defense chemical. Here, we describe an unusual tripartite marine symbiosis, in which an intracellular bacterial symbiont ("*Candidatus* Endobryopsis kahalalidefaciens") uses a diverse array of biosynthetic enzymes to convert simple substrates into a library of complex molecules (the kahalalides) for chemical defense of the host, the alga *Bryopsis* sp., against predation. The kahalalides are subsequently hijacked by a third partner, the herbivorous mollusk *Elysia rufescens*, and employed similarly for defense. "*Ca*. E. kahalalidefaciens" has lost many essential traits for free living and acts as a factory for kahalalide production. This interaction between a bacterium, an alga, and an animal highlights the importance of chemical defense in the evolution of complex symbioses.

One of the most efficient defense strategies in the ocean is chemical defense, in which organisms accumulate toxic small molecules that kill and/or deter predators (*1*). Because of the structural complexity of these defensive chemicals and their similarity to bacterial products, many have been proposed to originate from bacterial symbionts (*2*, *3*). Indeed, defensive molecules of substantial toxicity have been genetically linked to specific bacterial symbionts of several marine animals (*4*). The concept of chemical defense can also extend beyond a single animal, with the same chemicals benefiting a network of partners. This phenomenon is common in marine mollusks that prey on chemically defended organisms and sequester the organisms' toxic chemicals for their own defense (*5*). Although these two phenomena—production of defensive chemicals by bacterial symbionts and sequestration of defensive chemicals by predatory mollusks—have been separately described in marine systems, their integration in multipartite symbioses is rarely studied.

In this work, we investigated the marine alga *Bryopsis* sp. and its predator mollusk *Elysia rufescens* (Fig. 1). *Bryopsis* sp. harbors a toxic lipopetide molecule, kahalalide F (KF) (*6*), which was shown to be responsible for chemical defense against predators (*7*). Unlike other predators that are repelled by this toxin, *E. rufescens* not only feeds on KF-containing *Bryopsis* sp., but it also accumulates KF in its body at several times the concentration in *Bryopsis* sp. and employs KF for its own defense (*7*–*9*). Similarly to other species of *Elysia*, *E. rufescens* additionally maintains photosynthetically active algal chloroplasts in its digestive organs for several months, in a phenomenon known as kleptoplasty (*10*). Although KF has been isolated only from *Bryopsis* sp. and *E. rufescens*, it contains structural features that suggest a possible microbial origin, namely nonproteinogenic amino acids (e.g., ornithine, dehydrobutyrine, and several D-amino acids) and a fatty acid moiety (Fig. 1). These features motivated us to hypothesize that KF is produced by a cryptic third partner, such as a bacterial or fungal symbiont. To test this hypothesis and study the molecular and evolutionary details of toxin production in this unusual system, we employed a multidisciplinary approach that combines metagenomic, metatranscriptomic, and chemical analyses; microbial cultivation; fluorescence microscopy; and evolutionary genomics.

KF production by a bacterial symbiont

We collected fresh *Bryopsis* sp. specimens from an algal bloom at the exact location where KF was originally reported (Black Point Bay, Honolulu, Hawaii; sample designated *Bryopsis*-2015) (Fig. 1A) and analyzed them using a variety of chemical and molecular techniques. First, we chemically extracted *Bryopsis*-2015 as previously described (*6*), analyzed its resulting organic extract using high-performance liquid chromatography coupled with high-resolution tandem mass spectrometry (HPLC–HR-MS/MS), and confirmed that it harbors KF (fig. S1 and table S1). Second, to investigate the microbial community associated with *Bryopsis*-2015, we used high-throughput 16*S* ribosomal RNA (rRNA) gene amplicon sequencing (104,000 reads, on average, V4 region) on DNA isolated from multiple replicates of the same algal collection. Surprisingly, only three bacterial classes dominate the *Bryopsis*-2015 microbiome—Gammaproteobacteria (11.8% of the reads, on average), Alphaproteobacteria (9.3%), and Flavobacteriia (11.8%)—whereas most of the remaining reads (57.0%) map to the chloroplast-encoded 16*S* rRNA gene in *Bryopsis* sp. (Fig. 1C). A single sequence accounts for 74.1 to 92.0% of the fraction assigned to the class Flavobacteriia, making it the single most dominant species in *Bryopsis*-2015. Notably, four biological replicates collected on two consecutive days showed consistent microbiome profiles. These results reveal a relatively simple and uniform microbiome composition for KF-containing *Bryopsis* sp.

Owing to the cyclic lipopeptide nature of KF (Fig. 1D), this compound could be synthesized either by a nonribosomal peptide synthetase pathway (NRPS) or by a ribosomally synthesized and posttranslationally modified peptide pathway (RiPP) (*11*, *12*). In both cases, it is relatively straightforward to computationally link a biosynthetic gene cluster to its product. In NRPSs, the number of modules, the substrate specificity of the adenylation (A) domains, and the presence or absence of epimerization (E) domains are distinctive for a particular molecule. In RiPPs, the primary amino acid sequence of the final molecule is directly encoded on a precursor peptide within the biosynthetic gene cluster. To examine whether bacterial or fungal members of the *Bryopsis* sp. microbiome encode biosynthetic gene clusters consistent with the structure of KF, we deeply sequenced the *Bryopsis*-2015 metagenomic DNA using Illumina [48 million paired-end reads, 175 base pairs (bps)], performed multiple iterations of assemblies on the produced data, and queried the reads and resulting assemblies for possible biosynthetic gene clusters using several search strategies (see materials and methods).

First, for RiPP pathways, we searched for the expected primary amino acid sequence of KF (VTVVPRITIVFTV, where R replaces ornithine and T replaces dehydrobutyrine) against a database of algal metagenomic reads and contigs using the tBLASTn mode of the Basic Local Alignment Search Tool. Second, for NRPS pathways, we constructed a database of 36 phylogenetically diverse NRPS proteins and used it as a query for tBLASTn searches against the *Bryopsis*-2015 metagenomic assembly (18,308 scaffolds with length >5 kbps). Third, we subjected the same scaffolds to antiSMASH analysis, a stand-alone tool for the unbiased identification of small-molecule

[1]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. [2]Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA. [3]Center for the Physics of Biological Function, Princeton University, Princeton, NJ 08544, USA. [4]Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD 21202, USA.
*These authors contributed equally to this work. †Present address: National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910, USA.
‡Corresponding author. Email: donia@princeton.edu

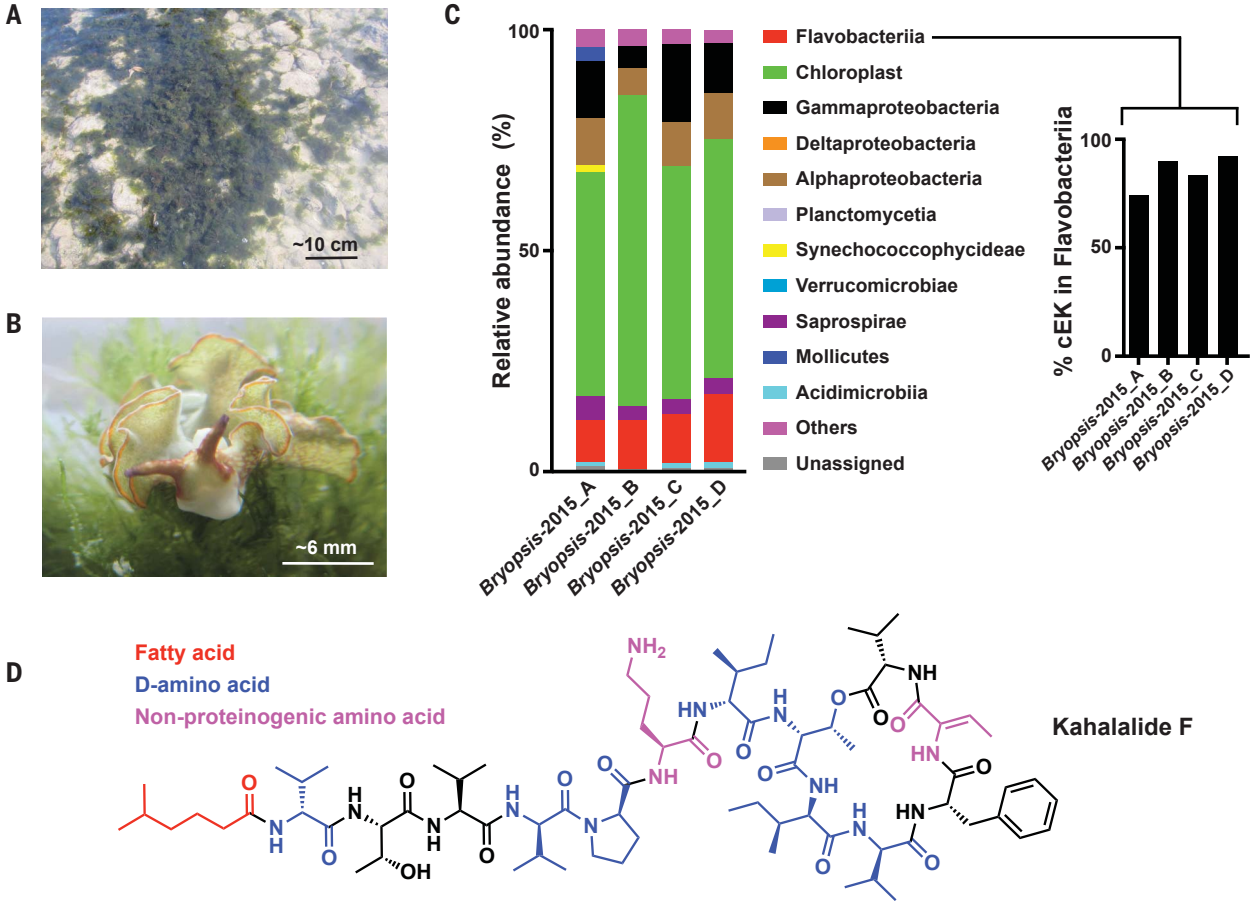


Fig. 1. Bacterial composition and chemical defense of the *Bryopsis*–*Elysia* symbiotic system. (**A**) The natural *Bryopsis* sp. bloom studied in this work. (**B**) *E. rufescens* feeding on *Bryopsis* sp. in the laboratory. (**C**) Composition of *Bryopsis*-2015–associated bacterial communities across four different replicates collected on two consecutive days (A and B: 29 March 2015; C and D: 30 March 2015) and grouped at the class level. Rare taxa (<1% of total sequences) are labeled as "Others." The inset bar graph represents the percentage of a single 16*S* rDNA sequence (hereafter designated "cEK") that dominates the class Flavobacteriia. (**D**) Molecular structure of kahalalide F (KF), the main defensive chemical in both *Bryopsis* sp. and *E. rufescens*. Structural features commonly associated with microbial biosynthesis are depicted in red, blue, and pink.

biosynthetic gene clusters (*13*, *14*). Although no matches to the primary KF sequence were retrieved using the RiPP search strategy, 130 scaffolds (5 to 394 kbps in length) that encode NRPS pathways were discovered using the NRPS search strategy. In addition, antiSMASH annotated a total of 25 scaffolds as NRPSs, only three of which were not detected with the BLAST strategy. Unfortunately, none of the 133 scaffolds containing possible NRPSs constituted a 13-domain pathway, as would be expected for KF, implying that the presumed KF biosynthetic gene cluster may be fragmented or unclustered. Careful analysis of the multimodular-NRPS–containing scaffolds revealed 33 scaffolds that have similar average GC content (~50%) and similar coverage in the metagenome (>300X), suggesting that they may be part of one large NRPS pathway or at least part of the same bacterial genome (fig. S2).

To obtain a better assembly for the fragmented NRPS scaffolds, we subjected *Bryopsis*-2015 metagenomic DNA to long-read, single-molecule real-time (SMRT) sequencing using the Pacific Biosciences platform (three SMRT cells, 325,000 reads, median insert size of 9 kbps). After assembly and error correction with the high-quality Illumina reads, we successfully closed a ~2.3-Mbp bacterial genome in which several of the partial NRPS sequences from the Illumina assembly were joined to produce a 55.6-kbp NRPS pathway (NRPS-8) that is consistent with the final KF structure (Fig. 2). Several lines of evidence link NRPS-8 to KF biosynthesis: (i) It begins with a starting condensation domain, as predicted by multiple biosynthetic analysis algorithms (NaPDoS and antiSMASH); these types of domains catalyze an amide bond formation between a fatty acid and the first amino acid of the growing peptide chain (*13*–*15*). Indeed, the KF structure starts with 5-methyl hexanoic acid conjugated to the first valine residue. (ii) It consists of 13 NRPS modules, as expected for the tridecapeptide KF. (iii) Modules 1, 4, 5, 7, 8, 9, 10, and 12 of NRPS-8 contain epimerization domains, which are responsible for converting the stereochemistry of the loaded amino acid from the L form to the D form (*11*). In agreement with this, KF harbors 7 D-amino acids at positions 1, 4, 5, 7, 8, 9, and 10, which correspond exactly to these epimerizing modules (threonine residue at position 12 is dehydrated to dehydrobutyrine, which is achiral). (iv) The predicted substrate specificity of almost all adenylation domains in NRPS-8 matches the amino acid residues observed in KF, in an analysis performed using published algorithms and further supported by a detailed phylogenetic analysis (see materials and methods and figs. S3 and S4) (*16*). Notably, at 40 kbps downstream of NRPS-8, the same genome encodes a 16*S* rRNA gene that is identical to the aforementioned, most abundant bacterium in the sample. A 16*S* rDNA-based phylogenetic tree of related organisms placed this bacterium as a previously unknown genus in the family Flavobacteriaceae, order Flavobacteriales, class Flavobacteriia, and

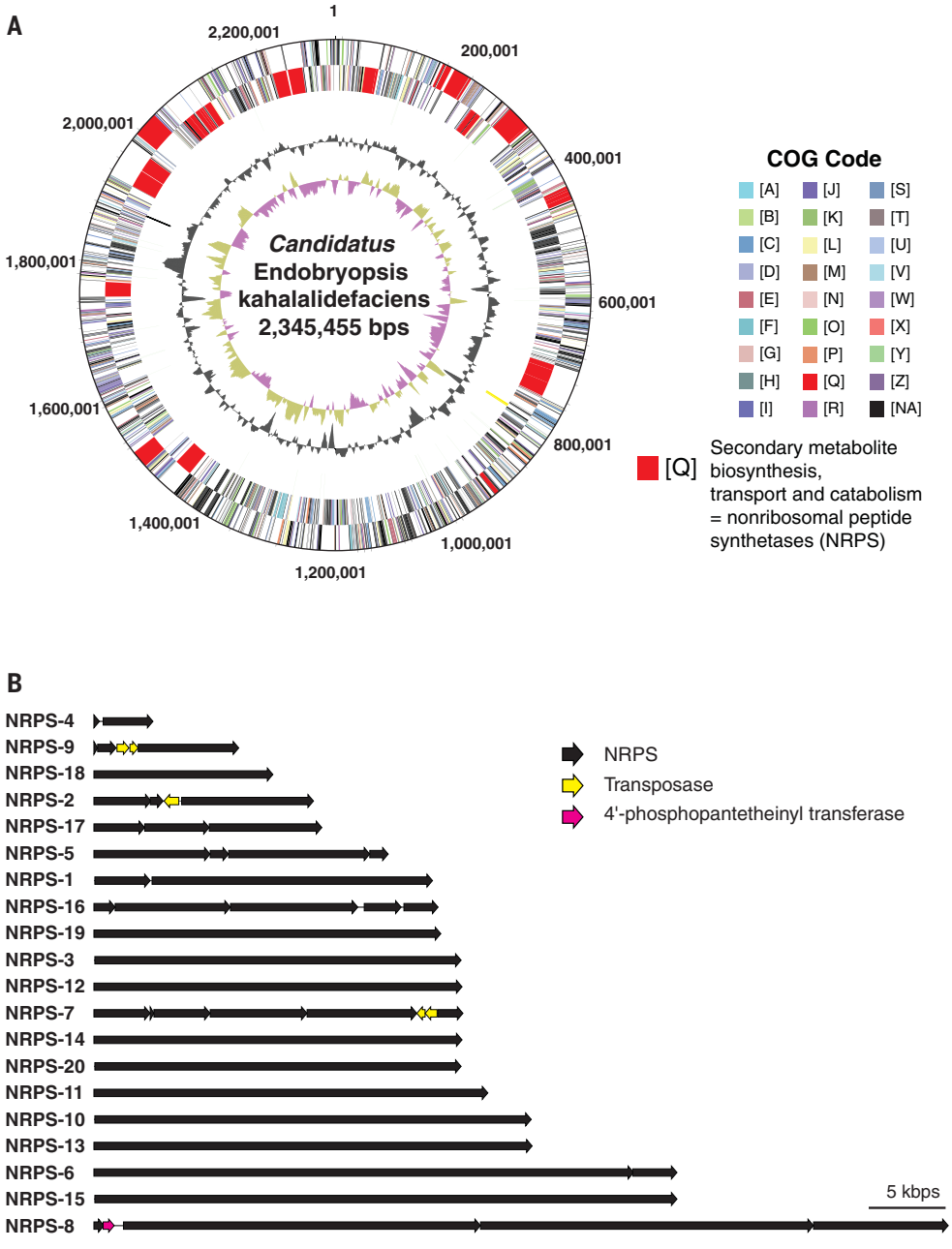


Fig. 2. The large biosynthetic capacity of "*Ca.* E. kahalalidefaciens." (**A**) Circular map of the "*Ca.* E. kahalalidefaciens" chromosome. Tracks (from outermost to innermost) represent genes on the forward frame, genes on the reverse frame, RNAs, GC content, and GC skew. Genes are color coded according to the Cluster of Orthologous Groups (COG) categories in the Integrated Microbial Genome platform (*21*) (A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division, and chromosome partitioning, E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination, and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, and chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport, and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; X, mobilome, prophages, and transposons; Y, nuclear structure; Z, cytoskeleton; and NA, not assigned). Note that 99.7% of the aggregate length of genes classified in the COG category Q in the "*Ca.* E. kahalalidefaciens" chromosome (red) correspond to NRPS pathways and occupy 20% of the genome coding capacity. (**B**) Genetic organization of the 20 NRPS pathways in "*Ca.* E. kahalalidefaciens" (ordered by size), where each arrow indicates a single gene.

phylum Bacteroidetes (fig. S5)—a taxonomical group not known as a prolific producer of complex small molecules. We termed this bacterium "*Candidatus* Endobryopsis kahalalidefaciens" (Latin for "kahalalide-making").

Symbiont biosynthetic capacity

Further analysis of the "*Ca.* E. kahalalidefaciens" genome revealed that it harbors 20 NRPS pathways, ranging in size from 3.9 (NRPS-4) to 55.6 (NRPS-8) kbps and accounting for all 33 high-coverage, ~50% GC NRPS fragments previously found in the Illumina assembly (Fig. 2 and fig. S2). Unusually for bacteria, these NRPS pathways occupy 20% of the genome. Since the initial discovery of KF from *Bryopsis* sp. and *E. rufescens*, other work has detected >15 cyclic and linear lipopeptides of different lengths and amino acid compositions from samples collected at the same site; these molecules are known collectively as the kahalalides (Fig. 3) (*8*, *9*, *17*–*20*). Because these molecules have been reported from the same species (sometimes at abundances similar to that of KF) and because of the structural similarities between KF and the rest of the kahalalides, we asked whether the remaining 19 NRPSs in the "*Ca.* E. kahalalidefaciens" genome encode other known kahalalides.

To answer this question, we compared the 19 remaining NRPSs encoded in the genome to the kahalalide structures. Overall, we were able to link eight additional NRPSs in the "*Ca.* E. kahalalidefaciens" genome to eight corresponding kahalalide chemotypes (a distinct chemotype can encompass several kahalalides sharing the same amino acid sequence and differing only in the length or hydroxylation of the fatty acid, hydroxylation of a proline residue in the molecule, or linearization of the cyclic peptide) (Fig. 3 and fig. S6). Although a genetic knockout (loss of function) or heterologous expression (gain of function) is typically needed to unequivocally connect biosynthetic gene clusters to their products, five lines of evidence indicate that "*Ca.* E. kahalalidefaciens" synthesizes

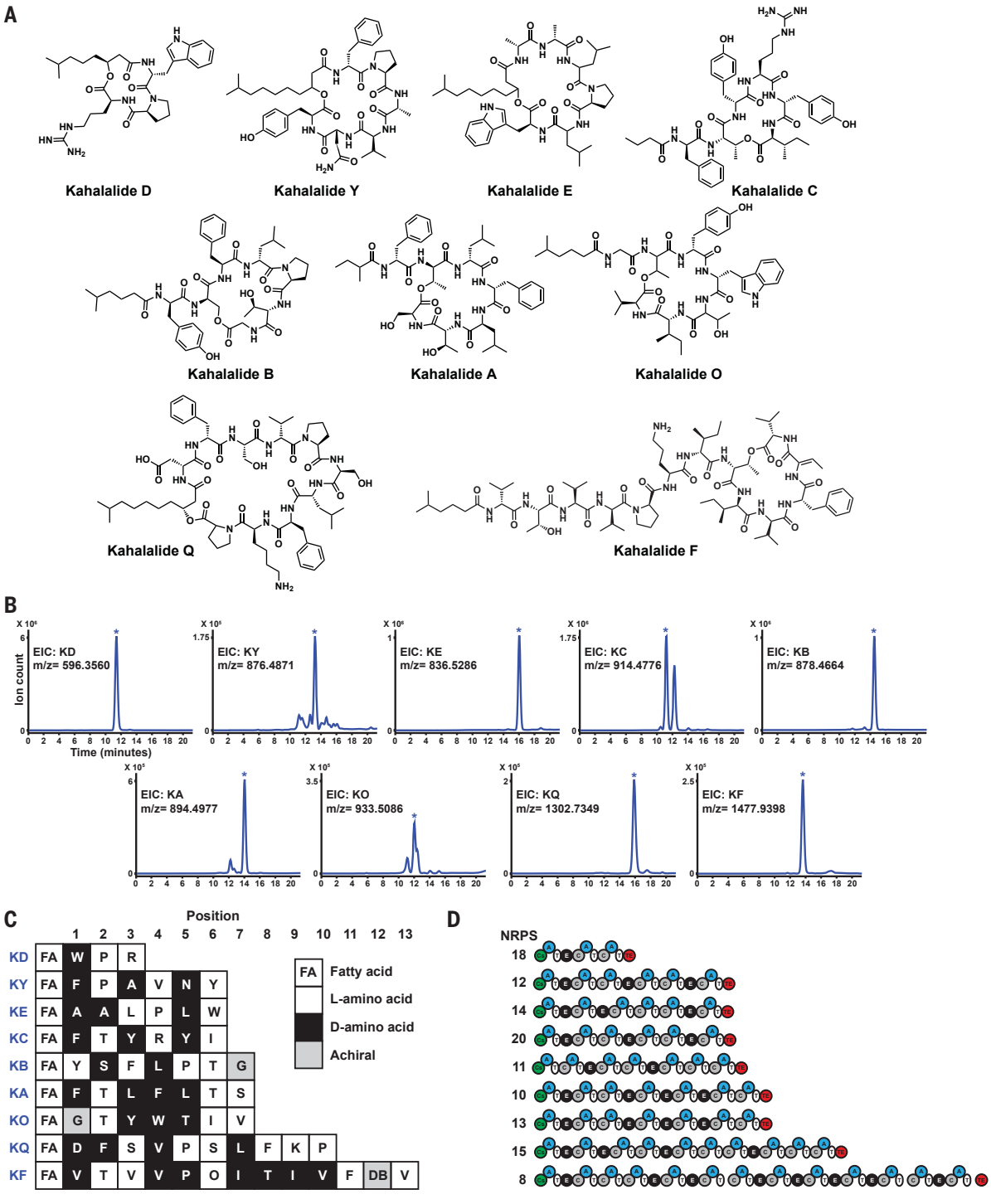


Fig. 3. Structural and biosynthetic diversity of kahalalides from *Bryopsis* sp. (**A**) Molecular structures of nine kahalalides that can be bioinformatically linked to specific NRPS pathways in "*Ca.* E. kahalalidefaciens" and chemically detected in *Bryopsis*-2015. (**B**) Extracted ion chromatograms (EICs) for the mass/charge ratios (m/z) corresponding to the $(M + H)^+$ ions of the molecules in (A), as detected in the chemical extract of *Bryopsis*-2015. An asterisk indicates the peak corresponding to the molecule of interest. (**C**) Amino acid composition of the nine kahalalides shown in (A), indicating the position and stereochemistry of each amino acid on a linear scale. Kahalalide names are represented with two-letter abbreviations (e.g., kahalalide D = KD), and amino acids are denoted by their canonical one-letter codes (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; Y, Tyr) except for ornithine (O) and dehydrobutyrine (DB). FA, fatty acid. (**D**) Domain organization of "*Ca.* E. kahalalidefaciens"–encoded NRPS pathways that were bioinformatically linked to corresponding kahalalides in C (Cs, starting condensation domain; A, adenylation; T, thiolation; E, epimerization; C, condensation; and TE, thioesterase). The number of modules in a given NRPS and the number and position of epimerization domains encoded in it (black) match exactly what is observed in its linked product (see also substrate specificity analyses in figs. S3 and S6 and detailed HR-MS/MS analyses in table S1 and figs. S1 and S7).

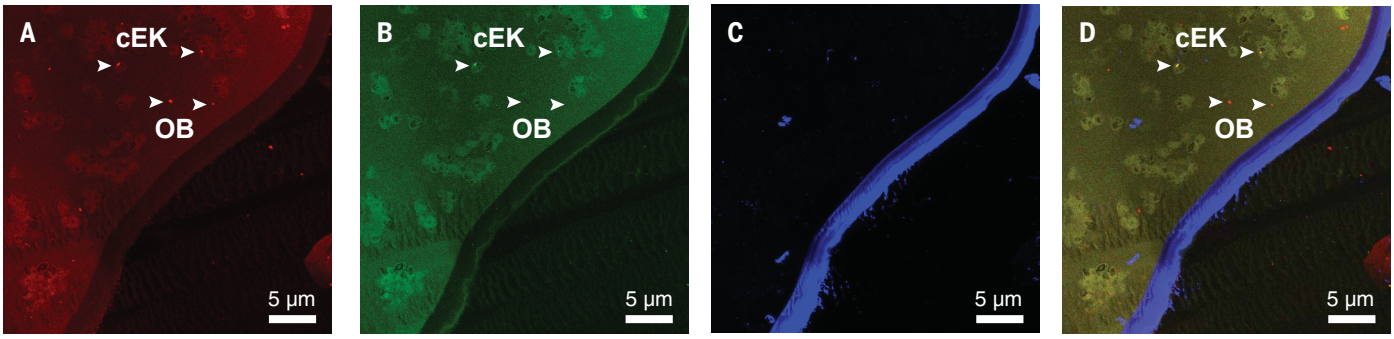


Fig. 4. Localization of "*Ca.* E. kahalalidefaciens" in *Bryopsis* sp. using FISH. Epifluorescent micrographs of a *Bryopsis* sp. section hybridized with (**A**) universal eubacterial probes EUB338 I to III (labeled with Cy3, red) and (**B**) "*Ca.* E. kahalalidefaciens"–specific probe JZP2 (labeled with 6-FAM, green). (**C**) Algal cell wall of the same section viewed under a 4′,6-diamidino-2-phenylindole (DAPI) channel using calcofluor counterstaining. (**D**) Composite of images (A) to (C), showing the colocalization of the red and green signals for "*Ca.* E. kahalalidefaciens" but not for other bacteria. Arrows indicate bacteria: cEK, "*Ca.* E. kahalalidefaciens"; OB, other bacteria.

all other kahalalides previously reported from this *Bryopsis* sp. First, consistent with the lipopeptide nature of the kahalalides, the matched NRPSs begin with a starting condensation domain. Second, the number of modules in the NRPSs and the position of epimerization domains in the modules match exactly the size and L- and D-amino acid positions in the corresponding molecules. Third, the predicted substrate specificity of adenylation domains in the NRPS (based on prediction algorithms and detailed phylogenetic analysis of the A domains) predominantly match the amino acid composition of the corresponding kahalalides (Fig. 3 and figs. S4 and S6). Fourth, no other NRPS pathways recovered from the metagenome are consistent with the structure of any of the kahalalides (97 scaffolds). Finally, and most importantly, in every case where we were able to bioinformatically match an NRPS in "*Ca.* E. kahalalidefaciens" to a corresponding kahalalide, we successfully confirmed the presence of the exact molecule in the chemical extract of *Bryopsis*-2015 using HPLC–HR-MS/MS analysis (Fig. 3, fig. S7, and table S1). These results establish "*Ca.* E. kahalalidefaciens" as a symbiotic microbial factory for at least nine complex molecules that are abundant enough to be isolated from the algal host. Of these nine molecules, at least one was shown to be essential for the host's chemical defense against predators.

An intracellular symbiont with a reduced genome

Apart from the NRPS-coding regions, the remaining coding capacity of the "*Ca.* E. kahalalidefaciens" genome is only 1.87 Mbps (average genome size in the Flavobacteriaceae family is ~3.7 Mbps), indicating that this bacterium is undergoing evolutionary genome reduction. Next, we annotated the "*Ca.* E. kahalalidefaciens" genome using the Integrated Microbial Genome platform (*21*) and compared it to the closest free-living and genome-sequenced bacterium: *Mangrovimonas* sp. ST2L12 (88% DNA sequence identity for the 16*S* rRNA gene) (*22*) (fig. S5). The genome of "*Ca.* E. kahalalidefaciens" encodes 873 protein-coding genes that can be functionally assigned into a Cluster of Orthologous Groups (COG) category and 449 that can be classified into a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, whereas the *Mangrovimonas* sp. ST2L12 genome encodes 1868 and 801 such genes, respectively. Most enzymes and pathways for DNA replication and protein synthesis, cofactor biosynthesis, and central and intermediary metabolism (glycolysis, tricarboxylic acid cycle, pentose phosphate pathway, electron transport, and fatty acid synthesis and degradation) are present in both "*Ca.* E. kahalalidefaciens" and *Mangrovimonas* sp. ST2L12. However, some DNA repair genes and pathways (damaged nucleotide removal and base excision repair, nucleotidyltransferase involved in DNA repair, and RNA helicase), as well as genes involved in chemotaxis, detoxification, and adaptations to atypical conditions, are mostly absent from "*Ca.* E. kahalalidefaciens," a finding that is commonly observed in obligate symbionts undergoing genome reduction (*23*) (table S2).

Most notably, no complete pathway to any of the 20 proteinogenic amino acids can be identified in the "*Ca.* E. kahalalidefaciens" genome, with most pathways missing all or the majority of the genes. The only partially complete series are in the early steps of aromatic amino acid biosynthesis leading to chorismate, but the downstream enzymes for phenylalanine, tyrosine, and tryptophan biosynthesis are mostly absent (table S2). Additionally, the "*Ca.* E. kahalalidefaciens" genome shows two features that are rarely found in reduced genomes. First, its GC content has not been reduced: The average GC content of "*Ca.* E. kahalalidefaciens" is 52%, higher than that of *Mangrovimonas* sp. ST2-L12 (38%). Second, it contains an unusually high number (84) of transposons, transposases, and transposable elements. With this enrichment of mobile elements, we wondered whether overt signs of horizontal gene transfer (HGT) can be detected in the genome. We searched for signs of genome heterogeneity by multiple parametric methods, including GC content (*24*), *k*-mer frequencies (k = 2 or 3) (*25*), and dinucleotide bias (*26*). We detected a total of seven genomic regions that strongly deviate in their composition from the genome average (fig. S8 and table S3). Six of these regions range in size from 0.1 to 2.6 kbps and encode hypothetical proteins. A seventh region of ~26 kbps (coordinates: 1821316 to 1847296) encodes several phage proteins, including capsid, portal, and tail sheath proteins, and likely constitutes a complete prophage. No other clear signs of HGT were detected across the entire genome, including the 20 NRPSs (fig. S8).

Next, we tested whether "*Ca.* E. kahalalidefaciens" is capable of free living by attempting its cultivation in various media and employing a sensitive, high-throughput sequencing–based method for its detection (see materials and methods and fig. S9). These efforts, however, were not successful, suggesting that "*Ca.* E. kahalalidefaciens" is an obligate symbiont. Indeed, most genome-reduced, obligate symbionts live intracellularly in their hosts (*23*). To determine whether "*Ca.* E. kahalalidefaciens" is an intracellular or extracellular symbiont of *Bryopsis* sp., we performed fluorescence in situ hybridization (FISH) on fixed *Bryopsis*-2015 specimens using fluorescent probes that target regions of the 16*S* rRNA. Two sets of probes were used: a combination of universal eubacterial probes EUB338 I, II, and III (*27*, *28*) and one that is specific for "*Ca.* E. kahalalidefaciens" (JZP2; see materials and methods). Both general and specific probes localized "*Ca.* E. kahalalidefaciens" to the cellular compartment of *Bryopsis*, where most detected signals were from dispersed cells within the algal cytosol (Fig. 4). Notably, we found that a partial 16*S* rRNA gene sequence (with 100% DNA sequence identity to that of "*Ca.* E. kahalalidefaciens") was previously cloned from a Brazilian sample of *Bryopsis* sp. (fig. S5) (*29*). This previous study suggested that the bacterium harboring this sequence is also intracellular and widespread in *Bryposis* sp. living in warm-temperate and tropical environments (*30*). Finally, we searched a recently published dataset composed of 16*S* rRNA gene amplicon sequences from 1194 sponges, 37 sediments, and 195 seawater samples for sequences that match "*Ca.* E. kahalalidefaciens" (*31*). At a cutoff of 97% DNA sequence identity, no matches were found, further supporting the specific, obligate, and intracellular symbiotic relationship between "*Ca.* E. kahalalidefaciens" and *Bryopsis* sp.

Given that there is no clear nutritional advantage provided by "*Ca.* E. kahalalidefaciens" and that 20% of the bacterial genome is dedicated to the biosynthesis of complex small molecules, we assume that the most important aspect of this relationship is chemical defense in exchange for a tolerant intracellular environment in which to live. Notably, "*Ca.* E. kahalalidefaciens" is defective in synthesizing all 16 amino acids that are incorporated as substrates during the biosynthesis of the kahalalides (Fig. 3). These substrates are provided by the autotrophic *Bryopsis* sp. host, although other intracellular bacteria may also contribute. These findings establish an unusual collaborative model for the biosynthesis of the kahalalides, where the host provides the initial substrates and the symbiont provides the biochemical machinery to transform these simple substrates into complex, biologically active molecules that may otherwise be unattainable in algal biochemistry.

A tripartite symbiosis

Through kleptoplasty, *E. rufescens* sequesters and maintains intact and functioning chloroplasts from the cells of its *Bryopsis* sp. diet. Can the mollusk likewise "steal" intact intracellular bacterial symbionts of *Bryopsis* sp. (e.g., "*Ca.* E. kahalalidefaciens")? To answer this question, we collected several specimens of *E. rufescens*, confirmed that they contain KF, and then used deep 16*S* rRNA gene amplicon sequencing from whole-animal metagenomic DNA of three individuals (an average of 57,000 reads per sample) to search for "*Ca.* E. kahalalidefaciens."

The *E. rufescens* microbiome is dominated by sequences matching the algal chloroplast (47.0%, on average), Mollicutes (29.0%), and Gammaproteobacteria (12.5%) (fig. S10), in agreement with a previous study (*32*). Despite the sequencing depth performed in our study, we barely detected 16*S* rRNA gene sequences that match "*Ca.* E. kahalalidefaciens" in *E. rufescens* (an average of 0.06% of the obtained reads). Additional sequencing of 16*S* rRNA gene amplicons from sections of two coarsely dissected *E. rufescens* individuals (an average of 81,000 reads per section) confirmed that there was no specific symbiosis organ in the mollusk and that "*Ca.* E. kahalalidefaciens" cells appear to be fully digested upon ingestion (fig. S10). Finally, though we could clearly observe signals for other bacteria using general eubacterial probes, negative FISH experiments on several *E. rufescens* sections using a specific probe for "*Ca.* E. kahalalidefaciens" corroborated our sequencing results (fig. S11). Taken together, our results establish chemical but not symbiont sequestration as a likely means of KF acquisition, although the mechanistic details of this step are unknown.

A model for the evolution of defense chemicals

Next, we asked what molecular mechanisms generate the observed diversity in kahalalide structures. Because we found no evidence of HGT-mediated acquisition of NPRS pathways in the "*Ca.* E. kahalalidefaciens" genome (fig. S8), we looked for signs of increased genome plasticity (*33*, *34*) and intragenome genetic exchange. Briefly, we segmented the genome into 150-bp fragments, produced a whole-genome identity dot plot by comparing all fragments, and searched for genomic loci that contain a high density of matched fragment pairs (>60% sequence identity). On average, NRPS loci showed a 50-fold higher density of matched fragments than the rest of the genome (Fig. 5A; see also materials and methods and fig. S12), in a largely indiscriminate (matched fragments within a single NRPS pathway are neither more frequent nor more similar than those between pathways) (fig. S13) and pervasive (matched fragments cover >98.8% of the combined NRPS sequences) manner. A notable outlier in this analysis was NRPS-8, where matched fragments between NRPS-8 and other pathways showed lower density and percent identity than between any other pairs of NRPSs (Fig. 5A and figs. S12 and S13).

Consecutive matched fragments between pairs of NRPSs form visually distinct stripes on identity dot plots, which represent alignments that start and end with abrupt changes in DNA sequence identity as a result of possible genetic exchange events (Fig. 5B and fig. S12). We identified more than 100 such events: Three represent full pathways and can therefore be explained by typical duplication and divergence events (one between NRPS-10 and NRPS-13; two between NRPS-3, NRPS-12, and NRPS-20), whereas the remaining ones represent only parts of the pathways and are best explained by separate recombination events (Fig. 5C and fig. S14). In most cases, lineage trees for different aligned sequences that appear in the same set of NRPSs do not follow the same path (Fig. 5D), indicating that frequent shuffling substantially disrupts the trajectories of pathway duplication. Finally, 9 of the 20 NRPSs contain or occur within a 5-kbp window of an annotated transposase or transposon (fig. S15), suggesting that transposition is at least partially responsible for the observed plasticity, as seen in other systems (*33*, *34*). These findings motivated us to propose a new model for the diversifying evolution of NRPSs in the "*Ca.* E. kahalalidefaciens" genome, where new pathways arise through duplication and divergence events that are concurrently accompanied by a high frequency of interpathway recombination, resulting in the observed diversity and an intertwined evolutionary history.

Why would a bacterial symbiont produce so much chemical diversity? In other words, are these molecules equally important in establishing a sustainable symbiotic relationship, especially given that maintaining and expressing their large biosynthetic pathways is a major metabolic cost? We reasoned that nonessential pathways should decay and eventually get lost, whereas essential ones should be intact and maintained. We looked for signs of genomic decay in the 20 NRPS pathways. Indeed, NRPS-4 and NRPS-16 show truncated domains and modules, and NRPS-2, NRPS-7, and NRPS-9 contain domains that have been interrupted by transposases (Fig. 2B). The remaining NRPS pathways appear to be pristine. We then compared the expression levels of the NRPS pathways under native conditions, reasoning that essential pathways would be highly expressed, whereas decaying ones would be expressed at a lower level or not expressed at all. Metatranscriptomic sequencing data from *Bryopsis*-2015 (40 million paired-end reads, 150 bps; 41 million single-end reads, 100 bps) were mapped to the "*Ca.* E. kahalalidefaciens" genome (see materials and methods), resulting in the successful alignment of 2.2 million reads (mostly to possible mRNAs) (Fig. 6A). We then ranked all genes according to their expression level (data S1), which produced a trimodal distribution: About 50% of the genes are almost silent or have a very low level of expression, 45% are at median level, and 5% are among the most highly expressed (Fig. 6B). Comparison of the expression level of the NRPS genes showed a similar distribution. Genes from NRPS-1, -2, -4, -6, -9, -16, and -17 (NRPS-2, -4, -9, and -16 were predicted to be decaying pathways) had the lowest expression levels, and genes from NRPS-8 had the highest. NRPS-8 genes were among the most highly expressed genes in the entire genome, along with cell division and transcription genes. NRPS-8 alone accounted for more than 12% of the transcriptional activity of "*Ca.* E. kahalalidefaciens," and in total, all of the NRPS pathways accounted for 26% (Fig. 6A). These results establish NRPS-8 (encoding the most toxic kahalalide, KF) as a potentially essential pathway, and NRPS-2, -9, and -16 as potentially on their way to being lost.

We thus propose a model for NRPS pathway generation, evolution, and selection in "*Ca.* E. kahalalidefaciens" (supplementary text). In our model, new pathways are created from an ancestral one through duplication and divergence, followed by extensive and continuous interpathway genetic exchange to generate more diversity. Newly generated pathways are then selected for or against, depending on unknown selection pressures, leading to their maintenance or loss, followed by a continuous repetition of the same cycle (fig. S16).

Discussion

KF is the most toxic of all known kahalalides and is likely the main contributor to *Bryopsis* sp. and *E. rufescens* chemical defense against predators (*7*). KF and more than 15 related kahalalides have been reported from this predator–prey system, yet the original source of this library of molecules has been unknown. In this study, we used metagenomics, metatranscriptomics, chemical analysis, microscopy, and evolutionary genomics to uncover the molecular details of kahalalide production by an elusive third partner, an intracellular bacterial symbiont of *Bryopsis* sp.

Chemically defended mollusks are widespread in the marine environment, but the source of their defense chemicals varies (*5*). Although chemical sequestration from diet has been proposed

in many cases (35), de novo biosynthesis has also been demonstrated through isotope labeling experiments (36). Here, we provide evidence that defense chemicals in *E. rufescens* do not originate in the algal diet itself but in intracellular bacterial symbionts within the alga. A similar indirect acquisition mechanism may be true for other chemicals found in mollusks, such as those acquired from dietary sponges and tunicates. Most work on symbioses in marine algae has focused on extracellular symbionts, as exemplified by the symbiosis between *Emiliania*

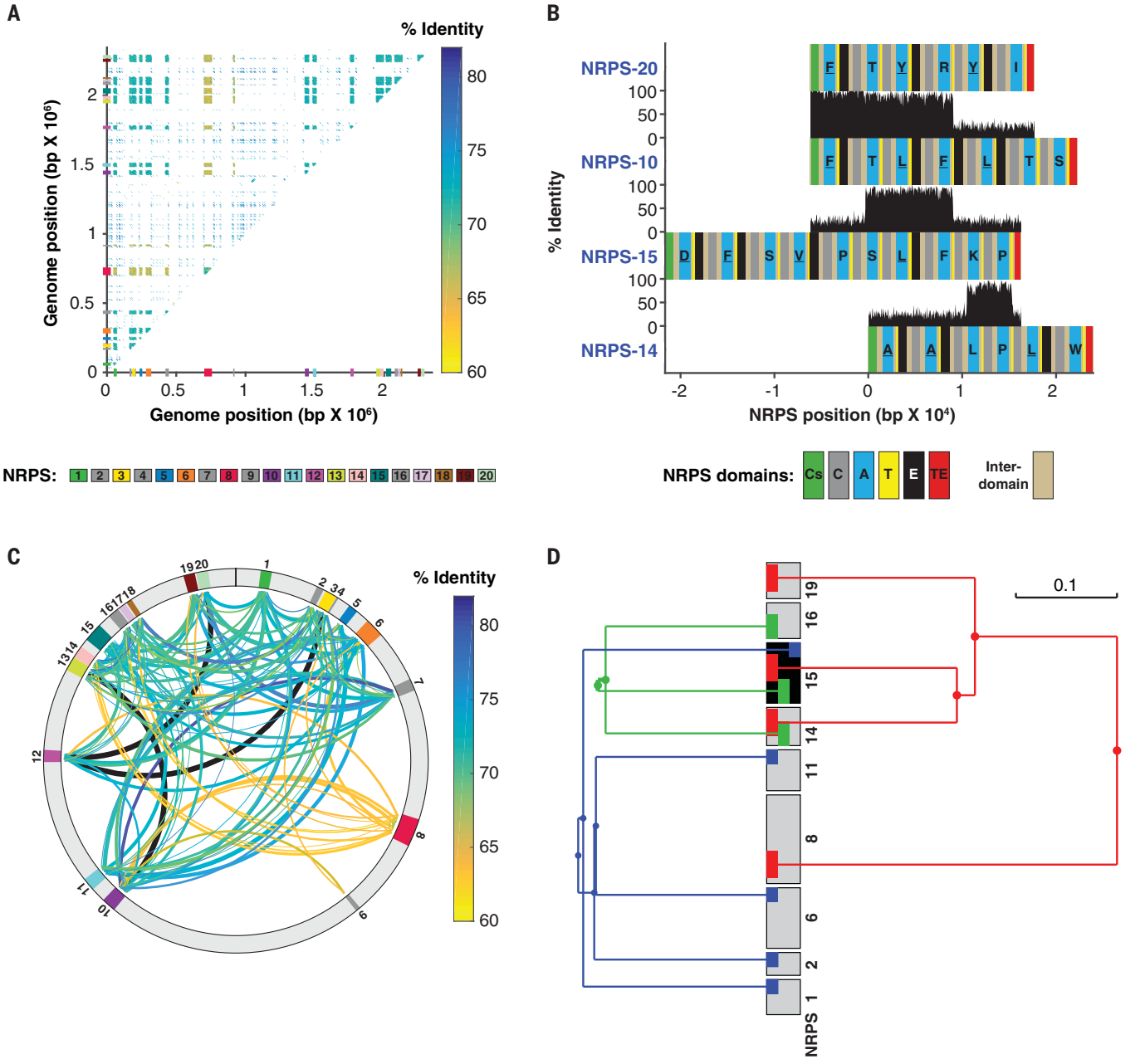


Fig. 5. Intensive genetic exchange in "*Ca.* E. kahalalidefaciens" NRPS pathways. (**A**) Pairs of sequence fragments sharing more than 60% identity across the whole "*Ca.* E. kahalalidefaciens" genome, shown on a linear scale on both the *x* and *y* axes. Diagonal matches were removed, and color indicates percent identity. Positions of the 20 NRPS pathways are indicated on the axes, matching the color code below. (**B**) Examples of aligned sequences between NRPS pathways. Note the abrupt rise and fall in sequence identity. Colors in pathways indicate different domains, matching the color code below. Encoded amino acids are indicated by their single-letter abbreviations, and underlined letters indicate D-amino acids. (**C**) Circular representation of "*Ca.* E. kahalalidefaciens" genome recombination events (coordinate one of the genome is indicated by a solid black line), in which NRPS pathways are shown at their respective positions around the genome and connecting colored lines indicate identified pairwise recombination events. Lines follow a yellow-to-blue color code that represents the percent identity between pairs of aligned sequences, and their thickness positively correlates with the length of the sequence. Sequences shared with NRPS-8 and NRPS-9 have a lower percent identity than ones shared with other pathways. Black lines indicate the three identified duplication events. (**D**) Phylogenetic trees for three groups of aligned sequences (red, green, and blue) that together cover the entire NRPS-15. Although the three aligned sequences are found in NRPS-15, they do not follow the same phylogenetic lineage and are likely a result of three independent recombination events. NRPS-15 is shown as a black rectangle, and NRPS pathways that share one or more of the three groups of aligned sequences with it are shown as gray rectangles. The scale bar at top right indicates 10% sequence divergence.

huxleyi and *Phaeobacter inhibens*, in which *P. inhibens* influences the biology of the host through alternating symbiotic and pathogenic cycles (*37*), and that between *Ulva* sp. and several bacteria, in which the symbionts trigger proper host growth and development (*38*). Here, we describe the functional importance of intracellular symbionts of marine macroalgae for the ecology of their host, which has been previously overlooked. We also establish macroalgae as an additional group of marine organisms in which bacterial symbionts produce complex natural products discovered in the host, as has been previously observed in marine tunicates, sponges, bryozoans, and crustaceans (*4*).

The diversifying evolution of NRPS pathways in the "*Ca.* E. kahalalidefaciens" genome is particularly noteworthy because it reveals a strategy for chemical innovation in bacteria in a HGT-independent manner. The closest examples are two cases reported in bacterial symbionts of marine ascidians: one in which the same polyketide synthase pathway has duplicated multiple times in the same genome, presumably resulting in higher expression levels (*39*), and another in which a cyanobactin pathway has duplicated once and accumulated strategic point mutations in its precursor peptide to produce novel molecules (*40*, *41*). The number of biosynthetic pathways (20 NRPSs) and resulting molecules discovered in "*Ca.* E. kahalalidefaciens" is substantial, especially given its reduced genome and strict intracellular lifestyle, and approaches the diversity encoded in the genome of *Entotheonella* sp., an extracellular symbiont of marine sponges and a talented producer of natural products (*42*, *43*).

We do not yet understand the nature of the selective pressures on the "*Ca.* E. kahalalidefaciens" genome. Assuming that all kahalalides have roles in chemical defense, we propose either that they contribute to a synergistic cocktail, where molecules function additively on diverse predators or pathogens and the selection acts to maintain diversity, or that each molecule has a specific role and is selected for independently. There is some support for the latter hypothesis in that KF has been shown to have a strong antipredatory activity (*7*), and the rest of the kahalalides have low to no cytotoxicity or antimicrobial activity and their natural role has not yet been determined (*9*). Furthermore, levels of different kahalalides produced in the same system vary considerably between times of collection (*17*), suggesting that they are selected for individually and not globally. We also do not understand how *E. rufescens* sequesters, concentrates, and resists the cytotoxicity of the kahalalides. This is complicated by the fact that their molecular target is still unknown, even for KF, which has been evaluated as a promising anticancer agent in several human clinical trials (*44*). We show that chemical and not symbiont sequestration is responsible for kahalalide acquisition in *E. rufescens*, but exactly how and where this occurs await further investigation. These outstanding questions are not specific to *E. rufescens* and the kahalalides but are true for almost all marine mollusks that sequester toxins from their diet.

Materials and methods summary

An expanded description of materials and methods can be found in the supplementary materials.

Bryopsis *sp. and* E. rufescens *collection and processing*

Bryopsis sp. and *E. rufescens* samples were collected from Black Point Bay, Honolulu, Hawaii (N21°15′34″; W157°47′24″). Small portions of *Bryopsis* sp. and individual *E. rufescens* specimens were preserved frozen for chemical analysis, preserved in RNAlater (ThermoFisher Scientific, USA) for DNA and RNA analysis, or fixed in paraformaldehyde for FISH analysis.

Metagenomic DNA and RNA extraction and sequencing

Total metagenomic DNA used for both Illumina and Pac Bio sequencing was extracted from lyophilized *Bryopsis* sp. and *E. rufescens* using the Mo Bio Power Biofilm DNA isolation kit (Mo Bio Laboratories, USA; now Qiagen, USA). Total RNA was extracted from RNAlater-preserved *Bryopsis* sp. using MasterPure Complete DNA and RNA Purification Kit (Epicentre, USA; now Illumina, USA). Illumina sequencing was performed on an Illumina HiSeq 2500 sequencer (Illumina, USA), and Pac Bio sequencing was performed on a PacBio RS II instrument (Pacific Biosciences, USA).

Fluorescence in situ hybridization (FISH)

Fixed specimens of *Bryopsis* sp. were embedded in LR White resin (Electron Microscopy Sciences, USA), cut into 2-μm-thick slices, hybridized with either universal eubacterial probes or a "*Ca.* E. kahalalidefaciens"–specific probe, counterstained using Calcofluor White Stain (Sigma Aldrich, USA), and imaged on a Leica SP8 confocal microscope.

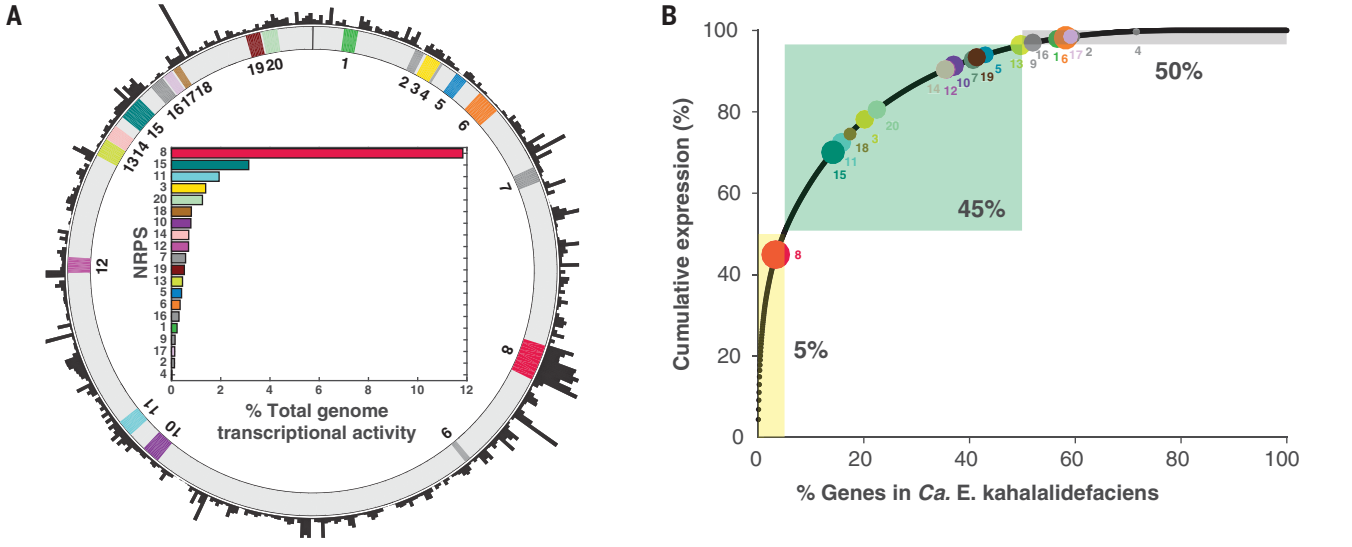


Fig. 6. Metatranscriptomic analysis of "*Ca.* E. kahalalidefaciens." (**A**) Genome-wide transcriptional activity of "*Ca.* E. kahalalidefaciens" using metatranscriptomic data collected from *Bryopsis*-2015. Black bars indicate total counts for each mapped position and averaged per 5 kbps of the genome. The inset bar graph indicates the ranked contribution of each NRPS pathway as a percentage of the total genome transcriptional activity. (**B**) "*Ca.* E. kahalalidefaciens" genes (*x* axis, shown in percentage of the total number of genes in the genome) plotted against their cumulative expression (*y* axis, shown in percentage of the total expression of all genes in the genome). Note the trimodal distribution of genes according to their expression and that genes from the NRPS pathways appear in all three modes.

Bacterial cultivation and screening for "Ca. E. kahalalidefaciens"

Freshly collected *Bryopsis* sp. (5 g) was homogenized in 9 ml of sterile 1X Artificial Sea Water (Instant Ocean Sea Salts, 35 g/liter, Instant Ocean, USA) using a mortar and pestle, and the algal homogenate was serially diluted in a 10-fold series. 100-µl aliquots of each dilution were plated on eight different media in triplicate, and plates were incubated at room temperature for 1 to 2 weeks. Colonies were scraped from one plate of each of the eight media, and DNA was extracted from the mixture and screened for the presence of "*Ca.* E. kahalalidefaciens" using high-throughput 16*S* rRNA gene amplicon sequencing.

Chemical extraction and analysis of Bryopsis-2015

Lyophilized *Bryopsis* sp. was extracted with methanol and analyzed by HPLC–HR-MS/MS using an Agilent 6500 Series Q-TOF LC/MS system (Agilent, USA).

Evolutionary analysis of the "Ca. E. kahalalide" genome and the 20 NRPSs

The "*Ca.* E. kahalalidefaciens" genome was segmented into fragments of 150 bps in length, with a 30-bp sliding window. Each fragment was aligned to all other fragments to obtain their percent identity; pairs of fragments with identities higher than 60% were recorded as "matched." Consecutive matched fragments between two genomic regions form stripes with a slope of 1 and varying intercepts on identity dot plots. These stripes were then used to infer genetic exchange events throughout the genome.

REFERENCES AND NOTES

1. V. J. Paul, K. E. Arthur, R. Ritson-Williams, C. Ross, K. Sharp, Chemical defenses: From compounds to communities. *Biol. Bull.* **213**, 226–251 (2007). doi: 10.2307/25066642; pmid: 18083964
2. J. Piel, Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **21**, 519–538 (2004). doi: 10.1039/b310175b; pmid: 15282634
3. J. Piel, Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **26**, 338–362 (2009). doi: 10.1039/B703499G; pmid: 19240945
4. M. Morita, E. W. Schmidt, Parallel lives of symbionts and hosts: Chemical mutualism in marine animals. *Nat. Prod. Rep.* **35**, 357–378 (2018). doi: 10.1039/C7NP00053G; pmid: 29441375
5. M. T. Davies-Coleman, in *Molluscs: From Chemo-ecological Study to Biotechnological Application*, G. Cimino, M. Gavagnin, Eds. (Springer, 2006), pp. 133–157.
6. M. T. Hamann, P. J. Scheuer, A Bioactive Depsipeptide from the Sacoglossan Mollusk *Elysia rufescens* and the Green Alga *Bryopsis* sp. *J. Am. Chem. Soc.* **115**, 5825–5826 (1993). doi: 10.1021/ja00066a061
7. M. A. Becerro, G. Goetz, V. J. Paul, P. J. Scheuer, Chemical defenses of the sacoglossan mollusk *Elysia rufescens* and its host Alga *bryopsis* sp. *J. Chem. Ecol.* **27**, 2287–2299 (2001). doi: 10.1023/A:1012287105923; pmid: 11817082
8. M. T. Hamann, C. S. Otto, P. J. Scheuer, D. C. Dunbar, Kahalalides: Bioactive Peptides from a Marine Mollusk *Elysia rufescens* and Its Algal Diet *Bryopsis* sp. *J. Org. Chem.* **61**, 6594–6600 (1996). doi: 10.1021/jo960877+; pmid: 11667527
9. J. Gao, M. T. Hamann, Chemistry and biology of kahalalides. *Chem. Rev.* **111**, 3208–3235 (2011). doi: 10.1021/cr100187n; pmid: 21480581
10. K. N. Pelletreau *et al.*, Sea slug kleptoplasty and plastid maintenance in a metazoan. *Plant Physiol.* **155**, 1561–1565 (2011). doi: 10.1104/pp.111.174078; pmid: 21346171
11. M. A. Fischbach, C. T. Walsh, Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006). doi: 10.1021/cr0503097; pmid: 16895337
12. P. G. Arnison *et al.*, Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013). doi: 10.1039/C2NP20085F; pmid: 23165928
13. T. Weber *et al.*, antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015). doi: 10.1093/nar/gkv437; pmid: 25948579
14. K. Blin *et al.*, antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017). doi: 10.1093/nar/gkx319; pmid: 28460038
15. N. Ziemert *et al.*, The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLOS ONE* **7**, e34064 (2012). doi: 10.1371/journal.pone.0034064; pmid: 22479523
16. D. Baranašić *et al.*, Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J. Ind. Microbiol. Biotechnol.* **41**, 461–467 (2014). doi: 10.1007/s10295-013-1322-2; pmid: 24104398
17. K. V. Rao *et al.*, Kahalalides V-Y isolated from a Hawaiian collection of the sacoglossan mollusk *Elysia rufescens*. *J. Nat. Prod.* **71**, 772–778 (2008). doi: 10.1021/np070508g; pmid: 18407693
18. Y. Kan, T. Fujita, B. Sakamoto, Y. Hokama, H. Nagai, Kahalalide K: A new cyclic depsipeptide from the hawaiian green alga *Bryopsis* species. *J. Nat. Prod.* **62**, 1169–1172 (1999). doi: 10.1021/np990053y; pmid: 10479330
19. G. Goetz, Y. Nakao, P. J. Scheuer, Two Acyclic Kahalalides from the Sacoglossan Mollusk *Elysia rufescens*. *J. Nat. Prod.* **60**, 562–567 (1997). doi: 10.1021/np970045m
20. F. D. Horgen *et al.*, A new depsipeptide from the sacoglossan mollusk *Elysia ornata* and the green alga *Bryopsis* species. *J. Nat. Prod.* **63**, 152–154 (2000). doi: 10.1021/np990402o; pmid: 10650101
21. A. Checcucci, A. Mengoni, The integrated microbial genome resource of analysis. *Methods Mol. Biol.* **1231**, 289–295 (2015). doi: 10.1007/978-1-4939-1720-4_18; pmid: 25343872
22. B. Dinesh *et al.*, Comparative genome analyses of novel *Mangrovimonas*-like strains isolated from estuarine mangrove sediments reveal xylan and arabinan utilization genes. *Mar. Genomics* **25**, 115–121 (2016). doi: 10.1016/j.margen.2015.12.006; pmid: 26795059
23. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011). doi: 10.1038/nrmicro2670; pmid: 22064560
24. S. Garcia-Vallvé, A. Romeu, J. Palau, Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* **10**, 1719–1725 (2000). doi: 10.1101/gr.130000; pmid: 11076857
25. F. Zhou, V. Olman, Y. Xu, Barcodes for genomes and applications. *BMC Bioinformatics* **9**, 546 (2008). doi: 10.1186/1471-2105-9-546; pmid: 19091119
26. S. Karlin, Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr. Opin. Microbiol.* **1**, 598–610 (1998). doi: 10.1016/S1369-5274(98)80095-7; pmid: 10066522
27. R. I. Amann *et al.*, Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990). pmid: 2200342
28. H. Daims, A. Brühl, R. Amann, K. H. Schleifer, M. Wagner, The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: Development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999). doi: 10.1016/S0723-2020(99)80053-8; pmid: 10553296
29. J. Hollants *et al.*, Who is in there? Exploration of endophytic bacteria within the siphonous green seaweed Bryopsis (Bryopsidales, Chlorophyta). *PLOS ONE* **6**, e26458 (2011). doi: 10.1371/journal.pone.0026458; pmid: 22028882
30. J. Hollants, F. Leliaert, H. Verbruggen, O. De Clerck, A. Willems, Host specificity and coevolution of Flavobacteriaceae endosymbionts within the siphonous green seaweed Bryopsis. *Mol. Phylogenet. Evol.* **67**, 608–614 (2013). doi: 10.1016/j.ympev.2013.02.025; pmid: 23499613
31. T. Thomas *et al.*, Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* **7**, 11870 (2016). doi: 10.1038/ncomms11870; pmid: 27306690
32. J. Davis *et al.*, Characterization of the bacterial community of the chemically defended Hawaiian sacoglossan *Elysia rufescens*. *Appl. Environ. Microbiol.* **79**, 7073–7081 (2013). doi: 10.1128/AEM.01568-13; pmid: 24014539
33. T. A. Hendry *et al.*, Ongoing Transposon-Mediated Genome Reduction in the Luminous Bacterial Symbionts of Deep-Sea Ceratioid Anglerfishes. *mBio* **9**, e01033-18 (2018). doi: 10.1128/mBio.01033-18; pmid: 29946051
34. W. Qi *et al.*, *Ca.* Endozoicomonas cretensis: A Novel Fish Pathogen Characterized by Genome Plasticity. *Genome Biol. Evol.* **10**, 1363–1374 (2018). doi: 10.1093/gbe/evy092; pmid: 29726925
35. J. R. Pawlik, M. R. Kernan, T. F. Molinski, M. K. Harper, D. J. Faulkner, Defensive chemicals of the Spanisch dancer nudibranch *Hexabranchus sanguineus* and its egg ribbons: Macrolides derived from a sponge diet. *J. Exp. Mar. Biol. Ecol.* **119**, 99–109 (1988). doi: 10.1016/0022-0981(88)90225-0
36. J. Kubanek, R. J. Andersen, Evidence for de novo biosynthesis of the polyketide fragment of diaulusterol A by the northeastern pacific dorid nudibranch diaulula sandiegensis. *J. Nat. Prod.* **62**, 777–779 (1999). doi: 10.1021/np9804839; pmid: 10346969
37. A. R. Bramucci *et al.*, The Bacterial Symbiont *Phaeobacter inhibens* Shapes the Life History of Its Algal Host *Emiliania huxleyi*. *Front. Mar. Sci.* **5**, 188 (2018). doi: 10.3389/fmars.2018.00188
38. F. Ghaderiardakani, J. C. Coates, T. Wichard, Bacteria-induced morphogenesis of *Ulva intestinalis* and *Ulva mutabilis* (Chlorophyta): A contribution to the lottery theory. *FEMS Microbiol. Ecol.* **93**, fix094 (2017). doi: 10.1093/femsec/fix094; pmid: 28810708
39. J. Lopera, I. J. Miller, K. L. McPhail, J. C. Kwan, Increased Biosynthetic Gene Dosage in a Genome-Reduced Defensive Bacterial Symbiont. *mSystems* **2**, e00096-17 (2017). doi: 10.1128/mSystems.00096-17; pmid: 29181447
40. M. S. Donia *et al.*, Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat. Chem. Biol.* **2**, 729–735 (2006). doi: 10.1038/nchembio829; pmid: 17086177
41. M. S. Donia, W. F. Fricke, J. Ravel, E. W. Schmidt, Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLOS ONE* **6**, e17897 (2011). doi: 10.1371/journal.pone.0017897; pmid: 21445351
42. M. C. Wilson *et al.*, An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014). doi: 10.1038/nature12959; pmid: 24476823
43. T. Mori *et al.*, Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1718–1723 (2018). doi: 10.1073/pnas.1715496115; pmid: 29439203
44. B. Miguel-Lillo, B. Valenzuela, J. E. Peris-Ribera, A. Soto-Matos, J. J. Pérez-Ruixo, Population pharmacokinetics of kahalalide F in advanced cancer patients. *Cancer Chemother. Pharmacol.* **76**, 365–374 (2015). doi: 10.1007/s00280-015-2800-1; pmid: 26093949

ACKNOWLEDGMENTS

We are indebted to J. Davidson and the late Ruth Gates at HIMB for laboratory space and assistance during the field work conducted in this study. We thank G. Laevsky, the Molecular Biology Confocal Microscopy Facility (a Nikon Center of Excellence), J. Yan, and B. Bassler for assistance with confocal microscopy; W. Wang and the Lewis Sigler Institute sequencing core facility for assistance with high-throughput Illumina and 16S rRNA gene amplicon sequencing; L. Tallon, L. Sadzewicz, and the Institute for Genome Sciences, Genomics Resource Center, University of Maryland, for Pac Bio sequencing; M. Cahn and J. Lopez for assistance with metagenomic data analysis; M. T. Hamann, W. L. Cheung-Lee, and A. J. Link for valuable scientific insights about the project; S. Chatterjee for general assistance; and the rest of the Donia lab for useful discussions. **Funding:** Funding for this project has been provided by Princeton University, M.S.D. is funded by an NIH Director's New Innovator Award (ID 1DP2AI124441), and Z.L. is supported by Princeton Center for Theoretical Science and Center for the Physics of Biological Function, the National Science Foundation Physics Frontier Center grant through the Center for the Physics of Biological Function (PHY-1734030), and an NSF

grant (PHY-1607612). **Author contributions:** M.S.D., J.Z., Z.L., and R.T.H. designed the study. J.Z., Z.L., M.D.T., J.D., and M.S.D. performed experiments and analyzed the data. M.S.D., J.Z., Z.L., and M.D.T. wrote the manuscript. J.D. and R.T.H. edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The "*Ca*. E. kahalalidefaciens" genome has been deposited to the Integrated Microbial Genomes (Joint Genome Institute, U.S. Department of Energy) public repository under IMG submission ID 115642.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/eaaw6732/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S3
References (*45*–*58*)
Data S1

15 January 2019; accepted 7 May 2019
10.1126/science.aaw6732

RESEARCH ARTICLE

MEMBRANES

Large-area graphene-nanomesh/carbon-nanotube hybrid membranes for ionic and molecular nanofiltration

Yanbing Yang[1]*, Xiangdong Yang[2]*, Ling Liang[3], Yuyan Gao[4], Huanyu Cheng[4], Xinming Li[5], Mingchu Zou[6], Anyuan Cao[6], Renzhi Ma[5], Quan Yuan[1,3]†, Xiangfeng Duan[7]†

Nanoporous two-dimensional materials are attractive for ionic and molecular nanofiltration but limited by insufficient mechanical strength over large areas. We report a large-area graphene-nanomesh/single-walled carbon nanotube (GNM/SWNT) hybrid membrane with excellent mechanical strength while fully capturing the merit of atomically thin membranes. The monolayer GNM features high-density, subnanometer pores for efficient transport of water molecules while blocking solute ions or molecules to enable size-selective separation. The SWNT network physically separates the GNM into microsized islands and acts as the microscopic framework to support the GNM, thus ensuring the structural integrity of the atomically thin GNM. The resulting GNM/SWNT membranes show high water permeance and a high rejection ratio for salt ions or organic molecules, and they retain stable separation performance in tubular modules.

Water desalination and purification is an attractive pathway to fresh water because seawater represents the largest fraction of water on Earth (*1*, *2*). An ideal water-treatment membrane should exhibit several critical characteristics: (i) minimal thinness to maximize permeance (*3*, *4*); (ii) sufficient mechanical strength to avoid breakage and solute leakage (*5*–*7*); and (iii) uniform and narrow pore size distribution for efficient separation (*8*, *9*). To this end, nanoporous two-dimensional (2D) materials of single- or few-atom thickness and with excellent mechanical strength have been considered as the ideal building blocks for constructing ultrathin membranes with minimum transport resistance and maximum permeance (*10*–*14*). Theoretical calculations have predicted that single-layer nanoporous 2D membranes can provide ultrafast water permeation and selective separation (*15*, *16*). Experimental studies of nanoporous graphene membrane have also demonstrated its exceptional performance in water desalination. However, the experimental studies to date are typically limited to proof-of-concept demonstrations on micrometer-scale graphene flakes (10^{-6} to 10^{-8} cm^2).

Although defect-free graphene exhibits exceptional mechanical performance, the inevitable in-plane grain boundaries in large-area graphene could seriously weaken the mechanical strength, and the introduction of pores could further compromise the structural integrity of single-layer graphene (*17*, *18*). Because the desalination process relies on the molecular-level separation of solute ions from water molecules, any slight tearing or cracking of the membrane could undermine the entire desalination system. As the stress of a selective membrane scales with $d^{-3/2}$ (where d is the membrane thickness), the atomically thin 2D membrane that is three orders of magnitude thinner than the commercial membrane would undergo substantially larger stress (*19*). Therefore, the applications of ultrathin 2D membranes for practical water treatment remain rather elusive because of the challenges in reliably producing large-area nanoporous 2D membranes of sufficient mechanical strength.

We report the design of an atomically thin nanoporous membrane with a single-layer graphene nanomesh (GNM) supported by an interwoven network of single-walled carbon nanotubes (SWNTs) (Fig. 1). In this structure, the mechanically strong, interconnected SWNT webs (*20*) feature a strong π–π interaction with the supported GNM, physically separate the GNM into microsized islands, and act as a microscopic framework to support the GNM (Fig. 1A). This construct can be viewed as Voronoi cells (*21*, *22*) of a Voronoi diagram (for example, as commonly seen in insect wings or plant leaves; fig. S1) defined in the mathematical structure models, thus ensuring the structural integrity of the atomically thin GNM over macroscopic scale. The large-area, ultrathin GNM/SWNT hybrid membrane can serve as an excellent size-exclusion nanofiltration membrane. In particular, the high-density subnanometer pores in the GNM layer allow efficient transport of water molecules with minimum resistance while blocking solute ions or molecules to enable selective separation (Fig. 1B), and the high mechanical strength of the GNM/SWNT hybrid membrane can prevent tear and solute leakage to ensure robust water treatment over large areas.

Fabrication and structural characterization of GNM/SWNT hybrid membranes

Figure 2A shows the fabrication process of the GNM/SWNT hybrid membranes. Briefly, chemical vapor deposition (CVD)–grown single-layer graphene on Cu foil was used as the starting material. A layer of SWNT membrane consisting

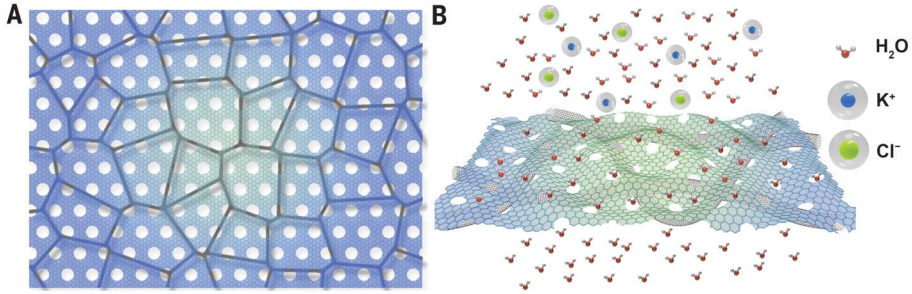


Fig. 1. Schematic illustration of the mechanically strong, large-area GNM/SWNT hybrid membrane for efficient water desalination. (**A**) Designed structural model of the GNM/SWNT hybrid membrane with single-layer GNM supported on SWNT networks. (**B**) Structural model of the GNM/SWNT hybrid membrane for size exclusion nanofiltration.

[1]College of Chemistry and Molecular Sciences, Key Laboratory of Analytical Chemistry for Biology and Medicine (Ministry of Education), Wuhan University, Wuhan 430072, China. [2]State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China. [3]Molecular Science and Biomedicine Laboratory, State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China. [4]Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA 16802, USA. [5]International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan. [6]Department of Materials Science and Engineering, College of Engineering, Peking University, Beijing 100871, China. [7]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: yuanquan@whu.edu.cn (Q.Y.); xduan@chem.ucla.edu (X.D.)

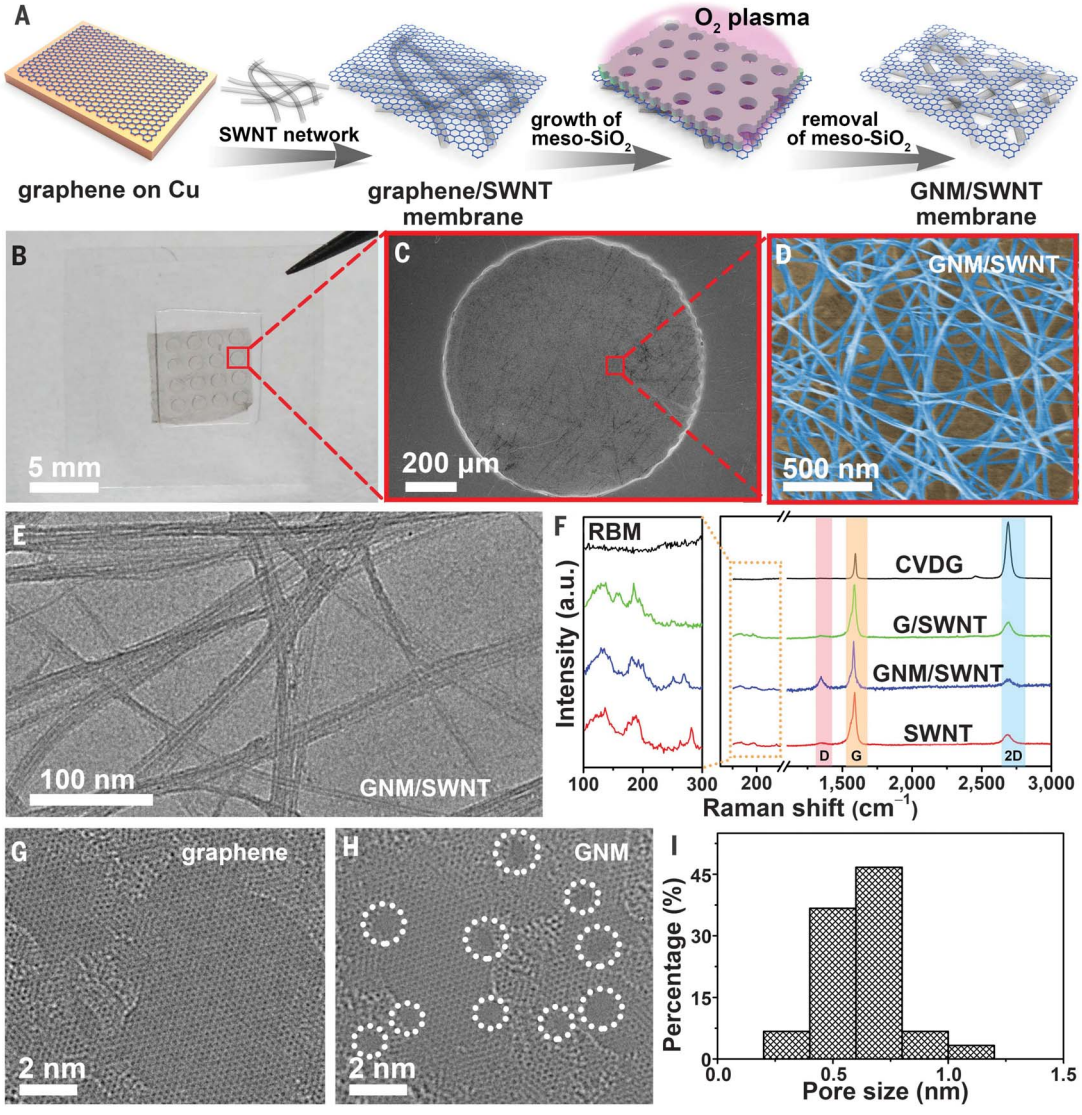


Fig. 2. Fabrication of GNM/SWNT hybrid membranes and structural characterizations. (**A**) Schematic illustration of the process used to fabricate GNM/SWNT hybrid membranes. (**B**) Photograph of the GNM/SWNT hybrid membrane suspended on a porous PDMS substrate with 16 holes (1 mm in diameter). (**C**) SEM image of a single hole in (B). (**D**) Magnified SEM image of the selected red area in (C). (**E**) TEM image of the GNM/SWNT hybrid membrane. (**F**) Raman spectra of CVDG, G/SWNT, GNM/SWNT, and SWNT membranes. (**G** and **H**) Aberration-corrected STEM image of graphene (G) and GNM (H) after 10 s of O_2 plasma etching. The white dashed circles highlight the pores present in GNM. (**I**) Pore-size distribution of the GNM prepared by 10 s of O_2 plasma etching.

of interconnected SWNTs was transferred on top of the graphene. After etching the Cu foils, a freestanding membrane with SWNT-supported graphene was obtained. To fabricate the GNM/SWNT membrane, we grew a layer of mesoporous SiO_2 (meso-SiO_2) film with uniform and perpendicular mesoporous channels (fig. S2) on the graphene surface as the porous template (*23*). By sequentially drilling pores in the graphene using O_2 plasma exposure and removing the meso-SiO_2 layer by hydrofluoric acid etching, we obtained a freestanding GNM/SWNT hybrid membrane, which can be suspended over a porous polydimethylsiloxane (PDMS) substrate. The suspended GNM/SWNT membrane retains the structural integrity with no obvious holes or tears (Fig. 2, B and C), as also confirmed by high-resolution scanning electron microscope (SEM) imaging (Fig. 2D). Transmission electron microscope (TEM) studies (Fig. 2E) further confirmed that the SWNT network intimately integrates with the GNM layer to form a mechanically strong, intact structure, whereas the pure graphene membrane readily cracks with obvious tears across the plane (fig. S3).

Raman spectroscopy was performed to evaluate the quality of the membranes. The pristine CVD-grown graphene (CVDG) showed clearly the characteristic 2D (2656 cm^{-1}) and G (1585 cm^{-1}) peaks with a high peak intensity ratio (I_{2D}/I_G~2) and no apparent D peak (1350 cm^{-1}) (Fig. 2F), suggesting that the pristine graphene exhibits a defect-free, single-layer characteristic. The presence of radial breathing modes at 100 to 300 cm^{-1} and the splitting of the G band confirm the existence of SWNTs in the graphene/SWNT (G/SWNT) membrane. After we introduced nanopores by O_2 plasma, the intensity of the D peak increased substantially to about one-third of that of the G peak (I_D/I_G~0.33) and the 2D peak intensity decreased, indicating the formation of defects in the GNM/SWNT membrane.

Scanning transmission electron microscopy (STEM) studies of the pristine graphene showed a honeycomb lattice of carbon atoms (Fig. 2G), whereas the STEM image of a GNM revealed the presence of subnanometer pores (Fig. 2H). The average pore size and pore density were

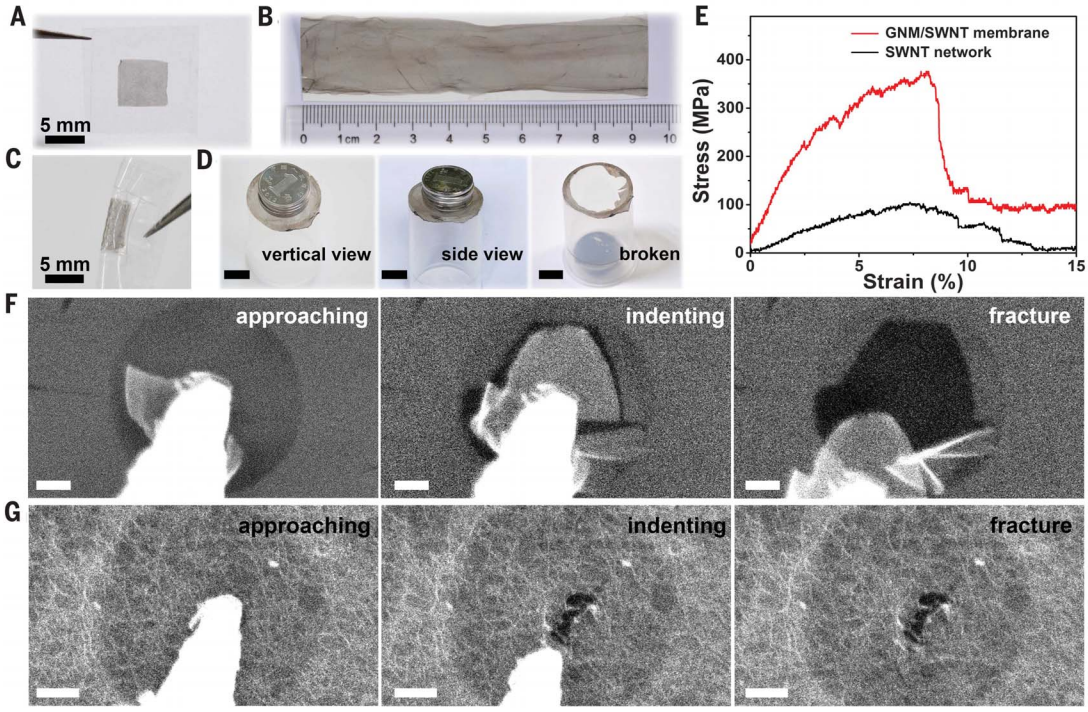


Fig. 3. Analysis of mechanical performance. (**A**) Photograph of the GNM/SWNT hybrid membrane suspended on a PET substrate with a 0.36-cm^2 hole. (**B**) Photograph of the large-area GNM/SWNT membrane. (**C**) Optical image of the GNM/SWNT hybrid membrane under bending conditions. (**D**) Photograph of the GNM/SWNT hybrid membrane suspended on a tube with six coins on the membrane. Scale bars, 1 cm. (**E**) Stress–strain curves of the SWNT membrane and the GNM/SWNT hybrid membrane under uniaxial tensile strain. (**F** and **G**) Rupture behavior of the (F) GNM and (G) GNM/SWNT hybrid membrane imaged by in situ SEM after poking with a micromanipulator. Scale bars, 0.5 μm (F) and 1 μm (G).

measured to be 0.63 nm and ~1.0 × 10^{12} cm^{-2}, respectively, in the GNM obtained with 10 s of O_2 plasma etching (Fig. 2I). The pore size matches well with the predicted optimal pore size for allowing water transport (0.32 nm) while effectively rejecting salt ions (~0.7 nm) (*17*). The pore size and pore density can be readily tailored by the O_2 plasma–etching time (fig. S4). A shorter etching time of 5 s results in an average pore size of ~0.55 nm, and an etching time of 20 s results in an average pore size of ~1.41 nm.

Mechanical characterization of GNM/SWNT hybrid membranes

The centimeter-scale GNM/SWNT hybrid membrane can be directly transferred onto a variety of substrates without a polymer support. A GNM/SWNT hybrid membrane suspended on a polyethylene terephthalate (PET) substrate with a 0.36-cm^2 square hole retains structural integrity without obvious cracks (Fig. 3A). The GNM/SWNT membrane may be readily scaled up (Fig. 3B) because there is no fundamental limitation in producing large-area graphene, the SWNT network, or meso-SiO_2 film up to meter scale. The flexible, freestanding SWNT network renders excellent mechanical strength and flexibility in the GNM/SWNT hybrid membrane, allowing it to endure large deformations without compromising the structural integrity (Fig. 3C). A hybrid membrane suspended on a tube edge is mechanically strong enough to support five coins (~16.0 g) without rupture (Fig. 3D).

We further investigated the tensile strength of the ultrathin GNM/SWNT hybrid membrane (*24*) (Fig. 3E and figs. S5 and S6). The SWNT and GNM/SWNT hybrid membranes showed similar fracture strains (8% and 9% for SWNT and GNM/SWNT membranes, respectively) (Fig. 3E). The pristine SWNT membrane could sustain a stress of 101.9 MPa (Fig. 3E) and the Young's modulus was calculated to be 2.6 GPa (fig. S7). By comparison, the GNM/SWNT membrane showed an enhanced mechanical strength to sustain a stress of 380.6 MPa (Fig. 3E) and had a considerably higher Young's modulus of 9.7 GPa (fig. S7). Additionally, atomic force microscope measurements indicated that the GNM/SWNT membrane has a modulus of ~5 to 10 GPa in the perpendicular direction (fig. S8). The mechanical strength is largely attributed to the Voronoi diagram structure of the GNM/SWNT network hybrid membrane, as confirmed by our mechanical simulations showing greatly enhanced tensile (2.4-fold) and bending stiffness (four orders of magnitude) of the GNM/SWNT membrane compared with the GNM membrane (figs. S9 to S12) (*24*).

In situ SEM imaging was conducted to visualize crack formation after punching a hole with a micromanipulator (Fig. 3, F and G). The GNM membrane quickly cracked into small pieces when a hole was punched. By contrast, the GNM/SWNT membrane maintained structural integrity during the entire process.

Water permeance and salt rejection under osmotic pressure

We explored the water-transport and salt-rejection properties of the GNM/SWNT hybrid membranes using three different configurations. The first configuration is based on cross-flow forward osmosis (FO) measurements in which the GNM/SWNT hybrid membrane suspended on a PET substrate (with a 0.16-cm^2 square aperture in the PET) separates two cavities, one filled with KCl solutions with different concentrations and the other with deionized water (Fig. 4A and fig. S13) (*24*). The osmotic pressure difference induced by the salt concentration gradient serves as the driving force for water transport. No apparent delamination between the SWNT network and the GNM membrane was observed in the cross-flow operation process or in the in situ SEM manipulations (Fig. 3G, fig. S14, and movie S1), confirming that the interaction between the SWNT and GNM is strong enough to resist the water flow under desalination conditions. On the basis of continuum mechanics, the maximum pressure (*P*) that a GNM/SWNT membrane with a pore size of 0.63 nm could withstand is calculated to be ~60 MPa (*24*) (Fig. 4B), consistent with previous theoretical calculations that the presence of a porous substrate (with openings <1 μm) would enable the GNM to sustain pressures exceeding 57 MPa (*19*). These analyses suggest that the GNM/SWNT hybrid membrane exhibits sufficient mechanical strength to be used as an effective semipermeable membrane for water desalination.

Water permeation studies showed that the G/SWNT membrane exhibits negligible water permeance (0.67 liter m^{-2} hour^{-1} bar^{-1}) (Fig. 4C and fig. S15), suggesting that the SWNT membrane maintains the structural integrity of graphene to prevent tearing and leakage. The water permeance for the GNM/SWNT hybrid membrane exhibits a clear dependence on the O_2 plasma–etching time (which correlates to the pore size). The GNM/SWNT membrane obtained with 5 s of etching (GNM/SWNT-5 s) showed a permeance of 7.5 liter m^{-2} hour^{-1} bar^{-1}. As the etching time was increased to 10 s (GNM/SWNT-10 s), the permeance increased to 20.6 liter m^{-2} hour^{-1}

bar^{-1} (5.7×10^{-16} g s^{-1} bar^{-1} per pore, assuming a pore density of 1.0×10^{12} cm^{-2}). This value is of the same order as that predicted by molecular dynamics simulations (*15*). A further increase of the etching time to 20 s (GNM/SWNT-20 s) resulted in a water permeance of 37.2 liter m^{-2} $hour^{-1}$ bar^{-1}, indicating reduced transport resistance with increasing pore size in the GNM.

Compared with the water flux of the commercial cellulose triacetate (CTA) membrane, which becomes "self-limiting" at high salt concentrations, the water flux of the GNM/SWNT hybrid membrane showed a linear dependence versus the salt concentration (Fig. 4D and fig. S16), indicating that the internal concentration polarization (ICP) was nearly eliminated in the freestanding ultrathin GNM/SWNT membrane. The reduced osmotic pressure loss caused by ICP leads to a higher osmotic driving force and enhanced permeance.

The salt-rejection behavior was evaluated by measuring the salt permeance through the GNM/SWNT hybrid membrane (fig. S17). The salt permeance through the graphene/SWNT membrane was essentially zero (no transport of ions; Fig. 4C), suggesting that the graphene/SWNT membrane remains intact over large areas. The salt permeance of the GNM/SWNT membrane varied from 5.5 to 16.2 mol m^{-2} $hour^{-1}$ when the O_2 plasma–etching time was increased from 5 to 20 s, indicating that the ion selectivity is highly dependent on the pore size. After 24 hours of permeation, the salt rejection for the GNM/SWNT membrane with 10 s of O_2 plasma–etching time remained >97% (Fig. 4E).

The separation performance of the GNM/SWNT membrane was also evaluated with NaCl as a seawater model. The measured salt rejection for NaCl was 98.1 ± 0.3% (fig. S18), which is comparable to that of KCl (97.1 ± 0.6%). The slightly improved salt rejection may be attributed to the relatively larger hydrated diameter of Na^+ (0.716 nm) compared with K^+ (0.662 nm). The minor salt leakage could be attributed to the presence of a small fraction of relatively large pores (~1 nm) in the GNM/SWNT membrane and the intrinsic defects or cracks formed during the graphene growth or transfer process. To this end, the presence of the SWNT network is particularly useful in separating the GNM into small domains and preventing the propagation of cracks and catastrophic failure.

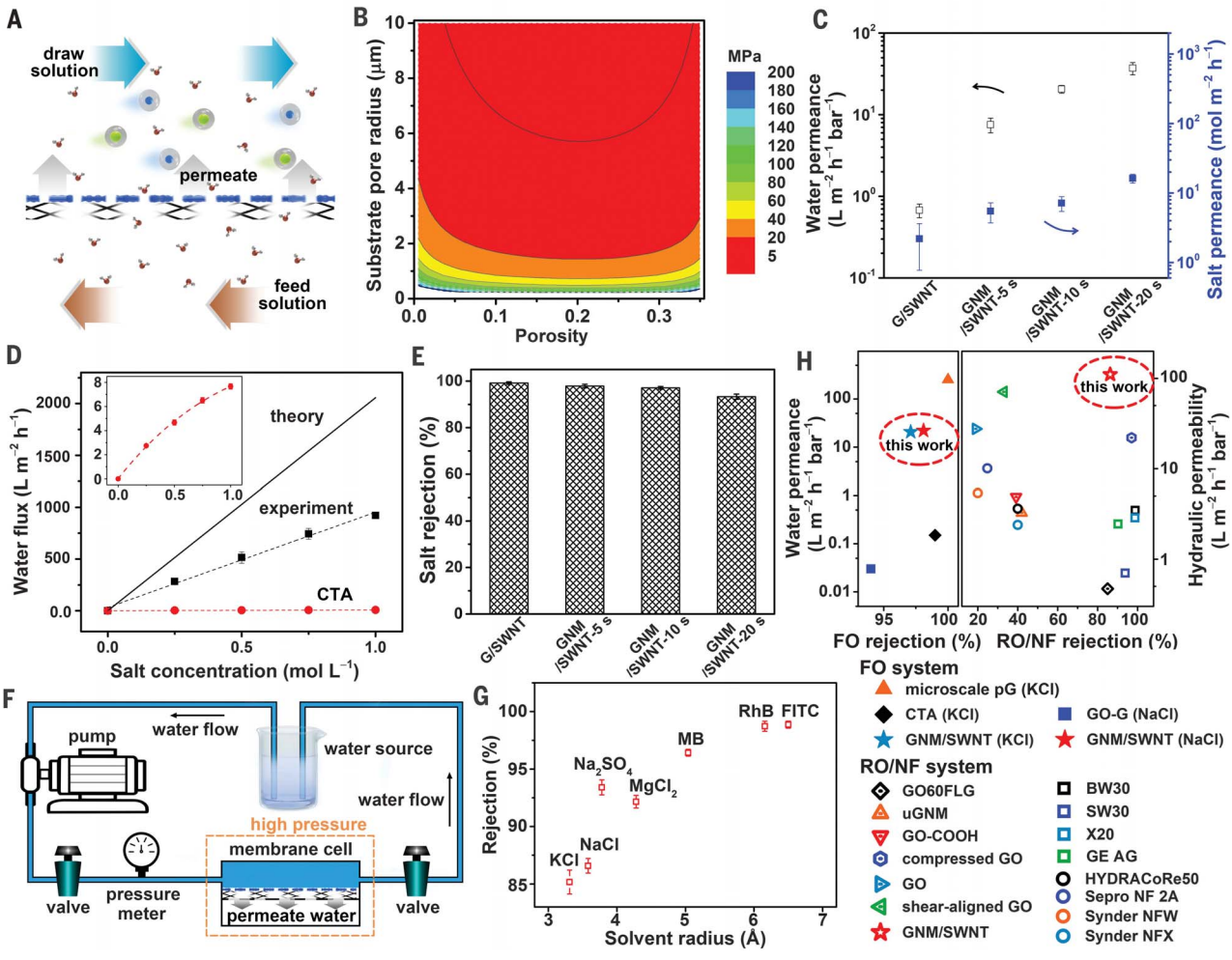


Fig. 4. Evaluation of water-desalination performance. (**A**) Sketch of water permeance through the GNM/SWNT membrane driven by osmotic pressure in the FO cross-flow system. (**B**) Contour plot of the maximum pressure versus GNM porosity and SWNT pore radius for GNM/SWNT-10 s. (**C**) Water and salt permeance through G/SWNT and GNM/SWNT membranes with O_2 plasma–etching times of 5, 10, and 20 s. (**D**) Water flux through the GNM/SWNT hybrid membrane and CTA membrane as a function of KCl concentration. Inset, the magnified view of the water flux of the CTA membrane. (**E**) Salt rejection of the G/SWNT and GNM/SWNT membranes prepared by O_2 plasma–etching times of 5, 10, and 20 s. (**F**) Schematic illustration of the RO cross-flow filtration apparatus. (**G**) Rejection of the GNM/SWNT membrane for KCl, NaCl, Na_2SO_4, $MgCl_2$, MB, RhB, and FITC. The error bars indicate the data acquired from three individual membranes. (**H**) Comparison of the water-permeability and salt-rejection performance of the GNM/SWNT hybrid membranes with commercial osmosis membranes and graphene-based separation membranes (*6*, *9*, *17*, *26*–*34*).

To further investigate the desalination performance of the GNM/SWNT membrane, we constructed a reverse osmosis (RO) cross-flow filtration apparatus (Fig. 4F) to measure the water- and salt-transport behaviors of the membrane toward Na_2SO_4, $MgCl_2$, NaCl, and KCl at a concentration of 2000 parts per million (ppm). As a nanofiltration membrane, the GNM/SWNT membrane could sustain a pressure in the range of ~2 to 4 MPa and ~8 to 10 MPa when supported by a stainless steel mesh (500 mesh, 30 μm pore size) and polycarbonate track etch membrane (0.2 μm pore size), suggesting that the GNM/SWNT membrane is mechanically strong enough to withstand typical commercial RO processes. The GNM/SWNT membrane showed high salt rejection between 85.2 and 93.4% (Fig. 4G), and the observed selectivity followed the order of Na_2SO_4 > $MgCl_2$ > NaCl > KCl. These observed salt-rejection ratios achieved in centimeter-scale GNM/SWNT membranes were generally smaller than those achieved in micrometer-scale nanoporous graphene flakes, which might be largely attributable to occasional defects or minor cracks in the large-area GNM/SWNT membranes. The hydraulic permeability of the GNM/SWNT membrane was maintained at 97.6 liters m^{-2} $hour^{-1}$ bar^{-1} for the NaCl solution compared with that of pure water (110.6 liters m^{-2} $hour^{-1}$ bar^{-1}) (fig. S19), further suggesting that the concentration polarization was effectively minimized. The increased membrane permeability could reduce the pressure necessary to drive efficient permeation and thus improve the energy efficiency of the desalination process. Additionally, the GNM/SWNT membrane exhibited higher water permeability and lower salt rejection in RO modules than in FO modules, and the exact reason for such differences is an intriguing topic for future investigations.

The GNM/SWNT membrane also exhibited a high rejection ratio for small charged or neutral molecules (96.4, 97.2, and 98.7% for methylene blue [MB], rhodamine B [RhB], and fluorescein isothiocyanate [FITC], respectively) with solvated diameters >1 nm (Fig. 4G and fig. S20), highlighting the merit of the GNM/SWNT membrane in rejecting small organic molecules, which is important for excluding small-molecule pharmaceuticals from drinking water, which typical commercial membranes usually fail to reject. The relatively high organic molecule exclusion performance and the decreased salt rejection with increasing Na_2SO_4 concentration (fig. S21) also suggest that the GNM/SWNT membrane is a nanofiltration membrane.

To gain further insight into the mechanism of salt rejection by the GNM/SWNT membrane, we also performed water-desalination measurements at different pH values. Little change was observed in salt rejection when the pH was varied from 7 to 3 (protonation of carboxylate groups at the pore edge), suggesting that the Gibbs–Donnan exclusion effect was negligible (figs. S22 to S24). Additionally, the GNM/SWNT membranes exhibited improved salt rejection for $MgCl_2$ (98.6%) compared with KCl (97.1%) (fig. S25) and an extremely low adsorption percentage (<0.5%; fig. S26) (*24*) for all of the investigated salts and organic molecules. These studies suggest that the salt-rejection performance of the GNM/SWNT membrane originates from (i) subnanometer-sized pores that facilitate effective separation by the size exclusion effect and (ii) the minimized concentration polarization due to the use of atomically thin nanoporous membrane in the cross-flow system.

We compared the permeability and selectivity of the GNM/SWNT water-desalination membranes with those reported in the literature (Fig. 4H and tables S1 and S2). The GNM/SWNT membrane exhibited a water permeance about one to two orders of magnitude higher than that of CTA, graphene oxide (GO)/graphene (*6*), and reduced GO membranes (*25*) at a comparable salt-rejection ratio. Nanoporous graphene (*17*) with high water permeance (252 liters m^{-2} $hour^{-1}$ bar^{-1}) has been reported previously, but only for micrometer-scale samples (10^{-6}–10^{-8} cm^2). Additionally, the GNM/SWNT membrane operated with the RO module showed about one to two orders of magnitude higher hydraulic permeability than that of commercial RO membranes, nanofiltration membranes, or graphene-based membranes (*9, 26–34*). The salt rejection of the GNM/SWNT membranes (>86% for NaCl) was lower than that for commercial RO membranes (90%~99%), but considerably higher than that for most graphene- or GO-based membranes (19 to 42%) (*26–29*) and commercial nanofiltration membranes (~24 to 40%) (*29, 33, 34*). The salt-rejection performance of the GNM/SWNT membranes could be further improved by optimizing the graphene growth conditions and the pore-formation process to minimize cracks and defects and to improve the pore size distribution.

The permeation and separation performance of our GNM/SWNT membranes was robust, and similar performance was observed in >100 studied membranes cut from large-area GNM/SWNT membranes. Compared with the CTA membrane, the GNM/SWNT membranes showed a high resistance to bacterial attachment for a long period of operation (figs. S27 and S28), indicating that the GNM/SWNT membranes exhibit excellent antibiofouling characteristics, which may be partly attributed to the smooth surface of the GNM (*35*) and the antibacterial performance of graphene (*36*).

Tubular water-desalination module

The commercial spiral-wound membranes are typically composed of stacked membranes scrolled into a tubular structure to enable a large contact area with feed solution and to further improve the water production output (*37*). Given the mechanical flexibility of the GNM/SWNT hybrid membranes, we investigated their water-desalination performance by bending the membrane to a specific curvature with a porous substrate (such as PDMS). The water-desalination device is composed of two stacked cylindrical silicone tubes, with the inner tube (0.16 cm^2 aperture present in the tube) incorporating a GNM/SWNT hybrid membrane (Fig. 5, A and B) (24). The control G/SWNT membrane exhibited negligible water and salt permeance that was nearly the same as that observed without bending (Fig. 5C and fig. S29), indicating that the G/SWNT membrane maintains the mechanical integrity well under bending conditions. The GNM/SWNT membrane showed a slight increase in permeance, by a factor of 1.2 and 1.6 for water and salt, respectively, indicating that the bending process may have induced the formation of a few small cracks. Despite the increased ion permeance, the GNM/SWNT membrane retained a salt rejection of up to 95.3% after 24 hours of osmotic operation (Fig. 5C), suggesting that the GNM/SWNT membrane is mechanically flexible enough to sustain

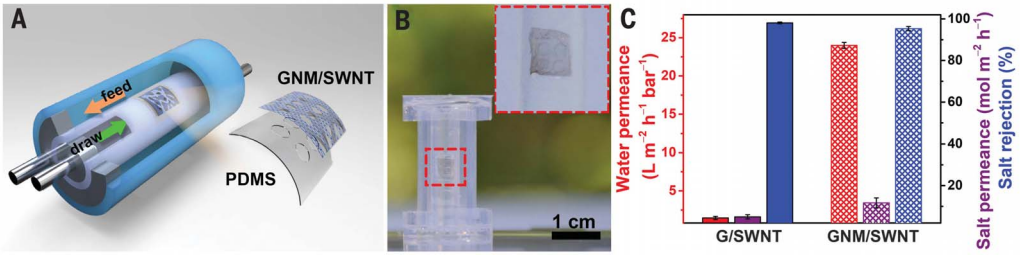


Fig. 5. Water-desalination performance of the bended membrane in a tubular module. (**A**) Schematic illustration and (**B**) photograph of the custom-assembled water-desalination cell for the measurement of permeance performance under bending conditions. Enlarged views show the corresponding structural model and photograph of the flexible GNM/SWNT membrane attached to the cylindrical silicone tube with a 0.16-cm^2 aperture. Scale bar, 1 cm. (**C**) Water and salt permeance and salt rejection of the G/SWNT and GNM/SWNT membranes under bending conditions. Error bars represent standard deviations of three independent measurements.

large deformation. The production efficiency can be further improved by packing the tubes incorporated with membranes into bundles with high packing density or scrolling the GNM/SWNT membrane into a spiral-wound structure. We further evaluated the water permeability and salt rejection of the GNM/SWNT membrane under different cross-flow velocities. The results indicated that the cross-flow velocity of 2 cm s^{-1} is the optimized condition for achieving high water permeance and salt rejection (fig. S30). At this velocity, the external concentration polarization is minimized while retaining sufficient membrane structural integrity.

We have designed a large-area, ultrathin GNM/SWNT hybrid membrane for highly efficient water purification. The macroscopic SWNT network helps to retain the structural integrity and improve the mechanical strength of monolayer GNM membrane, and the high-density subnanometer pores in the atomically thin GNMs ensure effective size-exclusion ionic/molecular nanofiltration and low permeation resistance. The GNM/SWNT hybrid membranes thus address the critical trade-off between water permeance and solute rejection in conventional desalination membranes. The high water permeance and excellent size selectivity combined with the excellent antifouling characteristics may make GNM/SWNT hybrid membranes highly attractive for energy-efficient and robust water treatment.

REFERENCES AND NOTES

1. A. Esfandiar *et al*., *Science* **358**, 511–513 (2017).
2. C. Klaysom, T. Y. Cath, T. Depuydt, I. F. J. Vankelecom, *Chem. Soc. Rev.* **42**, 6959–6989 (2013).
3. L.-C. Lin, J. C. Grossman, *Nat. Commun.* **6**, 8335 (2015).
4. S. Karan, Z. Jiang, A. G. Livingston, *Science* **348**, 1347–1351 (2015).
5. R. R. Nair, H. A. Wu, P. N. Jayaram, I. V. Grigorieva, A. K. Geim, *Science* **335**, 442–444 (2012).
6. J. Abraham *et al*., *Nat. Nanotechnol.* **12**, 546–550 (2017).
7. K. Goh *et al*., *Adv. Funct. Mater.* **25**, 7348–7359 (2015).
8. R. H. Tunuguntla *et al*., *Science* **357**, 792–796 (2017).
9. A. Morelos-Gomez *et al*., *Nat. Nanotechnol.* **12**, 1083–1088 (2017).
10. P. Sun, K. Wang, H. Zhu, *Adv. Mater.* **28**, 2287–2310 (2016).
11. J. Bai, X. Zhong, S. Jiang, Y. Huang, X. Duan, *Nat. Nanotechnol.* **5**, 190–194 (2010).
12. R. K. Joshi *et al*., *Science* **343**, 752–754 (2014).
13. B. Mi, *Science* **343**, 740–742 (2014).
14. L. Wang *et al*., *Nat. Nanotechnol.* **12**, 509–522 (2017).
15. D. Cohen-Tanugi, J. C. Grossman, *Nano Lett.* **12**, 3602–3608 (2012).
16. M. Heiranian, A. B. Farimani, N. R. Aluru, *Nat. Commun.* **6**, 8616 (2015).
17. S. P. Surwade *et al*., *Nat. Nanotechnol.* **10**, 459–464 (2015).
18. K. Celebi *et al*., *Science* **344**, 289–292 (2014).
19. D. Cohen-Tanugi, J. C. Grossman, *Nano Lett.* **14**, 6171–6178 (2014).
20. E. Shi *et al*., *Adv. Mater.* **27**, 682–688 (2015).
21. X. Lin *et al*., *Nat. Commun.* **4**, 2920 (2013).
22. J.-D. Boissonnat, F. Nielsen, R. Nock, *Discrete Comput. Geom.* **44**, 281–307 (2010).
23. Y. B. Yang *et al*., *Adv. Funct. Mater.* **27**, 1604096 (2017).
24. See supplementary materials.
25. H. Liu, H. Wang, X. Zhang, *Adv. Mater.* **27**, 249–254 (2015).
26. A. Akbari *et al*., *Nat. Commun.* **7**, 10891 (2016).
27. M. Hu, B. Mi, *Environ. Sci. Technol.* **47**, 3715–3723 (2013).
28. Y. Han, Z. Xu, C. Gao, *Adv. Funct. Mater.* **23**, 3693–3700 (2013).
29. Y. Q. Yuan *et al*., *Desalination* **405**, 29–39 (2017).
30. W. Li, W. Wu, Z. Li, *ACS Nano* **12**, 9309–9317 (2018).
31. W. Choi *et al*., *J. Membr. Sci.* **527**, 121–128 (2017).
32. Z. Jiang, S. Karan, A. G. Livingston, *Adv. Mater.* **30**, 1705973 (2018).
33. J. Y. Lin *et al*., *J. Membr. Sci.* **501**, 1–14 (2016).
34. Y.-J. Tang, Z.-L. Xu, S.-M. Xue, Y.-M. Wei, H. Yang, *J. Membr. Sci.* **498**, 374–384 (2016).
35. R. Yang, H. Jang, R. Stocker, K. K. Gleason, *Adv. Mater.* **26**, 1711–1718 (2014).
36. W. Hu *et al*., *ACS Nano* **4**, 4317–4323 (2010).
37. H. B. Park, J. Kamcev, L. M. Robeson, M. Elimelech, B. D. Freeman, *Science* **356**, eaab0530 (2017).

ACKNOWLEDGMENTS

Funding: We thank Y. Fang (National Center for Nanoscience and Technology, Beijing, China) for help with the synthesis of low-pressure CVD graphene. Q.Y. acknowledges support from the National Key Research and Development Program of China (2017YFA0208000), the National Natural Science Foundation of China (21675120), the Foundation for Innovative Research Groups of NSFC (21521063), the Ten Thousand Talents Program for Young Talents, the Start-up Research Fund (531107050973, 531109010053), and the State Key Laboratory of Chemo/Bio-Sensing and Chemometrics at Hunan University (734106172). H.C. is grateful for the start-up funds provided by The Pennsylvania State University. **Author contributions:** X.D., Q.Y., and A.C. proposed the research and supervised the project. Y.Y. and X.Y. designed and conducted the experiments and analyzed the experimental results. L.L. performed part of the structural characterization of the membranes. Y.G., H.C., X.L., and R.M. conducted the structural and simulation analysis of the membranes. M.Z. performed the growth of SWNT films and conducted the mechanical performance tests. Y.Y., X.Y., Q.Y., and X.D. cowrote the manuscript. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1057/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S30
Tables S1 and S2
References (*38*–*51*)
Movie S1

20 June 2018; resubmitted 25 November 2018
Accepted 3 May 2019
10.1126/science.aau5321

PHASE-CHANGE MEMORY

Femtosecond x-ray diffraction reveals a liquid–liquid phase transition in phase-change materials

Peter Zalden[1,2,3]*, Florian Quirin[4], Mathias Schumacher[5], Jan Siegel[6], Shuai Wei[7], Azize Koc[4,8], Matthieu Nicoul[4], Mariano Trigo[1,2], Pererik Andreasson[9], Henrik Enquist[9], Michael J. Shu[10], Tommaso Pardini[11], Matthieu Chollet[12], Diling Zhu[12], Henrik Lemke[12,13], Ider Ronneberger[5], Jörgen Larsson[9], Aaron M. Lindenberg[1,2,14], Henry E. Fischer[15], Stefan Hau-Riege[11], David A. Reis[1,2], Riccardo Mazzarello[5], Matthias Wuttig[7,16], Klaus Sokolowski-Tinten[4]*

In phase-change memory devices, a material is cycled between glassy and crystalline states. The highly temperature-dependent kinetics of its crystallization process enables application in memory technology, but the transition has not been resolved on an atomic scale. Using femtosecond x-ray diffraction and ab initio computer simulations, we determined the time-dependent pair-correlation function of phase-change materials throughout the melt-quenching and crystallization process. We found a liquid–liquid phase transition in the phase-change materials $Ag_4In_3Sb_{67}Te_{26}$ and $Ge_{15}Sb_{85}$ at 660 and 610 kelvin, respectively. The transition is predominantly caused by the onset of Peierls distortions, the amplitude of which correlates with an increase of the apparent activation energy of diffusivity. This reveals a relationship between atomic structure and kinetics, enabling a systematic optimization of the memory-switching kinetics.

The global amount of data grows exponentially (*1*) and alternative memory technologies are being considered to satisfy the resulting demands. Phase-change memory promises higher storage densities as well as faster retrieval rates due to the nondestructive readout compared with state-of-the-art memory technology (*2*–*5*). In phase-change memory, information storage occurs by cycling small volumes of the material between its glassy and crystalline states. An optically or electrically induced thermal stimulus allows switching between them. Crystallization is the time-limiting step as glass formation (vitrification) can be facilitated rapidly by a melt-quenching process. Nevertheless, even crystallization can take place within less than a nanosecond (*6*–*8*). Phase-change memory therefore relies on the kinetic contrast of the active materials: A high atomic mobility at elevated temperature allows quick crystallization, whereas low atomic mobility at ambient temperatures allows long-term data retention (*7*, *9*–*13*). Good glass-forming materials are generally characterized by a glass transition temperature T_g that is relatively close to the melting temperature T_m. This occurs for values of the Turnbull parameter $T_g/T_m \geq 2/3$ (*14*). Bad glass formers have typical values of $T_g/T_m \leq 1/2$, which enables rapid crystallization. Optimum performance of a phase-change material (PCM) cannot be achieved by focusing on any one of these criteria, and the temperature dependence of atomic mobility, i.e., its viscosity, must be understood in detail to enable a systematic optimization of kinetic properties suitable for phase-change memory devices.

The temperature dependence of viscosity differs among various glass-forming liquids as the temperature approaches the glass transition T_g. Although some liquids show an Arrhenius-like behavior and are classified as "strong" (e.g., silica), others display a range of non-Arrhenius behaviors and are classified as "fragile" (e.g., o-terphenyl) (*15*). In some anomalous liquids, a fragile-to-strong (FTS) crossover may occur, in which the high-temperature fragile liquid is transformed into a low-temperature strong liquid (*16*–*18*). The FTS crossover is usually associated with a maximum in thermodynamic response functions [e.g., heat capacity (C_p), thermal expansivity (α_P), and compressibility (κ_T)], which may be attributed to a phase transition between two liquid phases characterized by distinct physical properties (e.g., density and entropy) and different atomic structures. The FTS crossover was first proposed to explain the behavior of water, which is a fragile liquid down to the temperature of the homogeneous nucleation limit but behaves kinetically strong when heated from the solid amorphous state above T_g (*19*). Similar observations were made on the PCM Ag-In-Sb-Te in its liquid and solid amorphous states (*9*, *12*, *20*). However, in PCMs, liquid quenching is difficult owing to the rapid onset of crystallization. This limits the accessible supercooling range at common cooling rates (*21*), making observations of FTS crossovers or liquid–liquid phase transitions (LLPTs) difficult. The apparent FTS crossover in PCMs renders the hypothesis of LLPTs appealing and also implies anomalous structural, thermodynamic, and diffusive properties (*21*–*23*).

The microscopic mechanism underlying the crossover is unclear. Liquid $Ge_{15}Te_{85}$ shows an increase of medium-range order (MRO) upon supercooling, compatible with an increase of local atomic distortions (*24*), yet $Ge_{15}Te_{85}$ is a good glass former and thus cannot be used as a PCM. More generally, direct experimental evidence for LLPTs in systems that crystallize rapidly is rare because the time scales available to probe the supercooled liquid state before it crystallizes are short. Therefore, the name "no-man's land" was coined to refer to the highly supercooled region (150 to 236 K) (*25*) in the phase diagram of water, whose anomalous properties have long been debated (*26*). Only recently, ultrafast x-ray laser scattering has enabled directly probing supercooled water down to 227 K and revealed evidence of a continuous LLPT (*27*). From the computational side, LLPTs were observed previously in simulations of ST2 water (*28*, *29*), although the possibility of the existence of a true liquid–liquid critical point (LLCP) in the supercooled liquid has been challenged theoretically (*30*). Similar data are not yet available for PCMs, but an anomalous breakdown of the Stokes–Einstein relationship (SER) in the equilibrium liquid of $GeSb_2Te_4$ was recently observed and was speculated to be caused by an LLPT below T_m (*31*). We focused our attention on Sb-based PCMs because $Ag_4In_3Sb_{67}Te_{26}$ (AIST) (*9*, *12*, *20*) and Ge_9Sb_{91} (*32*) have crossovers in the apparent activation energy of viscosity. Both materials are used as PCMs in optical [AIST (*4*)] and electronic [$Ge_{15}Sb_{85}$ (*33*)] memory devices.

The liquid–liquid phase transition

We used hard x-ray laser pulses from an x-ray free electron laser (the Linear Coherent Light Source, LCLS) to overcome the temporal limitation of probing the atomic structure of highly supercooled PCMs. With the X-ray Pump Probe (XPP) instrument (*34*), optical laser pulses with

[1]Stanford PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Rd., Menlo Park, CA 94025, USA. [2]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, 2575 Sand Hill Rd., Menlo Park, CA 94025, USA. [3]European XFEL, Holzkoppel 4, 22869 Schenefeld, Germany. [4]Faculty of Physics and Center for Nanointegration Duisburg-Essen (CENIDE), University of Duisburg-Essen, Lotharstrasse 1, 47048 Duisburg, Germany. [5]Institut für Theoretische Festkörperphysik, JARA-FIT and JARA-HPC, RWTH Aachen University, Germany. [6]Instituto de Optica, CSIC, C/Serrano 121, 28006 Madrid, Spain. [7]I. Physikalisches Institut and JARA-FIT, RWTH Aachen, Sommerfeldstrasse 14, 52074 Aachen, Germany. [8]Institut für Physik und Astronomie, Universität Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany. [9]Department of Physics, Lund University, Professorsgatan 1, 223 62 Lund, Sweden. [10]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [11]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [12]Linac Coherent Light Source, SLAC National Accelerator Laboratory, 2575 Sand Hill Rd., Menlo Park, CA 94025, USA. [13]Paul Scherrer Institute, Forschungsstrasse 111, 5232 Villigen, Switzerland. [14]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [15]Institut Laue-Langevin, 71 Avenue des Martyrs, CS 20156, 38042 Grenoble Cedex 9, France. [16]PGI 10 (Green IT), Forschungszentrum Jülich, 52428 Jülich, Germany.
***Corresponding author. Email: klaus.sokolowski-tinten@uni-due.de (K.S.-T.); peter.zalden@xfel.eu (P.Z.)**

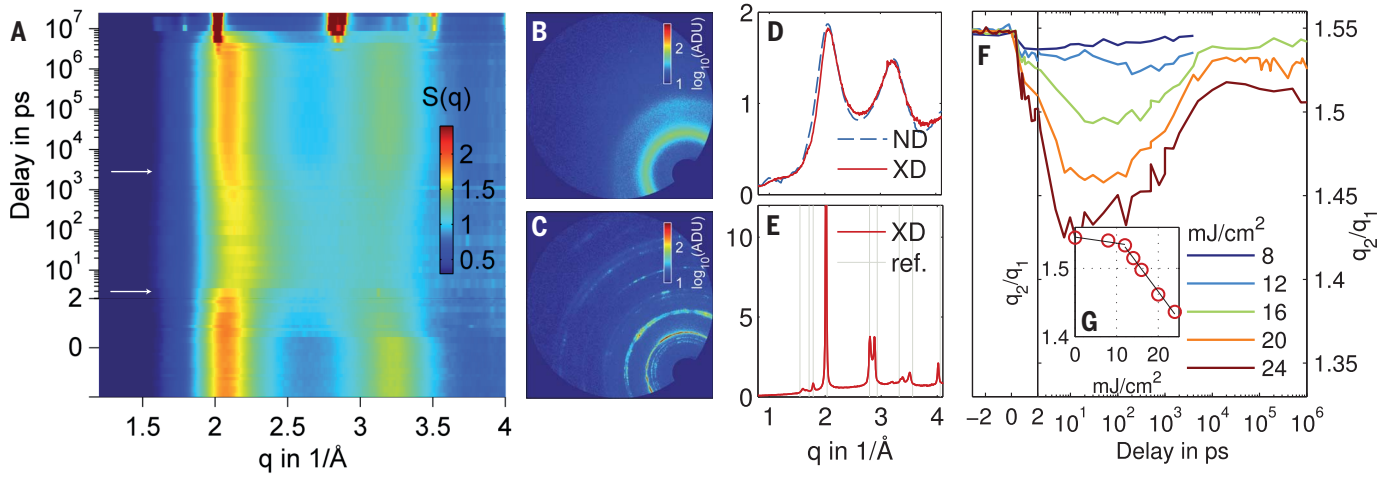


Fig. 1. Evolution of the atomic structure of AIST during the melt–quench cycle, resolved in reciprocal space. (**A**) Structure factor $S(q)$ of AIST after optical excitation with 24 mJ/cm^2 throughout the entire melt–quench cycle from the initial, as-deposited amorphous state (**B**) to the final crystalline state (**C**). The atomic structure factors of the initial and final states are in good agreement with reference data from total neutron scattering (ND) (*58*) and literature values for the crystalline structure (**D** and **E**). Femtosecond x-ray diffraction (XD) enables resolving an intermediate structural phase transition in the disordered state of AIST, as seen by the change in the ratio q_2/q_1 corresponding to a shift of reflections after a few picoseconds and again after a few nanoseconds (**F**). (**G**) Fluence dependence of the average q_2/q_1 after 10 to 100 ps showing the onset of the phase transition at 12 mJ/cm^2, also shown in (A) by white horizontal arrows.

800-nm center wavelength and 50-fs pulse duration are absorbed by a thin film of PCM at time t_0. Subsequently, electron–phonon coupling heats the PCM on the few-picosecond time scale (*35*, *36*), leading to melting at sufficiently high excitation fluences. The PCM is quenched by diffusive thermal transport into the supporting membrane of Si_3N_4, which is of similar thickness as the PCM. To probe the atomic structure during this melt–quench cycle, we collected the diffraction patterns of x-ray pulses with 1.305-Å center wavelength on a two-dimensional area detector (fig. S1) (*37*). We varied the delay Δt between the optical pump event at t_0 and the x-ray probe event at time t, $\Delta t = t - t_0$, between negative values and tens of microseconds. We performed each set of pump–probe events on a new spot of the sample, which was irreversibly modified as a result of the intense optical and x-ray pulses. Nevertheless, the x-ray probe pulse duration of 50 fs is sufficiently short to ensure that the atomic structure does not change during the probe interaction.

We found that the structure factor $S(q)$ of amorphous AIST was dominated by two broad diffraction rings centered at q_1 = 2.07 (1) Å^{-1} and q_2 = 3.21 (2) Å^{-1}, clearly visible in Fig. 1, A, B, and D. At $\Delta t < 0$, the diffraction pattern of the unpumped, as-deposited amorphous structure is recorded at the initial temperature of 298 K. At Δt of a few picoseconds, the peak intensity of both rings decreases and their radii approach each other. Consistently, the ratio q_2/q_1 depicted in Fig. 1F decreases within the first picoseconds, but almost recovers to the initial value of the amorphous state after a few nanoseconds. This behavior is strong evidence for a structural transition in a disordered state beyond just a thermally induced reduction of the scattering efficiency (Debye–Waller), which would leave the peak positions unaffected. Further evidence for a phase-transition behavior comes from the highly nonlinear scaling of the peak structural modification with the pump fluence (Fig. 1G). The finding that crystallization eventually occurs at fluences above 14 mJ/cm^2 (see fig. S4) enables us to derive a lower bound for the temperature jump several microseconds after optical excitation. At the time of crystallization (5 µs), the temperature in the PCM must be above the temperature at which crystallization sets in at calorimetric heating rates of a few kelvin per minute. In AIST, this temperature is 430 K (*38*). A comparison with literature data on AIST (vertical gray lines) shows good agreement of the momentum transfer associated with the strongest reflections (Fig. 1E).

To quantify the temperature evolution of our samples more accurately, we performed finite element simulations, which allow temperatures to be associated to the structural information that we obtained at various time delays and fluences. We show the resulting cooling behavior of the 50-nm–thick film of AIST on 50-nm–thick membranes (red curves) based on the normalized temperature $\Theta = \frac{T-T_0}{T_m-T_0}$, where T_m is the melting temperature, 810 K for AIST (*39*), and T_0 is the initial temperature of 298 K (Fig. 2). After the heating by electron–phonon coupling, the heat diffuses from the PCM into the Si_3N_4 membrane after a few hundreds of picoseconds. Their temperature equilibrates after a few nanoseconds. Subsequently, thermal transport into the frame supporting the membranes becomes the dominant cooling mechanism. The experimental data (Fig. 1F) reflect the impact of the two different length scales of thermal transport associated with the short out-of-plane and the long in-plane axes. The time scale of out-of-plane transport is in good agreement with the change of atomic structure in the disordered state at ~1 ns. We observed in-plane thermal transport only after delays longer than 10 µs, rendering it irrelevant for this work. We derived two temporal intervals ($\{t_a\}$ and $\{t_b\}$, indicated in Fig. 2), independent of the pump fluence, over which the temperature of the PCM was stable within 10%. During these intervals, we can predict the temperature of the PCM most accurately because it depends predominantly on the specific heat of the materials involved. We found the same intervals for 60-nm–thick $Ge_{15}Sb_{85}$ [T_m = 860 K (*40*)] films on 150-nm–thick membranes.

The atomic structure during $\{t_a\}$ and $\{t_b\}$ is described by the radii of the first (r_1) and second (r_2) coordination shells. We determined the radii from the pair-correlation functions $g(r)$ (Fig. 3A) and they revealed a pronounced increase of r_1 upon excitation. The ratio $R = r_2/r_1$ provides information about the structural ordering and we measured its evolution with time and fluence (Fig. 3, B and C).

On the basis of these numbers and in combination with the thermal simulations, we determined R for AIST and $Ge_{15}Sb_{85}$ as a function of inverse temperature T_m/T (Fig. 4A). For each fluence of the pump laser, we derived one value of R from the intervals $\{t_a\}$ and $\{t_b\}$. Additionally, we derived one average value from the data at negative delays. Direct evidence for a structural transition in the supercooled liquids of AIST and $Ge_{15}Sb_{85}$ comes from a change in slope of $R(T)$ in the range of T_m/T between 1 and 1.5. This resembles the $R(T)$ behavior in the good glass former $Ge_{15}Te_{85}$ that also has an FTS crossover and a structural phase transition that correlate (*24*), albeit at higher temperatures. We interpolated our data with an error function that allowed us to determine the temperature of the

structural transition by the maximum slope in r_2/r_1 occurring at 660 ± 20 K and 610 ± 20 K for AIST and $Ge_{15}Sb_{85}$, respectively. We estimate the experimental uncertainty as ±20 K SD in our experiment mostly caused by statistical fluctuations in the pump fluence. A normalized version of the same error function fits the prepeak intensity I_{pp} (Fig. 4B and fig. S3) in $S(q)$, which was located at q = 1.08 ± 0.02 and 1.06 ± 0.02 $Å^{-1}$ for AIST and $Ge_{15}Sb_{85}$, respectively. They correspond to the formation of periodic structures in real space with 5.8 and 5.9 Å, which is twice the radius of the first coordination shell r_1. This periodicity is due to the formation of alternating long and short bonds on opposite sides of a central atom, which is a characteristic fingerprint of a Peierls distortion (*41*). The pronounced increase of $g(r)$ around the second coordination shell indicates that the distribution of these interatomic distances, being the hypotenuses of ~90° bond angles, becomes narrower (figs. S7 and S8). The temperature dependence of R and I_{pp} matches the temperature dependence of the apparent activation energy of inverse diffusivity D^{-1} (Fig. 4C), which corresponds to the apparent activation energy of viscosity η, assuming that the SER is valid.

$$E_A = k_B \times \frac{\partial \ln(D_0/D)}{\partial(1/T)}.$$

We derived the diffusivities (Fig. 4D) from previous measurements of the viscosity above T_m for AIST (*42*) and $Ge_{15}Te_{85}$ (*43*). In the case of AIST, we also calculated diffusivities from the crystal growth velocity in the supercooled liquid state (*12*). Because of the uncertainty regarding the validity of the SER in the supercooled regime, we prefer to provide diffusivity data, which, unlike viscosity data, can be directly calculated from crystal growth velocities (*9*) and require the assumption of a valid SER only for the equilibrium liquid.

Two different temperatures characterized the correlation of apparent activation energies and atomic structure. The structural transition temperature (T_{LL}) occurs at the inflection points of R and I_{pp} (Fig. 4, A and B). In the non-PCM $Ge_{15}Te_{85}$, it occurs at T_{LL}/T_m = 1.05 (*24*, *44*), whereas the LLPT of the PCMs AIST and $Ge_{15}Sb_{85}$ takes place at T_{LL}/T_m = 0.81 ± 0.01 and 0.71 ± 0.01, respectively. T_{LL} coincides with the temperature at which the apparent activation energy reached ~20% of that for the low-temperature state, as shown in Fig. 4C for AIST and $Ge_{15}Te_{85}$. For $Ge_{15}Sb_{85}$, no kinetic data are available from literature. The FTS crossover temperature (T_{FTS}, Fig. 4, C and D) was reported to be 570 K in the case of AIST and in the vicinity of the melting temperature in $Ge_{15}Te_{85}$. It coincides with the temperature at which the atomic structure reaches ~90% of R for the low-temperature state in the case of AIST and $Ge_{15}Te_{85}$. In $Ge_{15}Sb_{85}$, this value of R would correspond to a temperature of 550 ± 20 K.

In this study, we provide direct evidence for the structural transition and emphasize its relevance for the kinetic properties, although the structural transition occurs at higher temperatures than the FTS crossover reported previously. The state for the kinetic data below the FTS crossover is still a matter of debate because of experimental difficulties in determining its kinetic properties. Orava *et al.* reported a strong liquid state of AIST based on T_g = 378 K for a standard cooling rate of 20 K/min (*20*). The kinetic data underlying this scenario, however, agree with data reported by Salinga *et al.* (*9*) for glass (Fig. 4D). In an alternative scenario in which the glass transition for similar standard rates is located at 450 K (*45*), the strong liquid is represented hypothetically by the dashed and dotted line in (Fig. 4D), similar to the case of $Ge_{15}Te_{85}$. In both scenarios, however, the structural transition observed in this study correlates with the kinetic crossover. To show that the structural transition at T_{LL} indeed occurs between two liquid states, it is necessary to discuss the role of the glass transition. Given the reported glass transition of AIST at 450 K (*45*) for calorimetric heating rates of ~1 K/s, an enormous cooling-rate dependence of T_g is required to explain a transition at 660 K. This requirement, however, is in contradiction to the invariant transition temperatures observed during the out-of-plane cooling during $\{t_a\}$ with a 10^{11} K/s cooling rate and the in-plane cooling during $\{t_b\}$ with a rate of 10^6 K/s (fig. S5c). Therefore, the structural transition observed here is an LLPT at temperature T_{LL}.

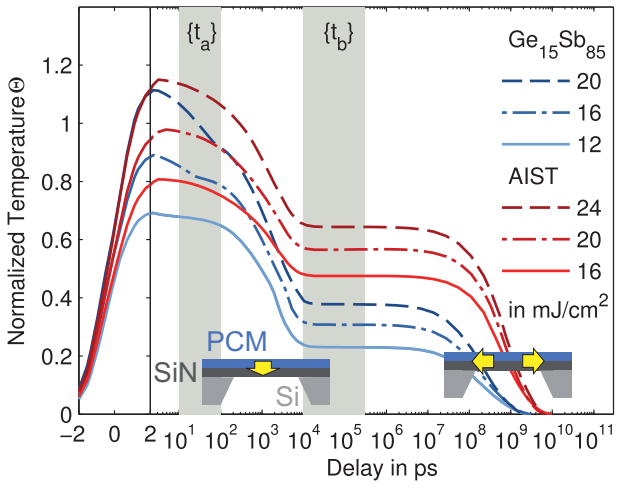


Fig. 2. Temporal evolution of the average temperature inside AIST (red curves) and $Ge_{15}Sb_{85}$ (blue curves) after optical excitation for different optical excitation conditions. We identified two intervals (gray areas), one between 10 and 100 ps $\{t_a\}$ and the other between 10 and 30 ns $\{t_b\}$, during which the temperature is stable within 10%. These time scales are associated with two thermal transport conditions, schematically depicted as insets (yellow arrows denote the main direction of heat flow) and are caused by the high aspect ratio of the samples.

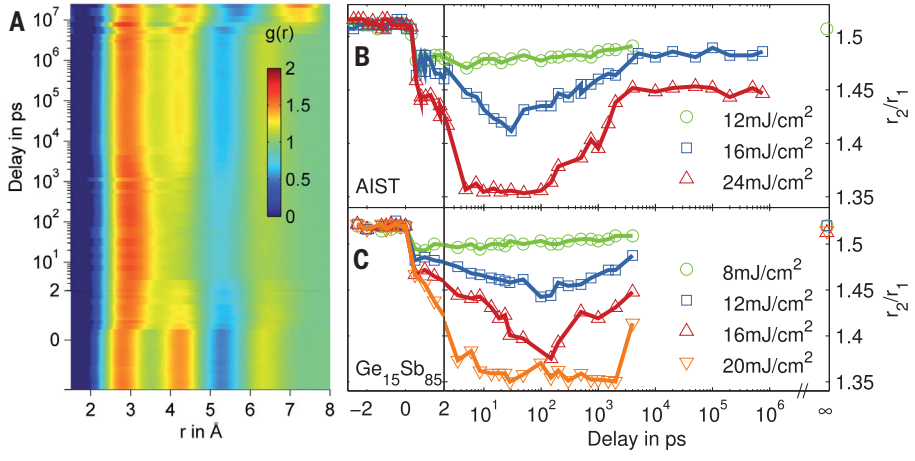


Fig. 3. Average local structure in AIST and $Ge_{15}Sb_{85}$ during the melt–quench process. (**A**) Atomic pair correlation functions $g(r)$ of AIST as a function of time after optical excitation with 24 mJ/cm^2. The structural transition is most clearly evidenced by a transient shift of the first coordination shell r_1 to longer distances between ≈1 ps and ≈5 ns, well before the onset of crystallization after ≈5 μs. Consequently, the structural parameter r_2/r_1 for AIST (**B**) and $Ge_{15}Sb_{85}$ (**C**) decreases and stays at the high-temperature value of 1.36 for several nanoseconds when the fluence is sufficiently high. If crystallization is avoided at low pump fluences below 14.5 mJ/cm^2, then r_2/r_1 eventually returns to its initial values (1.51 ± 0.01 in the case of AIST) in the final amorphous solid state for $t \to \infty$, when ambient temperature is restored.

Microscopic nature of the transition

Liquid PCMs are commonly found to be octahedrally coordinated (*46*), but their corresponding amorphous states contain several competing structural motifs. Local Peierls distortions of an octahedral environment cause the splitting of the first coordination shell into two subshells with unequal occupation of the first and second subshells (*47*). Although the Peierls distortion itself leaves R almost unchanged, the lower occupation of the second (larger) subshell will reduce r_1, thereby increasing R. This makes it impossible to distinguish this motif from an increase caused by the formation of tetrahedral sites, possibly together with homopolar bonds as reported for GeTe (*48*, *49*). Therefore, we require further information on the local structure to resolve the exact mechanism responsible for the LLPT. We performed ab initio molecular dynamics (AIMD) simulations and verified the simulations by comparing them with the temperature-dependent structure factors (fig. S6). We reproduced with our simulations the peak shift and the reduction of intensity. We found that R was less temperature dependent, which we attributed to the faster quench rate in the simulation (3×10^{12} K/s) (*13*). At higher quenching rates, the liquid falls out of equilibrium at higher temperatures, kinetically freezing-in an intermediate state. The lack of equilibration time at each temperature step means that the LLPT occurs over a wider temperature range. Nevertheless, the trend observed in AIMD was consistent with our experimental findings.

In the case of $Ge_{15}Sb_{85}$, 70% of the total number of bonds were formed between Sb atoms, which dominated the local order and the scattering signal due to their higher atomic weight. Histograms of interatomic distances from an Sb atom to its nth nearest Sb neighbor provide further insight. As Sb commonly is sixfold coordinated, the histograms reflect the inner structure within the first coordination shell. In the liquid state (Fig. 5B), all six histograms have equidistant peaks of similar width, indicating a regular octahedral environment with regular fluctuations due to the dynamics in the liquid state. After quenching (Fig. 5C), the n = 1 to 3 histograms shift to shorter distances, becoming narrower and higher. The n = 4 to 6 histograms get separated further and retain their width from the liquid state. This different behavior between the three shorter and three longer interatomic distances is evidence for the onset of a Peierls distortion, also evident from the angular limited bond correlation (ALTBC) plots (Fig. 5, E and F). The large variation of first-neighbor distances limits the discussion of R to a qualitative level, because the sixth-nearest neighbor distance in low-temperature $Ge_{15}Sb_{85}$ (Fig. 5C) is centered at 3.7 Å, which is closer to the center of mass for the second coordination shell (r_2 = 4.2 Å) than the first (r_1 = 2.9 Å). This means that the first and second coordination shells can no longer be resolved because of the inherent disorder. Although the Peierls distortion alone can explain the increase in R, the bond angles around Ge change from the octahedral value of 90° in the liquid state to a value of 105° at ambient conditions (fig. S14). This indicates the formation of tetrahedral sites (R = 1.61 for a purely tetrahedral structure), which also increase R. The conditions in AIST are comparable, but tetrahedral sites are formed during the quench only around a fraction of the 4% stoichiometric In atoms (*13*). Because AIST shows the same structural trend during the quench (Fig. 5A), a Peierls distortion must be the dominant mechanism responsible for the change in kinetic properties. With increasing Peierls distortion in supercooled liquid $Ge_{15}Sb_{85}$ and AIST, we also observed a tendency for a pseudogap (*50*) opening in the density of states (Fig. 5D and fig. S27). Supercooled liquids $Ge_{15}Te_{85}$ (*51*), As_2Se_3 (*52*), and As_2Te_3 (*53*) indeed undergo a metal–semiconductor transition during the quench, accompanied by a maximum of the thermodynamic response functions (e.g., C_p, α_P, and k_T), indicative of a fragile-to-strong crossover. Also, the pseudogap of liquid $Ge_2Sb_2Te_5$ exhibits an opening tendency during the quench, suggesting that it might also show an LLPT in the supercooled liquid regime (*54*).

Discussion and conclusion

We combined our observations and modeling to derive information on the change of the bonding mechanism underlying the LLPT. In the equilibrium liquid state, metallic bonds are nondirectional with a relatively low activation energy of viscosity. The formation of Peierls distortions during the quench localizes electronic charge between the atoms, constraining bond angles and increasing the activation energy of viscosity. This and the correlation between structure and kinetics imply that the LLPT constitutes the structural origin of the kinetic crossover. The tendency to form a Peierls distortion in the crystalline state

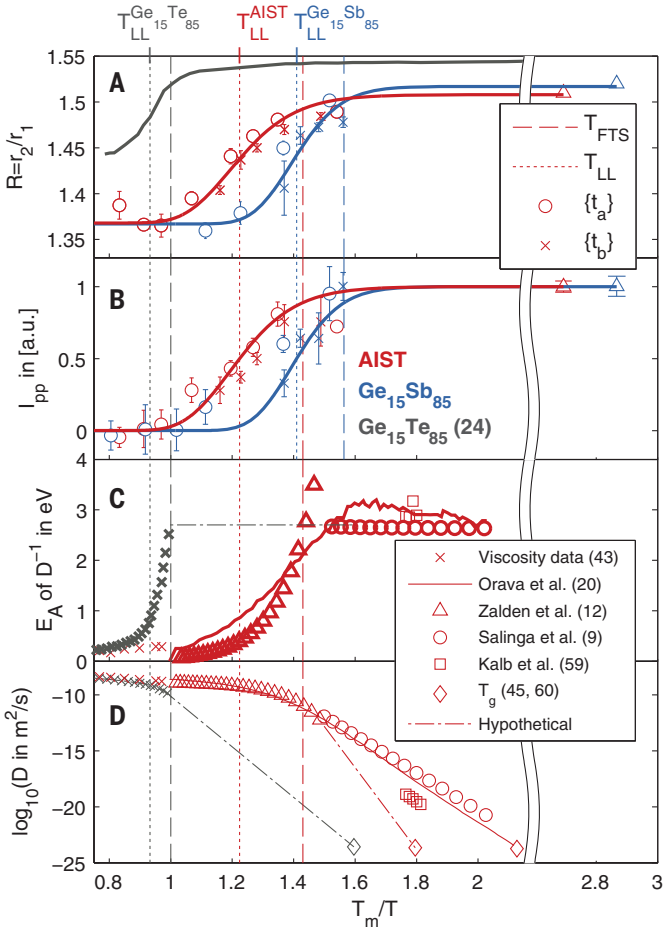


Fig. 4. Structural parameter $R = r_2/r_1$, the intensity of a pre-peak, and the apparent activation energy of diffusivity correlate in the case of AIST. (**A**) Ratio of second and first coordination shell radii, r_2/r_1, showing a transition for the two PCMs AIST (red) and $Ge_{15}Sb_{85}$ (blue) in the supercooled regime. Both were refined by an error function as a guide for the eye and to determine the structural transition temperatures T_{LL} (inflexion points) of 660 and 610 K, respectively, visualized by vertical dotted lines. Vertical dashed lines represent the temperature at which ~90% of the low-temperature structure is reached and coincide with the FTS crossover temperature T_{FTS}. Crosses and circles denote the data points obtained from $\{t_a\}$ and $\{t_b\}$, respectively, and triangles are obtained from the initial structures. A similar transition was reported for $Ge_{15}Te_{85}$ (gray) at the melting point (*24*). (**B**) Simultaneously, a prepeak is formed, which is related to additional MRO caused by a Peierls distortion. Both structural parameters are found to correlate with an increase of the apparent activation energy E_A of inverse diffusivity (**C**). The latter data are derived from measurements of viscosity above T_m [crosses (*43*)] and from crystal growth velocities in the supercooled liquid [triangles (*12*)] and glassy regime [circles (*9*) and squares (*59*)], as shown in (**D**). The solid line corresponds to the model proposed by Orava *et al*. describing the FTS crossover between two liquids (*20*). Glass transition temperatures (diamonds) are shown for $Ge_{15}Te_{85}$ (*60*) and for AIST (*45*). For the latter, literature values show a wide spread. Error bars in this figure correspond to the SD; errors on temperatures were omitted for clarity.

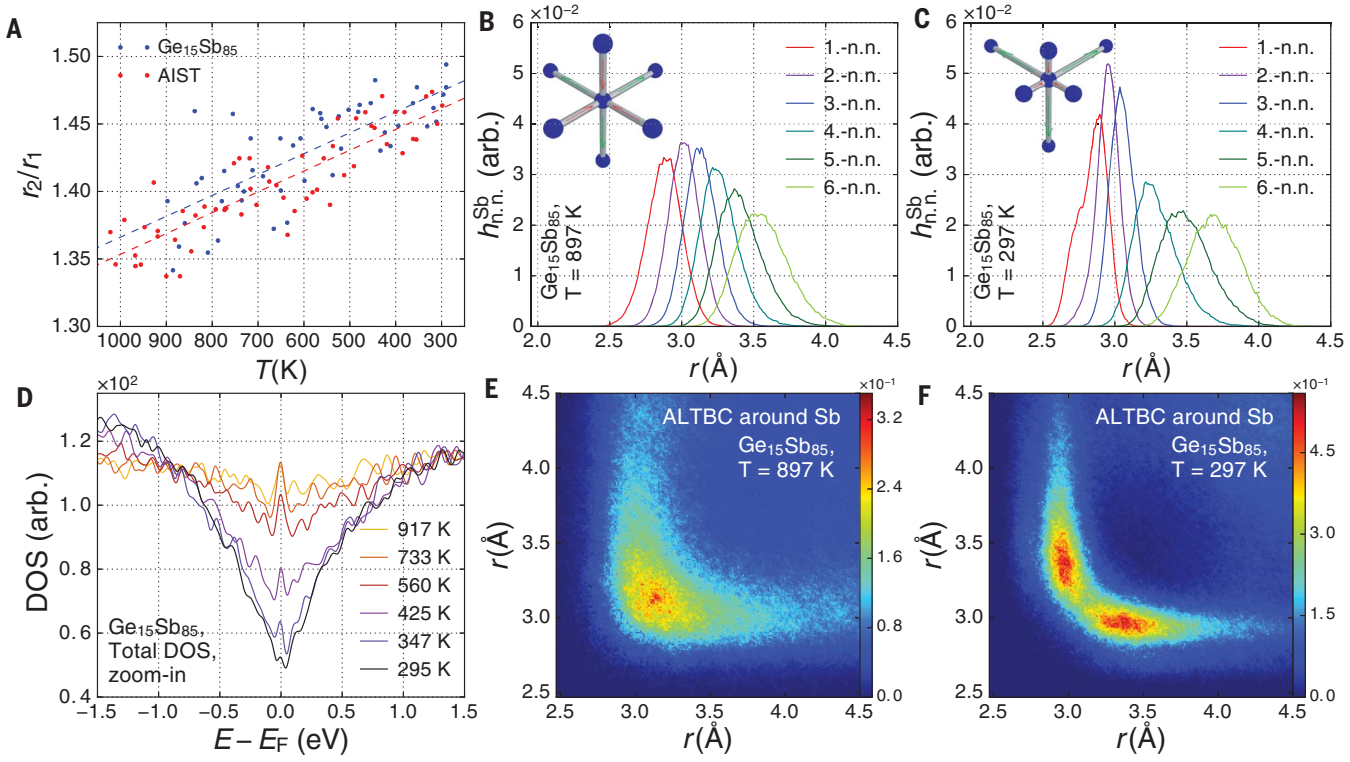


Fig. 5. AIMD simulations of melt–quenching $Ge_{15}Sb_{85}$ and AIST showing the onset of Peierls distortions upon cooling. (**A**) We observed that *R* increased continuously during the quench, with initial and final values similar to experimental values, except that the transition is wider in the simulation because of the higher quench rate. (**B** and **C**) Nearest-neighbor histograms of Sb atoms in $Ge_{15}Sb_{85}$ showing a continuous sixfold coordination in the high-temperature liquid state, whereas they split into three short and three more widely spaced distances in the quenched phase. (**D**) We observed the opening of a pseudogap during the quench in $Ge_{15}Sb_{85}$, but also in AIST (fig. S27). (**E** and **F**) ALTBC plots providing evidence for the formation of a Peierls distortion by including only the interatomic distances on opposite sides of Sb atoms.

is one of the characteristic attributes of PCMs, which are classified as incipient metals because of the sensitivity of their electronic localization on the distortion amplitude (*55*, *56*).

Ultrafast x-ray diffraction after short-pulse excitation provides access to the atomic structure of PCMs over the entire melt–quench cycle without interference by crystallization. The resulting diffraction data reveal a structural LLPT in AIST and $Ge_{15}Sb_{85}$. We used AIMD simulations to show that this LLPT in both PCMs is dominated by the formation of a Peierls distortion, forming distinct short and long bonds and opening a pseudogap in the electronic density of states. Our observation is consistent with earlier predictions that the atomic energy gain by a Peierls distortion $\Delta E_p < k_B T_m$ (*57*), because we show that the distortion is formed at T_{LL}, which implies $\Delta E_p \approx k_B T_{LL}$. The temperature-dependent structural parameters correlate with the apparent activation energy of inverse diffusivity, which suggests that the LLPT is responsible for the FTS crossover previously reported in AIST. The kinetic transition associated with the FTS crossover has important implications for the application of PCMs in memory devices: The low kinetic activation energy of the high-temperature liquid ensures that the high atomic mobility of the equilibrium liquid is available for crystallization, i.e., for temperatures between T_m and T_{LL}. At temperatures below the LLPT, the Peierls distortion localizes charge and stabilizes the amorphous atomic structure against crystallization, which increases the kinetic activation energy and rapidly reduces the atomic mobility during the quench, enabling the formation of a glass that is stable at ambient conditions. A material with lower T_{LL}/T_m therefore offers a wider temperature window between T_{LL} and T_m, where fast crystallization is possible, and explains why AIST and $Ge_{15}Sb_{85}$ are PCMs, whereas $Ge_{15}Te_{85}$ is not. Our results offer a new strategy for the design of improved PCMs for specialized memory applications based on the atomic-bonding mechanism in these materials.

REFERENCES AND NOTES

1. M. Hilbert, P. López, *Science* **332**, 60–65 (2011).
2. S. Raoux *et al.*, in *2007 Non-Volatile Memory Technology Symposium* (IEEE, 2007), vol. 5, pp. 30–35; https://ieeexplore.ieee.org/document/4389940/.
3. H. P. Wong *et al.*, *Proc. IEEE* **98**, 2201–2227 (2010).
4. M. Wuttig, N. Yamada, *Nat. Mater.* **6**, 824–832 (2007).
5. F. Xiong, A. D. Liao, D. Estrada, E. Pop, *Science* **332**, 568–570 (2011).
6. F. Rao *et al.*, *Science* **358**, 1423–1427 (2017).
7. D. Loke *et al.*, *Science* **336**, 1566–1569 (2012).
8. J. Siegel, C. N. Afonso, J. Solis, *Appl. Phys. Lett.* **75**, 3102–3104 (1999).
9. M. Salinga *et al.*, *Nat. Commun.* **4**, 2371 (2013).
10. R. Jeyasingh *et al.*, *Nano Lett.* **14**, 3419–3426 (2014).
11. J. Orava, A. L. Greer, B. Gholipour, D. W. Hewak, C. E. Smith, *Nat. Mater.* **11**, 279–283 (2012).
12. P. Zalden, A. von Hoegen, P. Landreman, M. Wuttig, A. M. Lindenberg, *Chem. Mater.* **27**, 5641–5646 (2015).
13. W. Zhang *et al.*, *Sci. Rep.* **4**, 6529 (2014).
14. D. Turnbull, *Contemp. Phys.* **10**, 473–488 (1969).
15. C. A. Angell, *Science* **267**, 1924–1935 (1995).
16. V. V. Vasisht, S. Saw, S. Sastry, *Nat. Phys.* **7**, 549–553 (2011).
17. I. Saika-Voivod, P. H. Poole, F. Sciortino, *Nature* **412**, 514–517 (2001).
18. S. Wei, P. Lucas, C. Austen Angell, *J. Appl. Phys.* **118**, 034903 (2015).
19. C. A. Angell, *J. Phys. Chem.* **97**, 6339–6341 (1993).
20. J. Orava, D. W. Hewak, A. L. Greer, *Adv. Funct. Mater.* **25**, 4851–4858 (2015).
21. S. Wei, G. J. Coleman, P. Lucas, C. A. Angell, *Phys. Rev. Appl.* **7**, 034035 (2017).
22. J. R. Errington, P. G. Debenedetti, *Nature* **409**, 318–321 (2001).
23. S. Hechler *et al.*, *Phys. Rev. Mater.* **2**, 085603 (2018).
24. S. Wei *et al.*, *Acta Mater.* **129**, 259–267 (2017).
25. P. G. Debenedetti, H. E. Stanley, *Phys. Today* **56**, 40–46 (2003).
26. D. T. Limmer, D. Chandler, *J. Chem. Phys.* **138**, 214504 (2013).
27. K. H. Kim *et al.*, *Science* **358**, 1589–1593 (2017).
28. P. H. Poole, F. Sciortino, U. Essmann, H. E. Stanley, *Nature* **360**, 324–328 (1992).
29. J. C. Palmer *et al.*, *Nature* **510**, 385–388 (2014).
30. K. Binder, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9374–9375 (2014).
31. S. Wei, Z. Evenson, M. Stolpe, P. Lucas, C. A. Angell, *Sci. Adv.* **4**, t8632 (2018).
32. G. Eising, T. Van Damme, B. J. Kooi, *Cryst. Growth Des.* **14**, 3392–3397 (2014).

33. Y. Chen, C. Rettner, Y. C. Chen, S. Raoux, G. Burr, Ultra-thin phase-change bridge memory device using GeSb (IEEE, 2006); https://ieeexplore.ieee.org/document/4154329.
34. R. Alonso-Mori *et al*., *J. Synchrotron Radiat.* **22**, 612–620 (2015).
35. M. J. Shu *et al*., *Appl. Phys. Lett.* **104**, 251907 (2014).
36. L. Waldecker *et al*., *Nat. Mater.* **14**, 991–995 (2015).
37. See supplementary materials.
38. W. K. Njoroge, M. Wuttig, *J. Appl. Phys.* **90**, 3816–3821 (2001).
39. J. Kalb, F. Spaepen, M. Wuttig, *J. Appl. Phys.* **93**, 2389–2393 (2003).
40. P.-Y. Chewvalier, *Thermochim. Acta* **155**, 227–240 (1989).
41. A. Chiba *et al*., *J. Phys. Conf. Ser.* **215**, 012077 (2010).
42. J. Orava, H. Weber, I. Kaban, A. L. Greer, *J. Chem. Phys.* **144**, 194503 (2016).
43. H. Neumann, W. Matz, W. Hoyee, M. Wobst, *Phys. Status Solidi* **90**, 489–498 (1985).
44. E. Barthélémy, S. Albert, C. Vigreux, A. Pradel, *J. Non-Cryst. Solids* **356**, 2175–2180 (2010).
45. J. A. Kalb, M. Wuttig, F. Spaepen, *J. Mater. Res.* **22**, 748–754 (2007).
46. C. Steimer *et al*., *Adv. Mater.* **20**, 4535–4540 (2008).
47. T. Matsunaga *et al*., *Nat. Mater.* **10**, 129–134 (2011).
48. J. Y. Raty *et al*., *Nat. Commun.* **6**, 7467 (2015).
49. S. Caravati, M. Bernasconi, T. D. Kühne, M. Krack, M. Parrinello, *Appl. Phys. Lett.* **91**, 171906 (2007).
50. N. F. Mott, R. S. Allgaier, *Phys. Status Solidi* **21**, 343–356 (1967).
51. Y. Tsuchiya, H. Saitoh, *J. Phys. Soc. Jpn.* **62**, 1272–1278 (1993).
52. S. Hosokawa, Y. Sakaguchi, K. Tamura, *J. Non-Cryst. Solids* **150**, 35–39 (1992).
53. Y. S. Tver'yanovich, V. M. Ushakov, A. Tverjanovich, *J. Non-Cryst. Solids* **197**, 235–237 (1996).
54. M. Schumacher *et al*., *Sci. Rep.* **6**, 27434 (2016).
55. M. Wuttig, V. L. Deringer, X. Gonze, C. Bichara, J.-Y. Raty, *Adv. Mater.* **30**, e1803777 (2018).
56. J. Y. Raty *et al*., *Adv. Mater.* **31**, e1806280 (2019).
57. J. P. Gaspard, *C. R. Phys.* **17**, 389–405 (2016).
58. H. E. Fischer *et al*., *Appl. Phys., A Mater. Sci. Process.* **74**, s160–s162 (2002).
59. J. Kalb, F. Spaepen, M. Wuttig, *Appl. Phys. Lett.* **84**, 5240–5242 (2004).
60. Y. Tsuchiya, *J. Non-Cryst. Solids* **312–314**, 212–216 (2002).
61. P. Zalden *et al*., Data for: Femtosecond x-ray diffraction reveals a liquid–liquid phase transition in phase-change materials (2019); https://doi.org/10.5281/zenodo.2649173.

ACKNOWLEDGMENTS

P.Z. thanks Christophe Bichara for valuable suggestions and discussions. Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-76SF00515. **Funding:** F.Q., A.K., M.N., and K.S.T. gratefully acknowledge financial support from the German Research Council through the Collaborative Research Center SFB 1242 project 278162697 ("Non-Equilibrium Dynamics of Condensed Matter in the Time Domain"), project C01 ("Structural Dynamics in Impulsively Excited Nanostructures"), and individual grant So408/9-1, as well as the European Union (7th Framework Programme, grant no. 280555 GO FAST). M.J.S., R.M., and M.W. acknowledge financial support from the German Research Council through the Collaborative Research Center SFB 917 ("Nanoswitches") and individual grant Ma-5339/2-1. M.J.S., I.R., and R.M. also acknowledge the computational resources granted by JARA-HPC from RWTH Aachen University under project nos. JARA0150 and JARA0183. M.T., A.M.L., and D.A.R. were supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, through the Division of Materials Sciences and Engineering under contract no. DE-AC02-76SF00515. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. J.L. acknowledges support from the Swedish Research Council. J.S. acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities through research grant UDiSON (TEC2017-82464-R). P.Z. gratefully acknowledges funding by the Humboldt Foundation. **Author contributions:** K.S.T. initiated the project. F.Q., K.S.T., and P.Z. conceived the experiment. P.Z. and M.W. were responsible for sample preparation. F.Q., P.Z., A.K., M.N., J.S., M.T., P.A., H.E., M.J.S., T.P., J.L., A.M.L., S.H.R., D.A.R., and K.S.T. performed the experiments. M.J.S., I.R., and R.M. conceived, performed, and evaluated the AIMD simulations. M.C., H.L., and D.Z. operated the XPP instrument. P.Z. analyzed the data with important input from K.S.T., F.Q., A.K., M.N., P.A., H.E., and J.S. H.E.F. and P.Z. conducted and evaluated the neutron total scattering measurements. P.Z. and K.S.T. wrote the manuscript with specific contributions from A.M.L., M.W., and S.W. and comments from all other authors. **Competing interests:** The authors declare no competing interests; **Data and materials availability:** The experimental data, scripts for data analysis, molecular dynamics trajectories, and the raw data from Figs. 1 to 5 are available from Zenodo (*61*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1062/suppl/DC1
Materials and Methods
Figs. S1 to S27
Tables S1 to S3
References (*62–82*)

26 November 2018; accepted 29 April 2019
10.1126/science.aaw1773

STRUCTURAL BIOLOGY

Cryo-EM structure of the mammalian ATP synthase tetramer bound with inhibitory protein IF1

Jinke Gu[1]*, Laixing Zhang[1]*, Shuai Zong[1]*, Runyu Guo[1]*, Tianya Liu[1], Jingbo Yi[1], Peiyi Wang[2], Wei Zhuo[1], Maojun Yang[1,3]†

The mitochondrial adenosine triphosphate (ATP) synthase produces most of the ATP required by mammalian cells. We isolated porcine tetrameric ATP synthase and solved its structure at 6.2-angstrom resolution using a single-particle cryo–electron microscopy method. Two classical V-shaped ATP synthase dimers lie antiparallel to each other to form an H-shaped ATP synthase tetramer, as viewed from the matrix. ATP synthase inhibitory factor subunit 1 (IF1) is a well-known in vivo inhibitor of mammalian ATP synthase at low pH. Two IF1 dimers link two ATP synthase dimers, which is consistent with the ATP synthase tetramer adopting an inhibited state. Within the tetramer, we refined structures of intact ATP synthase in two different rotational conformations at 3.34- and 3.45-Å resolution.

F_1F_O adenosine triphosphate (ATP) synthases are molecular motors that enable conversion of energy from a proton gradient to the high-energy terminal phosphodiester bond of ATP. The overall architecture of this multipart enzyme is largely conserved from bacteria to the mitochondria of eukaryotes (*1*–*3*). ATP synthases comprise the soluble F_1 hexagonal head plus central shaft and membrane-embedded F_O peripheral stalk and c-ring (*4*). The central shaft and c-ring form the well-known rotor, whereas the hexagonal head and peripheral stalk form the so-called stator. Because the number of protons required for one rotation, which generates three ATP molecules, is often not a multiple of three (*5*, *6*), it was proposed that torsion energy can be stored and released during the rotation cycle (*7*, *8*).

Oligomerization of ATP synthase is proposed to be involved in the formation of cristae in mitochondria (*9*, *10*). Two types of interfaces are proposed in oligomers: one between the two protomers in a dimer and another between dimers in row-like structures. A recent dimeric ATP synthase structure from yeast showed that subunit a and subunit i/j are major contributors for dimerization (*11*). It has been reported that the distance between protomers in a dimer of a given species is constant, but the distance between dimers within row-like structures is variable (*12*, *13*). This observation suggests that there are tight interactions within the dimer, but interactions between dimers are less constrained.

ATP synthase constitutes ~20% of the total membrane proteins present in mitochondria. Its ATP hydrolysis rate (>4000 molecules per second) is much higher than its ATP synthesis rate (~400 molecules per second), so it is critical to inhibit this enzyme when proton motive force across the inner membrane is lost (*14*). ATP synthase inhibitory factor subunit 1 (IF1) is an in vivo inhibitor of mammalian ATP synthase. As previously demonstrated (*15*, *16*), IF1 has two oligomeric states: dimer and tetramer. At pH 8.0, IF1 could form a tetramer through coiled-coil interactions involving the N-terminal inhibitory regions, thus preventing its inhibitory activity (*15*). However, when Δp level drops and pH plunges below 6.5 in matrix, IF1 disassembles into dimers, in which only the C-terminal regions form antiparallel coiled-coil interactions, and the N-terminal inhibitory regions are set free to insert into the F_1 heads (*17*).

In this study, we report a cryo–electron microscopy (cryo-EM) structure of a tetrameric ATP synthase from pig, *Sus scrofa domesticus*, bound with IF1 in a conformation that is consistent with an inhibited state. In our 6.2-Å tetrameric ATP synthase structure, we detected six sites of interactions between the two dimers. From this low-resolution complex, we extracted and solved substructures of two rotational states at 3.34 and 3.45 Å. The combination of low- and high-resolution information reveals mechanism for regulation of mammalian ATP synthase in which IF1 brings together dimers into a putatively nonproductive tetramer.

Structure determination

In our previous studies (*18*), when we analyzed the cryo-EM data of the mitochondrial electron transfer chain supercomplexes and megacomplexes, we invariably found a minor population of particles that had an H-shaped architecture in the two-dimensional (2D) classification processes (fig. S1). Further analyses of these particles revealed that they belonged to a tetrameric ATP synthase complex, which is corresponding to the high-molecular-weight bands in previous biochemical clear-native polyacrylamide gel electrophoresis (CN-PAGE) analyses (*19*). We optimized protein purification procedures to obtain larger amounts of the tetrameric ATP synthase from fresh porcine hearts and collected cryo-EM images (fig. S2 and supplementary materials). We refined the structure of this tetrameric porcine ATP synthase to an overall resolution of 6.2 Å (Fig. 1, figs. S3 and S4, and table S1). Two V-shaped ATP synthase dimers are positioned nearly perpendicular to form an architecture with C2 symmetry. Structural comparison indicates that the two protomers within each dimer are in E-state and DP-state (Fig. 1).

In order to obtain high-resolution structures of subcomponents of the tetramer, we first reconstructed the density map of the tetrameric porcine ATP synthase at an overall resolution of 10.5 Å in C1 symmetry using a total of 179,323 twice-binned particle images (fig. S5A). The diagonal protomers adopt similar conformations. To obtain the overall F_1F_O density, we performed masked 3D classification focused on the F_1 section followed by classification of the selected particles with an F_1F_O mask. After 3D autorefining, we obtained both the E-state and the DP-state densities at overall 3.34- and 3.45-Å resolutions, respectively (figs. S5, B to D, and S6).

To obtain the local density of F_O section, the selected particles were first refined by using F_1F_O mask and then refined by using F_O mask. E-state and DP-state F_O section densities were determined at 3.94- and 4.35-Å resolutions overall, respectively (fig. S7). Using a combination of structure docking and de novo modeling, we built structural models of all 30 subunits in COOT (*20*); 26 subunits could be built with clear side chains (table S2). The side chains of subunits e, g, k, and 6.8PL (6.8 kDa proteolipid) are not well resolved (fig. S8 and table S2). The "gold standard" Fourier shell correlation criteria between the models and maps have no significant differences, indicating the absence of overfitting (figs. S9 and S10).

Overall structure of mammalian ATP synthase

The two ATP synthase states, E and DP, differ in the direction of the central shaft subunit γ, the conformation of subunits $\alpha_3\beta_3$, and the position of the peripheral stalk (Fig. 2), similar to previous reports (*21*, *22*). The subunit compositions are identical in both states, including all 18 previously identified ATP synthase subunits (*23*, *24*) and a newly detected subunit k. In 1994, Walker defined the α and β subunits in the $\alpha_3\beta_3$ head as α_E, β_E, α_{DP}, β_{DP}, α_{TP}, and β_{TP}, respectively, with nucleotide binding site in the β_E being empty, in the β_{DP} containing adenosine diphosphate (ADP), and in the β_{TP} containing adenylyl-imidodiphosphate

[1]Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]SUSTech Cryo-EM Facility Center, Southern University of Science and Technology, Shenzhen 518055, China. [3]School of Pharmacy, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.
*These authors contributed equally to this work.
†Corresponding author. Email: maojunyang@tsinghua.edu.cn

(AMP-PNP). The α subunits all contain ATP (*17*, *25*, *26*) but are named according to the β subunit with which they form the intact catalytic site. Viewed from the N termini of the α and β subunits, β_{DP} is downstream of β_E, and β_{TP} is upstream of β_E, in a clockwise direction. In one state, the subunit α adjacent to the peripheral stalk is α_E, so we define this state as the E state for our structure. Likewise, in the other state, the subunit α adjacent to the peripheral stalk is α_{DP}, so we define this state as the DP state (Fig. 2) (*8*, *27*).

In this porcine ATP synthase structure, we identified a piece of density in the site occupied by subunit k in the yeast structure (Fig. 2 and figs. S8G and S11) (*11*, *27*). No subunit k is present in the bovine structure, and BLAST (Basic Local Alignment Search Tool) analysis indicated that no analog of yeast subunit k is present in mammalian ATP synthases. A recent report proposed that the human proteins DAPIT (diabetes-associated protein in insulin-sensitive tissue) and 6.8PL (6.8 kDa proteolipid) are analogs of subunits k and i/j in yeast, respectively (*23*). However, the density in our map corresponding to subunit i/j in the yeast structure is adequate to enable side chain building, and we assigned DAPIT to this density (fig. S8H). Besides, the N terminus of subunit b is assigned to a lateral helix over the membrane (Fig. 2), whereas this helix was assigned as subunit g in the yeast structure.

In addition to the density corresponding to yeast subunit k, we also identified a helical density in the center of the c_8-ring (fig. S12). Although we were unable to build the side chains of this subunit into the low-resolution density maps, we propose that this subunit may be the previously reported subunit 6.8PL for four reasons: (i) 6.8PL was found as a subunit of mammalian ATP synthase in previous biological and biochemical studies (*24*) and could not be assigned to the density corresponding to yeast subunit i/j because this density was built as subunit DAPIT in our structure. (ii) Secondary structure predictions clearly show that the N-terminal 45 amino acids of 6.8PL form a long helix, which is consistent with our density within the c_8-ring (fig. S12C). The density corresponding to subunit k contains a flexible N-terminal loop, suggesting that this density does not belong to 6.8PL (fig. S8G). (iii) The helix is close to the C terminus of subunit e (Fig. 2 and fig. S11), which is consistent with previous cross-linking studies (*28*). (iv) In a recent vacuolar H^+–adenosine triphosphatase (V-ATPase) structure (*29*), the elliptical c_{10}-ring is composed of eight c subunits, a subunit c′ and a subunit c″. Within the c_{10}-ring, two helical densities were identified and the authors assigned them to subunit Voa1 and the N terminus of subunit c″. Because our density map within the c_8-ring is not well resolved, we cannot exclude the possibility that this density belongs to a protein recruited into the c_8-ring when ATP synthase is inhibited. A structure with better defined density in this region will be needed to definitively establish the identity of this subunit.

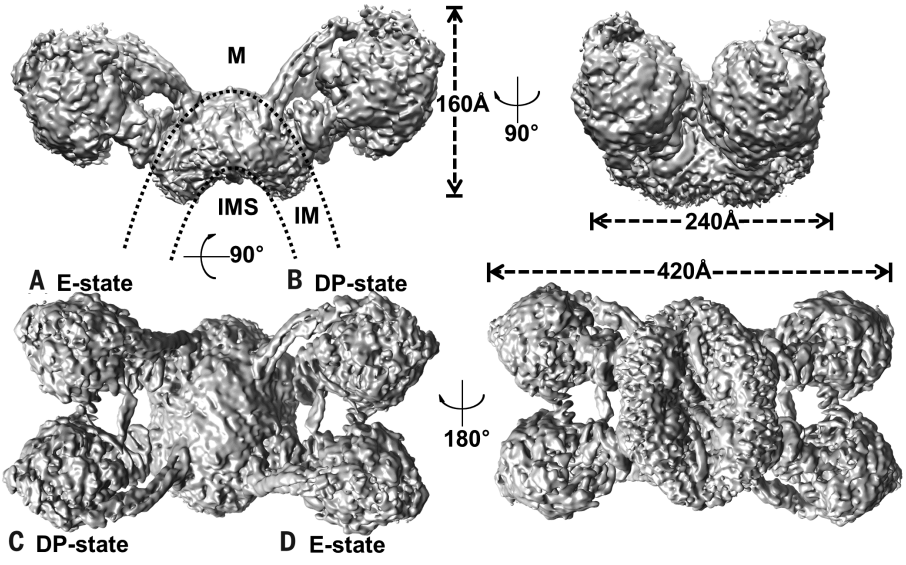


Fig. 1. Cryo-EM structure of a porcine ATP synthase tetramer. Cryo-EM map at 6.2-Å resolution displayed at 5σ contour level. M, matrix; IM, inner membrane; IMS, intermembrane space.

Conformational changes between two states of ATP synthase

Alignment of the E-state and DP-state models based on subunit a of the F_O section showed that the transmembrane helices of subunits b, e, f, g, k, A6L, and DAPIT match quite well between the two states (fig. S13, A and B), indicating that these helices are stably anchored in the membrane. Conversely, the c_8-ring and the peripheral stalk adopt different positions in the two states, indicating that these subunits move along with the rotation of the central shaft.

Alignment of the two states based on the F_1 section reveals that the conformation of the peripheral stalk and central shaft changes increasingly from top to bottom (fig. S13C). The $\alpha_3\beta_3$ heads, the top regions of the central shaft, and even the IF1 inhibitors are extremely well matched (fig. S13, C and D), suggesting that the conformation of the $\alpha_E\beta_E$ pair, $\alpha_{DP}\beta_{DP}$ pair, and $\alpha_{TP}\beta_{TP}$ pair is constant during each rotation step of the catalytic cycle. The lower part of the central shaft comprising subunits δ and ε is only slightly changed, whereas the c_8-ring is obviously rotated (fig. S13C).

Between the E and DP states, the central shafts have rotated by around 120°. We noticed that D194 in subunit γ can interact with R38 of one of the eight c subunits (fig. S14); thus, this R38 could potentially be used as a marker to distinguish a specific c subunit in the ring. Between the two states, the corresponding c subunit rotates ~140°, indicating that torsion must occur within the rotor during rotation. Considering that there are three αβ pairs but eight c subunits, the observed torsion is reasonable and consistent with previous reports (*21*, *30*, *31*). The conformations of the F_1 section and the c_8-ring are nearly unchanged on their own, so we conclude that torsion within the rotor between each catalytic step happens at the interaction surface between the central shaft and the c_8-ring. (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.)

The linkage between the F_1 and F_O sections

Consistent with the previous structural findings (*8*, *11*, *21*, *22*, *27*, *32*), in our E- and DP-state porcine ATP synthase structures, the F_1 and F_O sections are linked at two sites. One is at the top of the overall structure between the N-terminal domain of the $\alpha_3\beta_3$ head and the hook of the peripheral stalk (fig. S14, A to C). This link is mediated by the conserved subunit OSCP (oligomycin sensitivity conferral protein), which has N- and C-terminal domains. The C-terminal domain is in contact with subunit F6 and bends the subunit b C-terminal helix into a hook (figs. S14, A and B, and S15A). In agreement with the chloroplast ATP synthase structure, the N-terminal helix of the α_E subunit can interact with subunit b, and the N-terminal helix of F6 and the C-terminal domain of OSCP hitch the $\alpha_3\beta_3$ head onto the hook of the peripheral stalk (*8*, *27*). The N-terminal domain of OSCP forms stable interactions with the N-terminal helix of α_{TP}, but the N-terminal helix of α_{DP} is ambiguous in our structure (fig. S14, A and C). However, the density map of the N-terminal helix of α_{DP} is blurred, probably because of disorder in this helix. Another linking site is at the membrane surface between the central shaft and the c_8-ring. In subunit c, the loop linking the N- and C-terminal helices is responsible for interacting with residues from subunit γ, δ, and ε. According to our density map, it is likely that all eight c subunits are involved in the

interaction with the central shaft (fig. S14, A and D). Comparing these linking sites between the E-state and DP-state structures indicates that although the stator has rotated around 120°, the interacting pattern between c_8-ring and the central shaft remains largely unchanged (figs. S13A and S14D).

The ADP-inhibited state F_1 section

The E-state and DP-state models mainly differ in the position of the peripheral stalk, the orientation of the central shaft, and the corresponding conformational change of $\alpha_3\beta_3$ heads (Fig. 2 and fig. S13). Because the subunit composition and interactions between subunits are similar within both states, we focused on the E-state F_1 section. The classic N-terminal β-barrels of α and β subunits alternate, forming a crown on top of the head (Fig. 3, A and B, and fig. S15B). Alignment of the three αβ pairs based on the α subunit reveals shifts in β_E and β_{TP} relative to β_{DP} (Fig. 3B). The nucleotide binding position of β_E is disrupted, and the C-terminal region of β_E is forced to acquire an "open" conformation. On the basis of comparison of the E-state F_1 section with the previous bovine structure (without the IF1 protein), we propose that this disorder may be caused by the orientation of subunit γ and insertion of IF1 (fig. S16) (*15*, *33*, *34*).

In the mammalian central shaft, only the N- and C-terminal long helices of γ subunit protrude into the center of the $\alpha_3\beta_3$ head (Fig. 3A), whereas δ and ε subunits bind to the root of γ subunit and interact with the membrane-embedded c_8-ring (fig. S14, A and D). The N-terminal helix of γ subunit reaches the C-terminal domain of the $\alpha_3\beta_3$ head, whereas the C-terminal helix of γ subunit can reach the top of the nucleotide binding region of the $\alpha_3\beta_3$ head (Fig. 3A). The convex side of the N-terminal helix of γ subunit points toward the α_E and β_E subunits, and the C-terminal helix of γ subunit primarily contacts loops from the α_{TP} and β_{TP} subunits (Fig. 3A) (*14*, *35*–*40*).

According to the "binding change mechanism," during either ATP synthesis or ATP hydrolysis, the relative position of β_E, β_{DP}, and β_{TP} is constant. β_E is in an open conformation, with the nucleotide binding site being empty waiting for accepting new ADP or ATP; β_{DP} is in a "loose" conformation bound with ADP and Pi, ready for synthesizing or ejecting; and β_{TP} is in a "tight" conformation bound with ATP, waiting to be released or hydrolyzed (*4*, *31*, *41*). However, structural data suggests that in some conditions, β_E subunits are not empty, and the nucleotides bound in the β_{DP} or β_{TP} subunits do not strictly correspond to the DP and TP names (*1*, *26*). Correspondingly, not all β_{DP} subunits occur in the loose state, and not all β_{TP} subunits are in the tight state. These discrepancies can be interpreted

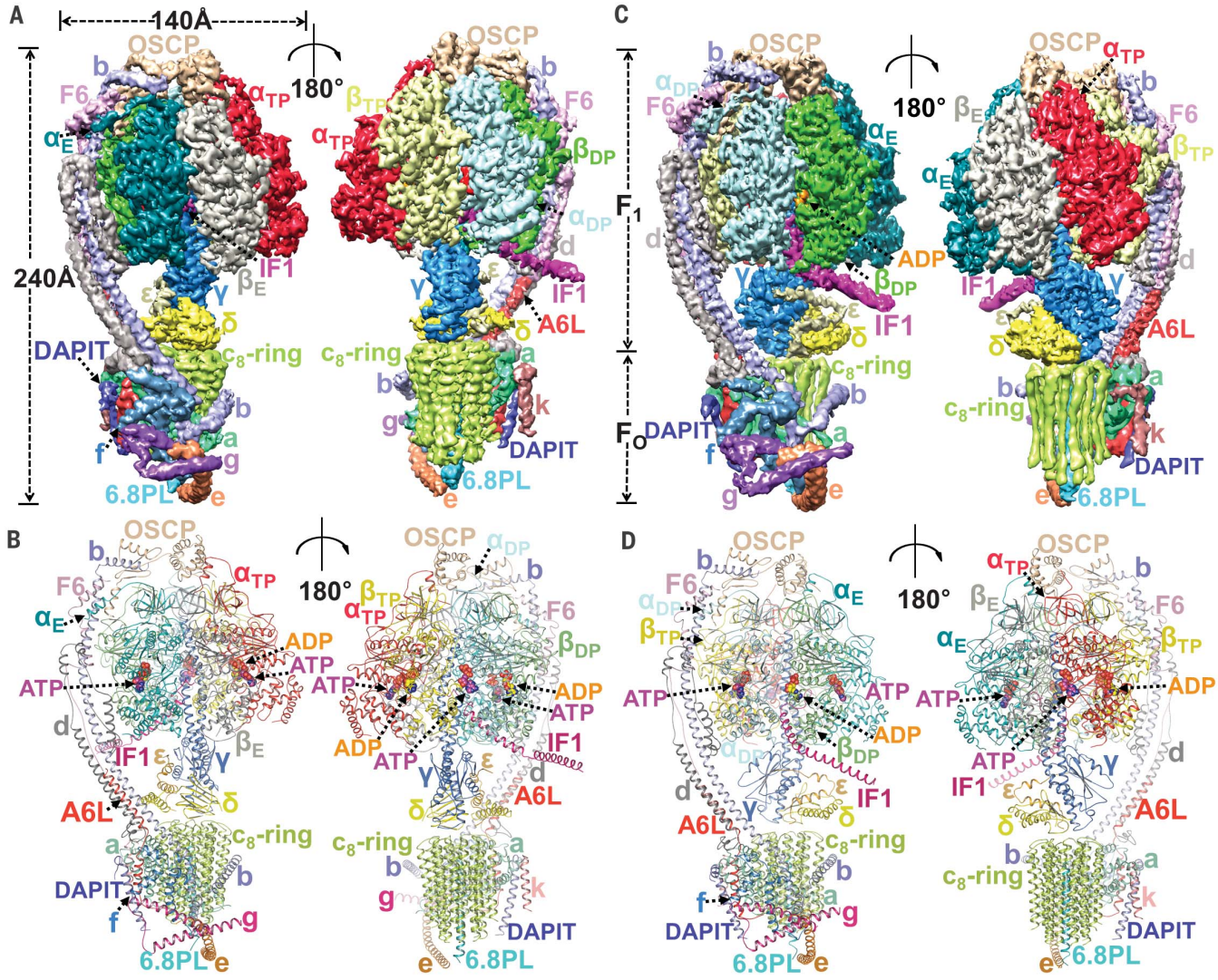


Fig. 2. Overall structures of E-state and DP-state ATP synthase. (**A** and **C**) Cryo-EM maps of (A) E-state ATP synthase at 3.45 Å (at 3σ contour level) and (C) DP-state ATP synthase at 3.34 Å (at 3σ contour level). (**B** and **D**) Cartoon representations of 3D reconstruction of substates extracted through focused alignment.

as relevant to the inhibited state or a catalytic intermediate. The only conclusion we can draw is that the structural differences between the β_{DP} and β_{TP} subunits are small. In our structure, all the α subunits are bound with MgATP, and all the nucleotide binding sites in the β_E subunits are empty. However, all the β_{DP} and β_{TP} subunits are tightly bound with MgADP without free Pi (Fig. 3, B to D). The β_{DP} and β_{TP} subunits in our structures may thus represent the non-competitive ADP-inhibited state (*14*).

The complete porcine F_O section

The F_O section is composed of a soluble matrix peripheral stalk and a membrane-embedded domain (Fig. 4A). The soluble portion of the subunit b C-terminal helix protrudes out from the membrane root and forms the backbone of the peripheral stalk (Fig. 4A). ATP synthases from chloroplasts and most bacteria contain two b subunits that protrude from the membrane: b and b′, with subunit b′ accompanying subunit b and inducing the latter to form a hook at the top region (*8*, *21*, *37*, *42*, *43*). By contrast, our work shows that the ATP synthase from porcine mitochondria only contains one b subunit, with the role of subunit b′ being replaced by subunit F6 at the top region, subunit d at the root region, and subunit A6L in the membrane (Fig. 4A) (*44*). Polar contacts between subunit F6 and subunit b bend the top region around 90° to form a hook (fig. S14, A and B). The C-terminal region of subunit d locks the root of subunit b, and two N-terminal helices of subunit d form a hairpin structure accompanying the middle region of subunit b to add extra plasticity and tenacity to this region (Fig. 4A and fig. S14A). The soluble portion of the subunit b C-terminal helix can obviously bend with its membrane portion, indicating that the peripheral stalk is capable of storing torsion energy. Consistent with previous studies (*28*, *30*), the C-terminal helix of the A6L subunit protrudes out from the membrane root and forms a three-helices bundle with subunits b and d, stabilizing the peripheral stalk (Fig. 4, A and B).

A total of 34 transmembrane helices can be detected within the membrane region (Fig. 4B and fig. S11), including a helix at the site of yeast subunit k and the helix inside the c_8-ring. We can clearly detect a piece of cylindrical density corresponding to a long α-helix containing more than 40 amino acids that extrudes from the top of the c_8-ring to the intermembrane space (IMS) (Fig. 4, A and B, and fig. S12), but we cannot determine whether this helix rotates with the c_8-ring or not. As stated above, we assigned subunit 6.8PL to this density.

A distinct feature of the F_O membrane domain is the five lateral helices (H2 to H6) of subunit a, which form a stable curved belt against the continuously rotating c_8-ring (Fig. 4, B and C). As in previous structures (*2*, *11*, *30*), the concave surface of H5 and H6 directly interacts with the c_8-ring. The tip of H5 and H6 is locked by the membrane portion of subunit b C-terminal helix. At the opposite side, the tip of H5 and H4 is locked by the k subunit (Fig. 4, A to C, and fig. S11). At the convex side of the belt, H3 and H4 are locked by subunit A6L and by the DAPIT subunit. The vertical H1 of subunit a is aligned with subunit A6L and subunit f (Fig. 4B and fig. S11). These vertical helices—including the membrane portion of the subunit b C-terminal helix, A6L, DAPIT, f, and k—are very likely to function as an anchor to stabilize the subunit a.

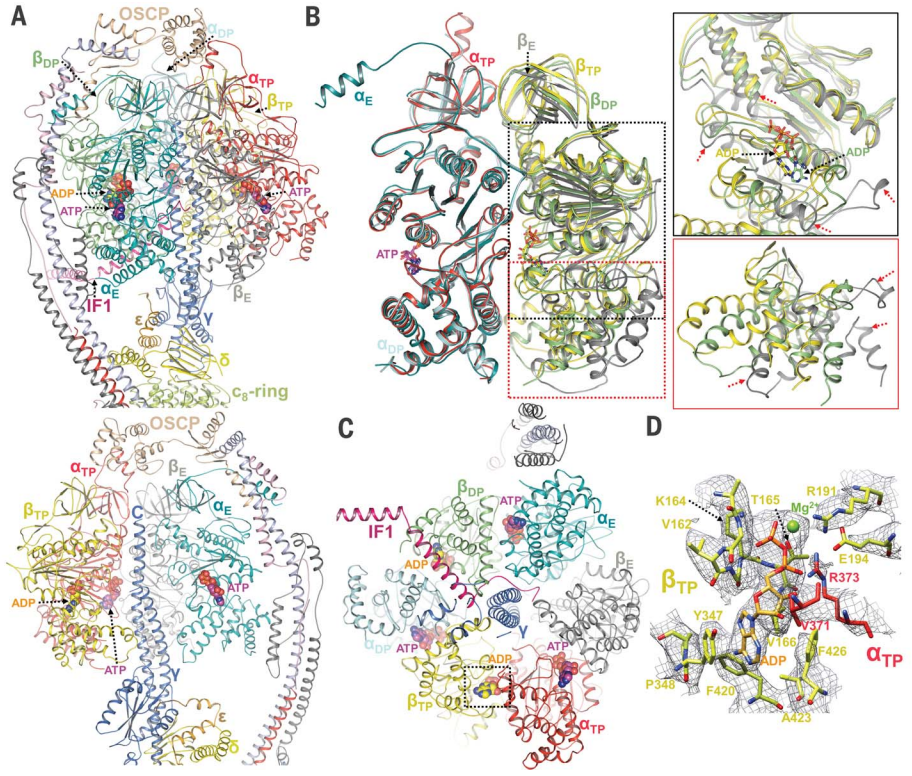


Fig. 3. The F_1 section of the IF1-bound E-state ATP synthase. (**A**) Side-on view of the F_1 section. The F_1 section structure was rotated 120° counterclockwise in the lower panel. (**B**) Alignment of the $\alpha_E\beta_E$, $\alpha_{TP}\beta_{TP}$, and $\alpha_{DP}\beta_{DP}$ pairs based on the α_{DP} subunit [root mean square deviation between the $\alpha_E\beta_E$ and $\alpha_{DP}\beta_{DP}$ pairs. ($RMSD_{E\ and\ DP}$) = 0.90 Å, RMSD between the $\alpha_{TP}\beta_{TP}$ and $\alpha_{DP}\beta_{DP}$ pairs. ($RMSD_{TP\ and\ DP}$) = 0.94 Å]. Black and red boxes indicate regions with major differences between the α/β pairs. (**C**) View of the IF1-inhibited F_1 head viewed from the IMS side. (**D**) The nucleotide binding site in the β_{TP} subunit is displayed at 4σ contour level.

Subunit e is a long, curved helix that crosses the membrane, subunit g comprises two helices that are embedded on the matrix side of the membrane, and the N-terminal helix of subunit b is positioned far away from the proton channel region (Fig. 4A and fig. S11). We propose that these helices can collectively account for the majority of the interactions occurring among the protomers in the tetramer. Given that subunit e and the C-terminal helix of subunit g are apparently tilted in the inner mitochondrial membrane, it is very likely that they are closely related to the bending known to occur for cristae.

The proton channels in the mammalian F_O section

Subunit c consists of two helices and adopts a hairpin shape. Eight c subunits form a circle, with the N-terminal helices arranged inward and the C-terminal helices arranged outward (Fig. 4B). The loops linking the N- and C-terminal helices interact with the central shaft of the F_1 section, including subunits γ, δ, and ε (fig. S14D). Surface electrostatic analysis showed that a circle of negative charges encircles the c_8-ring cylinder in the middle of the transmembrane helices (fig. S12D). These negative charges are mostly contributed by the conserved E58 residues of subunit c helices (Fig. 4D) because no charged residues exist in the vicinity of E58 in our porcine structure. According to previous structures of yeast and algal ATP synthase, these conserved E58 side chains could acquire two conformations (*27*, *45*, *46*). When protonated, these residues point inward to the c-ring in a "closed" conformation; when deprotonated, these residues point outward from the c-ring, acquiring an open conformation. It was previously proposed that only two E58 residues covered by subunit a are deprotonated in the open conformation, whereas the other six are protonated in the closed conformation, so when E58 encounters R159 in subunit a, the protons taken from E58 can be pumped across membrane (*46*). In our structure, all E58 residues are likely protonated in the closed conformation (Fig. 4, D and E) because no protons should translocate in the inhibited state.

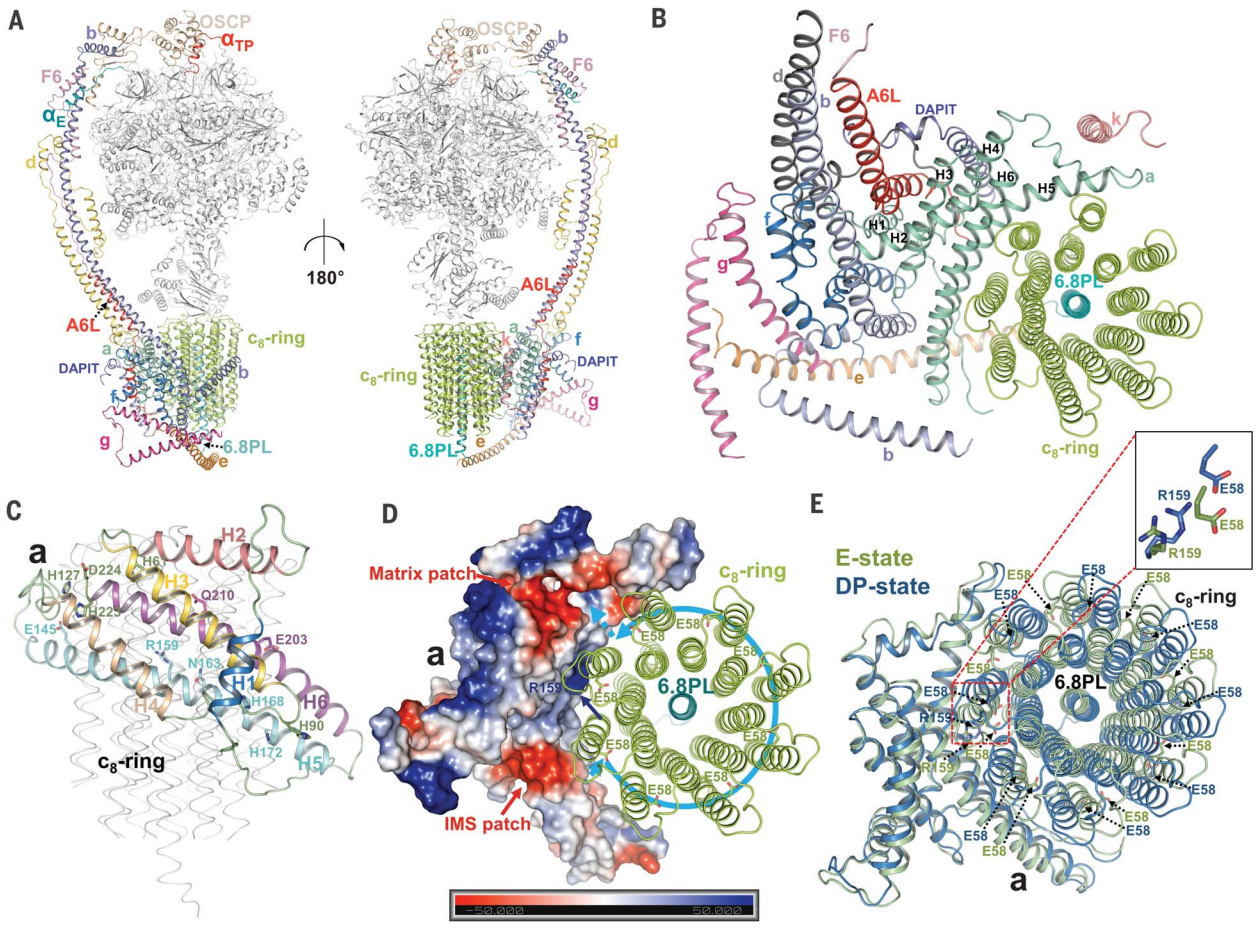


Fig. 4. The complete mammalian F_O section of ATP synthase. (**A**) Side views along the membrane highlighting the F_O section and OSCP. Other subunits are colored gray. (**B**) The F_O section membrane region viewed from the matrix side. (**C**) Subunit a residues forming the proton channel. (**D**) A charge-smoothed potential map of subunit a generated in PyMOL. (**E**) Comparison of subunit a and the c_8-ring of the E-state and DP-state. These two structures are aligned based on subunit a with a backbone root mean square deviation of 1.46 Å.

It was proposed that the two proton-pumping half channels of the F_O section are within the interface between subunit a and c_8-ring (*30*, *47*, *48*). In our structure, H5 and H6 of subunit a form a belt precisely holding the c_8-ring. Surface electrostatic analysis of subunit a showed three obviously charged patches on the belt (Fig. 4D). A positively charged subunit a residue R159 is located at the middle region of the belt, forming the central patch. Two negatively charged patches are separated at the two sides of the central patch. One negative patch is accessible to the matrix and forms an obvious half channel with the c_8-ring; we termed this the "matrix patch." The other negative patch is accessible to the IMS and forms a channel with subunit b on the opposite side of the c_8-ring, so we termed it the "IMS patch" (Fig. 4D). The negative charge of the matrix patch is mainly contributed by E145 and D224, and that of the IMS patch is mainly contributed by E203 (Fig. 4C). Compared with two recent yeast structures (*11*, *27*), the E203 side chain points to subunit b in our structure but points to subunit c in the yeast structures. When the side chain of E203 is pointing to subunit b, no proton can be transferred to E58 from the IMS, a situation that is consistent with the putatively IF1-inhibited state of ATP synthase tetrameric structure.

The relative position between subunit a and the c_8-ring is different in the two resolved states. When we aligned the two a subunits for the E-state and DP-state models, we observed a ~20° rotation in the position of the c_8-ring between the two states. In the E-state, the conserved E58 residue of subunit c is closest to the R159 residue of subunit a, even though its side chain is in the closed conformation. However, in the DP-state, clockwise rotation (viewed from the mitochondrial matrix) disrupts the interaction between E58 and R159, and R159 acquires a different orientation from that in the E-state (Fig. 4E), indicating that this DP-state structure may be in an intermediate c_8-ring rotation state.

Interactions within the ATP synthase tetramer

We docked the high-resolution models of the E-state and DP-state ATP synthase into the 6.2-Å map of the tetramer to obtain the complete tetrameric structure (fig. S17). The F_1 sections of adjacent protomers are linked with a long, bow-shaped, antiparallel IF1 homodimer (Fig. 5, A and B), which is consistent with a previous crystal structure of the IF1-inhibited F_1 head (*17*). The rotational axes between protomers in the dimeric subcomplex are ~97°, and the axis between protomers adjacent in the tetramer is ~47° (fig. S18). The four F_O sections form extensive contacts that stabilize the tetramer (Fig. 5 and fig. S19), and self-dimerization of the IF1 protein pulls the two dimers close together within the tetramer (Fig. 5B). In the IF1-mediated protomer pair, the C-terminal regions of α_E (DP state, Gln^{475}) and α_{DP} (E state, Ser^{470}) are close to each other; the closest distance is within 4 Å (Fig. 5B and fig. S17C). In the classic V-shaped dimer, the F_O membrane regions of the two protomers are loosely connected, and the F_1 sections are separated by a large gap (Fig. 5C and fig. S19).

We propose that the IF1-bound ATP synthase tetramer in our structure is an integral functional unit because the subunits in the protomers

interact with each other to stabilize the tetramer and regulate its function (Fig. 5 and fig. S19). The two dimers are linked to each other at six binding sites. For binding sites 1 and 6, two IF1 proteins form an antiparallel dimer that interacts with the F_1 heads within the mitochondrial matrix. Sites 2 to 4 are all above the membrane. Sites 2 and 4 include the N-terminal loop of subunit k and the N-terminal helix of subunit b from two adjacent dimers. Site 3 includes two N-terminal helices of g subunits from opposite protomers within the tetramer, positioned in parallel at the center of the structure (Fig. 6A). Site 5 is within the inner membrane, where two e subunits interact with each other (within the membrane), holding the tetramer together (Fig. 6B). Beyond these direct protein-protein interactions, the cryo-EM map clearly shows that there are additional densities in the membrane region, which we suspect may be filled by lipids.

In the middle region of the two protomers, the two curved domains formed by the e and g subunits with the N-terminal helix of subunit b are reminiscent of the previously reported BAR domain of the amphiphysin protein (Fig. 6 and fig. S19) (*11*, *49*); this BAR domain is able to bend the lipid bilayer through its amphipathic α helices. The two BAR-like domains in the center of our structure also form a dimer (Fig. 6). The membrane regions are likely bent by the perpendicular orientation of the ATP synthase dimers (fig. S18).

Discussion

Although ATP synthase structures from several species have been reported after the crystal structure of bovine F_1 subcomplex of ATP synthase (*26*), the complete mammalian ATP synthase structure has not been available. We obtained the tetrameric mammalian ATP synthase bound to IF1 and solved substructures in two different rotational states at overall resolutions of 3.34 and 3.45 Å. Among the 30 subunits in total (19 different types), b, k, DAPIT, and 6.8PL were differently assigned in our work as compared with previous reports (*11*, *23*, *30*). Alignment of the two rotational states showed that the central shaft has rotated around 120°, but the c_8-ring has rotated around 140°. We conclude from this result that torsional energy can be distributed over the central shaft and the peripheral stalk.

We identified two proton-pumping half channels that are accessible either to the IMS or to the matrix. Within the IMS half channel, a^{E203} points to subunit b in our structure rather than subunit c as in the yeast structure (fig. S20). The entire helix 6 of subunit a in our structure is shifted by one residue (one-third of a turn) compared with the yeast structure. This difference raises two possibilities. First, because we believe our ATP synthase may be in an inhibited state, the orientation of a^{E203} pointing to subunit b may be an inhibition mechanism to prevent proton transfer to subunit c. Second, we cannot exclude the possibility that conformational change of a^{E203} is required to pump protons in active state; in this condition, protons are absorbed into the membrane on one side of subunit a and expelled out of the membrane on the other side of subunit a. The conformational change of a^{E203} could transfer protons across subunit a. Further mutational studies of a^{E203} are needed to examine the function of this residue.

In the yeast dimer structure, the two protomers are linked by the long N-terminal loops of two a subunits that form an antiparallel β-sheet like a railing, and the long C-terminal loops of two i/j subunits contact each other. It was also proposed that the C-terminal regions of subunit k and subunit e can interact with each other in the IMS region (*11*). However, none of these interactions are evident in porcine ATP synthase. In mammals, the N-terminal sequence of subunit a is much shorter than in yeast (226 and 249 amino acids in pig and yeast, respectively), so the two a subunits in our structure do not form the railing that links the two protomers (figs. S19C and S21). Instead, the C-terminal loops of the

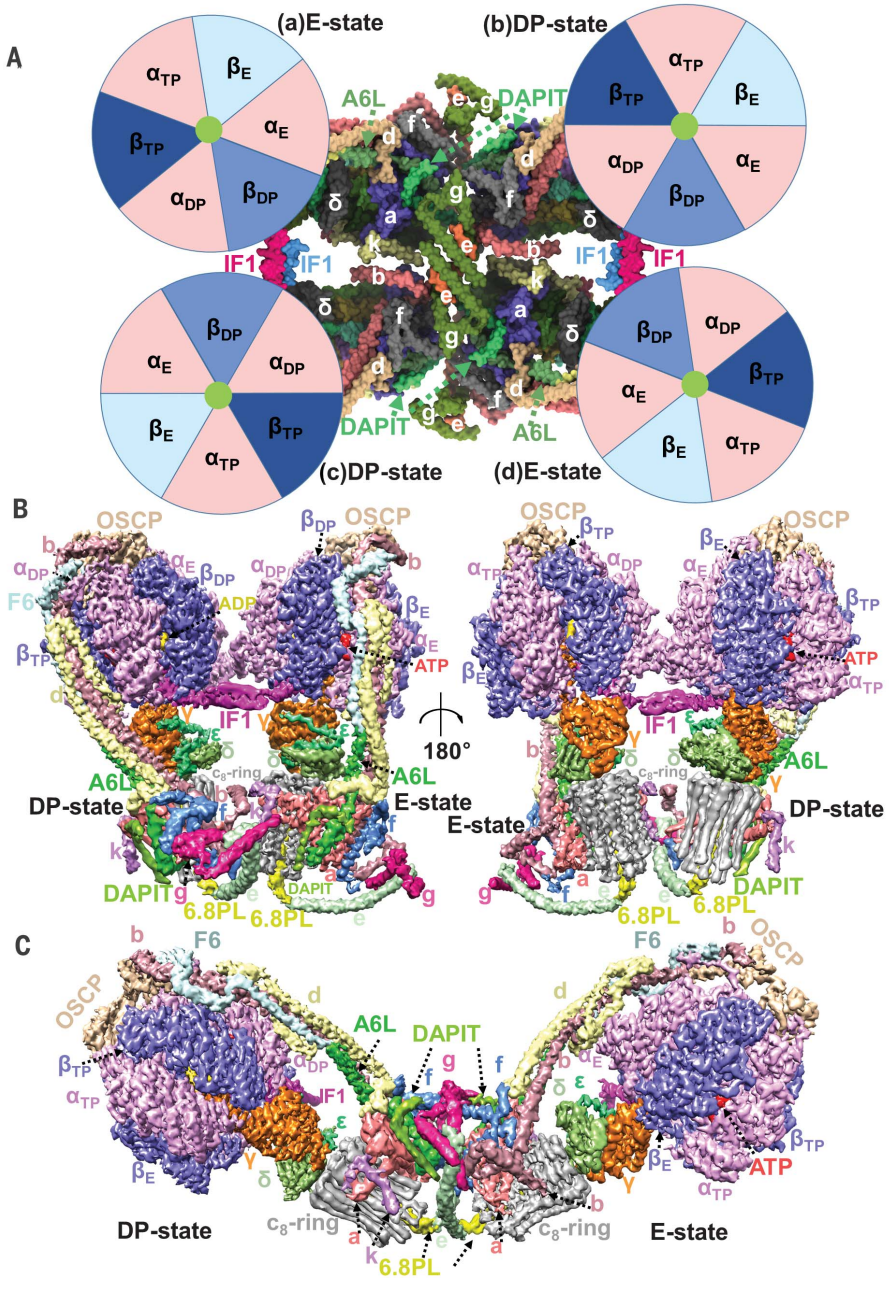


Fig. 5. Overall architecture of the porcine ATP synthase tetramer. (**A**) Schema of the porcine ATP synthase tetramer bound to IF1. (**B**) Cryo-EM map of the two IF1-linked ATP synthase protomers (at 3σ contour level). The maps of the extracted E-state and DP-state protomers were put together to generate the combined map. (**C**) Cryo-EM map of the V-shaped ATP synthase dimer (at 3σ contour level). The combined map was also generated by the extracted protomers.

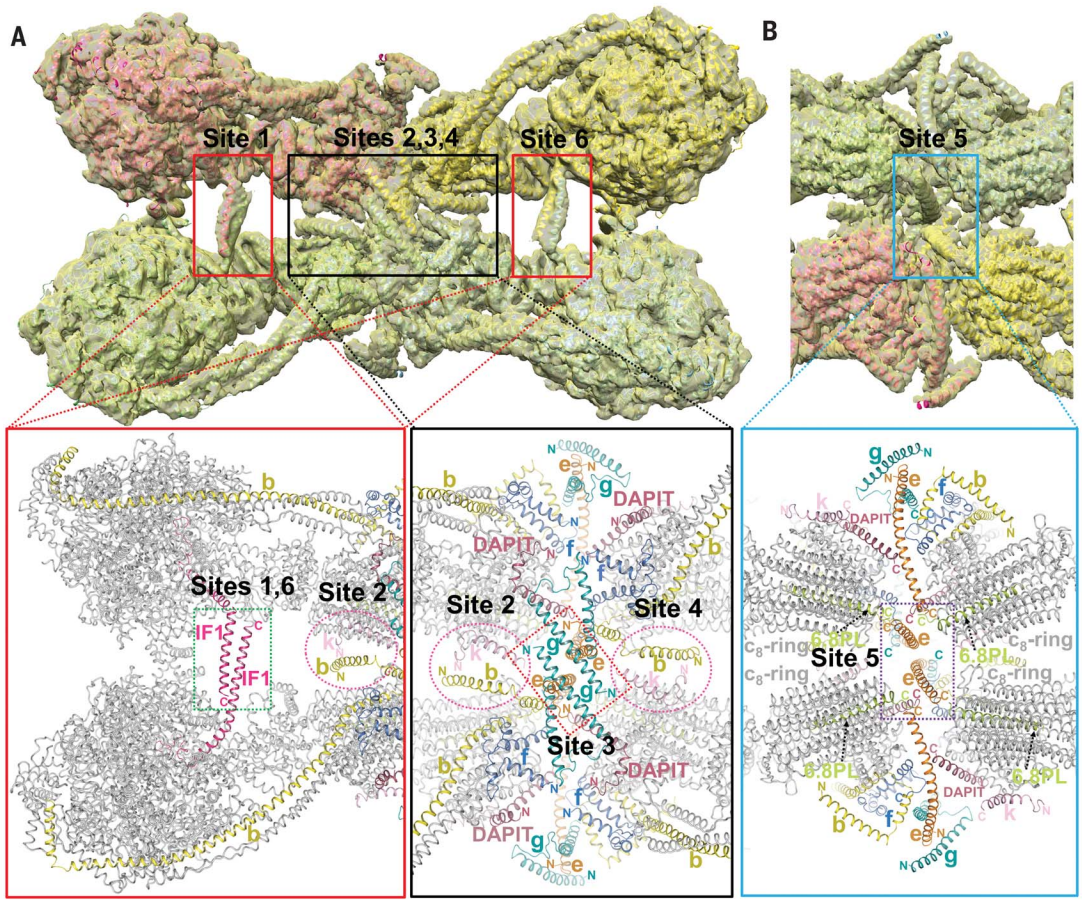


Fig. 6. The sites participating in the porcine ATP synthase tetramer formation. (**A**) Tetramer interaction sites viewed from the matrix side. High-resolution models were fit into the tetramer maps with a correlation coefficient of 0.85. (**B**) View from the IMS side.

DAPIT subunit and subunit f from the opposite protomer can form interactions. Presumably, the N-terminal helix of subunit k and the loop region of subunit g from the opposite protomer tend to form interactions (fig. S19C). On the basis of the increased variability in the angle between protomers in our porcine structure as opposed to the rigid structure in yeast (*11*), the interaction between two protomers may be weaker in mammals. Future functional studies are needed to show the relationship between the catalytic state of ATP synthase and the angle formed by the ATP synthase dimer.

Our structure suggests that ATP synthase tetramers can probably be inhibited by at least three mechanisms. The first mechanism is inhibition by IF1 protein dimers (fig. S22A). The long, stick-shaped IF1 dimer links the F_1 heads of protomers within different dimers, with its N-terminal regions inserted into the two $\alpha_{DP}\beta_{DP}$ pairs and contacting the two central γ subunits (fig. S22A). The second one is inhibition of the F_1 head by tightly bound MgADP (fig. S22B). As stated above, all the β_{DP} and β_{TP} subunits in the tetramer are tightly bound with an MgADP without free Pi, so these are in the widely recognized noncompetitive ADP-inhibited state (*14*). The third mechanism is inhibition by the long α-helical e subunits (fig. S22C). The C termini of the four e subunits are bent toward the four c_8-rings in each protomer, which—according to our unsharpened cryo-EM map—are able to interact with the C-terminal ends of the 6.8PL helices that fill the central hole inside the c_8-rings (fig. S23). We propose that these interactions may function as brake pads to block rotation of the c_8-rings in the inhibited state. Further studies are needed to confirm the identity of the protein within the c_8-ring. Whether it is 6.8PL or not, this protein could be involved in the regulation of ATP synthase activity.

Comparing the cryo-EM density maps of ATP synthase dimers from yeast and our structure (fig. S23), it is apparent that the C-termini of subunit e adopt different conformations. These helices of subunit e are straight in the yeast dimeric ATP synthase structure but are severely bent in our porcine tetramer structure (fig. S23D). The oligomerization of ATP synthase, and thus regulation of activity, should be intertwined with conformational changes in subunit e. Our assignment of the other interface subunits reveals intricate interactions in the tetrameric complex, supporting a functional role of this assembly in ATP synthase inhibition and suggesting roles for many supernumerary subunits with previously unknown or uncertain functions.

REFERENCES AND NOTES

1. W. Junge, N. Nelson, *Annu. Rev. Biochem.* **84**, 631–657 (2015).
2. H. Guo, J. L. Rubinstein, *Curr. Opin. Struct. Biol.* **52**, 71–79 (2018).
3. S. Cogliati, J. A. Enriquez, L. Scorrano, *Trends Biochem. Sci.* **41**, 261–273 (2016).
4. P. D. Boyer, *Annu. Rev. Biochem.* **66**, 717–749 (1997).
5. W. Kühlbrandt, K. M. Davies, *Trends Biochem. Sci.* **41**, 106–116 (2016).
6. K. M. Davies *et al.*, *J. Vis. Exp.* **91**, 51228 (2014).
7. H. Sielaff *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17760–17765 (2008).
8. A. Hahn, J. Vonck, D. J. Mills, T. Meier, W. Kühlbrandt, *Science* **360**, eaat4318 (2018).
9. P. Paumard *et al.*, *EMBO J.* **21**, 221–230 (2002).
10. T. B. Blum, A. Hahn, T. Meier, K. M. Davies, W. Kühlbrandt, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4250–4255 (2019).
11. H. Guo, S. A. Bueler, J. L. Rubinstein, *Science* **358**, 936–940 (2017).
12. K. M. Davies, C. Anselmi, I. Wittig, J. D. Faraldo-Gómez, W. Kühlbrandt, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13602–13607 (2012).
13. K. M. Davies *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14121–14126 (2011).
14. A. S. Lapashina, B. A. Feniouk, *Biochemistry* **83**, 1141–1160 (2018).
15. E. Cabezon, P. J. Butler, M. J. Runswick, J. E. Walker, *J. Biol. Chem.* **275**, 25460–25464 (2000).
16. E. Cabezón, M. J. Runswick, A. G. Leslie, J. E. Walker, *EMBO J.* **20**, 6990–6996 (2001).
17. E. Cabezón, M. G. Montgomery, A. G. Leslie, J. E. Walker, *Nat. Struct. Biol.* **10**, 744–750 (2003).
18. R. Guo, S. Zong, M. Wu, J. Gu, M. Yang, *Cell* **170**, 1247–1257. e12 (2017).
19. I. Wittig, H. Schägger, *Proteomics* **5**, 4338–4346 (2005).
20. P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
21. M. Sobti *et al.*, *eLife* **5**, e21598 (2016).
22. A. Zhou *et al.*, *eLife* **4**, e10180 (2015).

23. J. He *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2988–2993 (2018).
24. R. Chen, M. J. Runswick, J. Carroll, I. M. Fearnley, J. E. Walker, *FEBS Lett.* **581**, 3145–3148 (2007).
25. D. Stock, A. G. Leslie, J. E. Walker, *Science* **286**, 1700–1705 (1999).
26. J. P. Abrahams, A. G. Leslie, R. Lutter, J. E. Walker, *Nature* **370**, 621–628 (1994).
27. A. P. Srivastava *et al*., *Science* **360**, eaas9699 (2018).
28. J. Lee *et al*., *J. Biol. Chem.* **290**, 13308–13320 (2015).
29. S. H. Roh *et al*., *Mol. Cell* **69**, 993–1004.e3 (2018).
30. A. Hahn *et al*., *Mol. Cell* **63**, 445–456 (2016).
31. J. V. Bason, M. G. Montgomery, A. G. Leslie, J. E. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6009–6014 (2015).
32. A. W. Mühleip *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8442–8447 (2016).
33. M. Campanella, N. Parker, C. H. Tan, A. M. Hall, M. R. Duchen, *Trends Biochem. Sci.* **34**, 343–350 (2009).
34. M. E. Pullman, G. C. Monroy, *J. Biol. Chem.* **238**, 3762–3769 (1963).
35. L. R. Carrillo, J. E. Froehlich, J. A. Cruz, L. J. Savage, D. M. Kramer, *Plant J.* **87**, 654–663 (2016).
36. A. T. Jagendorf, *Photosynth. Res.* **73**, 233–241 (2002).
37. H. Sielaff, T. M. Duncan, M. Börsch, *Biochim. Biophys. Acta Bioenerg.* **1859**, 775–788 (2018).
38. C. M. Nalin, R. E. McCarty, *J. Biol. Chem.* **259**, 7275–7280 (1984).
39. M. Sekiya, R. K. Nakamoto, M. K. Al-Shawi, M. Nakanishi-Matsui, M. Futai, *J. Biol. Chem.* **284**, 22401–22410 (2009).
40. K. Y. Hara, Y. Kato-Yamada, Y. Kikuchi, T. Hisabori, M. Yoshida, *J. Biol. Chem.* **276**, 23969–23973 (2001).
41. W. Junge, H. Lill, S. Engelbrecht, *Trends Biochem. Sci.* **22**, 420–423 (1997).
42. J. E. Walker, *Biochem. Soc. Trans.* **41**, 1–16 (2013).
43. A. G. Stewart, L. K. Lee, M. Donohoe, J. J. Chaston, D. Stock, *Nat. Commun.* **3**, 687 (2012).
44. A. Dautant *et al*., *Front. Physiol.* **9**, 329 (2018).
45. J. Symersky *et al*., *Nat. Struct. Mol. Biol.* **19**, 485–491, S1 (2012).
46. D. Pogoryelov, O. Yildiz, J. D. Faraldo-Gómez, T. Meier, *Nat. Struct. Mol. Biol.* **16**, 1068–1073 (2009).
47. K. R. Vinothkumar, M. G. Montgomery, S. Liu, J. E. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12709–12714 (2016).
48. S. B. Vik, B. J. Antonio, *J. Biol. Chem.* **269**, 30364–30369 (1994).
49. B. J. Peter *et al*., *Science* **303**, 495–499 (2004).

ACKNOWLEDGMENTS

We thank the Cryo-EM Facility Center of Southern University of Science and Technology (Shenzhen) for providing the facility support. The computation was completed on the Yanglab GPU workstation. **Funding:** This work was supported by funds from the National Key R&D Program of China (2017YFA0504600 and 2016YFA0501100), the National Science Fund for Distinguished Young Scholars (31625008), the National Natural Science Foundation of China (21532004, 31570733 and 31800620), and the China Postdoctoral Science Foundation (2017M620040 and 2018T110091). **Author contributions:** M.Y. conceived, designed, and supervised the project; built the model; analyzed the data; and wrote the manuscript. J.G., T.L., J.Y., W.Z., and R.G. did the protein purification and detergent screening; J.G., P.W., and L.Z. performed EM sample preparation and data collection; J.G., L.Z., and S.Z. determined the structures. All authors discussed the manuscript data. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The atomic coordinates of the E-state F_O section, E-state, DP-state F_O section, DP-state, and tetramer of porcine ATP synthases have been deposited in the Worldwide Protein Data Bank with the accession codes 6J54, 6J5J, 6J5A, 6J5I, and 6J5K, respectively. The corresponding maps have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-0668, EMD-0669, EMD-0670, EMD-0677, and EMD-0667.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1068/suppl/DC1
Materials and Methods
Figs. S1 to S23
Table S1 and S2
References (*50*–*67*)

25 December 2018; accepted 23 May 2019
10.1126/science.aaw4852

REPORT

PHYSICS

Metastable ferroelectricity in optically strained $SrTiO_3$

T. F. Nova[1,2]*, A. S. Disa[1], M. Fechner[1], A. Cavalleri[1,2,3]*

Fluctuating orders in solids are generally considered high-temperature precursors of broken symmetry phases. However, in some cases, these fluctuations persist to zero temperature and prevent the emergence of long-range order. Strontium titanate ($SrTiO_3$) is a quantum paraelectric in which dipolar fluctuations grow upon cooling, although a long-range ferroelectric order never sets in. Here, we show that optical excitation of lattice vibrations can induce polar order. This metastable polar phase, observed up to temperatures exceeding 290 kelvin, persists for hours after the optical pump is interrupted. Furthermore, hardening of a low-frequency vibration points to a photoinduced ferroelectric phase transition, with a spatial domain distribution suggestive of a photoflexoelectric coupling.

Strontium titanate ($SrTiO_3$) is paraelectric and centrosymmetric at all temperatures. When cooled, it displays many anomalies that suggest its proximity to a ferroelectric phase, including a large rise in the dielectric function (*1*) and softening of a polar mode (*2*). This behavior is often referred to as incipient ferroelectricity, with quantum fluctuations of the ionic positions preventing long-range ordering (*3*, *4*). The proximity to a ferroelectric phase is underscored by the ease with which $SrTiO_3$ can be made ferroelectric, for example, by Ca-substitution [Sr → Ca; Curie Temperature (T_c) = 37 K] (*5*) or by isotope substitution ($^{16}O \rightarrow {}^{18}O$; T_c = 25 K) (*6*). Strain, as shown in the phase diagram of Fig. 1, has proven most effective in controlling the transition in $SrTiO_3$ (*7*), with reported ferroelectricity up to room temperature (*8*).

Strong lattice deformations can also be achieved by irradiation with strong mid-infrared optical pulses (see Fig. 1), which can be made resonant with a specific lattice vibration in $SrTiO_3$. Nonlinear phonon excitation (*9*–*11*) may induce a polar or ferroelectric phase in many ways. For example, dynamical softening (*12*, *13*) of the polar mode by cubic lattice nonlinearities (*14*) or the generation of transiently induced strain [see supplementary text S11 in (*15*)] may all provide routes to creating a ferroelectric order absent at equilibrium.

In the experiments reported here, the highest-frequency A_{2u} vibrational mode of $SrTiO_3$ was resonantly excited (at T = 4 K) with femtosecond mid-infrared optical pulses tuned to 15-μm wavelength (83-meV photon energy) derived from a 1-kHz repetition rate Ti:Sapphire laser, an optical parametric amplifier, and a difference frequency mixing crystal (fig. S1). The optical polarization was aligned along the [001] axis of a (110)-oriented crystal (Fig. 2A).

$SrTiO_3$ was monitored through second harmonic generation (*16*) of a 2.2-μm wavelength optical probe pulse, colinear and time-delayed with respect to the mid-infrared pump (Fig. 2A). As shown in Fig. 2B, a time delay–independent second harmonic signal, absent at equilibrium, indicated the appearance of a noncentrosymmetric phase after mid-infrared excitation.

The second harmonic signal was observed to accumulate with exposure to mid-infrared radiation and reach saturation after several minutes, with a maximum value determined by the pump fluence (30 mJ/cm^2 for the data of Fig. 2B; see fig. S8 and supplementary text S6). As all other contributions to the second harmonic were negligible, including surface and quadrupole terms of the nonlinear susceptibility tensor, this observation was a reliable reporter of a photoinduced phase with broken inversion symmetry (supplementary text S3C).

Structural symmetry information was obtained by continuously rotating the incoming 2.2-μm probe polarization and measuring the 1.1-μm second harmonic projected along the pseudocubic [1 −10] and [001] crystallographic directions with a second polarizer (Fig. 2C). These angular dependences indicate the formation of a polarization along the [1 −10] direction and are consistent with a (noncentrosymmetric) polar point group (C_{2v}) (supplementary text S3B).

The experiment was performed as a function of sample temperature and pump wavelength. The induced second harmonic signal was found to be maximum at low temperatures but remained detectable up to room temperature (fig. S4 and supplementary text S4). In addition, the polar state was found to be most efficiently created when pumping resonantly with the A_{2u} phonon, whereas for pump photons in the proximity of the 3.2-eV bandgap of $SrTiO_3$, the effect

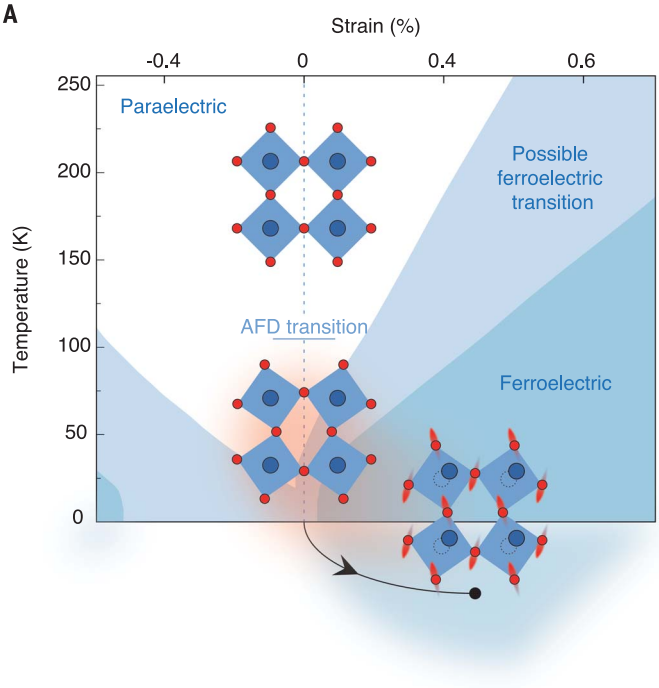


Fig. 1. Dynamical ferroelectricity in $SrTiO_3$. Bulk, unstrained $SrTiO_3$ is paraelectric at any finite temperature. At 105 K, it undergoes an antiferrodistortive (AFD) transition from cubic to tetragonal, although it retains a centrosymmetric structure. However, small amounts of strain cause the material to undergo a ferroelectric transition. The Curie temperature increases with growing values of applied strain, highlighting the paramount role of acoustic lattice deformations in controlling the transition. The strain-phase diagram shown here represents a thermodynamic analysis of the phase transition for a single-domain state, adapted from (*32*). The shaded cartoon explores the possibility of dynamically establishing a ferroelectric phase through vibrational excitation.

[1]Max Planck Institute for the Structure and Dynamics of Matter, Hamburg 22761, Germany. [2]The Hamburg Centre for Ultrafast Imaging, Hamburg 22761, Germany. [3]Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, United Kingdom.
*Corresponding author. Email: tobia.nova@mpsd.mpg.de (T.F.N.); andrea.cavalleri@mpsd.mpg.de (A.C.)

disappeared (Fig. 2A, inset, and supplementary text S7).

The photoinduced phase was found to be metastable when unperturbed, relaxing back to the nonpolar equilibrium paraelectric phase only several hours after the pump was turned off (fig. S5). We found that the optically induced polar state could be reverted to the paraelectric ground state by exposing it to above-bandgap photons (>3.2 eV) (fig. S6) or by thermal cycling.

Figure 3 displays ultrafast measurements of the same second harmonic signal discussed above. In addition to the time delay–independent background shown in Fig. 2, visible as an offset in the traces of Fig. 3A, the second harmonic signal exhibited ultrafast oscillations that resulted from the pump-driven, impulsive, inelastic excitation of low-frequency polar modes (see Fig. 3A for three representative measurements taken after 10, 25, and 105 min of illumination).

The frequency of these oscillations increased visibly with the baseline second harmonic, as shown more comprehensively in Fig. 3B. The detected mode-hardening is reminiscent of the

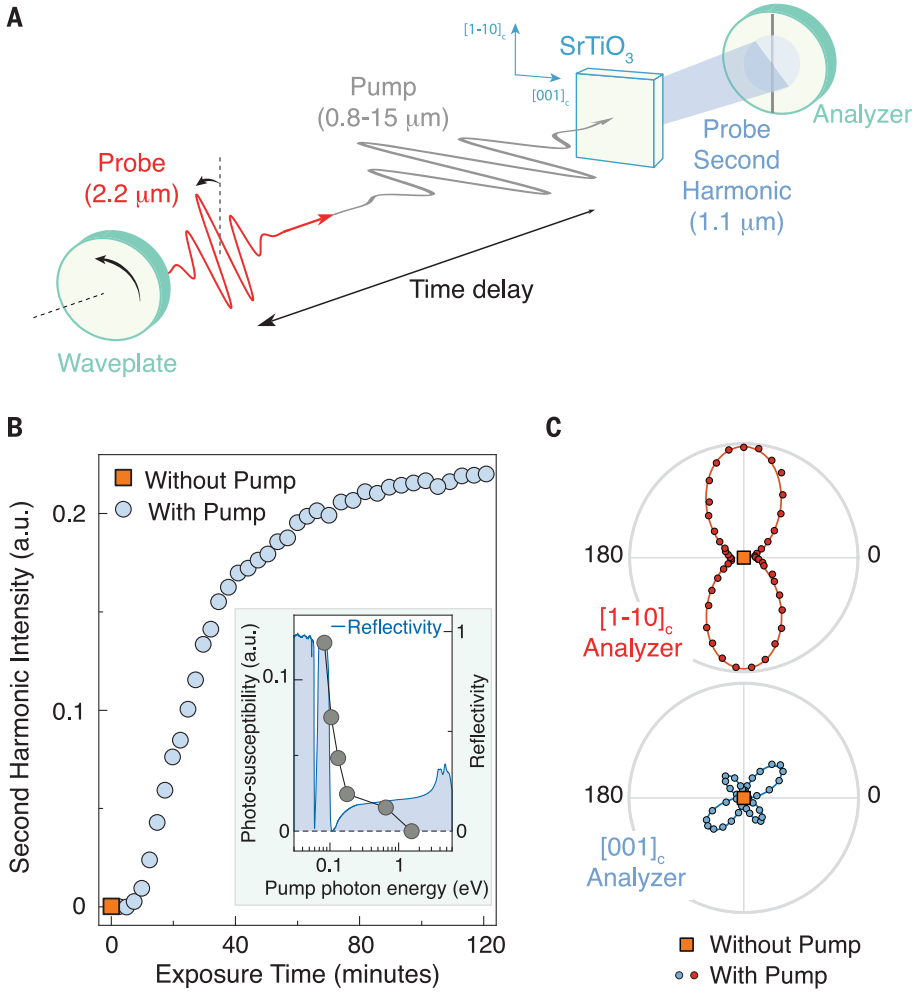


Fig. 2. A photoinduced polar state. (**A**) Sketch of the experimental setup. A (110)-oriented $SrTiO_3$ sample is coherently excited with tunable wavelength pulses (0.8 to 15 μm, gray line). Time-delayed colinear probe pulses (2.2-μm wavelength) impinge on the sample, with a half-waveplate controlling their polarization. The generated second harmonic (1.1 μm) is detected in transmission geometry. When needed, an analyzer (i.e., a polarizer) can be used to isolate orthogonal polarization components of the second harmonic. For a 15-μm wavelength pump, the full width at half maximum (FWHM) spot size was 72 μm. The probe was focused down to 52 μm FWHM. (**B**) Time delay–independent total second harmonic intensity impinging on the detector (without analyzer) as a function of exposure time to 15-μm pump pulses. a.u., arbitrary units. (**Inset**) Pump wavelength dependence. The gray dots represent the photosusceptibility of the effect, as defined in eq. S6.1 (see supplementary text S6 and S7). The data are compared with the static reflectivity of $SrTiO_3$ (blue line). The low-energy reflectivity data (< 0.2 eV) were measured by Fourier transform infrared spectroscopy (FTIR) with a commercial spectrometer. The high-energy data (>1 eV) were adapted from (*33*). (**C**) Second harmonic intensity in the saturated state as a function of the probe polarization for two orthogonal analyzer configurations.

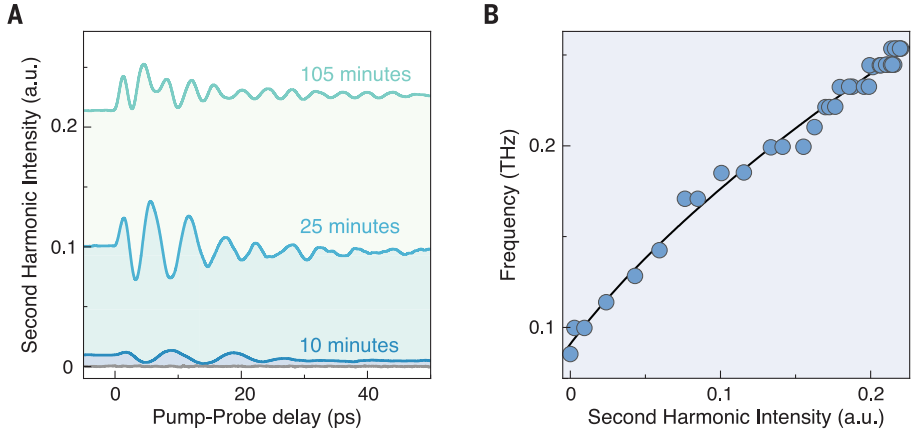


Fig. 3. Ferroelectric-like mode-hardening. (**A**) Pump-induced ultrafast modulation of the second harmonic intensity as a function of the pump-probe delay for different exposure times. The increasing offset corresponds to the delay-independent second harmonic of Fig. 2B. The gray line represents a measurement taken without mid-infrared pumping. No static second harmonic could be detected. (**B**) Frequency of the detected oscillations as a function of baseline second harmonic. The blue dots represent the frequency at which the Fourier spectrum of each ultrafast trace is maximum. The solid line is a guide to the eye.

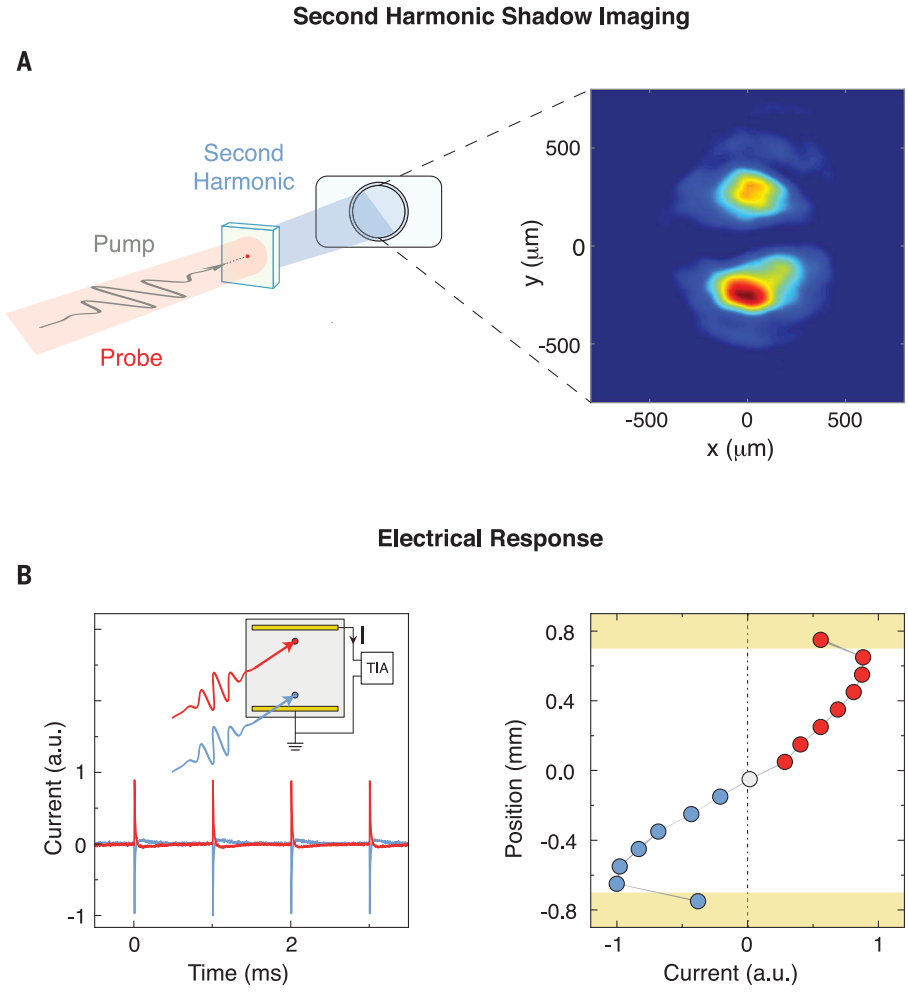


Fig. 4. A bipolar domain structure. (**A**) Second harmonic spatial dependence. (Left) Second harmonic shadow imaging setup. A collimated probe beam (1200 μm, much bigger than the 70-μm pump spot) uniformly illuminates the sample. The emitted second harmonic spatial profile is recorded with a CMOS camera (details in supplementary text S8). (Right) Measured second harmonic spatial dependence. (**B**) Electrical response. (Left) Electrical currents induced in $SrTiO_3$ by successive pump pulses (1-kHz repetition rate). The red and blue traces represent measurements taken with the pump impinging closer to the upper and lower contact, respectively. (**Inset**) Sketch of the transport measurements setup. Gold contacts were deposited on the sample surface, and the electrical response was recorded in short-circuit conditions without any applied bias. TIA, transimpedence amplifier (materials and methods S2). (Right) Maximum amplitude of the current pulses as a function of the pump position with respect to the contacts (yellow).

behavior of a ferroelectric soft mode across a paraelectric-to-ferroelectric phase transition (*17*–*19*).

To determine the spatial distribution of the photoinduced polar phase, we repeated the experiment with a collimated 2.2-μm probe beam that was made larger than the photoexcited spot. The second harmonic shadow of the transformed volume was projected onto a complementary metal-oxide semiconductor camera (Fig. 4A, left). Two oval bright regions with a dark area in the center were observed (Fig. 4A, right), indicative of an inhomogeneous polar state.

This inhomogeneity was further evidenced by electrical measurements (materials and methods S2). Gold contacts were deposited onto the sample surface, and the optical pump spot was kept much smaller (70-μm diameter) than the distance between the electrodes (1.5 mm). Measurements of the pump-induced electrical response were conducted under short-circuit conditions, without any applied bias (inset, Fig. 4B, left). Because the optically transformed region was far smaller than the gap between the electrodes, and because the electrodes themselves were not irradiated, the only coupling between the light-induced polar domains and the external circuit was capacitive. The ultrafast response was transmitted to the external circuit, in the form of short bursts of electrical currents at the 1-kHz repetition rate of the pump laser (Fig. 4B, left panel, red line).

The sign and amplitude of these current pulses were observed to be dependent on the position of the pump beam. No current pulses were detected when the pump pulse spot was in the middle of the gap (Fig. 4B, right), whereas currents of opposite sign (red and blue, Fig. 4B) were observed as the spot was moved toward either of the contacts.

The observations reported in Fig. 4 are indicative of the creation of two oppositely oriented polar domains in the pumped spot, which can be understood as follows.

First, oppositely oriented domains are expected to generate second harmonic light with equal intensities but with a relative π shift in their respective optical phases. Hence, as the two second harmonic beams propagated toward the detector, they destructively interfered in the center (Fig. 5C and fig. S11), resulting in the observed pattern (Fig. 5B). Second, upon excitation, each of the two domains is expected to draw currents in opposite directions from the electrodes because of the induced polarizations. For a pair of domains aligned at equal distances from the electrodes, no net current is anticipated. However, when these are moved in either direction, the two capacitances connecting each domain to the electrodes would become unequal, drawing a net current in either direction (Fig. 5D).

We next turn to a possible explanation for the formation of the two domains. We consider the dynamical equivalent of the ionic contribution to electrostriction. An optically pumped phonon mode coordinate Q_{IR} generates prompt stress (pressure) σ. Only hundreds of picoseconds after the femtosecond mid-infrared pulse, the stress is expected to evolve into transient strain (supplementary text S11). According to our estimates, pump peak field strengths of 18 MV/cm, predicted to induce ionic oscillations of 5-pm amplitude, result in a peak stress of 0.5 GPa. A 0.2% upper bound for the resulting transient strain is estimated (supplementary text S11). If we consider a strain profile with the same Gaussian shape of the pump beam, the pump-induced strain can generate a polarization in two ways. One can exclude a direct, strain-induced ferroelectric order, which would result in a single domain with polarization either up or down (Fig. 5A), a homogeneous second harmonic shadow, and monopolar current pulses (supplementary text S10). Rather, the coupling between the spatial gradient of the optical strain and polarization (Fig. 5B), arising from the flexoelectric effect (*20*, *21*), would result in the creation of two distinct, opposing polarizations and would readily explain the data of Fig. 4 (see Fig. 5, B to D, and supplementary text S10).

Although photoflexoelectricity appears to be a plausible explanation for the transient generation of polar domains, this hypothesis does not, per se, clarify the accumulation or metastability of the ferroelectric state.

It is possible that an ultrafast photoinduced polarization, triggered by the rapid formation of surface strain and subsequent flexoelectricity, would act on preexisting randomly distributed

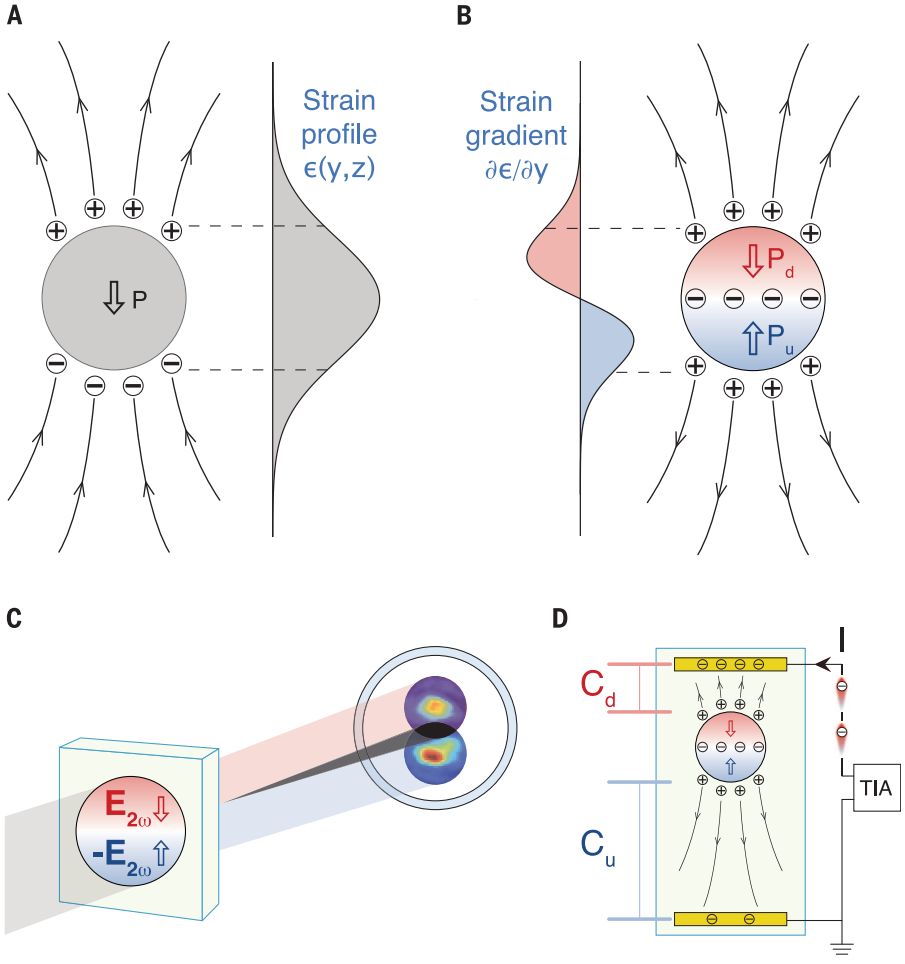


Fig. 5. Flexoelectric polarizations. (**A**) Phonon pumping results in a strain profile that follows the Gaussian profile of the pump beam (72-μm spot size FWHM). The figure depicts how a direct strain-induced transition to a ferroelectric monodomain state would look. (**B**) Inhomogeneous strain results in two distinct flexoelectric polarizations of opposite polarity that follow the strain gradient. (**C**) Two opposing polarizations generate antiphase second harmonic optical fields that interfere destructively on the CMOS camera (supplementary text S8 and S10). (**D**) The position of the flexoelectric domains with respect to the contacts determines the current amplitude and sign (supplementary text S10).

polar nanoregions (PNRs) (*22*). These may then align along a preferential direction. Owing to the transient nature of the aligning field and the intrinsically slow response time of the PNR, their alignment would only be partial after a single pulse. Subsequent pulses may facilitate the growth of macroscopic domains over a long incubation period, requiring tens of thousands of pulses. This hypothesis implies that in between two pulses, the induced partial alignment would not relax entirely to the unperturbed state—an assumption compatible with the known long relaxation times of PNRs (*22*).

An alternative explanation starts from photorefractivity (*23*, *24*). Impurity centers, which may lie in the energy range of the mid-infrared wavelengths used in this experiment, could be ionized concomitantly with the optically induced strain. The transient flexoelectric polarizations would force such photogenerated charges to slowly diffuse toward the edge of the excited area, where they eventually get trapped. However, all reported observations of photorefractivity take place for pump photon energies in the vicinity of the optical gap. In our experiment, on the contrary, the second harmonic signal progressively disappears for pump light approaching the optical gap of $SrTiO_3$. Overall, we regard this second hypothesis as less likely.

Our experimental results should motivate a new class of optical control experiments that rely on the perturbation of $SrTiO_3$ in functional materials. Because $SrTiO_3$ is used commonly as a substrate for the growth of oxide heterostructures, one could conceive of new ways to drive functional properties at interfaces (*25*, *26*)—including magnetic, electronic, and even superconducting states. Furthermore, the applicability of the optical control of flexoelectric polarizations extends far beyond the specific case of $SrTiO_3$, because flexoelectricity is allowed in materials of any symmetry (*20*, *21*).

More generally, our findings are a rare example of symmetry breaking by light and can thus be considered in the context of other types of light-induced phase transitions (*27*–*29*) in the presence of fluctuating order above the thermodynamic transition temperature T_c, as in the case of photoinduced superconductivity (*30*, *31*).

REFERENCES AND NOTES

1. T. Sakudo, H. Unoki, *Phys. Rev. Lett.* **26**, 851–853 (1971).
2. A. Yamanaka *et al.*, *Europhys. Lett.* **50**, 688–694 (2000).
3. K. A. Müller, H. Burkard, *Phys. Rev. B* **19**, 3593–3602 (1979).
4. K. A. Müller, W. Berlinger, E. Tosatti, *Z. Phys. B Condens. Matter* **84**, 277–283 (1991).
5. J. G. Bednorz, K. A. Müller, *Phys. Rev. Lett.* **52**, 2289–2292 (1984).
6. M. Itoh *et al.*, *Phys. Rev. Lett.* **82**, 3540–3543 (1999).
7. D. G. Schlom *et al.*, *Annu. Rev. Mater. Res.* **37**, 589–626 (2007).
8. J. H. Haeni *et al.*, *Nature* **430**, 758–761 (2004).
9. M. Först *et al.*, *Nat. Phys.* **7**, 854–856 (2011).
10. T. F. Nova *et al.*, *Nat. Phys.* **13**, 132–136 (2017).
11. A. Subedi, A. Cavalleri, A. Georges, *Phys. Rev. B* **89**, 220301 (2014).
12. A. Subedi, *Phys. Rev. B* **95**, 134113 (2017).
13. A. P. Itin, M. I. Katsnelson, *Phys. Rev. B* **97**, 184304 (2018).
14. A. Cartella, T. F. Nova, M. Fechner, R. Merlin, A. Cavalleri, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12148–12151 (2018).
15. Supplementary materials.
16. M. Fiebig, V. V. Pavlov, R. V. Pisarev, *J. Opt. Soc. Am. B* **22**, 96–118 (2005) review.
17. M. Takesada, M. Itoh, T. Yagi, *Phys. Rev. Lett.* **96**, 227602 (2006).
18. U. Bianchi, W. Kleemann, J. G. Bednorz, *J. Phys. Condens. Matter* **6**, 1229–1238 (1994).
19. H. Uwe, T. Sakudo, *Phys. Rev. B* **13**, 271–286 (1976).
20. P. Zubko, G. Catalan, A. K. Tagantsev, *Annu. Rev. Mater. Res.* **43**, 387–421 (2013).
21. P. Zubko, G. Catalan, A. Buckley, P. R. L. Welche, J. F. Scott, *Phys. Rev. Lett.* **99**, 167601 (2007).
22. H. W. Jang *et al.*, *Phys. Rev. Lett.* **104**, 197601 (2010).
23. S. Orlov, A. Yariv, M. Segev, *Appl. Phys. Lett.* **68**, 1610–1612 (1996).
24. Y. Pochi, G. Claire, Eds., *Landmark Papers on Photorefractive Nonlinear Optics* (World Scientific, 1995).
25. A. Ohtomo, H. Y. Hwang, *Nature* **427**, 423–426 (2004).
26. C. H. Ahn, K. M. Rabe, J.-M. Triscone, *Science* **303**, 488–491 (2004).
27. L. Stojchevska *et al.*, *Science* **344**, 177–180 (2014).
28. M. Fiebig, K. Miyano, Y. Tomioka, Y. Tokura, *Science* **280**, 1925–1928 (1998).
29. A. Cavalleri *et al.*, *Phys. Rev. Lett.* **87**, 237401 (2001).
30. M. Mitrano *et al.*, *Nature* **530**, 461–464 (2016).
31. D. Fausti *et al.*, *Science* **331**, 189–191 (2011).
32. Y. L. Li *et al.*, *Phys. Rev. B* **73**, 184112 (2006).
33. P. Dore, G. De Marzi, A. Paolone, *Int. J. Infrared Millim. Waves* **18**, 125–138 (1997).

ACKNOWLEDGMENTS

We are grateful to T. Gebert for providing the beam propagation simulations. We thank A. Cartella for the help in building the OPA and K.Beyerlein for useful discussions. **Funding:** The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC [grant agreement no. 319286 (QMAC)]. **Author contributions:** A.C. conceived the project together with T.F.N. T.F.N. and A.S.D designed the experiments and acquired and analyzed the data. M.F. developed the theoretical model and performed the DFT calculations. The manuscript was written by T.F.N. and A.C., with feedback from all coauthors. **Competing interests:** The authors declare no competing interests. **Data availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1075/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S3
References (*34*–*51*)

27 December 2018; accepted 20 May 2019
10.1126/science.aaw4911

PHYSICS

Terahertz field–induced ferroelectricity in quantum paraelectric $SrTiO_3$

Xian Li[1], Tian Qiu[2], Jiahao Zhang[2], Edoardo Baldini[3], Jian Lu[1]*, Andrew M. Rappe[2], Keith A. Nelson[1]†

"Hidden phases" are metastable collective states of matter that are typically not accessible on equilibrium phase diagrams. These phases can host exotic properties in otherwise conventional materials and hence may enable novel functionality and applications, but their discovery and access are still in early stages. Using intense terahertz electric field excitation, we found that an ultrafast phase transition into a hidden ferroelectric phase can be dynamically induced in quantum paraelectric strontium titanate ($SrTiO_3$). The induced lowering in crystal symmetry yields substantial changes in the phonon excitation spectra. Our results demonstrate collective coherent control over material structure, in which a single-cycle field drives ions along the microscopic pathway leading directly to their locations in a new crystalline phase on an ultrafast time scale.

In recent years, important advances have been made in the search for materials with complex multiphase landscapes that host photoinduced metastable collective states, or "hidden phases." These phases are rarely accessible on equilibrium phase diagrams and may persist long after the external stimuli that induced them have ceased. Recent experiments (*1*–*6*) have illustrated some of the possibilities and expanded our understanding of nonequilibrium material properties and dynamics. In some cases, ultrafast resonant excitation of crystal lattice vibrations (phonons) has played the key role in reaching hidden metallic and superconducting phases (*1*, *2*). Here, we extend this capability in the discovery of a hidden ferroelectric (FE) phase in the paradigmatic material $SrTiO_3$ (STO). We accessed the hidden phase by selectively exciting the "soft" phonon mode that serves as a collective reaction coordinate along which ions move from their initial positions toward their positions in the new phase (Fig. 1). The resulting ultrafast control over ferroelectricity may find rich applications in memory devices (*7*), STO-based heterostructures (*8*), and high-T_c superconductivity (*9*, *10*). In other recent experiments (*11*), ionic displacements along soft-mode coordinates have been driven through nonlinear coupling between the soft modes and other phonon modes that were excited by long-wavelength infrared pulses. In the present case, we excite the soft mode directly, using a terahertz (THz) light field to move the ions into their positions in the incipient crystalline phase. This case was foreshadowed by molecular dynamics (MD) simulations of THz field–induced switching between different FE domain orientations, a closely related type of "collective coherent control" (*12*).

STO is a widely used dielectric material that has a cubic perovskite structure at room temperature. Many members of this crystal family (e.g., $PbTiO_3$) undergo transitions into FE phases in which the transition metal ions occupy positions that are displaced from the unit cell center so that the material has a macroscopic electric polarization. The collective pathway between the cubic, paraelectric phase and the FE phase

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. *Present address: Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. †Corresponding author. Email: kanelson@mit.edu

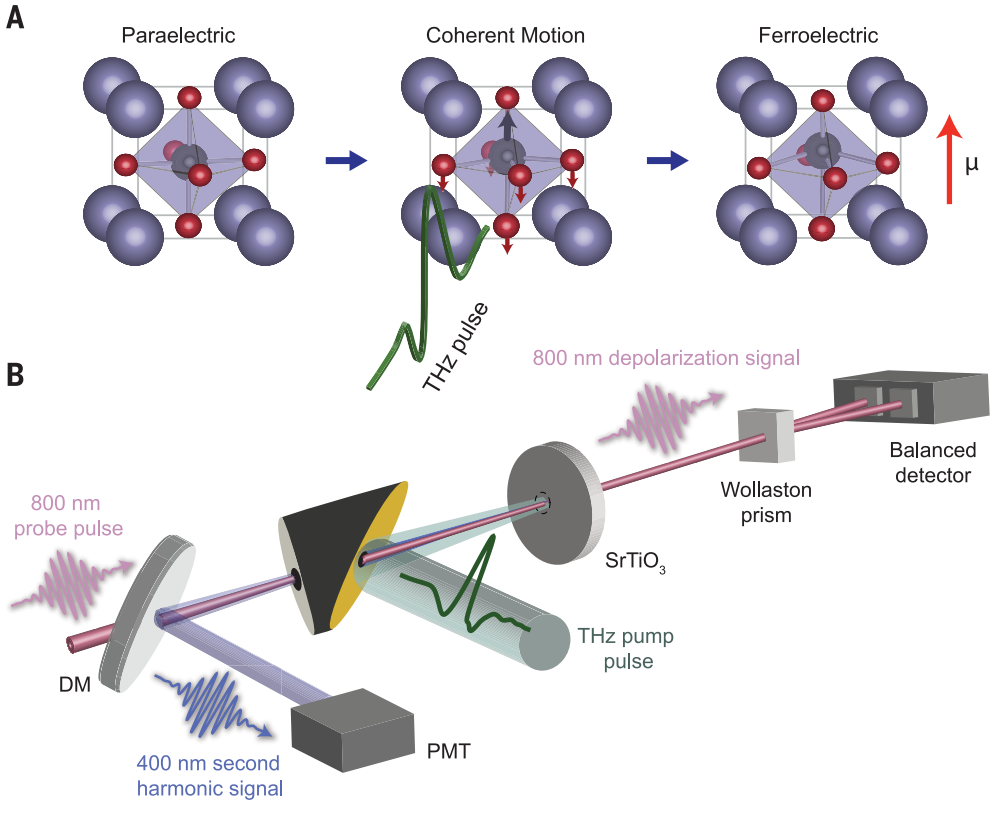


Fig. 1. Hidden ferroelectric phase accessed through THz field excitation. (**A**) Collective coherent control over material structure. A single-cycle THz-frequency electric field moves all the ions it encounters toward their positions in a new crystalline phase. In STO, the initial high-symmetry configuration around each Ti^{4+} ion has no dipole moment and the crystal is paraelectric. The incident field drives the "soft" lattice vibrational mode, moving the ions along the directions indicated into a lower-symmetry geometry with a dipole moment. Long-range ordering of dipole moments in the same direction yields a FE crystalline phase. (**B**) Experimental setup. THz field–induced lowering of the STO crystal symmetry is observed using 800-nm probe pulses that are partially depolarized (terahertz Kerr effect, or TKE) and which are partially converted to the second harmonic frequency (THz field–induced second harmonic, or TFISH). STO crystal cut is (100). The 800-nm probe pulses are polarized at 45° relative to the vertical THz polarization in the TKE experiments and 0° in the TFISH experiments, respectively. The reflected 400-nm signal is not polarization-resolved. DM, dichroic mirror; PMT, photomultiplier tube.

involves motions of the ions along the soft phonon coordinate illustrated in Fig. 1A. In contrast, upon reduction of the temperature to 105 K, STO undergoes an antiferrodistortive (AFD) structural phase transition into a second paraelectric phase of tetragonal symmetry (*13*, *14*). Further cooling reveals mode softening (decrease in frequency ω) in the usual Curie-Weiss form $\omega \propto (T - T_c)^{1/2}$ with critical temperature $T_c = 36$ K (*15*), but at such a low temperature, the zero-point quantum uncertainties in ion positions prevent long-range FE ordering of their locations. Thus, STO is a textbook example of a so-called quantum paraelectric (QPE) phase (*15*), in which dipole correlation lengths do not extend beyond nanometer length scales (*16*). Recently, studies have shown that the QPE state in STO is a result of a more complex competition among three driving forces (*17*, *18*): quantum fluctuations, AFD structural distortions (rotations of neighboring oxygen octahedra in opposite directions), and ferroelectric ordering. As a result, even subtle perturbations such as ^{18}O isotope substitution (*19*) are able to turn STO ferroelectric.

Here, we show that intense coherent THz excitation of the FE soft modes in STO can lead to highly nonlinear phonon responses that overcome the quantum fluctuations and yield clear signatures of an ultrafast QPE-to-FE phase transition. The observed signals reveal a substantial rise in ferroelectric ordering and restructuring of phonon spectra beyond a threshold THz field strength, indicating the emergence of the collective FE phase.

We carried out two complementary experiments with single-cycle THz pump pulses and time-delayed optical probe pulses (Fig. 1B). THz field–induced second harmonic (TFISH) generation spectroscopy (*20*) was conducted to observe signals that arise from inversion-symmetry breaking due to coherent soft-mode lattice vibrational motion away from the initially centrosymmetric structure of the QPE phase. THz field–induced optical birefringence (THz Kerr effect, or TKE) spectroscopy (*21*) was performed to characterize Raman-active phonon responses that were driven nonlinearly by the THz-initiated soft-mode lattice vibrations. Figure 2 shows TFISH measurement results from STO and their Fourier transforms at several temperatures and THz field amplitudes. At temperatures above 30 K, a single mode that softens with decreasing temperature, consistent with the FE soft mode, is observed (fig. S6) (*14*, *22*). The coherent vibrational displacements in either

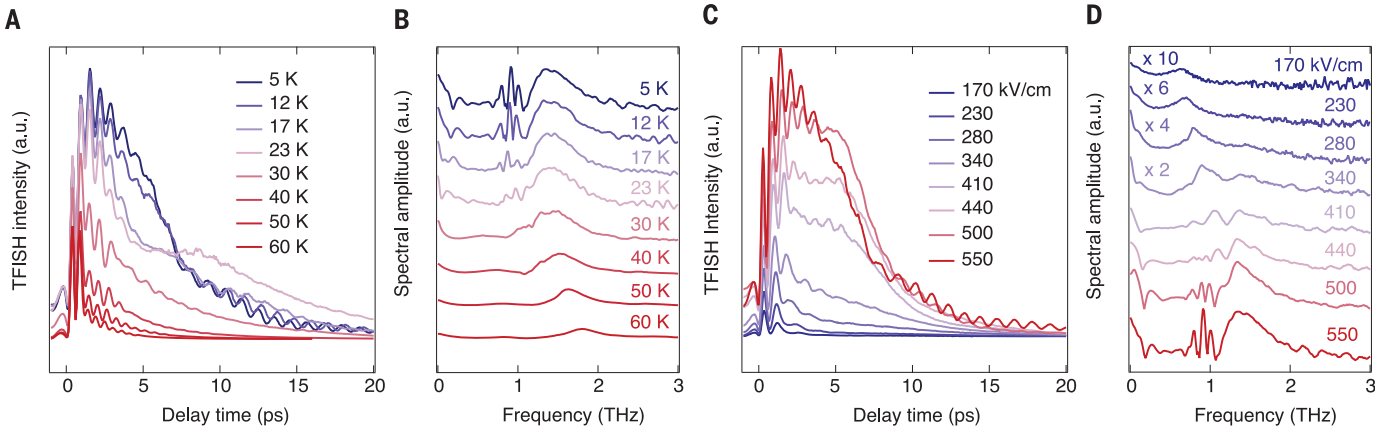


Fig. 2. STO symmetry breaking measured by optical second harmonic generation (TFISH). (**A** and **B**) Temperature-dependent TFISH signals recorded at 550 kV/cm field amplitude from STO (A) and their Fourier transforms (B). The FE soft mode is observed above 30 K, and new phonon peaks as well as nonoscillatory signals appear at lower temperatures. (**C** and **D**) THz field strength–dependent TFISH signals measured at 5 K (C) and their Fourier transforms (D). Signals at low field strengths are magnified by the amounts indicated in (D) for better visibility. Pronounced changes in the nonoscillatory signal components and the phonon spectra occur when the THz field level is increased above 340 kV/cm. The numerical first derivatives of the time-domain signals were calculated before Fourier transformation to reduce the relative amplitude of the nonoscillatory components.

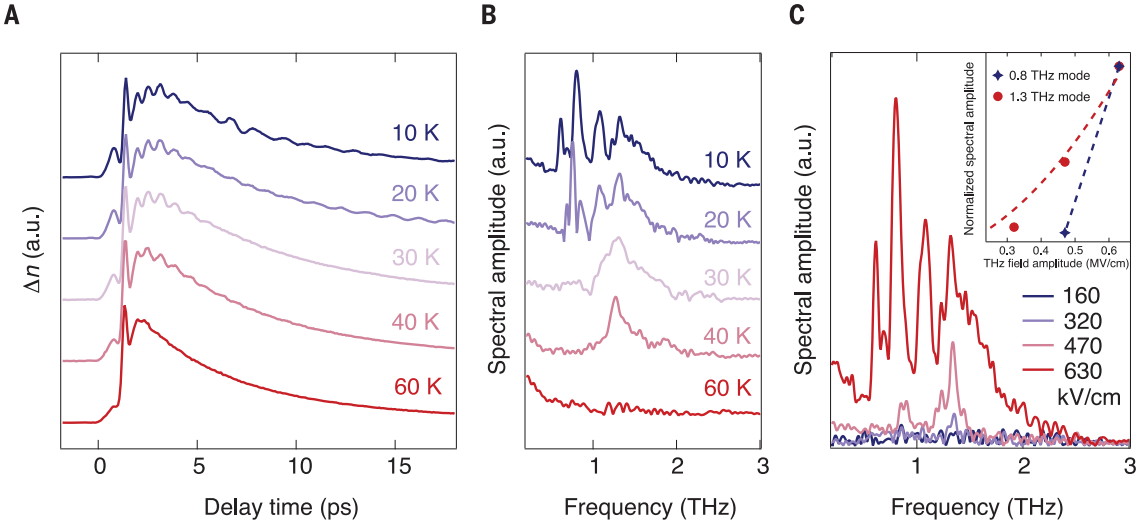


Fig. 3. Strongly nonlinear phonon responses appear in the low-symmetry STO phase. (**A** and **B**) Temperature dependence of THz-induced optical depolarization (TKE) signals recorded with 630 kV/cm THz pump field amplitude (A) and their Fourier transforms (B). The numerical first derivatives of the time-domain signals were calculated before Fourier transformation to reduce the relative amplitude of the nonoscillatory components. At temperatures of 60 K and above (*22*), no oscillatory signal is observed after THz excitation. The 1.3-THz peak and additional low-frequency modes appear at low temperatures. (**C**) THz field strength dependence of the TKE spectra at 10 K. New peaks grow in sharply as the THz field level is increased from 470 to 630 kV/cm. Inset: Quadratic fit to the 1.3-THz A_{1g} mode. The 0.8-THz mode shows faster than quadratic scaling in the THz field strength.

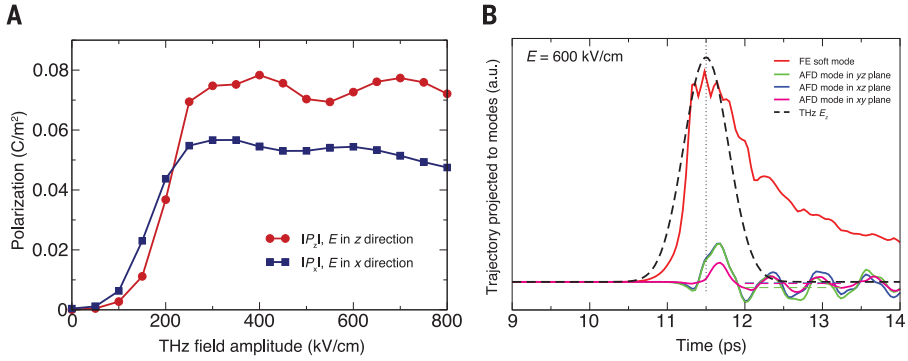


Fig. 4. MD simulation of response to STO THz excitation. (**A**) The peak global polarization induced by excitation with a THz field along different crystallographic axes. A threshold electric field amplitude of about 300 kV/cm is needed in order to fully polarize the crystal. (**B**) MD simulation trajectory projection onto different vibrational mode coordinates. The FE soft-mode response is driven directly and peaks at the same time as the *z*-polarized THz field (dashed vertical line). The antiferrodistortive (AFD) modes are driven through coupling to the FE soft mode and reach their maximum displacements after a delay. A steady-state AFD mode displacement (dashed green line shows time-averaged value) as well as FE soft-mode displacement remain well after the THz driving field has ceased.

direction break the symmetry, resulting in optical second harmonic signals that oscillate at twice the soft-mode frequency. There is also a nonoscillatory signal component due to THz-induced orientation of dipoles, whose decay becomes slower as T is reduced because of the increasing dipolar correlation (*23*). In the QPE phase at $T <$ 36 K, two features in the signals change substantially at high THz field amplitudes: (i) The nonoscillatory signal component grows in a highly nonlinear fashion as a function of the field strength, which indicates a pronounced growth in the extent of steady-state (nonoscillatory) dipole ordering (*24*); (ii) additional phonon signatures appear with amplitudes that also increase in a highly nonlinear fashion with THz field strength. These features reveal additional ionic displacements that take place as the FE crystal structure is formed. At soft-mode amplitudes sufficient to reach the new phase, collective displacements of other phonon modes (coupled nonlinearly to the soft mode) are induced. The THz-induced ordered structure is noncentrosymmetric, so oscillations about this structure produce changes in the second harmonic signal level that oscillate at the phonon frequencies, not twice the frequencies. It is noteworthy that the three distinct low-frequency peaks in the TFISH response at high field strength harden gradually as T is reduced (fig. S9), as is known to occur for the soft modes in $SrTi^{18}O_3$ below its FE phase transition temperature (*19*). It is likely that we are observing these modes, with frequencies altered slightly as a result of the nonequilibrium transient crystal structure in which we are observing them, and that their sharp onset at high fields indicates their displacements associated with the FE crystal structure. We also observe a broad phonon feature at 1.3 THz whose frequency does not appear to change with temperature and whose signal strength does not increase as sharply as the lower-frequency peaks. We believe this behavior to be consistent with a Raman-active A_{1g} mode (we retain the "A_{1g}" label even though the crystal symmetry has been changed; fig. S1B shows the AFD mode coordinate) (*25*) that is coupled anharmonically to the soft mode. Similar nonlinear coupling has been observed in room-temperature STO using femtosecond x-ray diffraction (*26*).

Figure 3 shows TKE data recorded at several sample temperatures and THz field strengths. Although the optical and THz pulses propagate with very different velocities in STO (*21*, *27*), the strong THz absorption (*28*) ensures that this does not affect the time-dependent signals. At high temperatures (Fig. 3A), only nonoscillatory signals are observed. Unlike such signals in the TFISH data, these signals show only weakly T-dependent decay kinetics and they do not increase substantially as functions of either temperature or THz field amplitude (figs. S7 and S8). They are associated with dipole alignment rather than FE orientation or polarity (*22*). We also observe the A_{1g} mode at 1.3 THz, which increases quadratically with THz field strength, indicating ordinary anharmonic coupling to the FE soft mode as suggested above. By far most striking is the emergence of several low-frequency phonon features whose strengths depend in a highly nonlinear fashion on the THz field strength, clearly similar to what we observed in TFISH measurements. We conclude from all the experimental evidence that at sufficiently large soft-mode amplitudes, an ultrafast FE phase transition is triggered. The strong nonoscillatory TFISH signals reveal the associated increase in FE ordering. The modes that grow in sharply as the THz field amplitude is increased reveal collective displacements of ions along multiple vibrational modes that are coupled nonlinearly to the soft mode and also reveal the change in lattice symmetry. It has been suggested that excitation of the Raman modes may provide constructive feedback to the FE soft mode that drives them by disrupting the balance between AFD and FE structural distortions (*17*, *18*, *29*), thereby dynamically destabilizing the paraelectric ground state on a multidimensional energy landscape.

To reach a clearer understanding of THz-induced effects, we conducted classical MD simulations for a supercell of 20 × 20 × 20 unit cells in an isothermal-isobaric ensemble, with the interatomic interaction described by the bond valence model and with external pressure applied (*22*). For each MD simulation, the system was first relaxed for 100 ps to reach equilibrium at 5 K, and then a Gaussian-profile electric field pulse (duration 0.66 ps, full width at half maximum) was applied in either the z or x crystallographic direction. Trajectories of the system were collected for 50 ps, with the electric field reaching its maximum at 11.5 ps. To evaluate whether the electric field could induce ferroelectricity in STO, we performed simulations with different field amplitudes, and in each case we calculated the global polarization of the system from the collected trajectories. As shown in Fig. 4A, the induced polarization rises sharply over a narrow range of applied field amplitudes, saturating at around 300 kV/cm along both the x and z directions. The important result of the simulations is the confirmation that a single-cycle THz field can induce a substantial global FE polarization when the field is above a threshold level on the order of 200 kV/cm. By calculating the projections of the MD simulation trajectory along different lattice vibrational mode coordinates, we also confirmed that the key displacements occur along the FE soft mode and the coupled AFD mode coordinates [see (*22*) for details of the mode assignments], whose calculated time-dependent responses are shown in Fig. 4B. The soft-mode response is driven directly by the THz field and reaches its peak at the same time as the peak field. The AFD modes are driven indirectly through their anharmonic coupling to the FE soft mode, and their peak displacements are delayed as a result. The soft mode and the AFD modes show steady-state displacements that persist well after the THz field is gone, indicating relaxation of the coupled system into the FE structure.

The experimental data and MD simulations together demonstrate a THz-induced ultrafast QPE-to-FE phase transition in STO. The THz field drives the soft mode, and additional coupled-mode displacements occur to reach the FE structure. Our results demonstrate collective coherent control of material structure that may be applicable to a wide range of classical and quantum phase transitions in which soft phonon modes play key roles in the collective structural transformations.

REFERENCES AND NOTES

1. M. Rini *et al.*, *Nature* **449**, 72–74 (2007).
2. D. Fausti *et al.*, *Science* **331**, 189–191 (2011).
3. L. Stojchevska *et al.*, *Science* **344**, 177–180 (2014).
4. Y. H. Wang, H. Steinberg, P. Jarillo-Herrero, N. Gedik, *Science* **342**, 453–457 (2013).
5. H. Ichikawa *et al.*, *Nat. Mater.* **10**, 101–105 (2011).
6. T. Morimoto *et al.*, *Phys. Rev. Lett.* **118**, 107602 (2017).

7. J. F. Scott, C. A. Paz de Araujo, *Science* **246**, 1400–1405 (1989).
8. Y.-Y. Pai, A. Tylan-Tyler, P. Irvin, J. Levy, *Rep. Prog. Phys.* **81**, 036503 (2018).
9. J.-F. Ge *et al*., *Nat. Mater.* **14**, 285–289 (2015).
10. A. Stucky *et al*., *Sci. Rep.* **6**, 37582 (2016).
11. R. Mankowsky, A. von Hoegen, M. Först, A. Cavalleri, *Phys. Rev. Lett.* **118**, 197601 (2017).
12. T. Qi, Y. H. Shin, K. L. Yeh, K. A. Nelson, A. M. Rappe, *Phys. Rev. Lett.* **102**, 247603 (2009).
13. G. Shirane, Y. Yamada, *Phys. Rev.* **177**, 858–863 (1969).
14. H. Vogt, *Phys. Rev. B* **51**, 8046–8059 (1995).
15. K. Müller, H. Burkard, *Phys. Rev. B* **19**, 3593–3602 (1979).
16. S. E. Rowley *et al*., *Nat. Phys.* **10**, 367–372 (2014).
17. U. Aschauer, N. A. Spaldin, *J. Phys. Condens. Matter* **26**, 122203 (2014).
18. A. Yamanaka *et al*., *EPL* **50**, 688–694 (2000).
19. M. Takesada, M. Itoh, T. Yagi, *Phys. Rev. Lett.* **96**, 227602 (2006).
20. D. J. Cook, J. X. Chen, E. A. Morlino, R. M. Hochstrasser, *Chem. Phys. Lett.* **309**, 221–228 (1999).
21. M. C. Hoffmann, N. C. Brandt, H. Y. Hwang, K.-L. Yeh, K. A. Nelson, *Appl. Phys. Lett.* **95**, 231105 (2009).
22. See supplementary materials.
23. Y. Uesu *et al*., *J. Phys. Soc. Jpn.* **73**, 1139–1142 (2004).
24. A. Subedi, *Phys. Rev. B* **95**, 134113 (2017).
25. W. G. Nilsen, J. G. Skinner, *J. Chem. Phys.* **48**, 2240–2248 (1968).
26. M. Kozina *et al*., *Nat. Phys.* **15**, 387–392 (2019).
27. M. Sajadi, M. Wolf, T. Kampfrath, *Opt. Express* **23**, 28985–28992 (2015).
28. W. G. Spitzer, R. C. Miller, D. A. Kleinman, L. E. Howarth, *Phys. Rev.* **126**, 1710–1721 (1962).
29. D. M. Juraschek, S. F. Maehrlein, *Phys. Rev. B* **97**, 174302 (2018).

ACKNOWLEDGMENTS

We thank T. Egami, A. Steinbacher, Y. Wang, and E. Demler for stimulating discussions. **Funding:** The work at MIT was supported in part by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, under award DE-SC0019126, and by Swiss National Science Foundation fellowship P2ELP2-172290 (E.B.). The work at Penn (T.Q., J.Z., and A.M.R.) was supported by the DOE, Office of Science, Office of Basic Energy Sciences, under grant DE-FG02-07ER46431. We acknowledge computational support from the NERSC of the DOE. **Author contributions:** K.A.N. conceived the project and the experiments together with X.L. and J.L.; the time-resolved THz setup was built by X.L., who performed the TFISH and TKE measurements and analyzed the data with support from E.B.; T.Q., J.Z., and A.M.R. provided MD calculations of the pump thresholds and mode displacements for the THz-driven ferroelectricity and the time- and mode-resolved responses; the manuscript was written by K.A.N., X.L., and E.B. with input from all authors. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1079/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Table S1
References (*30*–*47*)

27 December 2018; accepted 20 May 2019
10.1126/science.aaw4913

NEUROSCIENCE

Long-duration hippocampal sharp wave ripples improve memory

Antonio Fernández-Ruiz[1]*, Azahara Oliva[1,2]*, Eliezyer Fermino de Oliveira[1,3], Florbela Rocha-Almeida[1,4], David Tingley[1], György Buzsáki[1,5]†

Hippocampal sharp wave ripples (SPW-Rs) have been hypothesized as a mechanism for memory consolidation and action planning. The duration of ripples shows a skewed distribution with a minority of long-duration events. We discovered that long-duration ripples are increased in situations demanding memory in rats. Prolongation of spontaneously occurring ripples by optogenetic stimulation, but not randomly induced ripples, increased memory during maze learning. The neuronal content of randomly induced ripples was similar to short-duration spontaneous ripples and contained little spatial information. The spike content of the optogenetically prolonged ripples was biased by the ongoing, naturally initiated neuronal sequences. Prolonged ripples recruited new neurons that represented either arm of the maze. Long-duration hippocampal SPW-Rs replaying large parts of planned routes are critical for memory.

Sharp wave ripples (SPW-Rs) in the hippocampus are considered a key mechanism for memory consolidation and action planning (*1*–*17*). SPW-Rs are composed of the sharp wave, a negative polarity deflection in the CA1 apical dendritic layer, which reflects the magnitude of the CA3 excitatory input, and the ripple, a short-lived fast oscillatory pattern (140 to 200 Hz) of the local field potential (LFP) in the CA1 pyramidal layer (*1*, *18*–*20*). The duration of ripples exhibited a skewed distribution (Fig. 1, A to C). A minority of long-duration ripples was interspersed among a majority of short- and intermediate-duration events (*9*, *11*). We hypothesized that long-duration ripples are related to mnemonic demand and, therefore, examined ripple statistics across a variety of tasks in rats. The duration of ripples that occurred during immobility periods on a novel maze was longer (Fig. 1B) ($P < 10^{-93}$; rank-sum test; $n = 14$ rats), the fraction of long ripples (>100 ms) was increased ($P < 10^{-3}$; rank-sum test), and the proportion of ripple doublets and triplets increased significantly compared with ripples on a familiar maze in the same animals ($P < 10^{-3}$ and 10^{-2}; rank-sum test) (fig. S1). In addition, the duration of awake ripples in tasks requiring memory (T-maze and M-maze delayed alternation, cheeseboard maze) was longer compared with tasks without a memory-demand requirement (open field exploration, linear or circular track running) (Fig. 1C) ($P < 10^{-90}$), as were both the proportion of long ripples ($P < 10^{-9}$) and the proportion of ripple doublets and triplets (fig. S1) ($P < 10^{-10}$ and 10^{-4}). Correct trials in the M-maze were associated with significantly longer ripples compared with error trials (Fig. 1D) ($P < 10^{-5}$ for correct versus error trials in the same session; sign-rank test), and both ripple duration and proportion of long ripples decreased as learning progressed (fig. S2) ($P < 0.001$ and 0.05, Kruskal-Wallis test). In contrast, the frequency and amplitude of ripples were largely similar across conditions (fig. S1). These observations suggest that the hippocampus generates longer-duration ripple events when memory demands are high.

Next, we examined the neuronal mechanisms underlying the differences between short (≤100 ms) and long-duration (>100 ms) ripples. Confirming the LFP measures, the spiking activity of both putative pyramidal cells and putative interneurons during ripples lasted longer in a novel environment (compared with a familiar one) and in memory tasks (compared with nonmemory tasks) (fig. S1) ($P < 0.01$ and 0.001; rank-sum test; $n = 11$ and 19 rats, respectively). The baseline firing rate of single neurons, calculated from the entire sleep-wake session, was positively correlated with the neuron's participation probability in SPW-Rs (Fig. 1E) ($P < 10^{-100}$ and 10^{-83}; Pearson's correlation; for $n = 3080$ CA1 pyramidal cells and $n = 508$ interneurons in 19 rats). The average rank order of a neuron's within-ripple sequence negatively correlated with that neuron's baseline firing rate (Fig. 1F) ($P < 10^{-82}$ and 10^{-53}) and its participation probability in SPW-Rs (Fig. 1G) ($P < 10^{-53}$ and 10^{-52}). Longer-duration ripples were characterized by the recruitment of more CA1 neurons with lower baseline firing rates (Fig. 1H) ($P < 10^{-35}$ and 10^{-21}).

To test the hypothesis that longer ripples are beneficial for memory, we optogenetically prolonged them. We virally expressed channelrhodopsin 2 (ChR2) in dorsal CA1 pyramidal cells bilaterally (Fig. 2A and fig. S3). At 3 weeks after virus injection, we implanted recording electrodes in area CA1 and placed optic fibers above the pyramidal layer (three per hemisphere along the septotemporal axis) (Fig. 2B). Before behavioral testing, the magnitude of optogenetic stimulation was adjusted to induce ripple-like oscillations with similar frequency and duration to spontaneously occurring ripples (Fig. 2C and fig. S4) (*18*, *21*). To prolong spontaneous ripples, online detection of ripples (fig. S5) triggered a tapered-onset 100-ms-long light stimulus (Fig. 2C), which approximately doubled the duration of each ripple and associated neuronal firing (Fig. 2, D and E, and fig. S4) ($P < 10^{-86}$ for ripple duration and $P < 10^{-19}$ for pyramidal cell firing in spontaneous and prolonged ripples; rank-sum test). In control sessions, optogenetic stimulation followed the ripple by a random delay (400 to 1000 ms) (Fig. 2C).

We tested the consequence of closed-loop ripple prolongation on memory performance in a hippocampus-dependent M-maze task (Fig. 3A) (*22*). Rats were rewarded on the M-maze each time they visited the end of one of the three maze arms in the correct task sequence (center-left-center-right-center, and so on) (Fig. 3A). This task has two parts: In the "outbound" alternation part, the rat starts in the central arm and, after a 20-s delay, must chose the side arm opposite the most recently visited one to receive a sugar-water reward (*7*). In the "inbound" part, the rat has to return to the center arm from either of the side arms without a delay. The outbound component of the task has been shown to be impaired with SPW-R disruption but not the inbound part (*7*). The effect of closed-loop elongation of ripples on memory performance was compared to two control conditions: sessions in which optogenetic stimulation was delivered at randomly delayed intervals (Fig. 2C) after the detected ripple ("random" condition) and sessions with a "no-stimulation" control condition. Two sessions (30 min each) were conducted on each of 10 days (a.m. and p.m.), comparing two conditions in a given cohort (ripple prolongation versus no stimulation or random stimulation, and random stimulation versus no stimulation; five rats in each of three cohorts, plus another "sham" cohort of five rats with no stimulation) (fig. S7C). The majority of both short and long ripples occurred in the delay period, followed by the reward-side areas (Fig. 3B and fig. S6). Because performance of the no-stimulation and sham groups did not differ (fig. S6C) [$P > 0.05$; repeated measures analysis of variance (ANOVA); main effect of group], subjects were combined into a single no-stimulation control group. Closed-loop ripple prolongation significantly increased performance on the outbound component of the task when compared with no-stimulation or random-stimulation controls (Fig. 3C) (main effect of group; $P < 10^{-10}$; repeated measures ANOVA), whereas no difference was found between the no-stimulation and random-stimulation sessions ($P > 0.05$) (fig. S7C). These differences were present both when comparing all animals per condition across days (Fig. 3C) and different conditions in the same animal and day (figs. S7 and S8). The number of days to reach the 80% correct choice criterion was

[1]New York University Neuroscience Institute, New York University, New York, NY 10016, USA. [2]Department of Neuroscience and Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA. [3]Center for Mathematics, Computing and Cognition, Universidade Federal do ABC, São Bernardo do Campo, São Paulo, Brazil. [4]Division of Neurosciences, University Pablo de Olavide, 41013 Seville, Spain. [5]Center for Neural Science, New York University, New York, NY 10016, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: gyorgy.buzsaki@nyumc.org

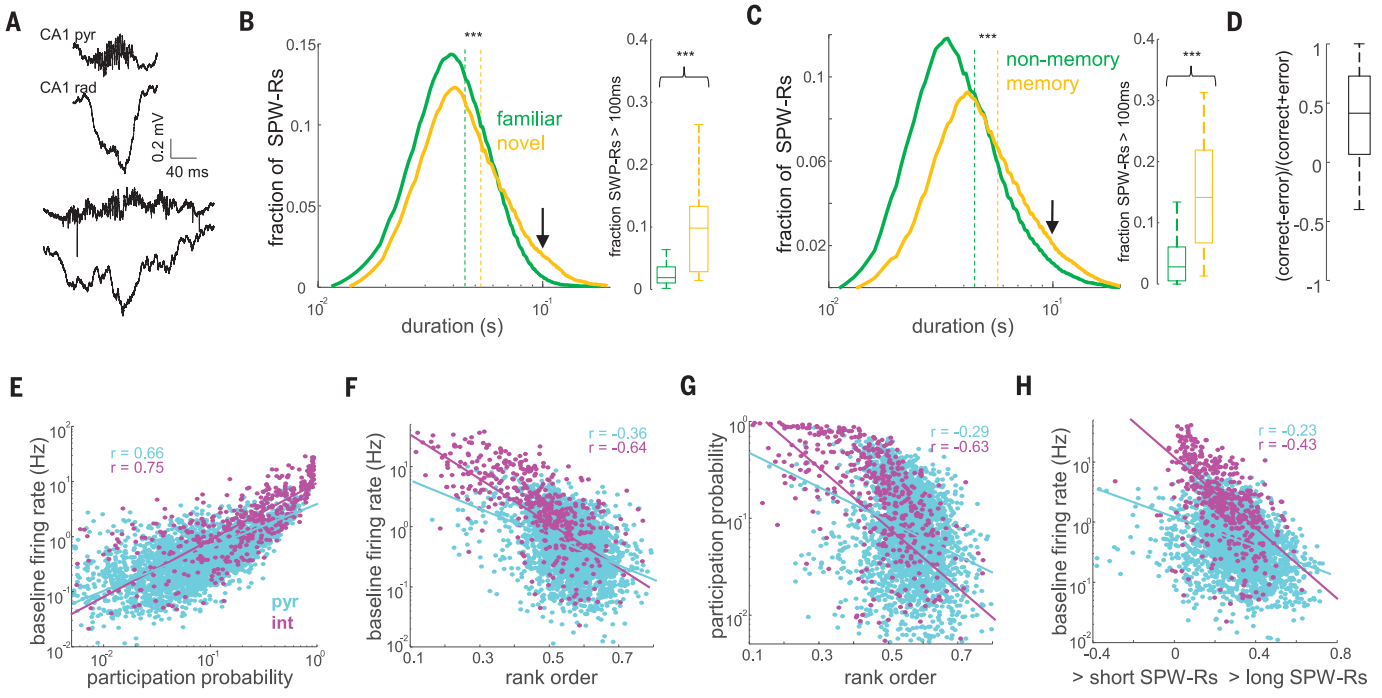


Fig. 1. Longer-duration waking SPW-Rs occur more frequently under increased memory demand. (**A**) Examples of short (top) and long (bottom) SPW-Rs during waking state. pyr, pyramidal; rad, stratum radiatum. (**B**) Distribution of the durations of SPW-Rs in familiar and novel environments (n = 16,732 and 7495 SPW-Rs, respectively, from 14 rats). Dashed lines indicate medians. The fraction of long SPW-Rs (>100 ms, indicated with an arrow) is shown in the right panel. ***P < 0.001; rank-sum test. (**C**) Distribution of the durations of SPW-Rs and the fraction of long SPW-Rs in memory-demanding and nonmemory tasks (see supplementary materials and methods) (n = 33,690 and 31,637 SPW-Rs, respectively, from 31 rats). (**D**) Proportion of SPW-Rs >100 ms in the delay box of the M-maze prior to correct versus error trials (ratio per session; n = 45 sessions from 10 rats). (**E**) Correlation between the baseline firing rate and mean participation probability in SPW-Rs for CA1 pyramidal cells ($P < 10^{-100}$) and interneurons (int) ($P < 10^{-83}$). (**F**) Correlation between the baseline firing rate and the mean rank order in SPW-Rs (pyr, $P < 10^{-82}$; int, $P < 10^{-53}$). (**G**) Correlation between the mean participation probability and the mean rank order in SPW-Rs (pyr, $P < 10^{-52}$; int, $P < 10^{-52}$). (**H**) Firing rate as a function of that neuron's probability of firing in long versus short SPW-Rs [(long − short)/(long + short)] (pyr, $P < 10^{-35}$; int, $P < 10^{-29}$).

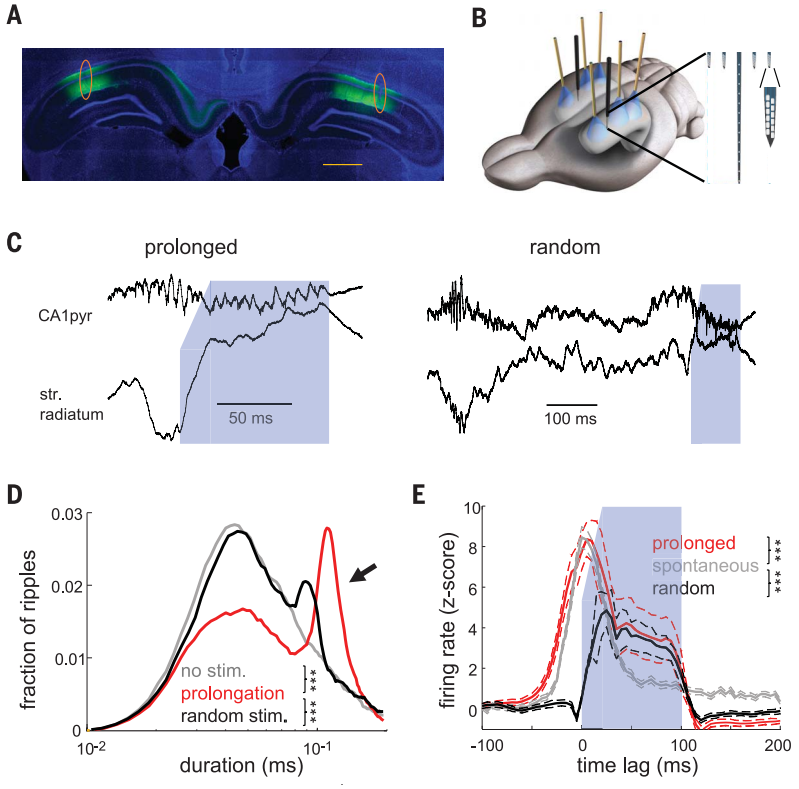


Fig. 2. Closed-loop optogenetic prolongation of ripples. (**A**) CamKII-ChR2-EYFP expression (green) in the dorsal hippocampus. Blue is 4′,6-diamidino-2-phenylindole (DAPI) staining. Orange circles show electrode tracks. Scale bar, 1 mm. (**B**) Three optic fibers and silicon probe arrays were implanted above the dorsal-posterior CA1 pyramidal layer. (Inset) Detail of silicon probe tips. (**C**) Examples of closed loop–prolonged ripple (left) and induced ripple at a random delay after the SPW-R (right). Blue shading indicates duration of light activation (100 ms). Sharp wave is absent in stratum radiatum (str. radiatum) during optogenetic stimulation. (**D**) Distribution of ripple durations in no-stimulation, ripple-prolongation, and random-stimulation sessions (n = 10 sessions of each type from 10 rats). (**E**) Z-scored firing ± SEM rates of pyramidal neurons during spontaneous and induced ripples (n = 1116 units for prolonged and spontaneous ripples and n = 705 for random stimulation from 5 rats). Onset of stimulation or peak of ripple power at 0 ms. ***P < 0.001; rank-sum test.

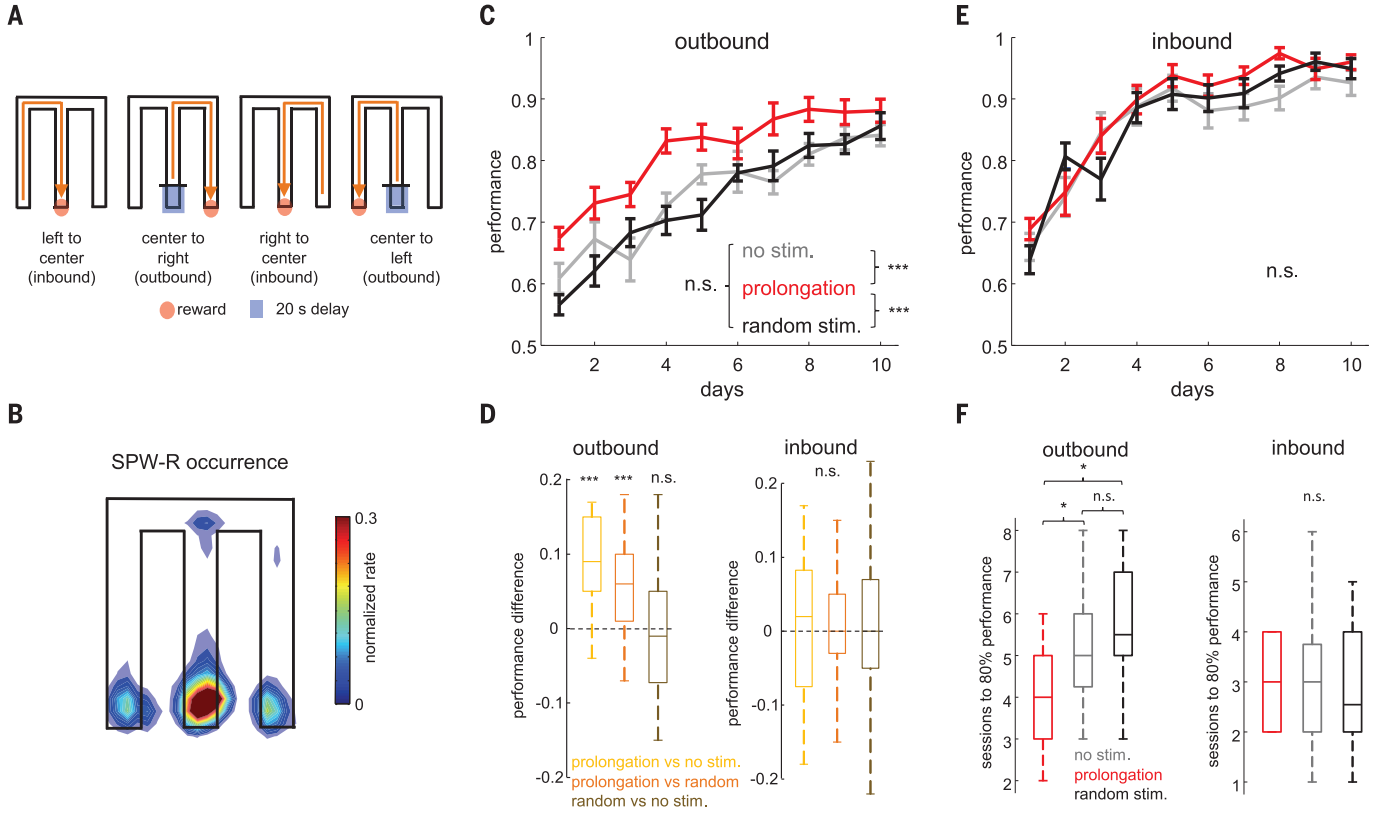


Fig. 3. Ripple prolongation improves memory. (**A**) Schematic of the M-maze task. Red dot, reward area; blue square, start area (20-s delay). (**B**) Distribution of SPW-R occurrence in the M-maze (n = 33,179 SPW-Rs from 10 rats). (**C** and **D**) Fraction of correct trials (mean ± SEM) in the M-maze task during outbound (C) and inbound (D) runs in no-stimulation, ripple-prolongation, and random-stimulation sessions (total of 20 rats). ***P < 0.001; Tukey's post-hoc test. n.s., not significant. (**E**) Performance ratios for the same day and animal for ripple-prolongation versus no-stimulation sessions, ripple-prolongation versus random-stimulation sessions, or random-stimulation versus no-stimulation sessions in all animals (n = 20). ***P < 0.001; two-sided Wilcoxon signed-rank test. (**F**) Session number in which animals reached performance criterion of 80% correct trials. *P < 0.05; rank-sum test.

reached significantly earlier in the closed-loop condition when compared with either the no-stimulation or random-stimulation conditions (Fig. 3F) (P < 0.05; rank-sum test). Ripple prolongation had no impact on the inbound component of the task (Fig. 3, D to F) (main effect of group; P > 0.05; repeated-measures ANOVA). The number of ripples per session, the number of trials in the 30-min sessions, mean running speed, theta frequency, and theta/delta band ratio did not significantly differ across the different groups (fig. S6).

In five of the AAV-CaMKII-ChR2–injected rats, we tested how truncation of SPW-Rs affected performance after a 1-week rest period. Short-duration (10 ms), high-intensity pulses were used to abort ripples. Preventing the occurrence of long-duration ripples resulted in significant deterioration of performance in the outbound, but not inbound, trials compared with random-stimulation controls (fig. S9), similar to results obtained with electrical disruption of ripples (*7*).

Searching for a physiological mechanism of the memory-improving effect of elongated ripples, we compared properties of neurons during spontaneous SPW-Rs with those recruited by optogenetic stimulation. The participating neurons active in spontaneous and optogenetically prolonged ripples were similar, as demonstrated by the significant correlations of the ripple-participation index (i.e., the probability to fire within ripples) [spontaneous versus random, $P < 10^{-38}$ and 10^{-36} for pyramidal cells and interneurons (Fig. 4A); spontaneous versus prolonged, $P < 10^{-25}$ and 10^{-7} (Fig. 4B)] and firing rates between conditions (fig. S11) (*18*). Within an elongated ripple, a given pyramidal cell typically fired either in the early (spontaneous) or the optogenetically prolonged (late) part of the ripple (Fig. 4C) (0.51 ± 0.01 and 0.43 ± 0.01 fractions) but rarely in both parts (0.06 ± 0.03%; $P < 10^{-100}$; sign-rank test). The diversity of neurons (fraction of different cells participating) in prolonged ripples was thus significantly higher compared with randomly induced ripples and spontaneous short ones (Fig. 4D) ($P < 10^{-60}$ and 10^{-100}; rank-sum test). Fast firing neurons tended to fire in the early part of the prolonged ripples, whereas neurons added to the ongoing sequence in the prolonged part were recruited mainly from the lower firing rate population (Fig. 4, E and F) ($P < 10^{-6}$ and 10^{-6} for pyramidal cells and interneurons, respectively). In addition, in prolonged ripples, a larger proportion of neurons had place fields on the maze compared with spontaneous or random ripples (P < 0.01 and 0.001; rank-sum test). Pyramidal cells recruited in prolonged and spontaneously long SPW-Rs also contained more spatial information and were more spatially selective and less sparse than those in short SPW-Rs and randomly induced ripples (fig. S12). Closed loop–prolonged ripples are thus not simply the summation of spontaneous and randomly induced events but reflect a recruitment of a different population of neurons with relevant spatial information, reminiscent of the more diverse neuronal population of long-duration spontaneous ripple events (Fig. 4D) (short versus long spontaneous ripples; $P < 10^{-100}$) (figs. S10 to S12).

We examined the relationship between spike content of ripples and place-related firing in the maze. First, we considered place cells that had the highest participation probability (first quartile) in either the early or late parts of the different types of ripples (spontaneous short, spontaneous long, prolonged, and random) and plotted the distribution of their place fields on the maze. Neurons that preferred to fire in the late part of spontaneously long or prolonged ripples had more place fields on the side arms of the maze and fewer in the center arm during

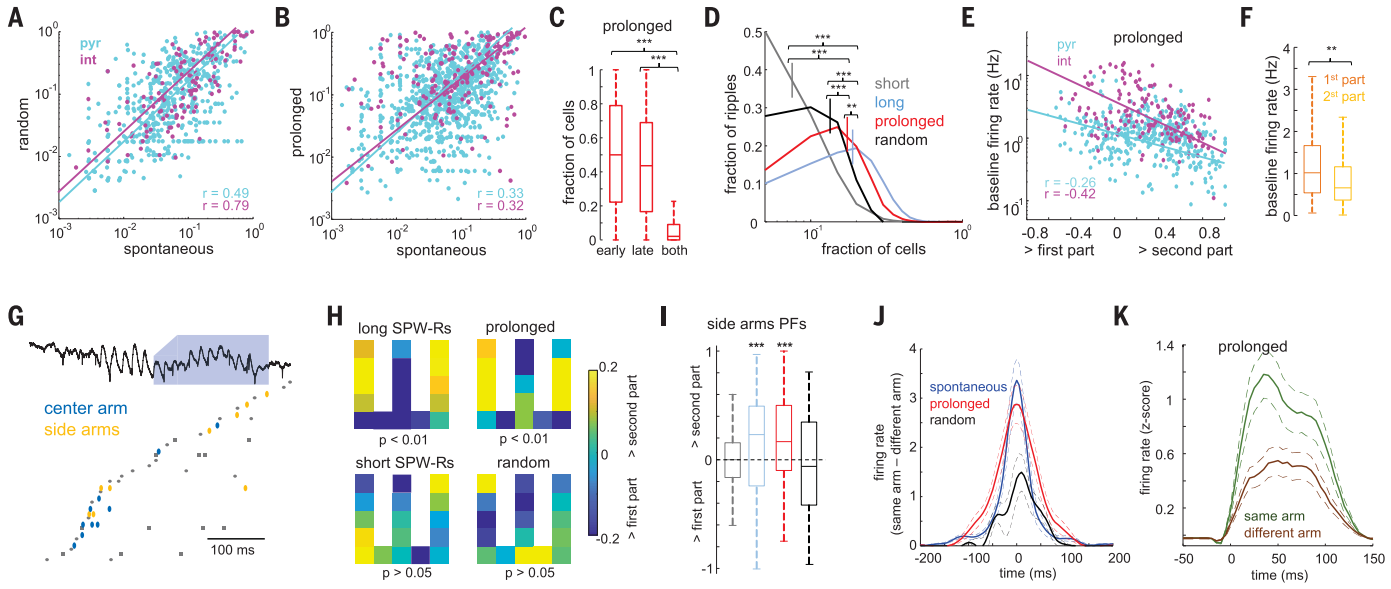


Fig. 4. Spontaneously long and optogenetically prolonged ripples contain more-diverse task-related activity. (**A** and **B**) Correlation of neuron participation probability between spontaneous and random delay–induced ripples (A) and between spontaneous and prolonged ripples (B). (**C**) Probability of firing in the early-only, late-only, or both parts of prolonged ripples. ***P < 0.001; sign-rank test. (**D**) Distribution of the fraction of all recorded neurons participating in each ripple type. Short and long (>100 ms) spontaneous SPW-Rs were analyzed separately. **P < 0.01; ***P < 0.001; rank-sum test. (**E**) Firing rate as a function of that neuron's probability of firing in the spontaneous versus closed loop–prolonged part of the ripple ($P < 10^{-7}$). Only cells with participation probability of at least 0.05 were included. (**F**) Baseline firing rate of neurons with maximum participation in the spontaneous and closed loop–prolonged parts of the ripple. **P < 0.001; rank-sum test. (**G**) Example replay sequence of travel trajectory during a prolonged ripple. Neurons representing the central and side arms fire mainly during the spontaneous and prolonged parts of the ripple, respectively. Ellipsoids indicate pyramidal cells; squares indicate interneurons. (**H**) Distribution of place fields of neurons with higher participation probability in either the early or late part of different types of ripples. P values correspond to Fisher's exact test. (**I**) Proportion of place cells with side arm–specific fields that fired with higher probability in either the first or the second part of ripples. ***P > 0.001; rank-sum test. (**J**) Differences of spike cross-correlations between z-scored firing rates of neurons representing the same or different side arms during ripples (P < 0.05 between spontaneous and prolonged ripples versus random ones; rank-sum test). (**K**) Cross-correlations between neurons' z-scored firing rates in the spontaneous and induced parts of prolonged ripples, shown separately for neurons representing the same or different side arms (P < 0.05; signed-rank test).

outbound travels compared with cells that fired in the early part (Fig. 4H). Second, we identified neurons with side arm–unique place fields during outbound travels and calculated their spike occurrence in the first and second halves of spontaneous and induced ripples. Side arm–unique place cells fired at higher probability during the second part of closed loop–prolonged and spontaneous long ripples (Fig. 4I) (P < 0.01 and 0.001; sign-rank test) but not during spontaneous short or random-delay optogenetically induced ripples (P > 0.05). There was an inverse correlation between place cells active in the left or right side arm in different ripples. During both spontaneous and prolonged ripples, the likelihood of neurons encoding the same side arms firing together was higher than that of neurons encoding opposite arms (Fig. 4J) ($P < 10^{-4}$ and 10^{-5}; rank-sum test). The difference between these likelihoods was significantly lower for random ripples (P < 0.05 and 0.05 for spontaneous and prolonged against random ripples; rank-sum test). When we compared spikes occurring in the spontaneous and optogenetically prolonged parts of closed loop–induced ripples, the cross-correlation was significantly stronger for neurons representing the same side arms than those representing opposite side arms (Fig. 4K) (P < 0.05; sign-rank test) and was similar to spontaneous SPW-Rs (fig. S12).

Our findings demonstrate that a simple measure, such as the duration of SPW-Rs, can provide valuable information about the underlying neuronal computations. Learning and correct recall in spatial memory tasks were associated with extended SPW-Rs (Fig. 1). Closed-loop optogenetic prolongation of CA1 ripples improved working memory performance (Fig. 3), whereas aborting the late part of ripples decreased performance (*7*). Prolongation of CA1 ripples did not induce repeated spiking of the already active neurons but instead recruited spikes from the low-firing population of pyramidal cells and increased the diversity of the participating neurons. Neurons recruited during the artificially elongated portion of ripples had place fields preferentially in the side arms of the maze (Fig. 4). This feature of the optogenetically prolonged ripples resembled those in spontaneous long ripple events. In contrast, randomly induced ripples recruited largely the same neurons as in short spontaneous ripples (*18*) (Fig. 4). We hypothesize that the neuronal trajectory of optogenetically induced events depends on the contemporaneous brain state (Fig. 4, D and E) (*13*, *23*, *24*), with SPW-Rs and inter-SPW-R periods referring to different network states. Once a trajectory has been selected at the onset of a SPW-R, artificial recruitment of additional neurons via optogenetic perturbation is constrained by the attractor dynamic of the CA1 network (*11*, *25*). This observation supports the idea that CA1 neuronal sequences are computed locally in CA1 (*18*, *26*) rather than being fully inherited from upstream regions (*1*, *11*, *27*). Alternatively, SPW-Rs are embedded in larger networks, including entorhinal and neocortical areas (*21*, *28*–*30*), and the differential effects of closed-loop and random stimulation might be determined by the network state in areas upstream from CA1. In either case, diversification of neurons, covering large segments of planned routes, may explain the memory-improving effect of both spontaneous long and closed loop–prolonged ripples.

REFERENCES AND NOTES

1. G. Buzsáki, *Hippocampus* **25**, 1073–1188 (2015).
2. K. Diba, G. Buzsáki, *Nat. Neurosci.* **10**, 1241–1242 (2007).

3. D. J. Foster, M. A. Wilson, *Nature* **440**, 680–683 (2006).
4. H. Xu, P. Baracskay, J. O'Neill, J. Csicsvari, *Neuron* **101**, 119–132.e4 (2019).
5. A. S. Gupta, M. A. A. van der Meer, D. S. Touretzky, A. D. Redish, *Neuron* **65**, 695–705 (2010).
6. S. Takahashi, *eLife* **4**, e08105 (2015).
7. S. P. Jadhav, C. Kemere, P. W. German, L. M. Frank, *Science* **336**, 1454–1458 (2012).
8. D. Dupret, J. O'Neill, B. Pleydell-Bouverie, J. Csicsvari, *Nat. Neurosci.* **13**, 995–1002 (2010).
9. T. J. Davidson, F. Kloosterman, M. A. Wilson, *Neuron* **63**, 497–507 (2009).
10. C. Drieu, R. Todorova, M. Zugaro, *Science* **362**, 675–679 (2018).
11. B. E. Pfeiffer, D. J. Foster, *Science* **349**, 180–183 (2015).
12. M. P. Karlsson, L. M. Frank, *Nat. Neurosci.* **12**, 913–918 (2009).
13. G. Dragoi, S. Tonegawa, *Nature* **469**, 397–401 (2011).
14. H. F. Ólafsdóttir, D. Bush, C. Barry, *Curr. Biol.* **28**, R37–R50 (2018).
15. M. A. Wilson, B. L. McNaughton, *Science* **265**, 676–679 (1994).
16. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, *Nat. Neurosci.* **12**, 1222–1223 (2009).
17. T. Nakashiba, D. L. Buhl, T. J. McHugh, S. Tonegawa, *Neuron* **62**, 781–787 (2009).
18. E. Stark, L. Roux, R. Eichler, G. Buzsáki, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10521–10526 (2015).
19. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
20. A. Oliva, A. Fernández-Ruiz, G. Buzsáki, A. Berényi, *Neuron* **91**, 1342–1355 (2016).
21. A. Oliva, A. Fernández-Ruiz, E. Fermino de Oliveira, G. Buzsáki, *Cell Rep.* **25**, 1693–1700.e4 (2018).
22. S. M. Kim, L. M. Frank, *PLOS ONE* **4**, e5494 (2009).
23. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsáki, *Science* **321**, 1322–1327 (2008).
24. M. Tsodyks, T. Kenet, A. Grinvald, A. Arieli, *Science* **286**, 1943–1946 (1999).
25. N. Maingret, G. Girardeau, R. Todorova, M. Goutierre, M. Zugaro, *Nat. Neurosci.* **19**, 959–964 (2016).
26. R. M. Memmesheimer, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11092–11097 (2010).
27. M. F. Carr, M. P. Karlsson, L. M. Frank, *Neuron* **75**, 700–713 (2012).
28. J. Yamamoto, S. Tonegawa, *Neuron* **96**, 217–227.e4 (2017).
29. J. O'Neill, C. N. Boccara, F. Stella, P. Schoenenberger, J. Csicsvari, *Science* **355**, 184–188 (2017).
30. S. P. Jadhav, G. Rothschild, D. K. Roumis, L. M. Frank, *Neuron* **90**, 113–127 (2016).

ACKNOWLEDGMENTS

We thank J. Makara, S. Mckenzie, G. Kozak, P. Petersen, L. Sjulson, R. Swanson, V. Varga, and M. Valero for insightful comments and M. Soula for technical assistance. **Funding:** This work was supported by a Sir Henry Wellcome Postdoctoral Fellowship (A.F.-R.), EMBO Postdoctoral Fellowship ALTF 120-2017 (A.O.), FAPESP grant 2017/03729-2 (E.F.d.O.), NIH (MH107396 and NS074015 to G.B. and U19NS104590), and NSF 1707316 (NeuroNex MINT). **Author contributions:** A.F.-R. and G.B. conceived of and designed the experiments; A.F.-R. and E.F.d.O. performed experiments; F.R.-A. performed immunohistological experiments; A.F.-R. and A.O. analyzed data; and G.B., A.F.-R., A.O., and D.T. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials. Part of the dataset included in this study is already available in the CRCNS.org and in the buzsakilab.com databases. The remaining data will be deposited in the same databases and are immediately available upon reasonable request. Custom Matlab scripts can be downloaded from https://github.com/buzsakilab/buzcode.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1082/suppl/DC1
Materials and Methods
Figs. S1 to S12
Table S1
References (*31–40*)

19 February 2019; accepted 1 May 2019
10.1126/science.aax0758

METASURFACES

Phase-only transmissive spatial light modulator based on tunable dielectric metasurface

Shi-Qiang Li, Xuewu Xu, Rasna Maruthiyodan Veetil, Vytautas Valuckas, Ramón Paniagua-Domínguez, Arseniy I. Kuznetsov*

Rapidly developing augmented reality, solid-state light detection and ranging (LIDAR), and holographic display technologies require spatial light modulators (SLMs) with high resolution and viewing angle to satisfy increasing customer demands. Performance of currently available SLMs is limited by their large pixel sizes on the order of several micrometers. Here, we propose a concept of tunable dielectric metasurfaces modulated by liquid crystal, which can provide abrupt phase change, thus enabling pixel-size miniaturization. We present a metasurface-based transmissive SLM, configured to generate active beam steering with >35% efficiency and a large beam deflection angle of 11°. The high resolution and steering angle obtained provide opportunities to develop the next generation of LIDAR and display technologies.

Spatial light modulators (SLMs) have widespread applications ranging from three-dimensional video projection (*1*) and additive manufacturing (*2*) to quantum (*3*) and adaptive optics. (*4*) The most versatile phase-only SLMs are capable of reconfiguring the phase retardation of light transmitted through or reflected from each pixel without changing its intensity. Typically, SLMs achieve this using liquid crystals (LCs). The LC molecules, called directors, align with each other, endowing the LC medium with a large uniaxial anisotropy of refractive index, $\Delta n = n_e - n_o$, ranging from 0.2 to 0.4 in the visible spectral range (n_e and n_o being the extraordinary and ordinary refractive indices, respectively). Moreover, these molecules rotate under the influence of an applied electric field, providing the means to dynamically control the refractive index along a given direction. One limitation of LC-based SLMs is their large pixel size, which is above 3 μm for the best reflective SLM devices and tens of micrometers for the transmissive ones. This limits the field of view (FOV) of the SLM, defined as the angular coverage of the first diffractive order. Further downsizing of the pixels while keeping the required thickness of the LC layer is limited by their mutual cross-talk (*5*). A typical commercial transmissive SLM (HOLOEYE LC 2012) has a pixel pitch of 36 μm, giving a maximum FOV of 0.7°, severely limiting its potential applications.

Recently, a new class of flat optical elements, called metasurfaces, have been developed. They abruptly modify the phase of light by designated amounts using subdiffractive optical elements called nanoantennas (*6*, *7*). Although metasurfaces have been successfully applied to realize static optical components (*8*–*11*), for various applications it is important to modify the phase dynamically (*12*–*19*). If each nanoantenna could be individually tuned by applying electrical voltages, the metasurface would act as an SLM with a subwavelength pixel size. One way to achieve tunability is by combining nanoantennas with LCs. Initial results on the integration of nanoantennas inside LC cells (*14*, *20*, *21*) showed the possibility of obtaining spectral shifts of resonances using thermally (*20*) and electrically (*21*) switched LCs, as well as on-off switchable devices (*14*). However, a universal active metasurface capable of arbitrary beam shaping through phase-only manipulation is yet to be realized.

Here, we demonstrate how to achieve this active metasurface by integrating the nanoantennas into an LC-SLM device. Modifying the LC orientation around the nanoantennas changes their local environment and resonances. In this case, the main phase accumulation happens inside the nanoantennas rather than in the LC layer, thus uncoupling it from the LC thickness. This helps to reduce the LC cell thickness, which helps to solve the cross-talk issue and to reduce the pixel size, without requiring major modifications to the existing SLM technology.

We use the Huygens' metasurface concept, realized via dielectric nanoantennas supporting spectrally overlapped electric and magnetic dipole resonances (*22*, *23*). They provide full-range phase shift from 0 to 2π around the resonances peaks (*22*) and, when the induced dipole moments have equal amplitudes and phases, suppression of the back-scattering (*24*), resulting in a close-to-unity transmission (*22*, *23*, *25*). To achieve high efficiencies in the visible spectral range, we design nanoantennas made of TiO_2 (*9*, *26*), with negligible absorption and sufficiently high refractive index ($n \sim 2.5$) to obtain resonances inside the LC environment (E7 from Merck; $n_o \sim 1.5$ and $n_e \sim 1.7$) (see fig. S1). To find the optimum dimensions, we perform full-wave simulations (*27*), starting with a square lattice of cylindrical nanoantennas. The embedding LC (thickness $h_{LC} = 1500$ nm) is sandwiched between two glass plates with its director oriented in-plane ($\theta_{LC} = 0°$) (see Fig. 1A). The incident light impinges normally onto the device, with its electric field polarized parallel to the LC director, experiencing its extraordinary refractive index. We show maps of transmission and phase for different radii of the TiO_2 nanoantennas, sweeping around the optimized dimensions and targeted wavelength of 660 nm (Fig. 1, A and B). The period in all cases is $p = 360$ nm, so that the unit cell is subdiffractive, and the nanoantenna height is $h_a = 205$ nm. Huygens' condition (high transmission and 2π-phase coverage) is achieved in a wide range of wavelengths from 650 to 675 nm, which is optimized for the chosen nanoantenna height, becoming narrower at different heights (figs. S2 and S3).

Next, we study the phase retardation of light transmitted through the optimized nanoantenna array when the LC director orientation is switched from in-plane to out-of-plane (Fig. 1C). This can be realized by applying an electrical bias between two electrodes sandwiching the LC (*28*). We see that above 670 nm the relative phase difference at different LC orientations does not exceed 0.8π. This is the spectral region in which the nanoantennas are not resonant and the phase retardation is attributed to mere propagation in the LC layer. Below 670 nm, strong phase variations are generated by spectral shifts of the nanoantenna resonances induced by the refractive index change of the surrounding LC. In Fig. 1C, we highlight three series of results corresponding to LC director orientations of 0°, 45°, and 90°. The phase in these three series is evenly spaced at around 2π/3 to each other, in the wavelength region between 660 and 670 nm. Furthermore, they have similar transmittance, ranging between 60 and 90%, thus satisfying the requirements for a three-phase-level transmissive SLM—evenly spaced, full-phase coverage and uncoupled, high transmission (see fig. S4 for the corresponding field distributions).

We then simulate a beam-steering SLM device having the unit cell shown in the inset of Fig. 1D. Instead of one nanoantenna per pixel, we use three. This gives the individually addressable (bottom) electrodes sufficient width to accommodate the fringing electric field and phase-broadening effects (*29*). Further increasing the number of particles per pixel does not strongly improve performance (see table S1). A supercell comprising three pixels (with different LC rotations and corresponding phase levels, in accordance with Fig. 1C) is periodically repeated to produce an infinite gradient metasurface. The numerically calculated diffraction efficiencies of the device (Fig. 1D) show a clear increase of the transmission into the –1 diffraction order at the wavelength of 665 nm, reaching a value of ~48%, together with a strong reduction of efficiencies into the 0 and +1 orders. This translates into the

Institute of Materials Research and Engineering, A*STAR (Agency for Science, Technology and Research), 138634 Singapore.
*Corresponding author. Email: arseniy_kuznetsov@imre.a-star.edu.sg

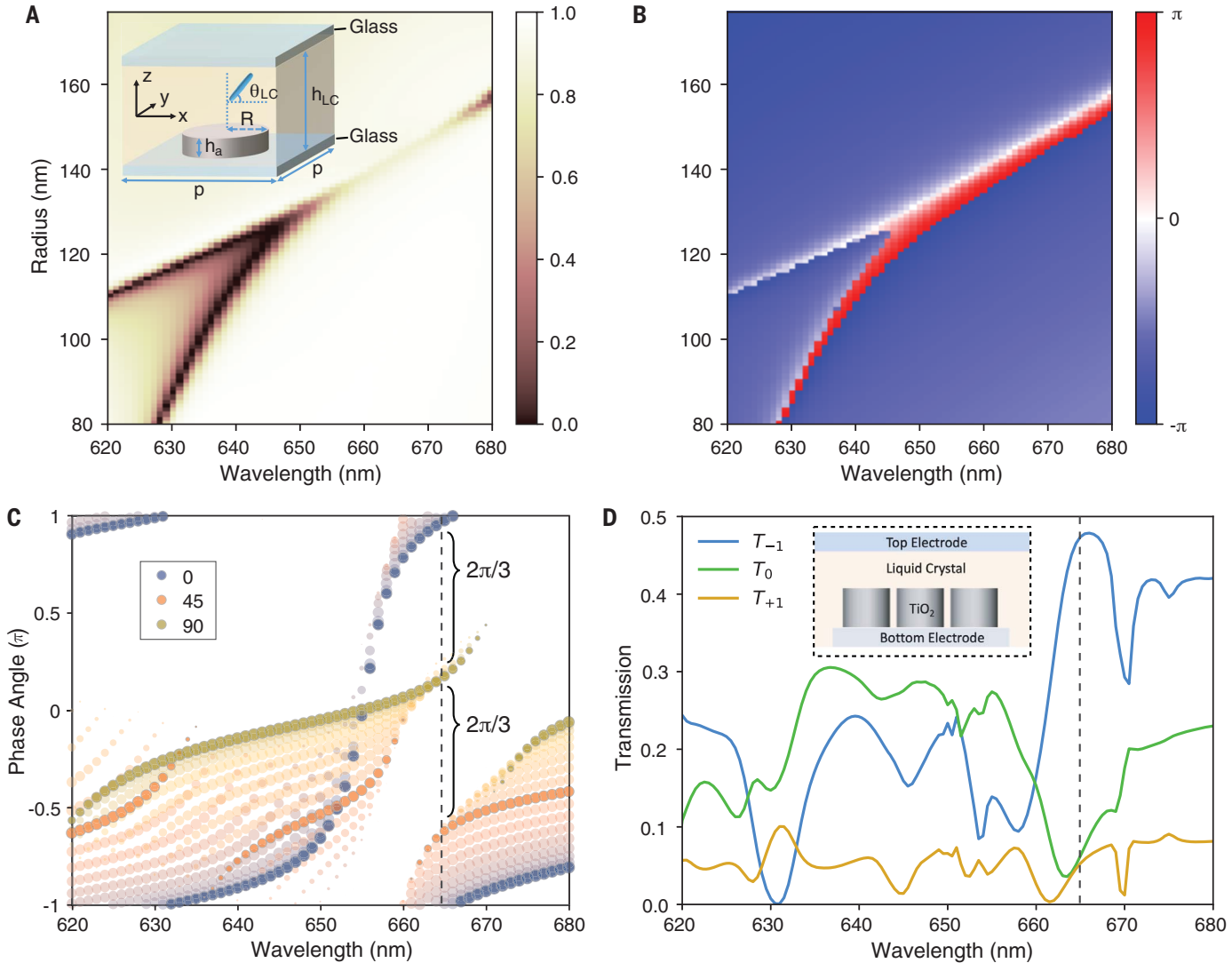


Fig. 1. Optimization of the nanoantenna geometry through simulations. (**A**) Amplitude and (**B**) phase spectra of the zero-order transmission of nanoantenna arrays with different particle radii. The inset shows the schematic of the simulated unit cell. A square array is formed by translating the unit cell in both the *x* and *y* directions. (**C**) Calculated relative phase retardation experienced by a normally incident wave passing through the optimized unit cell (h_a = 205 nm, radius R = 135 nm, p = 360 nm) as a function of the LC director rotation. The color of the marker indicates the rotation—from blue (in-plane) to orange to yellow-green (out-of-plane)—and the size indicates the associated transmittance, with respect to the bare LC layer, the ones in the legend representing 50% transmittance. (**D**) Transmission spectra of the three main diffraction orders (T_{-1}, T_0, and T_{+1}) of a wave passing through the beam-deflecting configuration. A unit cell, containing three nanoantennas, is depicted in the inset. The individual bottom electrodes are separated by an additional gap g = 60 nm.

expected beam deflection, consistent with the phase analysis of homogeneous arrays (Fig. 1C and fig. S4).

To prove our concept, we fabricate a device comprising 28 individually addressable electrodes (*27*). An optical image of the finalized device and its computer-generated design, including the printed circuit board, are shown in Fig. 2, A and B. Scanning electron microscopy (SEM) images, showing the nanoantennas on top of the electrodes before LC cell assembly, are shown in Fig. 2C (see the device transmission spectra in fig. S5).

To test the device, we first use a two-level scheme, in which we address alternating electrodes at ground voltage (keeping the in-plane LC alignment) and elevated voltage (6 to 8 V, inducing vertical LC alignment). In this situation, the device acts as a simple diffraction grating in which the periodicity, and thus the diffraction angles, can be tuned by choosing different electrode-addressing configurations. The configurations and their corresponding measured diffraction intensities are shown in Fig. 2D. The optical performance of the device in this and all subsequent cases is characterized using the spectrally resolved back focal plane imaging technique (*8*, *27*).

Next, we introduce the intermediate-voltage-level electrode inducing the partial rotation of the LC. In this way, we obtain a three-level-addressing scheme in which we can realize the beam deflection configuration analyzed theoretically above. By choosing different addressing schemes, we can tune the deflection angle, as demonstrated in Fig. 2E. The diffraction efficiencies into the three main diffraction orders, –1, 0, and +1, are shown in Fig. 2, F and G, for two characteristic cases. They correspond, respectively, to a case in which the beam deflects at around 4° and the one giving the maximum deflection angle of 11°. Both plots show a clear region at around 650 nm in which the 0 order and the +1 order are suppressed and the –1 order is enhanced, reaching values >15%. In both cases, the optimum voltage for the intermediate-level

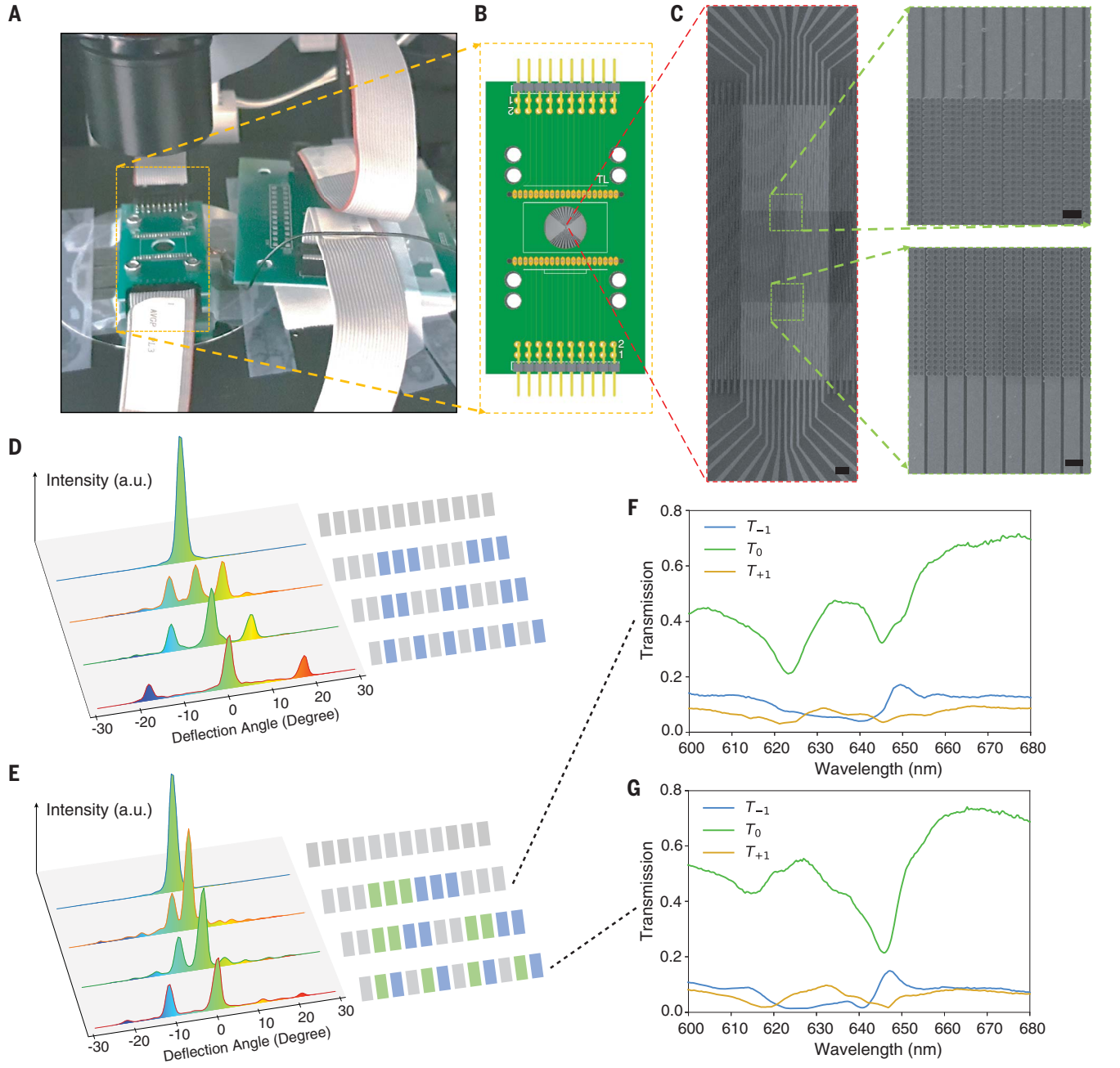


Fig. 2. Metasurface-based SLM: Demonstration of a dynamic beam deflection device. (**A**) A photo of the device mounted for measurements. (**B**) Computer-generated design consisting of the printed circuit board and the nanofabricated part (gray, central area). "TL" stands for top-left (this notation is used to keep the correct sample orientation). (**C**) SEM images of the fabricated device, showing the individual electrodes with the nanoantennas on top. Scale bars: left panel, 5 μm; right panels, 1 μm. (**D**) Two-level addressing of the device, which acts as a grating with tunable periodicity. The left side shows the experimentally measured diffraction intensities as a function of the deflection angle for different electrode-addressing configurations, shown in the corresponding right side. Gray patches represent grounded electrodes, and blue patches represent biased ones (at 8 V). a.u., arbitrary units. (**E**) As in (D), but for the three-level addressing of the device, which acts as a beam deflector with tunable deflection angle. Gray patches represent grounded electrodes, blue patches represent biased ones inducing total rotation of the LC, and green patches represent intermediate-voltage-level ones inducing partial rotation of the LC. (**F** and **G**) Measured transmission efficiencies (transmission intensity normalized to the incident intensity) of the three main diffraction orders for the case in which the beam deflects at ~4° (F) and at the maximum achievable angle of ~11° (G).

electrode was found to be 3.5 V, whereas the high-level voltage was 7 and 8 V for 4° and 11° deflection, respectively.

The measured efficiencies are noticeably lower than the theoretical predictions (fig. S6). We attribute this to perturbation of the LC alignment at the edge of the active region and the limited number of electrodes fabricated. Both are related to the small sample size and can be mitigated upon further process improvement and integration of active-matrix electrodes, which can be addressed at a large scale. To corroborate this, we designed and fabricated a simpler but larger device in which two of three neighboring bottom electrodes are connected to the two terminals of an

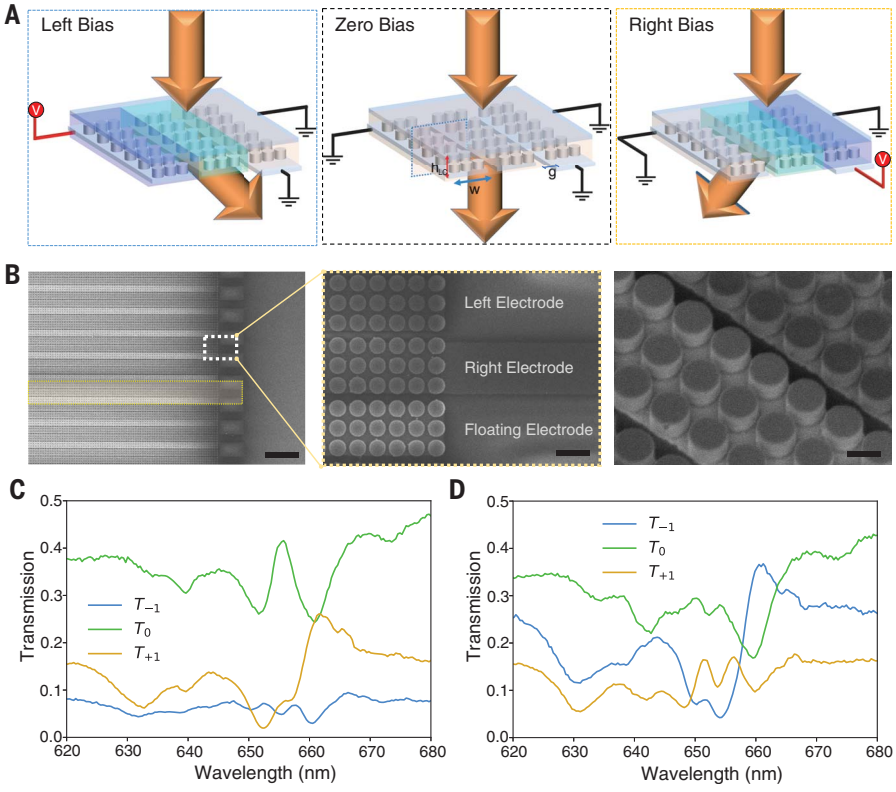


Fig. 3. Large aperture, simplified metasurface-based SLM. (**A**) Schematic drawings of the SLM and its operation concept. h_{LC} denotes the thickness of the LC layer, *w* the width of the electrode, and *g* the gap between the electrodes. (**B**) SEM images showing the fabricated device. The unconnected (floating) electrodes appear brighter because of induced charging. Each device is 120 μm by 100 μm. Scale bars: left panel, 5 μm; central panel, 600 nm; right panel (tilted view), 250 nm. The area outlined in yellow in the left panel highlights a supercell of the device. (**C**) Measured transmission efficiencies (transmission intensity normalized to the incident intensity) of the three main diffraction orders for the case of 12-V left bias. (**D**) As in (C), but for the 13-V right bias.

external voltage source. One of the connected electrodes and the top electrode are put to the ground state, and the other is biased to induce the out-of-plane switching of the LC directors (Fig. 3A). The third electrode is left floating (unconnected) to pick up an electric potential in between the biased and the grounded electrodes, so that the LC directors can attain the intermediate level of rotation leading to deflection of the incoming light beam. This design allows the reversal of the deflection direction by switching the applied voltages, as shown in Fig. 3A, but does not allow a change in deflection angle. On the other hand, it allows us to fabricate a device with a much larger number of electrodes and thus a much larger aperture size. SEM images of the fabricated device before the LC infiltration are shown in Fig. 3B.

In Fig. 3, C and D, we plot the measured diffraction efficiency spectra of the device at the optimum beam deflection conditions for each case (see figs. S7 and S8 for the voltage optimization). The efficiencies and the overall trends in this device show very good agreement with the simulated results plotted in Fig. 1D. In the best case, the experimental beam deflection efficiency at 660 nm reaches 36%, compared with 48% obtained from simulations. There are several effects not considered in the simulations, which may all contribute to this slight discrepancy, such as the fringing fields at the gaps between the electrodes, nonuniformities in nanoantenna sizes, and imperfections in the LC director alignment. Nevertheless, the results indicate that the moderate efficiencies obtained for the fully dynamic device (Fig. 2) can be improved by realizing a larger device size.

In this study, we demonstrate a one-dimensional phase-only nanoantenna-based transmissive SLM device that reaches an experimental efficiency of 36% with a pixel size of only ~1 μm and a FOV of 22°. The presence of the nanoantennas allows a reduction of the LC layer thickness required to achieve the necessary phase modulation by more than half compared with traditional SLMs. This, in turn, alleviates the phase broadening and fringing field effects, which are the main limiting factors preventing pixel-size reduction in traditional SLM devices. Our design can be further extended to two-dimensional, phase-only LC SLMs, providing opportunities to make ultra-high-resolution devices, which may perform faster and have a larger FOV, while keeping high efficiencies.

REFERENCES AND NOTES

1. H. Sasaki *et al.*, *Sci. Rep.* **4**, 6177 (2014).
2. M. Shusteff *et al.*, *Sci. Adv.* **3**, eaao5496 (2017).
3. K. H. Kagalwala, G. Di Giuseppe, A. F. Abouraddy, B. E. A. Saleh, *Nat. Commun.* **8**, 739 (2017).
4. N. Ji, *Nat. Methods* **14**, 374–380 (2017).
5. D.-K. Yang, S.-T. Wu, *Fundamentals of Liquid Crystal Devices* (Wiley, 2014).
6. P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad, R. Devlin, *Optica* **4**, 139–152 (2017).
7. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, B. Luk'yanchuk, *Science* **354**, aag2472 (2016).
8. R. Paniagua-Domínguez *et al.*, *Nano Lett.* **18**, 2124–2132 (2018).
9. M. Khorasaninejad *et al.*, *Science* **352**, 1190–1194 (2016).
10. D. Lin, P. Fan, E. Hasman, M. L. Brongersma, *Science* **345**, 298–302 (2014).
11. A. Arbabi, Y. Horie, M. Bagheri, A. Faraon, *Nat. Nanotechnol.* **10**, 937–943 (2015).
12. A. Howes, W. Wang, I. Kravchenko, J. Valentine, *Optica* **5**, 787–792 (2018).
13. R. Sarma *et al.*, *Appl. Phys. Lett.* **113**, 061108 (2018).
14. A. Komar *et al.*, *ACS Photonics* **5**, 1742–1748 (2018).
15. E. Arbabi *et al.*, *Nat. Commun.* **9**, 812 (2018).
16. G. Kafaie Shirmanesh, R. Sokhoyan, R. A. Pala, H. A. Atwater, *Nano Lett.* **18**, 2957–2963 (2018).
17. A. She, S. Zhang, S. Shian, D. R. Clarke, F. Capasso, *Sci. Adv.* **4**, eaap9957 (2018).
18. A. L. Holsteen, S. Raza, P. Fan, P. G. Kik, M. L. Brongersma, *Science* **358**, 1407–1410 (2017).
19. Q. Wang *et al.*, *Nat. Photonics* **10**, 60–65 (2016).
20. J. Sautter *et al.*, *ACS Nano* **9**, 4308–4315 (2015).
21. A. Komar *et al.*, *Appl. Phys. Lett.* **110**, 071109 (2017).
22. M. Decker *et al.*, *Adv. Opt. Mater.* **3**, 813–820 (2015).
23. Y. F. Yu *et al.*, *Laser Photonics Rev.* **9**, 412–418 (2015).
24. M. Kerker, D. S. Wang, C. L. Giles, *J. Opt. Soc. Am.* **73**, 765–767 (1983).
25. I. Staude *et al.*, *ACS Nano* **7**, 7824–7832 (2013).
26. E. Khaidarov *et al.*, *Nano Lett.* **17**, 6267–6272 (2017).
27. Materials and methods are available as supplementary materials.
28. N. Mukohzaka, N. Yoshida, H. Toyoda, Y. Kobayashi, T. Hara, *Appl. Opt.* **33**, 2804–2811 (1994).
29. P. G. de Gennes, J. Prost, *The Physics of Liquid Crystals*, vol. 83 of *International Series of Monographs on Physics* (Oxford Univ. Press, 1995).

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Y. H. Fu, Y. F. Yu, Z. Pan, and S. T. Ha, and also the fabrication support provided by K. Goh, S. Yap, A. Huang, and Y. T. Toh. **Funding:** The authors acknowledge financial support from National Research Foundation of Singapore (grant NRF-NRFI2017-01), IET A F Harvey Engineering Research Prize 2016, A*STAR SERC Pharos program (grant 152 73 00025) (Singapore), and AME Programmatic Grant A18A7b0058 (Singapore). **Author contributions:** S.-Q.L. performed simulations, sample nanofabrication, device design, electro-optical characterization, and wrote the first draft; X.X. performed simulations and optical characterization; R.M.V. performed liquid crystal cell fabrication and characterization; V.V. performed SEM measurements; R.P.-D. conceived the idea and contributed to simulations and overall supervision of the project; A.I.K. conceived the idea and supervised the whole project. All authors discussed the results and worked on the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1087/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1
Reference (*30*)

15 January 2019; accepted 22 May 2019
10.1126/science.aaw6747

ELECTROCHEMISTRY

Atomically dispersed Fe^{3+} sites catalyze efficient CO_2 electroreduction to CO

Jun Gu[1], Chia-Shuo Hsu[2], Lichen Bai[1], Hao Ming Chen[2]*, Xile Hu[1]*

Currently, the most active electrocatalysts for the conversion of CO_2 to CO are gold-based nanomaterials, whereas non–precious metal catalysts have shown low to modest activity. Here, we report a catalyst of dispersed single-atom iron sites that produces CO at an overpotential as low as 80 millivolts. Partial current density reaches 94 milliamperes per square centimeter at an overpotential of 340 millivolts. Operando x-ray absorption spectroscopy revealed the active sites to be discrete Fe^{3+} ions, coordinated to pyrrolic nitrogen (N) atoms of the N-doped carbon support, that maintain their +3 oxidation state during electrocatalysis, probably through electronic coupling to the conductive carbon support. Electrochemical data suggest that the Fe^{3+} sites derive their superior activity from faster CO_2 adsorption and weaker CO absorption than that of conventional Fe^{2+} sites.

Electrochemical reduction of carbon dioxide (CO_2) is a promising approach to store intermittent renewable solar and wind energy in carbon-based fuels and chemicals, leading to reduced anthropogenic CO_2 emission (*1*). To achieve high energy efficiency and scalability, the reaction must occur rapidly and selectively at low overpotentials. Numerous electrocatalysts have been developed for CO_2 reduction (*2*), among which gold (Au) and, to a lesser degree, silver (Ag) are the most efficient at low overpotentials. For example, the Faradaic efficiency of carbon monoxide (CO) formation can exceed 90% with Au- (*3*–*5*) and Ag-based (*6*, *7*) catalysts. On certain Au nanostructures, the partial current density of CO (denoted as j_{CO}) reached 10 mA cm^{-2} at overpotentials even lower than 300 mV (*4*, *5*). Catalysts composed solely of Earth-abundant elements typically have low selectivity for CO_2 reduction (*8*–*12*). Recently, many single-atom catalysts (*13*, *14*) have been developed, in which numerous catalytic metal sites separated from each other were chemically and electronically constrained on solid supports. These catalysts exhibit properties and activity distinct from both nanoparticles and molecular complexes of the same metal elements. Among them, iron (Fe) (*15*, *16*), cobalt (Co), (*17*) and nickel (Ni) (*18*–*20*) catalysts were reported to exhibit Faradaic efficiency of CO formation comparable with those of Au and Ag catalysts. However, with these non–precious metal catalysts, much larger overpotentials were required to obtain the same j_{CO}. Here, we report a catalyst with dispersed single-atom Fe sites with ultrahigh activity for CO_2 electroreduction to CO.

The Fe catalyst (Fe^{3+}–N–C) was prepared through the pyrolysis of Fe-doped zinc (Zn) 2-methylimidazolate framework (ZIF-8) (*21*) under N_2 at 900°C. The precursor adopts the same crystal structure as that of undoped ZIF-8 (fig. S2A), with a mole ratio of Fe:Zn of 4:96 (fig. S2E). Fe ions occupy Zn sites and are coordinated by four pyrrolic-type nitrogens (N), as revealed by the fitting of the Fe K-edge extended x-ray absorption fine structure (EXAFS) spectrum (fig. S2, G to I, and table S1). Fe^{3+}–N–C is porous, with a Brunauer-Emmett-Teller surface area of 772 m^2 g^{-1} (fig. S3A) and an electrochemical (double-layer) surface area of 554 m^2 g^{-1} (fig. S3D). The porosity was confirmed by means of high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) (Fig. 1A). Inductively coupled plasma optical emission spectrometry (ICP-OES) analysis showed the weight fractions of Fe and Zn to be 2.8 and 3.4%, respectively, corresponding to a nearly equal mole ratio of Fe:Zn. A similar Fe:Zn mole ratio was found by means of x-ray photoelectron spectrometry (XPS) (fig. S3E) and energy dispersive x-ray spectroscopy (EDS) (fig. S4F) measurements. The majority of Zn ions in the ZIF-8 precursor were presumably reduced to Zn particles, which then evaporated during pyrolysis. The x-ray diffraction (XRD) pattern of Fe^{3+}–N–C (fig. S3G) showed a broad feature at ~25° corresponding to the interlayer distance of the carbon matrix (with a *d* value of ~0.35 nm). No diffraction peaks of any crystalline species of Fe and Zn were observed. Likewise, no nanoparticles were found in the transmission electron microscopy (TEM) and high-resolution TEM (HRTEM) images; only curved fringes of the layered carbon matrix were observed (fig. S4, D and E). The EDS mappings (Fig. 1, B and C) revealed the homogeneous distributions of Fe and N in the carbon matrix. In the aberration-corrected HAADF-STEM image with atomic resolution (Fig. 1D), the bright spots with size of ~0.2 nm correspond to atomically dispersed Fe and Zn sites. The Fe 2p3/2 XPS spectrum (fig. S3F) and the Fe K-edge x-ray absorption near-edge structure (XANES) spectrum (Fig. 1F) showed binding and edge energies close to those of Fe_2O_3 and Fe^{3+}-tetraphenylporphyrin-Cl (Fe^{3+}TPPCl), indicating that the Fe ions in the as-synthesized Fe^{3+}–N–C were in the +3 oxidation state. Thus, the Fe ions were oxidized from +2 to +3 during the pyrolysis, which is in agreement with previous reports of pyrolysis of Fe-containing organic precursors (*15*, *22*). The oxidants might be of the same species as protons or residual oxygens that oxidized the carbon skeleton of the ZIF precursor. Fe K-edge EXAFS (Fig. 1H) supported the atomic dispersion of Fe sites in Fe^{3+}–N–C. The fitting of the spectrum (table S1) indicated that the Fe center adopts a planar Fe–X_4 (X = N or C) structure. The average coordination numbers of Fe–N and Fe–C paths were 3.4 and 0.5, respectively. No Fe–Fe bond was detected.

As shown by the linear sweep voltammetry (LSV) curve of Fe^{3+}–N–C in CO_2-saturated 0.5 M potassium bicarbonate ($KHCO_3$) electrolyte (fig. S5), the onset potential was more positive than −0.20 V versus reversible hydrogen electrode (RHE). Compared with Fe^{3+}–N–C, the current density of the Fe-free control sample, Zn–N–C (prepared by pyrolysis of undoped ZIF-8), was negligible, indicating that the electrocatalytic activity of Fe^{3+}–N–C originates from Fe sites. We first tested the electrocatalytic activity in CO_2 reduction using carbon paper electrodes in an H-cell (fig. S6A). CO and H_2 were the only gas-phase products, and no solution-phase product was detected (fig. S6, B to D). CO was detected after electrolysis at −0.19 V versus RHE, equivalent to an overpotential of 80 mV (fig. S6, E and F). The Faradaic efficiency of CO was higher than 80% between −0.2 and −0.5 V versus RHE (Fig. 2A). The j_{CO} reached 20 mA cm^{-2} at −0.47 V versus RHE (overpotential of 360 mV) (Fig. 2B). The rate of CO_2 reduction might be limited by mass transport in an H-cell (*23*). Thus, we deposited Fe^{3+}–N–C on a gas diffusion electrode (GDE) (*24*). The electrolysis was then conducted in a flow cell with N_2-saturated 0.5 M $KHCO_3$ as the catholyte, and CO_2 gas was fed behind the GDE (fig. S7A). Ni–Fe layered double hydroxide (LDH) nanosheets (*25*) were used as the electrocatalyst for the anodic reaction (oxygen evolution). At −0.45 V versus RHE (overpotential of 340 mV), j_{CO} reached 94 mA cm^{-2} (corresponding to 1.75 $mmol_{CO}$ $hour^{-1}$ cm^{-2}) (Fig. 2B), with Faradaic efficiency of CO on the cathode higher than 90% (Fig. 2A). Good reproducibility was observed in measurements of three independently prepared samples, giving a small standard deviation (Fig. 2, A and B).

We compare the j_{CO} of Fe^{3+}–N–C to other state-of-the-art catalysts in Fig. 2C and fig. S8. The j_{CO} of Fe^{3+}–N–C measured in an H-cell between −0.2

[1]Laboratory of Inorganic Synthesis and Catalysis, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), EPFL-ISIC-LSCI, BCH 3305, Lausanne CH 1015, Switzerland. [2]Department of Chemistry, National Taiwan University, Taipei 10617, Taiwan.
*Corresponding author. Email: haomingchen@ntu.edu.tw (H.M.C.); xile.hu@epfl.ch (X.H.)

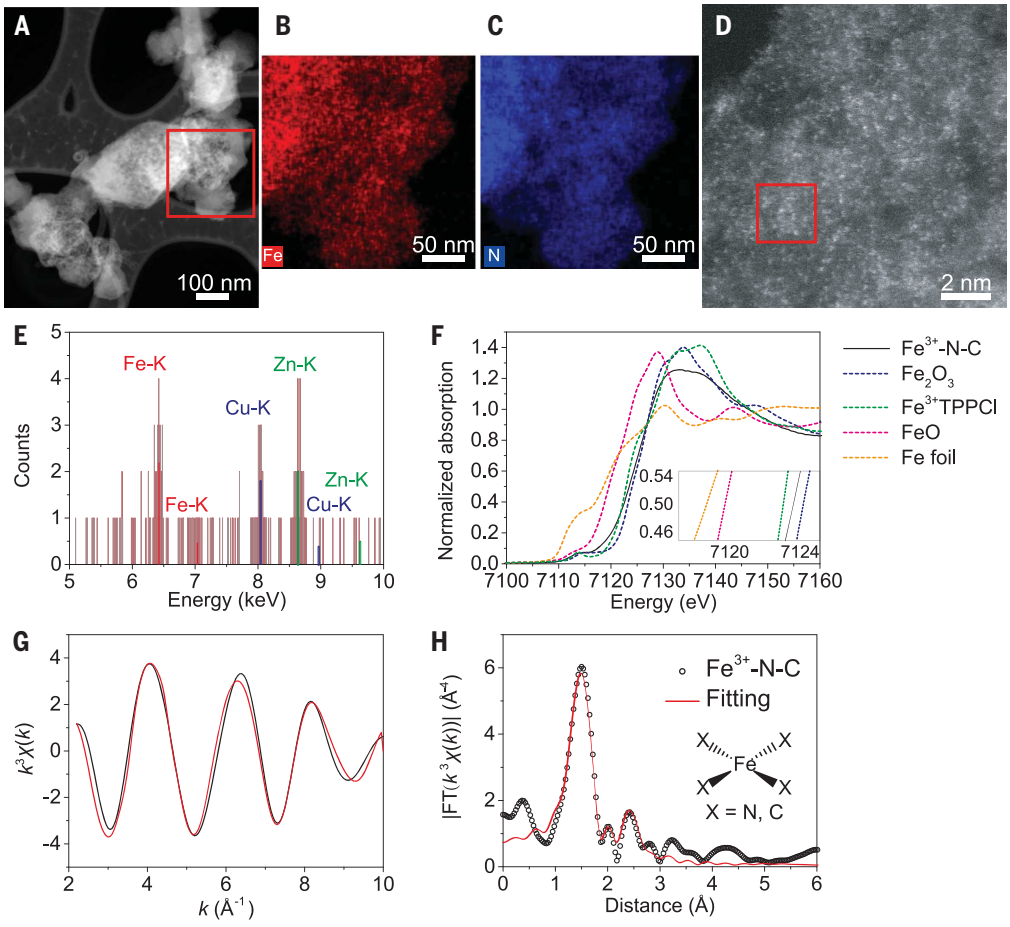


Fig. 1. Characterizations of Fe^{3+}–N–C. (**A**) HAADF-STEM image and the corresponding EDS mappings of (**B**) Fe and (**C**) N of the region enclosed by the red square. (**D**) Aberration-corrected HAADF-STEM image and (**E**) EDS spectrum of the red square region. (**F**) Fe K-edge XANES spectra of Fe^{3+}–N–C (black), Fe_2O_3 (blue dashed), Fe^{3+}TPPCl (green dashed), FeO (pink dashed), and Fe foil (orange dashed). (Inset) The enlargement of the main edges. (**G**) *k*-space and (**H**) R-space Fe K-edge EXAFS spectra. Shown are data (black) and fitting curves (red).

and –0.5 V versus RHE is considerably higher than that attained by other non–precious metal catalysts and even Ag catalysts (*7*), reaching comparable levels with those of oxide-derived Au catalysts (*3*). The mole-normalized current of Fe^{3+}–N–C is significantly higher than that of Au catalysts in this potential range (fig. S8B). Assuming all Fe atoms to be catalytically active, the apparent turnover frequencies (TOFs) of Fe^{3+}–N–C (Fig. 2D) are comparable with those of Au catalysts (*3*, *4*, *26*) and greatly exceed those of other non–precious metal catalysts (*15*, *16*, *19*). Unlike for catalysts based on copper (Cu), Ag, and Au (*27*), ultrapure electrolyte solutions were not necessary for Fe^{3+}–N–C. When $KHCO_3$ with a purity of 99.5% was used to prepare electrolyte, or even tap water was used in place of deionized water (18.2 megohm cm), the Faradaic efficiency and j_{CO} show no obvious change, and the performance was stable for at least 12 hours (Fig. 2E). After 12 hours of electrolysis, the weight fraction of Fe in Fe^{3+}–N–C (measured with ICP-OES) was 2.6%, indicating no appreciable leaching of Fe ions from the catalyst. No aggregation of Fe or Zn species was detected in TEM images, and a high density of discrete Fe and Zn atoms was still observed (fig. S9).

The performance of Fe^{3+}–N–C was stable between –0.2 and –0.5 V versus RHE, although at potentials more negative than –0.5 V versus RHE, the activity became unstable (fig. S7B). As shown in fig. S7C, the current density at –0.41 V was stable during a 28-hour chronoamperometry test. Whereas at –0.51 V versus RHE, the initial j_{CO} was much higher, it decreased rapidly to a value similar to that obtained at –0.41 V versus RHE. This result indicates some changes of Fe^{3+}–N–C around –0.5 V versus RHE. To explore the nature of this change, we conducted operando XAS measurements in the CO_2-saturated 0.5 M $KHCO_3$ catholyte. Fe K-edge spectra were obtained on dry samples and on samples that were loaded on glassy carbon electrodes and immersed in the electrolyte at open circuit potential (OCP) as well as at –0.1 to –0.6 V versus RHE (Fig. 3A). For Fe^{3+}–N–C, the Fe K-edge showed no obvious shift between the dry powder and the in situ sample at –0.4 V versus RHE. The edge energy was close to that of Fe^{3+}TPPCl, indicating that the Fe ions in Fe^{3+}–N–C remained in the +3 oxidation state during CO_2 electroreduction at potentials as negative as –0.4 V versus RHE. When the applied potential was shifted further negative, to –0.5 V versus RHE and beyond, the Fe K-edge shifted to lower energies, which were comparable with that of FeO, suggesting the reduction of Fe^{3+} to Fe^{2+}. This reduction process occurred at the same potential as the above-mentioned deactivation of Fe^{3+}–N–C, implying that Fe^{3+} sites are more active for generating CO. Moreover, the fitting of EXAFS spectra (fig. S10C) indicates that the reduction of Fe^{3+} sites is accompanied by a change of local structure around the Fe ions. Before the reduction of Fe^{3+} sites, the first shell coordination number of Fe (Fe–N and Fe–C) was about 4, whereas as the Fe^{3+} sites were reduced to Fe^{2+} sites, the first shell coordination number of Fe decreased to about 3.

To investigate the origins of the improved activity of Fe^{3+}–N–C as compared with previously reported single-atom Fe catalysts, we measured in situ Fe K-edge XANES spectra of Fe0.5d (fig. S11, E and F) (*15*). Under potentials between –0.2 and –0.5 V versus RHE, the energy of the Fe K-edge was close to that of FeO, indicating a +2 rather than +3 oxidation state for the Fe sites during CO_2 reduction. This difference in oxidation state might be due to different ligand environments, particularly with regard to the N atoms. For Fe0.5d (*15*) and Fe–N–C (*16*), Fe ions coordinated with four pyridinic N were proposed as the active sites. For Fe^{3+}–N–C, XANES and XPS spectra of N indicate that the Fe ions were coordinated to pyrrolic N. In the N K-edge XANES spectrum (fig. S12A), π* and σ* features of Fe^{3+}–N–C were similar to those of a metal-porphyrin derivative (*28*). In the N 1s XPS spectrum (fig.

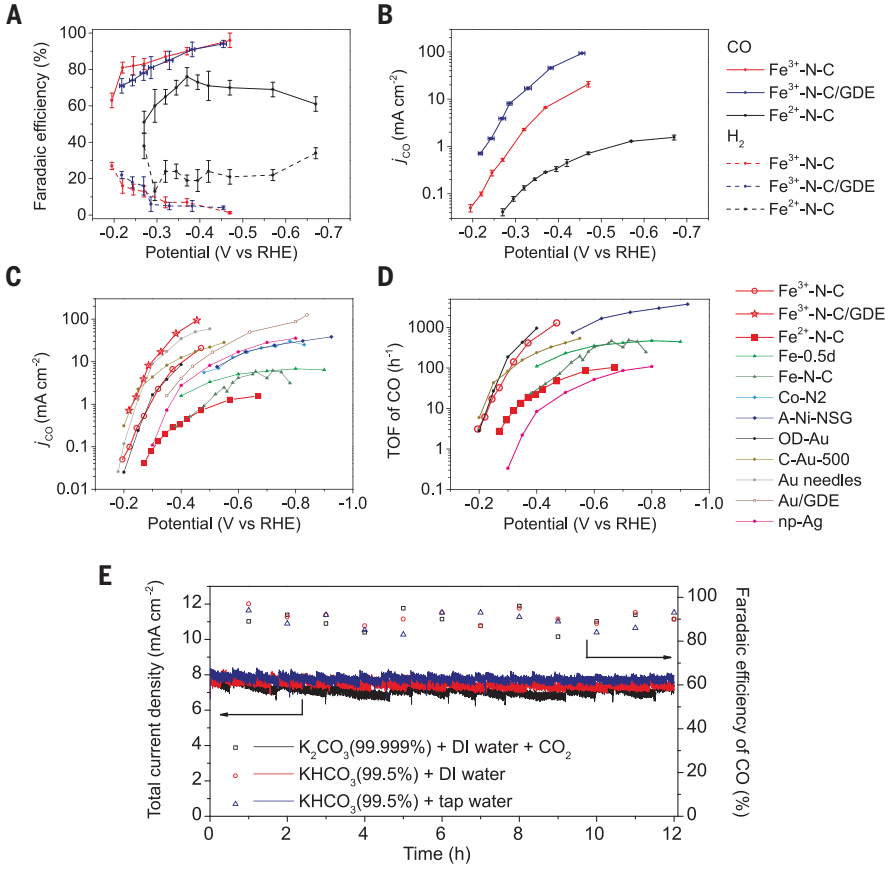


Fig. 2. CO_2 electroreduction performance. (**A**) Faradaic efficiency of CO (solid lines) and H_2 (dashed lines) production and (**B**) j_{CO} of Fe^{3+}–N–C in an H-cell (red) and on a GDE (blue), and of Fe^{2+}–N–C in an H-cell (black). Data from the H-cell were obtained by means of chronoamperometry, whereas data from the GDE were obtained by means of chronopotentiometry. Each error bar was the standard deviation determined based on tests of three individual electrodes. Loading was 0.6 mg cm^{-2} for Fe^{3+}–N–C and Fe^{2+}–N–C; 2.5 mg cm^{-2} for Fe^{3+}–N–C/DGE. (**C** and **D**) Comparison of (C) j_{CO} and (D) apparent TOFs of CO production of Fe^{3+}–N–C in an H-cell (red circles) and on a GDE (red stars) and of Fe^{2+}–N–C in an H-cell (red squares), to that of other reported catalysts: other Fe-N-C catalysts [Fe-0.5d (*15*) and Fe–N–C (*16*)], a Co–N–C catalyst with two coordinating nitrogen atoms (Co–N2) (*17*), atomically dispersed Ni on nitrogen-sulfur codoped graphene (A–Ni–NSG) (*19*), oxide-derived Au electrode (OD-Au) (*3*), carbon black supported Au nanowires with a length of 500 nm (C–Au-500) (*4*), needle-shape Au nanostructures (Au needles) (*5*), nanoporous Ag electrode (np-Ag) (*7*), and Au-polymer-multiwall carbon nanotubes composite loaded on GDE (Au/GDE) (*34*) in bicarbonate electrolytes. (**E**) Chronoamperometry curve and Faradaic efficiency of CO production (dots) by Fe^{3+}–N–C in H-cell at −0.37 V versus RHE. The electrolytes were prepared from potassium carbonate (K_2CO_3) (99.999%) and deionized water (18.2 megohms cm) (black), $KHCO_3$ (99.5%) and deionized water (red), and $KHCO_3$ (99.5%) and tap water (blue), respectively.

S12B), the major peak at 398.6 eV was attributed to pyrrolic N coordinated to Fe, which is in agreement with the spectrum of Fe^{3+}TPPCl. This assignment is consistent with the atomic fractions of N and metals and the coordination number of metal-N (table S4).

To further test the above hypothesis, we directly compared the Fe^{3+}–N–C catalyst with an analogous Fe–N–C catalyst in which the Fe ions were coordinated by pyridinic N atoms. Considering that pyrolysis of Fe precursors containing either pyridinic ligands or pyrrolic ligands seemed to conserve the pyridinic or pyrrolic nature of the N atoms, we prepared the reference sample (Fe^{2+}–N–C) by pyrolysis of a composite containing a Fe-phenanthroline complex at 700°C (*22*). Aberration-corrected HAADF-STEM (fig. S13D) and Fe K-edge EXAFS (fig. S14E and table S1) confirmed the single-atom nature of the Fe sites in Fe^{2+}–N–C. In the N 1s XPS spectrum (fig. S12C), the major feature at 399.7 eV was assigned to pyridinic N coordinated to Fe, which is in agreement with the assignments of the spectra of Fe-phenanthroline complexes and previously reported metal-N-C catalysts with pyridinic N ligands (*16*). This assignment is also consistent with the percentage of coordinated N measured with other methods (table S4). Thus, Fe^{2+}–N–C contained Fe ions coordinated by pyridinic N atoms. A XANES spectrum (fig. S14D) showed that initially, the energy of Fe K-edge of Fe^{2+}–N–C was considerably higher than that of FeO. The Fe 2p XPS (fig. S14C) spectrum showed that the binding energy of Fe ion was similar to that of Fe_2O_3. These data suggested an important number of Fe^{3+} sites in the as-prepared sample of Fe^{2+}–N–C. The in situ XANES (Fig. 3B) showed that Fe^{3+} in the as-prepared Fe^{2+}–N–C started to be reduced to Fe^{2+} at −0.1 to −0.2 V versus RHE. During CO_2 electroreduction (at potentials more negative than −0.2 V versus RHE), the energy of the Fe K-edge was slightly lower than that of FeO. Thus, the Fe sites in Fe^{2+}–N–C under reaction conditions had an oxidation state of +2 or lower. The TOF of CO production of Fe^{2+}–N–C is more than an order of magnitude lower than that of Fe^{3+}–N–C under the same potential (Fig. 2D). The current density of Fe^{2+}–N–C decreased markedly during 2-hour choronoamperometry tests (fig. S6G), indicating its lower stability as compared with the Fe^{3+}–N–C catalyst. These data suggest that pyrrolic type ligands are important to keep Fe sites in the +3 oxidation state during CO_2 electroreduction and consequently maintain the high activity and stability of Fe^{3+} sites. This hypothesis is further supported by the different reactivity of Fe^{3+}–N–C and Fe^{2+}–N–C toward $NaBH_4$. The Fe^{3+} ions coordinated by pyridinic N ligands in Fe^{2+}–N–C could be reduced by $NaBH_4$, whereas those coordinated by pyrrolic N ligands in Fe^{3+}–N–C could not (fig. S15).

For Fe^{2+}–N–C, the j_{CO} at a fixed potential versus the standard hydrogen electrode (SHE) is largely independent of the concentration of the proton donor (fig. S16, A and B), indicating that the 1-electron reduction (adsorption) of CO_2 is decoupled from a proton transfer (*29*, *30*). At modest overpotentials, the j_{CO} of Fe^{2+}–N–C has a Tafel slope of 117 mV/decade (fig. S16E), suggesting that CO_2 adsorption is slow and rate-limiting (supplementary materials, kinetic and mechanistic analysis). On the other hand, the j_{CO} of Fe^{3+}–N–C is approximately first-order in the concentration of HCO_3^- (fig. S16D) and has a Tafel slope of 64-71 mV/decade at low overpotentials (fig. S16E). These kinetic data suggest that for Fe^{3+}–N–C, the 1 electron reduction of CO_2 is also decoupled from a proton transfer. Moreover, CO_2 adsorption is fast, and the rate-limiting step is the protonation of the adsorbed CO_2^- to form an adsorbed COOH intermediate (supplementary materials, kinetic and mechanistic analysis). These results indicate CO_2 adsorption as a descriptor of catalytic activity at

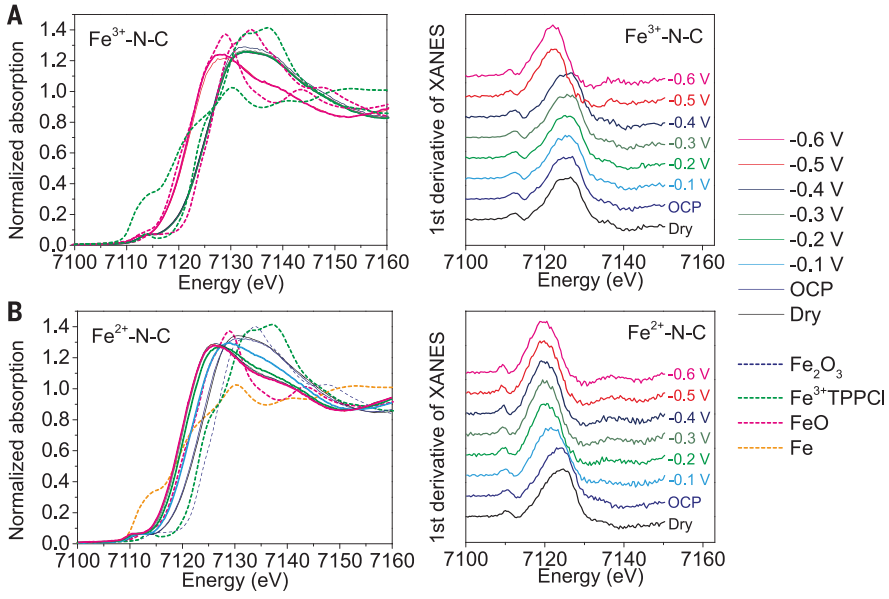


Fig. 3. Operando XAS characterization. (**A** and **B**) Fe K-edge XANES spectra (left) and the first derivative of the spectra (right) of (A) Fe^{3+}–N–C and (B) Fe^{2+}–N–C as dry powder (black) and loaded on glassy carbon electrodes at open circuit potential (OCP) (blue), −0.1 V (light blue), −0.2 V (green), −0.3 V (dark green), −0.4 V (dark blue), −0.5 V (red), and −0.6 V (pink) versus RHE, with the spectra of Fe_2O_3 (blue dashed), Fe^{3+}TPPCl (green dashed), FeO (pink dashed), and Fe foil (orange dashed) as references.

low to modest overpotentials. They also reveal a faster CO_2 adsorption in Fe^{3+}–N–C than in Fe^{2+}–N–C, which explains why Fe^{3+}–N–C has a lower onset overpotential. The CO_2 electroreduction was conducted in the presence of CO (0.2 atm) (fig. S17). External CO did not influence the activity of Fe^{3+}–N–C but largely decreased the activity of Fe^{2+}–N–C. This result suggests that at high overpotentials, CO desorption becomes rate limiting for Fe^{2+}–N–C. Because CO desorption is a non-Faradaic step, once it becomes rate limiting, the rate will hardly increase with increasing overpotentials. At higher overpotentials, the Tafel slope of Fe^{2+}–N–C becomes enormous (546 mV/decade) (fig. S16E), and j_{CO} cannot exceed 2 mA cm^{-2}. On the other hand, the reaction at Fe^{3+} sites was not limited by CO desorption and could reach a very high current density. Thus, the higher activity of Fe^{3+}–N–C compared with Fe^{2+}–N–C at high overpotentials can be rationalized by a weaker CO binding at an Fe^{3+} center than at an Fe^{2+} center.

The spectroscopic data indicate that Fe^{3+}–N–C comprises pyrrolic N ligands, whereas Fe^{2+}–N–C comprises pyridinic N ligands. The pyrrolic N ligands may stabilize Fe^{3+} relative to Fe^{2+}, whereas the pyridinic N ligands have the opposite effect. Thermodynamically, the respective reduction potentials support this hypothesis: The standard reduction potential of $[Fe(phen)_3]^{3+/2+}$ is 1.06 V versus SHE, whereas that of $Fe^{3+/2+}$ couple in Fe-porphyrin complexes can reach as low as −0.4 V versus SHE (*31*). Preservation of the +3 oxidation state during CO_2 electroreduction is counterintuitive because the formation of highly reduced, low-valent Fe species is necessary for molecular Fe catalysts (*32*). Once conjugated to a conductive carbon matrix, however, the Fe^{3+} site is electronically coupled to the conductive support so that the $Fe^{3+/2+}$ reduction potential moves to the same degree as the Fermi level of the carbon support when applying an external bias (fig. S18, A and B). The $Fe^{3+/2+}$ reduction potential remains more negative than the Fermi level of the carbon support, leading to the stabilization of the Fe^{3+} ions. An analogous "strong-coupling" effect was formulated to explain the lack of redox chemistry on conjugated molecular sites during potential cycling (*33*). The reduction of Fe^{3+} sites in Fe^{3+}–N–C at −0.5 V versus RHE is probably enabled by a change of their coordination environment. In situ EXAFS (fig. S10C) revealed that the Fe ion lost one pyrrolic N ligand at this potential, possibly because of protonation or hydrogenation of the ligand driven by the electric field. The new coordination environment increases the $Fe^{3+/2+}$ reduction potential to be more positive than the Fermi level of the carbon support, resulting in conjugated Fe^{2+} ions (fig. S18C).

REFERENCES AND NOTES

1. S. Chu, Y. Cui, N. Liu, *Nat. Mater.* **16**, 16–22 (2016).
2. D. D. Zhu, J. L. Liu, S. Z. Qiao, *Adv. Mater.* **28**, 3423–3452 (2016).
3. Y. Chen, C. W. Li, M. W. Kanan, *J. Am. Chem. Soc.* **134**, 19969–19972 (2012).
4. W. Zhu *et al.*, *J. Am. Chem. Soc.* **136**, 16132–16135 (2014).
5. M. Liu *et al.*, *Nature* **537**, 382–386 (2016).
6. B. A. Rosen *et al.*, *Science* **334**, 643–644 (2011).
7. Q. Lu *et al.*, *Nat. Commun.* **5**, 3242 (2014).
8. T. Zheng, K. Jiang, H. Wang, *Adv. Mater.* **30**, e1802066 (2018).
9. P. P. Sharma *et al.*, *Angew. Chem. Int. Ed.* **54**, 13701–13705 (2015).
10. H. Won *et al.*, *Angew. Chem. Int. Ed.* **55**, 9297–9300 (2016).
11. M. Asadi *et al.*, *Nat. Commun.* **5**, 4470 (2014).
12. J. Gu, F. Héroguel, J. Luterbacher, X. Hu, *Angew. Chem. Int. Ed.* **57**, 2943–2947 (2018).
13. A. Wang, J. Li, T. Zhang, *Nat. Rev. Chem.* **2**, 65–81 (2018).
14. Y. Chen *et al.*, *Joule* **2**, 1242–1264 (2018).
15. T. N. Huan *et al.*, *ACS Catal.* **7**, 1520–1525 (2017).
16. W. Ju *et al.*, *Nat. Commun.* **8**, 944 (2017).
17. X. Wang *et al.*, *Angew. Chem. Int. Ed.* **57**, 1944–1948 (2018).
18. C. Zhao *et al.*, *J. Am. Chem. Soc.* **139**, 8078–8081 (2017).
19. H. B. Yang *et al.*, *Nat. Energy* **3**, 140–147 (2018).
20. T. Zheng *et al.*, *Joule* **3**, 265–278 (2019).
21. K. S. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10186–10191 (2006).
22. W. Liu *et al.*, *J. Am. Chem. Soc.* **139**, 10790–10798 (2017).
23. M. L. Perry, J. Newman, E. J. Cairns, *J. Electrochem. Soc.* **145**, 5–15 (1998).
24. H.-R. M. Jhong, F. R. Brushett, P. J. A. Kenis, *Adv. Energy Mater.* **3**, 589–599 (2013).
25. F. Song, X. Hu, *Nat. Commun.* **5**, 4477 (2014).
26. Z. P. Jovanov *et al.*, *J. Catal.* **343**, 215–231 (2016).
27. Y. Hori *et al.*, *Electrochim. Acta* **50**, 5354–5369 (2005).
28. G. Polzonetti *et al.*, *Chem. Phys.* **296**, 87–100 (2004).
29. A. S. Varela *et al.*, *ACS Energy Lett.* **3**, 812–817 (2018).
30. A. Wuttig, Y. Yoon, J. Ryu, Y. Surendranath, *J. Am. Chem. Soc.* **139**, 17109–17113 (2017).
31. N. Silva *et al.*, *J. Electroanal. Chem.* **819**, 502–510 (2018).
32. C. Costentin, S. Drouet, M. Robert, J.-M. Savéant, *Science* **338**, 90–94 (2012).
33. M. N. Jackson *et al.*, *J. Am. Chem. Soc.* **140**, 1004–1010 (2018).
34. S. Verma *et al.*, *ACS Energy Lett.* **3**, 193–198 (2018).
35. J.Gu, C.-S. Hsu, L. Bai, H. M. Chen, X. Hu, X-ray absorption data and microscopic images of "Atomically dispersed iron(3+) sites catalyze efficient CO_2 electroreduction to CO". *Zenodo* (2019); doi:10.5281/zenodo.2671703

ACKNOWLEDGMENTS

We thank C. Corminboeuf and K.-h. Lin (EPFL) for discussion of theoretical analysis. We thank J. Luterbacher and F. Héroguel (EPFL) for their help in physical adsorption experiments. **Funding:** This work was supported by the GAZNAT SA and the European Research Council (grant 681292). We also acknowledge support from the Ministry of Science and Technology, Taiwan (contract MOST 107-2628-M-002-015-RSP). **Author contributions:** J.G. performed the majority of the synthesis, characterization, and electrochemical tests. L.B. contributed to the initial synthesis of catalysts and TEM measurement. C.-S.H. performed the in situ x-ray absorption experiments. J.G., C.-S.H., H.M.C., and X.H. analyzed the data. J.G. and X.H. wrote the paper, with input from all other co-authors. H.M.C. and X.H. directed the research. **Competing interests:** European priority patent applications (nos. 18156529.2 and 18193304.5) titled "Fe-N-C Catalyst, method of preparation and uses thereof" were filed by GAZNAT SA with J.G. and X.H. as inventors. **Data and materials availability:** All results are reported in the main text and supplementary materials. EXAFS, XANES, and microscopy data files are deposited in Zenodo (*35*). Unless restricted by patent, the materials are available for the purpose of reproducing or extending the analysis.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1091/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S4
References (*36*–*43*)

22 January 2019; accepted 20 May 2019
10.1126/science.aaw7515

PLANT SCIENCE

Mutation of a bHLH transcription factor allowed almond domestication

R. Sánchez-Pérez[1,2,3]*†, S. Pavan[4,5]*†, R. Mazzeo[2,3,4], C. Moldovan[2,3], R. Aiese Cigliano[6], J. Del Cueto[1,2,3,7], F. Ricciardi[2,3,8], C. Lotti[8], L. Ricciardi[4], F. Dicenta[1], R. L. López-Marqués[9], B. Lindberg Møller[2,3]

Wild almond species accumulate the bitter and toxic cyanogenic diglucoside amygdalin. Almond domestication was enabled by the selection of genotypes harboring sweet kernels. We report the completion of the almond reference genome. Map-based cloning using an F_1 population segregating for kernel taste led to the identification of a 46-kilobase gene cluster encoding five basic helix-loop-helix transcription factors, bHLH1 to bHLH5. Functional characterization demonstrated that bHLH2 controls transcription of the P450 monooxygenase–encoding genes *PdCYP79D16* and *PdCYP71AN24*, which are involved in the amygdalin biosynthetic pathway. A nonsynonymous point mutation (Leu to Phe) in the dimerization domain of bHLH2 prevents transcription of the two cytochrome P450 genes, resulting in the sweet kernel trait.

Almond [*Prunus dulcis* Miller (D. A. Webb), syn. *Prunus amygdalus* L.] is the main tree nut species worldwide, with a cultivated area of about 1.9 million ha and an annual in-shell production exceeding 2.2 million metric tons (FAOSTAT 2017, www.fao.org/faostat/en/#data/QC). The spatiotemporal framework of almond domestication is still controversial, although archaeobotany and genetic studies suggest it occurred in the Fertile Crescent—the "Cradle of Civilization"—during the first half of the Holocene (*1*–*4*). Cultivation in the Mediterranean Basin is documented by the presence of almonds among the artifacts found in historical places such as the tomb of Tutankhamen (*5*) and the Franchthi Cave in Greece (*6*). Most recently, almond was introduced in suitable habitats in America (primarily California) and the Southern Hemisphere. In addition to almond, the botanic family of Rosaceae encompasses an extraordinary variety of species of great economic importance in temperate regions, including apple (*Malus domestica*), apricot (*Prunus armeniaca*), Japanese apricot (*Prunus mume*), peach (*Prunus persica*), strawberry (*Fragaria vesca*), and sweet cherry (*Prunus avium*) (*7*, *8*). Among these, almond is unique in that the kernel (seed), rather than the flesh (mesocarp), represents the edible commercial product.

Kernels of wild almond species are bitter and highly toxic to humans and predators because the cyanogenic diglucoside amygdalin accumulates in the cotyledons (*9*–*12*). The key event enabling almond domestication was the selection of sweet kernel genotypes. Genetic studies showed that edible sweet almond kernels originate from a dominant mutation (*13*, *14*) within the almond linkage group (LG) 5, at a locus referred to as *Sweet kernel* (*Sk*) (*15*, *16*). The nature of the *Sk* gene has remained elusive.

Amygdalin is produced from the cyanogenic monoglucoside prunasin (*9*), which is biosynthesized in the seed coat (tegument). Recently, it was shown that the first two genes in the prunasin biosynthetic pathway, *PdCYP79D16* and *PdCYP71AN24*, are poorly expressed in the tegument of sweet genotypes relative to bitter genotypes (*12*). The almond genome harbors an additional CYP79-encoding gene, named *PdCYP79A68* (*12*), a putative ortholog of the *P. mume* gene *PmCYP79A68* (*17*). *PmCYP79A68* was shown to have weak activity toward tryptophan, and it did not catalyze conversion of phenylalanine into the corresponding oxime, as required for a function in prunasin and amygdalin synthesis (*17*).

Despite the economic importance of almond, genomic resources in almond are limited relative to those available for other Rosaceae species (*18*–*22*). Here, we report the assembly of the draft genome of almond. The sequence information was used to unravel the genetic differences between bitter and sweet kernel genotypes. The results provide insights into almond domestication history, as well as a sequence inventory for studies of other almond agronomic traits such as flowering time, drought tolerance, and resistance to disease.

The genome ($2n = 2x = 16$, haploid genome size = 246 Mb) of the sweet homozygous almond cultivar Lauranne was sequenced with a combination of Illumina (paired-end and 5-kb mate pairs) and PacBio technologies. The long PacBio reads were used to perform the genome assembly, whereas the Illumina reads were used to perform gap filling and scaffolding. The scaffolds obtained were organized in pseudomolecules using the *P. persica* reference genome, as the genomes across the *Prunus* genus are essentially collinear (*23*–*25*). The final assembly constitutes 4078 scaffolds, of which 2572 are organized in eight pseudomolecules, with a final N50 of 21.8 Mb (where N50 is the minimum contig length needed to cover 50% of the genome) and a L90 of 306 (where L90 is the smallest number of contigs whose length sum makes up 90% of genome size) (table S1). A high level of collinearity between the almond genomic assembly and the recently reported single-nucleotide polymorphism (SNP)/short sequence repeat (SSR) genetic linkage maps (*25*–*27*) was demonstrated (fig. S1).

By combining ab initio, RNA sequencing (RNA-seq), and transcript sequences from *P. persica* and the Viridiplantae Swissprot database, the

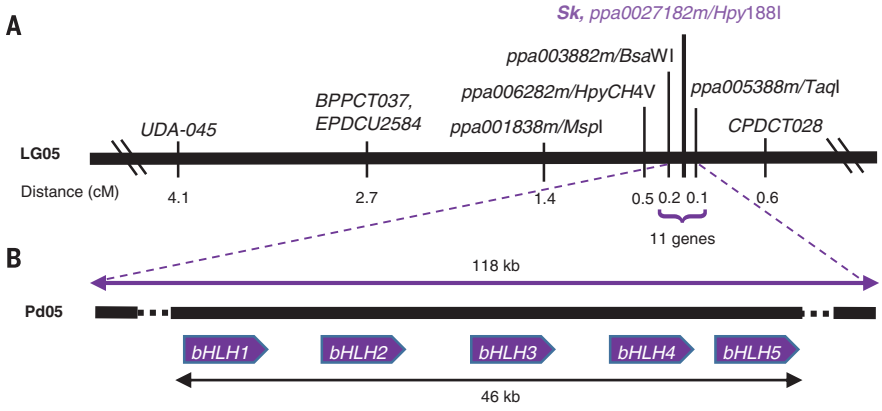


Fig. 1. Map-based cloning of the *Sk* gene. (**A** and **B**) Comparison of linkage group 5 (LG05) (A) and pseudomolecule 5 (Pd05) (B) in the region where the *Sk* locus is mapped. Approximate positions of the five *bHLH* genes present within that region are shown.

[1]Department of Plant Breeding, CEBAS-CSIC, Campus Universitario de Espinardo, 30100 Espinardo, Spain. [2]Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark. [3]VILLUM Research Center for Plant Plasticity, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark. [4]Department of Soil, Plant and Food Science, University of Bari "Aldo Moro," 70126 Bari, Italy. [5]Institute of Biomedical Technologies, National Research Council (CNR), 70126 Bari, Italy. [6]Sequentia Biotech SL, Campus de la UAB, 08193 Bellaterra, Barcelona, Spain. [7]Arboriculture Research Group, Agroscope, Conthey, Switzerland. [8]Department of the Sciences of Agriculture, Food and Environment, University of Foggia, 71100 Foggia, Italy. [9]Transport Biology Section, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark.
***Corresponding author. Email: rsanchez@cebas.csic.es (R.S.-P.); stefano.pavan@uniba.it (S.P.)** †These authors contributed equally to this work.

P. dulcis genome was predicted to contain 27,817 genes (table S2), of which 16,747 had an annotation edit distance (AED) ≤ 0.3. A total of 1080 conserved plant genes were identified among the annotated proteins with the BUSCO tool (*28*) (table S3). Of these, 95% were complete and present in a single copy, confirming the quality of the assembly and annotations. Functional annotation of the genes was then performed by searching for orthologous sequences in *P. persica*, *Arabidopsis thaliana*, and *P. mume*. In this way, a Gene Ontology term was assigned to 12,777 genes, whereas a description was given to 24,035 genes. About 34.6% of the genome was found to be repetitive, including 11.1% of long-term repeat elements (table S4).

Using a set of *Sk*-linked cleaved amplified polymorphic sequence (CAPS) and SSR markers within the almond LG5 (*29*), together with two *Sk*-linked CAPS markers developed in this study (table S5 and fig. S2), a large mapping population of 475 F_1 individuals segregating at the *Sk* locus was genotyped. The resulting map displayed the two markers ppa003882m/*Bsa*WI and ppa005388m/*Taq*I flanking the *Sk* locus, separated by 0.2 cM and 0.1 cM, respectively (Fig. 1A). Alignment of the two marker sequences against the assembled almond genome sequence allowed the definition of a physical interval of 118 kb, harboring 11 genes (Fig. 1B). A 46-kb cluster of five genes encoding putative basic helix-loop-helix (bHLH) transcription factors (*bHLH1* to *bHLH5*) was positioned within this interval. We investigated these *bHLH* genes for their ability to control amygdalin accumulation because (i) sweet and bitter almond genotypes are associated with markedly different transcriptional levels of the amygdalin biosynthetic genes *PdCYP79D16* and *PdCYP71AN24* (*12*), as confirmed by our RNA-seq experiment (Fig. 2); (ii) bHLH transcription factors are known to regulate the biosynthesis of several bioactive natural products (*30*); and (iii) we identified several bHLH-binding DNA matrices in the *PdCYP79D16* and *PdCYP71AN24* promoter regions (data S1).

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) time-course experiments on tegument tissues of the sweet cultivar Lauranne (*Sk/Sk*) and the bitter genotype S3067 (*sk/sk*) did not reveal differential expression for *bHLH1*, *bHLH2*, and *bHLH4* (fig. S3 and data S2); expression of *bHLH3* and *bHLH5* was not detected. A scan of a 20-kb genomic region associated with the *Sk* locus revealed higher heterozygosity in the bitter genome than in the sweet genome (fig. S4).

The expressed *bHLH* genes (*bHLH1*, *2*, and *4*) were sequenced in Lauranne and S3067 (fig. S5). In *bHLH1*, this highlighted an indel polymorphism of 161 bp leading to a truncated protein (262 versus 401 amino acids) in the sweet genotype. In *bHLH2*, a C/T SNP and a 3-bp indel were observed, resulting in a nonsynonymous Leu^{346} → Phe (L346F) substitution and an Asn^{412} insertion in the sweet genotype. In *bHLH4*, no polymorphism was found.

Functional characterization of *bHLH1* and *bHLH2* was carried out using a factorial transient β-glucuronidase (GUS) expression assay in *Nicotiana benthamiana*. The alleles were used in combination with the two *PdCYP79D16* promoter regions cloned from Lauranne and S3067 (henceforth referred to as *p79sweet* and *p79bitter*) and the three promoter regions cloned from *PdCYP71AN24* (henceforth referred to as *p71sweet1*, *p71sweet2*, and *p71bitter*) (Fig. 3). No significant differences were observed in GUS activity when expressing the two *bHLH1* alleles (Fig. 3A). In contrast, a significant induction of GUS activity was observed for the *bHLH2* bitter allele in combination with all versions of *p79* and *p71* (Fig. 3B). This indicated that *bHLH2* could be the *Sk* gene. DNA sequencing of *bHLH2* in 56 almond genotypes demonstrated cosegregation between the *Sk*/*sk* alleles and the C/T polymorphism in all cases except one (table S6), represented by the cultivar Atocha. However, Atocha displayed a different SNP positioned nearby, resulting in a Leu^{330} → Arg (L330R) nonsynonymous change (fig. S6).

The L346F and L330R nonsynonymous mutations observed in sweet genotypes were both predicted to occur in the dimerization interphase typical of bHLH transcription factors (fig. S5). To study the possible functional consequences of the L346F mutation, we carried out direct mutagenesis of the coding sequence of the two *bHLH2* alleles, substituting the Leu^{346} residue with Phe and vice versa. In addition, the last 151 amino acids of the two *bHLH2* alleles (henceforth called the "tail") were also exchanged to investigate whether the extra Asn residue occurring at position 412 in the sweet genotype of bHLH2 influenced protein activity. Induction of GUS activity was observed only in association with the Leu^{346} residue. Indel polymorphisms at the protein tail did not affect GUS activity (Fig. 3B). bHLH2 thus controls transcription of the P450 monooxygenase–encoding genes *PdCYP79D16* and *PdCYP71AN24* in the amygdalin biosynthetic pathway.

To understand how a change from Leu to Phe at position 346 affects activity of the bHLH2 transcription factor, we generated a structural model of the protein sequence surrounding this amino acid (Fig. 4, A to C). When dimerized, the corresponding Leu residues of each monomer are placed facing each other, allowing hydrophobic interactions between the amino acids. Change of this Leu into a Phe might have two consequences. The first is that steric hindrances would prevent dimer formation. This would be augmented by the presence of two additional phenylalanines in the near vicinity in the model. Alternatively, the dimer would be formed, but the ability of the aromatic rings to form π-stacks (*31*) would result in a dimer with an aberrant, nonfunctional conformation. To discriminate between these two possibilities, we expressed the *bHLH2* bitter and sweet transcription factor variants in *Escherichia coli* as a fusion to glutathione *S*-transferase (GST) or His, and performed pull-down assays (Fig. 4D) as well as polyacrylamide gel electrophoresis under native conditions (fig. S7). Both protein versions migrated as a band with a molecular mass corresponding to the dimer, indicating that the L346F polymorphism does not affect the ability of bHLH2 proteins to form dimers. To test the DNA-binding ability of the bHLH2 dimers, we cloned a triple repeat of the six-nucleotide motif (CATGTG) present in all *p79* and *p71* promoter variants (data S1) upstream of a β-galactosidase gene in a yeast one-hybrid–compatible plasmid. Only the bHLH2 variants carrying Leu^{346} were able to induce β-galactosidase activity (Fig. 4E), indicating that the L346F mutation generates a nonfunctional bHLH2 dimer.

The history of almond domestication is marked by the selection of sweet kernel genotypes, although other traits such as thinner endocarp and increased seed size might also have been objects of early selection (*32*). The knowledge of the

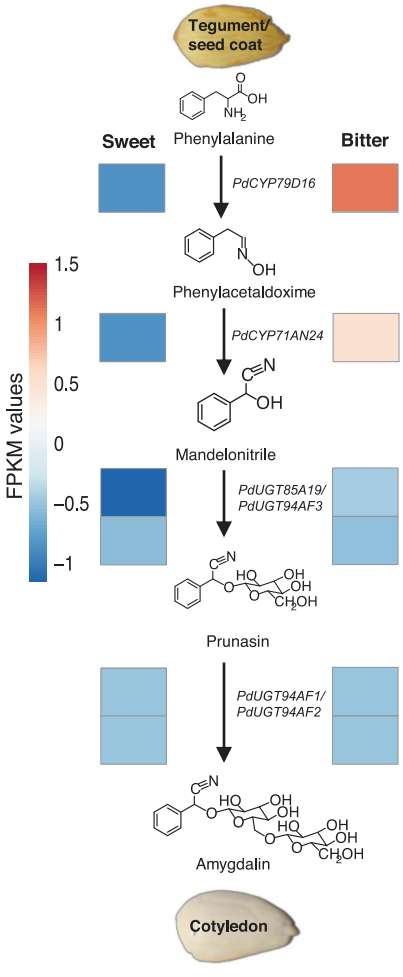


Fig. 2. The amygdalin biosynthetic pathway and transcript levels of the encoding genes during fruit development. Expression patterns (standardized FPKM values) of the genes of the amygdalin pathway are shown during tegument growth in the sweet genotype Lauranne and the bitter genotype S3067.

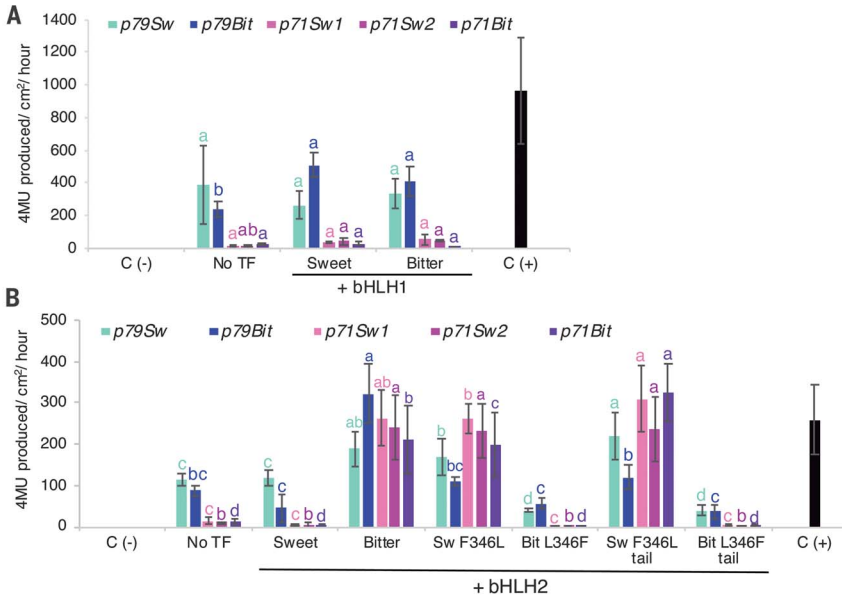


Fig. 3. Trans-activation analysis of putative downstream gene promoters by the bHLH transcription factors localized in the *Sk* locus and expressed in the tegument during kernel development. (**A** and **B**) The bar plots show the capacity of bHLH1 (A) and bHLH2 (B) variants of sweet (Sw) and bitter (Bit) genotypes to trans-activate the sweet and bitter promoters of *PdCYP79D16* (*p79*; light/dark blue colors) and *PdCYP71AN24* (*p71*; pink/lilac colors), respectively. The effects of mutations in bHLH2 bitter and sweet at position 346 as well of their C-terminal tails are also shown. Control experiments: p19 silencing suppressor [C(−)] and no transcription factor (no TF) as negative controls, and pCaMV35S:APL3:GUS as positive control [C(+)]. For each promoter, shown in different colors, data were compared using analyses of variance (ANOVA) least significant differences test with $P < 0.05$. Therefore, significant differences in the multiple combinations of promoters with or without TF are shown by different letters but with the same color.

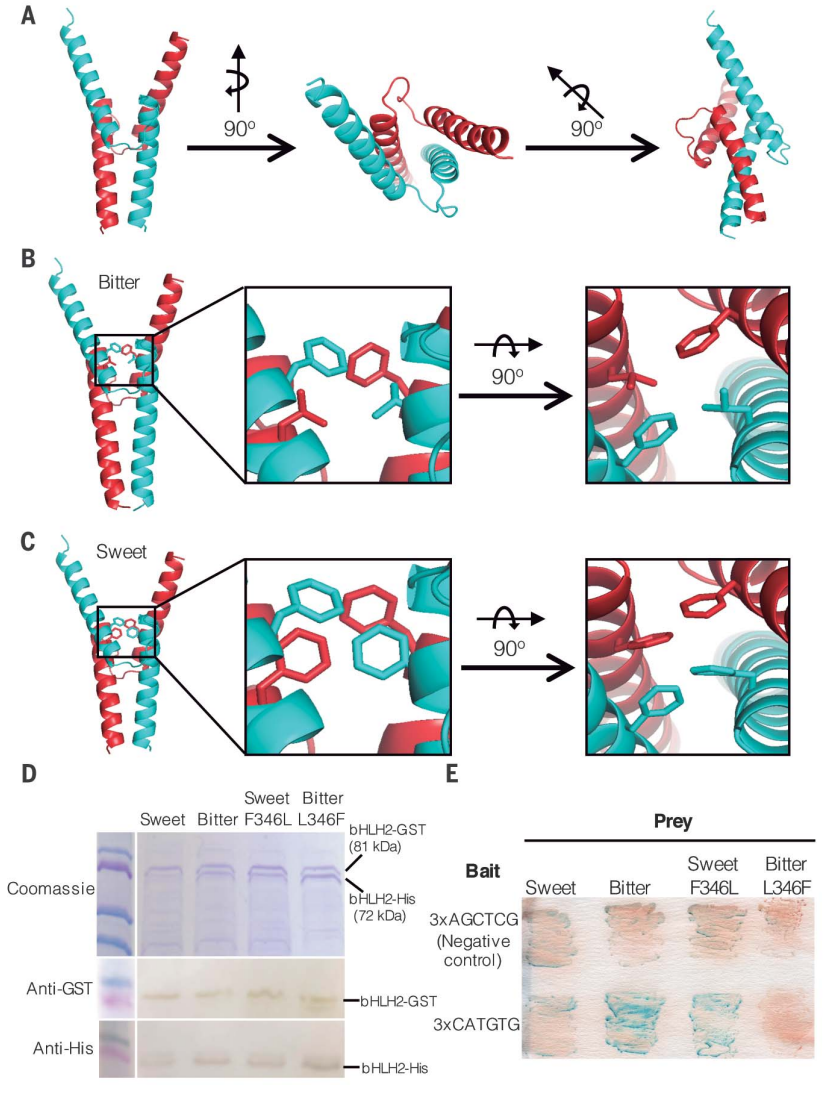


Fig. 4. A Leu → Phe substitution is responsible for formation of a nonfunctional bHLH2 dimer. (**A**) Visualization of the relative position of two bHLH2 monomers (red and cyan) after dimer formation. (**B** and **C**) A single amino acid substitution (Leu to Phe) is present in the sweet variant, possibly interfering with formation of a functional dimer. (**D**) *E. coli*–expressed bHLH2 proteins were subjected to pull-down assays with a His-tag resin, transferred to polyvinylidene difluoride membranes, and detected with an antibody recognizing the His- and GST-tag engineered at the N-terminal end of the proteins. Precision Plus Protein Standard (Dual Color) was used as a protein standard, in which the pink-colored marker protein corresponds to 75 kDa. (**E**) The sweet and bitter variants of bHLH2 were tested in a yeast one-hybrid screening for their activation of a synthetic promoter including a triple repetition of a CATGTG motif, a typical binding sequence for the basic β-helix-loop-helix transcription factor family. As a negative control, a triple repetition of a random hexanucleotide sequence (AGCTCG) was used. The effects of the L346F mutation in bHLH2 bitter and sweet are also shown.

toxicity of bitter almond and peach kernels was already realized in ancient Egypt in the second millennium BCE, where traitorous priests in the capital cities of Memphis and Thebes were poisoned to death with pits of peaches—a practice referenced in hieroglyphics as "death by peach" (*33*, *34*). Caius Plinius Secundus (Pliny the Elder) wrote in his 37-volume encyclopedia *Naturalis Historia* that the Romans were proud of knowing how to remove the bitterness and toxicity from bitter almond kernels (*35*). In Basil of Caesaria's *Hexameron* from the 4th century CE, it is stated that Greek agriculturists in Cappadocia had discovered that piercing of the trunk of a highly bitter almond tree near the root, so as to introduce a fat plug of pine into the middle of the pith, resulted in production of delicious sweet almonds (*36*).

Today, almond breeding often introduces desirable traits from genotypes heterozygous or homozygous for the bitter *sk* allele, thus requiring the selection of sweet kernel individuals in segregating populations. With the identification of the polymorphisms causally related to the kernel taste, such a selection could be efficiently performed early at the seedling stage.

Besides contributing to the advancement of almond genomics and breeding, the knowledge obtained in this study offers a route toward neodomestication of other plants (*37*), for controlling accumulation of not only other cyanogenic glucosides, such as linamarin and lotaustralin in cassava tubers (*38*), but also other specialized metabolites, such as anthocyanins in strawberry (*39*) saponins in the sweet/bitter quinoa lines (*40*) and formation of gossypol-containing glands in cotton (*41*). Finally, we envisage that DNA sequencing of the *bHLH* gene in almonds found at ancient archaeological sites may be used to delineate the roots of cultivation (*42*) or how ancient human populations moved across Asia, Europe (through the Mediterranean Basin), and the Americas (*43*).

REFERENCES AND NOTES

1. G. Willcox, S. Fornite, L. Herveux, *Veg. Hist. Archaeobot.* **17**, 313–325 (2008).
2. D. Rivera Núñez, G. Matilla Séiquer, C. Obón de Castro, F. J. Alcaraz Ariza, in *Plants and Humans in the Near East and the Caucasus: Ancient and Traditional Uses of Plants as Food and Medicine, a Diachronic Ethnobotanical Review (Armenia, Azerbaijan, Georgia, Iran, Iraq, Lebanon, Syria, and Turkey)* (Edit.um, 2012), vol. 2, pp. 886–944.
3. M. Delplancke *et al*., *Evol. Appl.* **5**, 317–329 (2012).
4. M. Delplancke *et al*., *Mol. Ecol.* **22**, 1092–1104 (2013).
5. D. Zohary, M. Hopf, E. Weiss, *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford Univ. Press, 2012).
6. J. Hansen, J. M. Renfrew, *Nature* **271**, 349–352 (1978).
7. T. Yamamoto, S. Terakami, *Breed. Sci.* **66**, 148–159 (2016).
8. S. Farinati, A. Rasori, S. Varotto, C. Bonghi, *Front. Plant Sci.* **8**, 1247 (2017).
9. R. Sánchez-Pérez, K. Jørgensen, C. E. Olsen, F. Dicenta, B. L. Møller, *Plant Physiol.* **146**, 1040–1052 (2008).
10. R. Sánchez-Pérez, K. Jørgensen, M. S. Motawia, F. Dicenta, B. L. Møller, *Plant J.* **60**, 894–906 (2009).
11. R. Sánchez-Pérez *et al*., *Plant Physiol.* **158**, 1916–1932 (2012).
12. S. Thodberg *et al*., *Plant Physiol.* **178**, 1096–1111 (2018).
13. M. J. Heppner, *Genetics* **8**, 390–391 (1923).
14. F. Dicenta, J. E. García, *Heredity* **70**, 308–312 (1993).
15. R. Sánchez-Pérez *et al*., *Tree Genet. Genomes* **6**, 237–245 (2010).
16. T. Joobeur *et al*., *Theor. Appl. Genet.* **97**, 1034–1041 (1998).
17. T. Yamaguchi, K. Yamamoto, Y. Asano, *Plant Mol. Biol.* **86**, 215–223 (2014).
18. R. Velasco *et al*., *Nat. Genet.* **42**, 833–839 (2010).
19. V. Shulaev *et al*., *Nat. Genet.* **43**, 109–116 (2011).
20. Q. Zhang *et al*., *Nat. Commun.* **3**, 1318 (2012).
21. International Peach Genome Initiative *et al*., *Nat. Genet.* **45**, 487–494 (2013).
22. K. Shirasawa *et al*., *DNA Res.* **24**, 499–508 (2017).
23. E. Dirlewanger *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9891–9896 (2004).
24. S. Jung *et al*., *BMC Genomics* **10**, 76 (2009).
25. S. N. Goonetilleke *et al*., *G3* **8**, 161–172 (2018).
26. R. Sánchez-Pérez, F. Dicenta, P. Martínez-Gómez, *Tree Genet. Genomes* **8**, 379–389 (2012).
27. R. Sánchez-Pérez, W. Howad, F. Dicenta, P. Arús, P. Martínez-Gómez, *Plant Breed.* **126**, 310–318 (2007).
28. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, *Bioinformatics* **31**, 3210–3212 (2015).
29. F. Ricciardi *et al*., *Genes* **9**, 385 (2018).
30. W. R. Chezem, N. K. Clay, *Phytochemistry* **131**, 26–43 (2016).
31. G. B. McGaughey, M. Gagné, A. K. Rappé, *J. Biol. Chem.* **273**, 15458–15463 (1998).
32. D. Velasco, J. Hough, M. Aradhya, J. Ross-Ibarra, *G3* **6**, 3985–3993 (2016).
33. R. H. Davis, in *Toxic Substances in Crop Plants* (Royal Society of Chemistry, 1991), pp. 202–225.
34. D. Blum, *The Poisoner's Handbook* (Penguin, 2010).
35. Pliny the Elder, in *Naturalis Historia. Liber XV* (77 AD), p. 26.
36. Basil of Caesaria, in *Hexameron*, Homily 5, paragraph 7.
37. P. Gepts, in *Encyclopedia of Agriculture and Food Systems*, N. K. Van Alfen, Ed. (Elsevier, 2014), vol. 2, pp. 474–486.
38. K. Jørgensen *et al*., *Plant Physiol.* **155**, 282–292 (2011).
39. F. Zhao *et al*., *Sci. Rep.* **8**, 2721 (2018).
40. D. E. Jarvis *et al*., *Nature* **542**, 307–312 (2017).
41. D. Ma *et al*., *Nat. Commun.* **7**, 10456 (2016).
42. G. Willcox, *Science* **341**, 39–40 (2013).
43. C. L. Scheib *et al*., *Science* **360**, 1024–1027 (2018).

ACKNOWLEDGMENTS

We thank T. Cremades Rosado and M. Dahl for technical help; the Biology of Stress and Plant Pathology Department and Biotechnology Fruit Trees Group at CEBAS-CSIC for use of their facilities; J. A. Egea for statistical analysis advice; and T. Sicheritz-Ponten, B. Petersen, and A. Dhingra for their initial contributions and efforts regarding the genome sequencing. L. M. De Luca, Johns Hopkins University, is acknowledged for providing the reference to Basil of Caesaria's *Hexameron*. **Funding:** Supported by VILLUM Foundation grants to the VILLUM Center for Plant Plasticity (VKR023054) (B.L.M. and R.S.-P.) and VILLUM project 13234 (R.L.L.-M.); a University of Copenhagen Excellence Program for Interdisciplinary Research grant to the Center for Synthetic Biology (B.L.M.); European Research Council Advanced Grant ERC-2012-ADG_20120314 (B.L.M.); VILLUM Young Investigator Program grant VKR023124 (R.S.-P.); and the projects "Mejora Genética del Almendro" (MINECO-Spain), Séneca Foundation grants for "Molecular Biology of Cyanogenesis in Almond" (11937-PI-09) (R.S.-P.), and "Breeding stone fruit species assisted by molecular tools" (19879/GERM/15) (F.D., J.D.C., and R.S.-P.). **Author contributions:** R.S.P., J.D.C., and F.D. designed crosses, generated and evaluated plant material; R.S.-P., R.M., C.M., J.D.C., F.R., R.A.C., R.L.L.-M., and S.P. performed the research; R.S.-P., S.P., C.L., L.R., F.D., R.L.L.-M., and B.L.M. designed the research; and R.S.-P., S.P., R.A.C., R.L.L.-M., and B.L.M. wrote the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** The following sequences can be found at GenBank database with the following codes: pCYP71AN24sweet2, MK041085; pCYP71AN24sweet1, MK041086; pCYP71AN24bitter, MK041087; pCYP79D16sweet, MK041088; pCYP79D16bitter, MK041089; bHLH1sweetLauranne, MK041090; bHLH1bitterS3067, MK041091; bHLH2sweetLauranne, MK041092; bHLH2bitterS3067, MK041093; bHLH4, MK041094.The almond genome sequence is available at the DDBJ database: AP019297-AP019304. All data are available in the main text or the supplementary materials or in the databases previously named.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6445/1095/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S9
Data S1 and S2
References (*44*–*89*)

30 November 2018; accepted 23 May 2019
10.1126/science.aav8197

Cold Trap

Incorporating a stainless-steel chamber with mechanically refrigerated walls, Vapor Trap by SP Scientific provides highly efficient vapor trapping without the time-consuming annoyance, expense, and hazard caused by expendable coolants such as liquid nitrogen or dry ice. Designed to run with minimal operator attention, Vapor Trap is easy to use and requires a minimal 26.7 cm × 52.1 cm × 27.9 cm lab space. Available in a choice of 2-L, 4-L, and 8-L condenser configurations, and operating at temperatures down to –90°C, Vapor Trap can be used in vapor streams with vacuum pumps up to 250 L (9 ft^3) per min. A new external type T thermocouple offers labs the convenience of remote temperature display and monitoring. Vapor Trap is available as an entry-level manual unit with an on/off switch. Alternatively, Titan Trap features an optional microprocessor-controlled version that can automate operation, including system alerts, vacuum pump activation, auto defrost routines, and digital temperature and vacuum displays.

SP Scientific
For info: 845-255-5000
www.spscientific.com

Automated Endotoxin Testing

Lonza offers a fully automated and integrated plate-based PyroTec PRO Robotic Solution for endotoxin testing. Combining the speed and reproducibility of a robotic liquid handling platform with the power of Version 6.0 of Lonza's market-leading WinKQCL software, the system simplifies and accelerates endotoxin testing of parenteral pharmaceuticals regardless of sample complexity. Conventional endotoxin detection methods for parenteral pharmaceuticals are labor intensive and error prone, as they involve manual data entry and template setup steps. Consequently, the reliability of results can be compromised and the retest rate increased, while the absence of metadata associated with the manual steps limits traceability and the laboratory's audit capability. PyroTec PRO Automated Robotic Solution reduces manual intervention considerably, minimizing the potential for human error and the need for repeat test runs, while greatly enhancing laboratory efficiency. As a fully automated system, it captures preparation metadata, adding traceability into tracking, trending, and audit controls.

Lonza
For info: 800-638-8174
www.lonza.com/endotoxin-automation

Secondary Antibodies for VHH Discovery

Jackson ImmunoResearch announces a new range of products for scientists developing single-domain VHH (variable domains of camelid heavy chain–only) antibodies. The Anti-Alpaca secondary antibody product line targets the immunoglobulin G (IgG) repertoire of camelid immunoglobulins, including whole-molecule, subclass, and VHH domain specificities. They are raised against alpaca and cross-react with llama. Anti-Alpaca IgG subclasses 2+3–specific antibodies enable scientists to monitor immune response and pinpoint the optimal B-cell harvest timeframe. Anti-Alpaca VHH domain antibodies detect single-domain antibodies across all immunotechniques and can be used in diagnostics and drug development.

Jackson ImmunoResearch
For info: 800-367-5296
www.jacksonimmuno.com/technical/products/vhh-discovery

Human Tissue Samples

OriGene's Tissue Biorepository has over 140,000 samples, comprising human cancer and normal tissue products, ready for your research needs. These ethically sourced, high-quality tissues are pathologist-verified with hematoxylin and eosin (H&E) images available on our website for your review, selection, and quick, easy ordering. OriGene's comprehensive tissue products provide researchers ready-to-use tools for studies in biomarker discovery and validation, cancer drug target identification and validation, companion diagnostic assay development, and personalized medicine for the cancer patient. Tissues are banked under strict collection protocols and undergo rigorous quality control to ensure each source block's exceptional quality.

OriGene
For info: 888-267-4436
www.origene.com

EV Isolation System

ExoQuick ULTRA for Serum and Plasma from System Bioscience allows for extracellular vesicle (EV) preps with high levels of yield and purity. Important for intercellular communication in both normal physiology as well as disease states such as cancer, EV biology is a rapidly growing field. However, obtaining EVs for in vivo and ex vivo studies can be challenging. Ultracentrifugation has been considered the gold standard for exosome isolation, but the method is time-consuming and requires large sample-volume inputs and access to specialized equipment. With ExoQuick ULTRA, you no longer have to make the trade-off between yield, purity, speed, a simple protocol, and price. From as little as 250 µL of serum or plasma and our ExoQuick ULTRA kit, you can isolate high-quality EVs for a wide range of downstream applications, such as Western blotting, mass spectrometry, NGS sequencing, exosome labeling, and in vivo/ex vivo exosome delivery. Each kit comes with enough serum/plasma reagent and convenient, prepacked columns to complete 20 reactions.

System Biosciences
For info: 888-266-5066
www.systembio.com

Low-Adhesion Microplates

Lipidure-coated plates from AMS Biotechnology use MPC (2-methacryloyloxyethyl phosphorylcholine), a biocompatible synthetic polymer with the same structure as the phosphorylcholine polar bases that form cell membranes. Due to the high biocompatibility of this coating, cells will not adhere to the surface of Lipidure-coated plates or dishes. Different well shapes (U, V, and flat-bottom) can be used to control the size, position, and density of organoid culture. Studies have shown that Lipidure-coated plates promote the formation of a single spheroid in concave (U- or V-shaped) wells, while other plate brands lead to formation of satellite spheroids—indicating a higher degree of cell adhesion. Formation of spheroids, tumorspheres, embryoid bodies, and neurospheres using Lipidure plates has been demonstrated for cell types including embryonic stem cells and induced pluripotent stem cells (human and mouse), NIH3T3, preadipocytes, HepG2, and other cancer cell lines as well as primary neuronal cells.

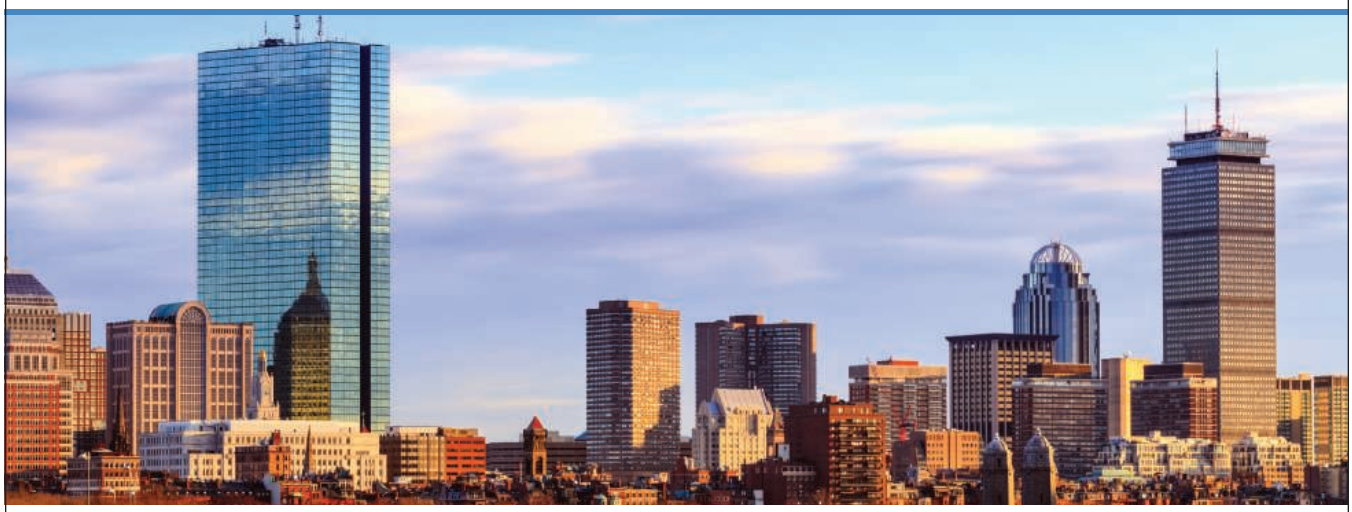
AMS Biotechnology
For info: 617-945-5033
www.amsbio.com/lipidure-coat.aspx

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Science Careers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG



Science Career Fair at Harvard University & FAS – PDA Poster Session

Date: Thursday, June 21, 2019
Location: Harvard Campus, Northwest Building, B100, Cambridge, MA
Time: 11:00 am – 6:30 pm EST

JOB SEEKERS! *Science* Careers has partnered with Harvard University and the Faculty of Arts and Sciences Postdoctoral Association (FASPDA) to produce a unique career fair open to all. Join us for a chance to meet with top scientific organizations and get important advice from career experts. The combination of valuable career development sessions, company presentations, poster sessions, and employer exhibits makes this a free "must-attend" event for scientists.

Learn more at https://scim.ag/3055DRK

Icahn School of Medicine at Mount Sinai

The Department of Microbiology at the Icahn School of Medicine at Mount Sinai seeks highly qualified individuals for **faculty positions in virology**. Appointments can be at the **Assistant, Associate, or Professor level**.

All positions require outstanding academic credentials and scholarly productivity. Candidates will be expected to sustain an active research and publication portfolio and supervise graduate students and postdoctoral scholars.

Interested applicants are asked to submit a brief one page summary of their research interests along with a CV or NIH biosketch. All material can be sent to the Chair of the search committee, Professor Benjamin tenOever at **Benjamin.tenOever@mssm.edu**.

Stanford University School of Medicine
Assistant Professor
Department of Developmental Biology

Stanford University is seeking applicants for a tenure-track Assistant Professor position in the Department of Developmental Biology in the Beckman Center for Molecular and Genetic Medicine in the School of Medicine. We are particularly interested in candidates who will establish a vigorous and innovative research program studying fundamental biological processes in any experimental system. We also expect the successful candidate to participate in teaching advanced courses for graduate and medical students.

We view Developmental Biology in the broadest sense, encompassing Microbes to Humans and employing a wide-variety of molecular and genetic approaches as well as systems level biology, engineering, and computational science. Current faculty in the department of Developmental Biology use a diverse range of genetic, genomic, cell biological, biochemical, and computational approaches to study many different organisms, including microbes, worms, flies, fish, mice, and humans. For more information on our department see: **https://devbio.stanford.edu**.

The predominant criterion for appointment in the University Tenure Line is a major commitment to research and teaching. Individuals appointed as Assistant Professors in the UTL will have completed housestaff training (where applicable) and, additionally, one or two years of postdoctoral research experience. Their accomplishments during graduate and postgraduate training should already have stamped them as creative and promising investigators. If these individuals have not had formal teaching experience, they should have demonstrated during their postdoctoral training a commitment to develop the skills necessary for first-rate teaching. In short, the successful candidate must have demonstrated true distinction (or the promise of achieving true distinction) in research, and the capability of sustaining first-rate performance (or the promise of this) in teaching, and excellence in patient care (if applicable) appropriate to the programmatic need upon which the appointment is based.

Stanford is an Equal Employment Opportunity and Affirmative Action Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, protected veteran status, or any other characteristic protected by law. Stanford welcomes applications from all who would bring additional dimensions to the University's research, teaching and clinical missions.

Interested applicants should send a *curriculum vitae*, a brief statement of research interests in a single pdf file, and, arrange to have three letters of recommendation sent to: **devbiofacultysearch@stanford.edu**. Consideration of application files will begin September 15, 2019, and will be accepted until the position is filled.

EGL Charitable Foundation

The EGL Charitable Foundation invites you to apply to the

Gruss Lipper Post-Doctoral Fellowship Program

Eligibility

- Israeli citizenship
- Candidates must have completed PhD and/or MD/PhD degrees in the Biomedical Sciences at an accredited Israeli University/Medical School or be in their final year of study
- Candidates must have been awarded a postdoctoral position in the U.S. host research institution.

Details regarding the fellowship are available at www.eglcf.org

Application Deadline is September 29, 2019

The Sarah and Daniel Hrdy Fellowship in Conservation Biology
Department of Organismic and Evolutionary Biology
Harvard University

Application deadline: October 1, 2019

The Department of Organismic and Evolutionary Biology at Harvard University invites applications or nominations for the Sarah and Daniel Hrdy Visiting Fellowship in Conservation Biology. The Hrdy Fellowship is open to researchers at any rank; we encourage applications from both established scientists as well as those who have recently received their Ph.D. The fellowship supports visits of either one or two semesters.

The Hrdy Fellowship is awarded to an individual who will engage in scientific study in the Department of Organismic and Evolutionary Biology. Recipients of this fellowship are expected to have a strong and transformative effect on the study of conservation biology at Harvard University. Applicants from any research field within conservation biology are welcome to apply. Research of previous Hrdy Fellows has included conservation paleobiology, marine evolution and conservation, conservation biology of amphibians and reptiles, and the impact of human activities on the environment. Information about previous fellows is available here: http://oeb.harvard.edu/hrdy-fellowship.

The Sarah and Daniel Hrdy Fellowship in Conservation Biology is made possible by the generosity of Sarah and Daniel Hrdy.

The Hrdy Fellow is primarily expected to engage in leading-edge research, where possible in collaboration with members of the Harvard community. Additional responsibilities include a public lecture by the Fellow in any area of conservation biology. Finally, the Fellow is required to teach a one-semester, seminar-style course aimed at upper-level undergraduates. For more information on teaching, contact OEB Chair Elena Kramer, at ekramer@oeb.harvard.edu.

Application Process: Fellowships are awarded through a competitive review process that will begin on October 1, 2019 for appointments to begin no earlier than August 1, 2020. To be considered for a fellowship, please submit materials through the ARiES portal at **http://academicpositions.harvard.edu/postings/9057**.

Harvard University 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.

By Sally G. Hoskins

Teaching ingenuity

After a fulfilling career as a college biology professor, I'm retiring. "What will you miss most?" a colleague asked. My answer was something that, 30 years ago, I would never have expected myself to say: "I will miss the creativity of teaching." When I was a new faculty member, I considered teaching a necessary evil that took me away from the lab bench. I wanted to focus on research, guiding graduate students in what I hoped would be groundbreaking studies on nerve growth. I believed imagination lived not in the classroom, but in the laboratory—to be used for inventing techniques, designing experiments, and interpreting data. But when my life took an unexpected turn, I realized how wrong I had been.

I was 10 years into my career, happily plugging away at my research as a tenured professor, when my teenage niece was orphaned and I became her guardian and single parent. After taking some time to adjust, I decided that I wouldn't be able to manage a full-fledged neuroscience lab and give my niece the attention that she needed. So, I decided to shift my focus to teaching mostly undergraduate classes. Teaching made it easier for me to get home at the same time each evening and spared me the stress and time required to manage people and projects in the lab.

It was hard to drop a research program that—up to that point—had defined my career and fueled my passions. To stay close to the research world, I began to assign journal articles in my upper-level undergraduate course, anticipating lively discussions about the latest discoveries. This failed miserably. My students would skim the papers, but they'd rarely dive into them fully. Many wouldn't even look at the figures, which I had expected them to focus on.

A clue to the problem came when I took a look at the introductory biology textbooks they had studied in earlier classes. There were abundant illustrations of scientific facts—the array of bones in a bird's wing, the structure of a bacterial flagellum—but hardly any of the figures looked like the data presented in scientific papers. Equally problematic, the books had vanishingly few illustrations of *how* key findings had been made, or of *who* did the work. Now it made sense: My students were comfortable memorizing facts, but they lacked insight into how those facts were generated and how the conclusions were drawn. The ingenuity of research—what I loved most about being a scientist—was lost on them.

This epiphany changed the way I used the primary literature in my teaching; I started to go for depth over breadth.

"I wanted [my students] to think deeply about the research process."

I spent multiple class sessions deconstructing a single paper with my students, analyzing each figure and table. I then asked, "If you had coauthored the paper we just studied, what would you do next?"

Some balked. "I'm not creative," they'd say. But I asked them to give it a try—adding a sense of urgency by announcing that, in a later class, we'd form "grant panels" that would rank their proposed studies and decide where to invest an imaginary pool of research funds.

After taking part in the panels, the students changed their tunes. They were amazed by the variety of follow-up studies their classmates had thought up. They argued passionately about which ideas were superior, expressing surprise when other panels made different choices: "Isn't it *obvious* that No. 6 is best?" It was a thrill to see each student commit to an idea, in the process discovering something about their own powers of invention. Afterward, one whip-smart woman told me that—for the first time—she realized that it was OK to come up with her own scientific ideas.

Could I have conveyed more information per minute by talking at my students? Sure. But that's not how I wanted to teach. My students already knew how to learn facts. I wanted them to think deeply about the research process and to develop their own inventiveness. I wanted them to tap into their imaginations.

In a famous lyric, Stephen Sondheim writes, "Look, I made a hat—where there never was a hat." To my decades of students, I tip my hat—hoping that what they learned about their own creativity is the knowledge that lasts. ■

Sally G. Hoskins is a professor emeritus at the City College of New York. Send your story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER